Parking on random trees

Alice Contat

▶ **To cite this version:**

Alice Contat. Parking on random trees. Probability [math.PR]. Université Paris-Saclay, 2023. English. NNT : 2023UPASM022 . tel-04229230

HAL Id: tel-04229230
https://theses.hal.science/tel-04229230v1

Submitted on 5 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

université PARIS-SACLAY

Parking on random trees

Parking sur des arbres aléatoires

Thèse de doctorat de l'université Paris-Saclay

École doctorale de mathématiques Hadamard n° 574 (EDMH)
Spécialité de doctorat : Mathématiques fondamentales
Graduate School : Mathématiques. Référent : Faculté des sciences d'Orsay

Thèse préparée au **Laboratoire de mathématiques d'Orsay** (Université Paris-Saclay, CNRS), sous la direction de **Nicolas CURIEN**, Professeur

Thèse soutenue à Paris-Saclay, le 29 Septembre 2023, par

Alice CONTAT

Composition du jury

Membres du jury avec voix délibérative

Jean-François LE GALL Professeur, Université Paris-Saclay	Président
Christina GOLDSCHMIDT Professeure, University of Oxford	Rapportrice & Examinatrice
Jean-François MARCKERT Directeur de recherche, Université de Bordeaux	Rapporteur & Examinateur
Thomas DUQUESNE Professeur, Sorbonne Université	Examinateur
Bénédicte HAAS Professeure, Université Sorbonne Paris Nord	Examinatrice
Yueyun HU Professeur, Université Sorbonne Paris Nord	Examinateur

THESE DE DOCTORAT
NNT : 2023UPASM022

université PARIS-SACLAY | ÉCOLE DOCTORALE de mathématiques Hadamard (EDMH)

Titre: Parking sur des arbres aléatoires

Mots clés: Arbres aléatoires, percolation, algorithmes, parking

Résumé: Cette thèse porte sur l'étude de modèles de parking dans un sens large sur de grands graphes et arbres aléatoires. Dans un premier temps, nous étudions deux algorithmes permettant d'obtenir un grand ensemble indépendant d'un graphe, c'est-à-dire un sous-ensemble de sommets du graphe qui ne contient pas de paire de sommets voisins. Le premier utilise une stratégie gloutonne pour construire un ensemble indépendant maximal pour l'inclusion. L'ensemble ainsi obtenu a en général une densité positive et nous donnons des exemples de grands graphes (aléatoires) où l'on peut calculer exactement la loi de la taille de l'ensemble indépendant que l'on obtient. Le second algorithme est l'algorithme de Karp–Sipser, qui est optimal au sens où il existe un ensemble indépendant de taille maximale qui contient le sous-ensemble de sommet produit par l'algorithme. Cependant, cet algorithme s'arrête lorsque le sous-graphe inexploré ne contient plus de feuilles et on appelle alors ce sous-graphe le cœur de Karp–Sipser. Nous donnons la localisation de la transition de phase pour l'existence d'un cœur de Karp–Sipser géant pour un modèle de configuration avec des sommets de degrés 1, 2 et 3, et nous analysons précisément la taille de ce cœur au point critique.

Dans un second temps, nous nous intéressons au modèle de parking (dynamique) introduit par Konheim et Weiss dans le cas de la ligne. Dans cette version, on se place sur un arbre enraciné où chaque sommet représente une place de parking et les arêtes sont orientées vers la racine. Les voitures arrivent sur les sommets, se garent dès que possible en suivant les arêtes orientées et sortent de l'arbre par la racine si elles ne trouvent pas de place. On s'attend naturellement à observer une transition de phase. En effet, si peu de voitures arrivent, la plupart d'entre elles va pouvoir se garer tandis que si la densité de voitures est trop élevée, une proportion positive de voitures ne trouvera pas de place disponible et sortira par la racine de l'arbre. Nous formalisons d'abord l'existence de cette transition de phase pour une suite d'arbres qui converge vers une limite locale sous des hypothèses assez légères. Ensuite, nous en donnons la localisation pour des arbres de Bienaymé–Galton–Watson critiques en utilisant à nouveau la limite locale, et sur l'arbre binaire infini via des techniques combinatoires. Pour les arbres critiques, nous montrons également le caractère abrupte de la transition de phase. De plus, pour un modèle particulier d'arbres et d'arrivées de voitures, un couplage entre le modèle de parking et le modèle de graphe d' Erdős–Rényi nous permet notamment d'étudier la fenêtre critique de la transition de phase et fournit des informations sur les arbres de voitures garés. Enfin, nous établissons un lien entre le modèle de parking et les cartes planaires via une décomposition combinatoire issue du parking vis-à-vis de la dernière voiture.

université PARIS-SACLAY | ÉCOLE DOCTORALE de mathématiques Hadamard (EDMH)

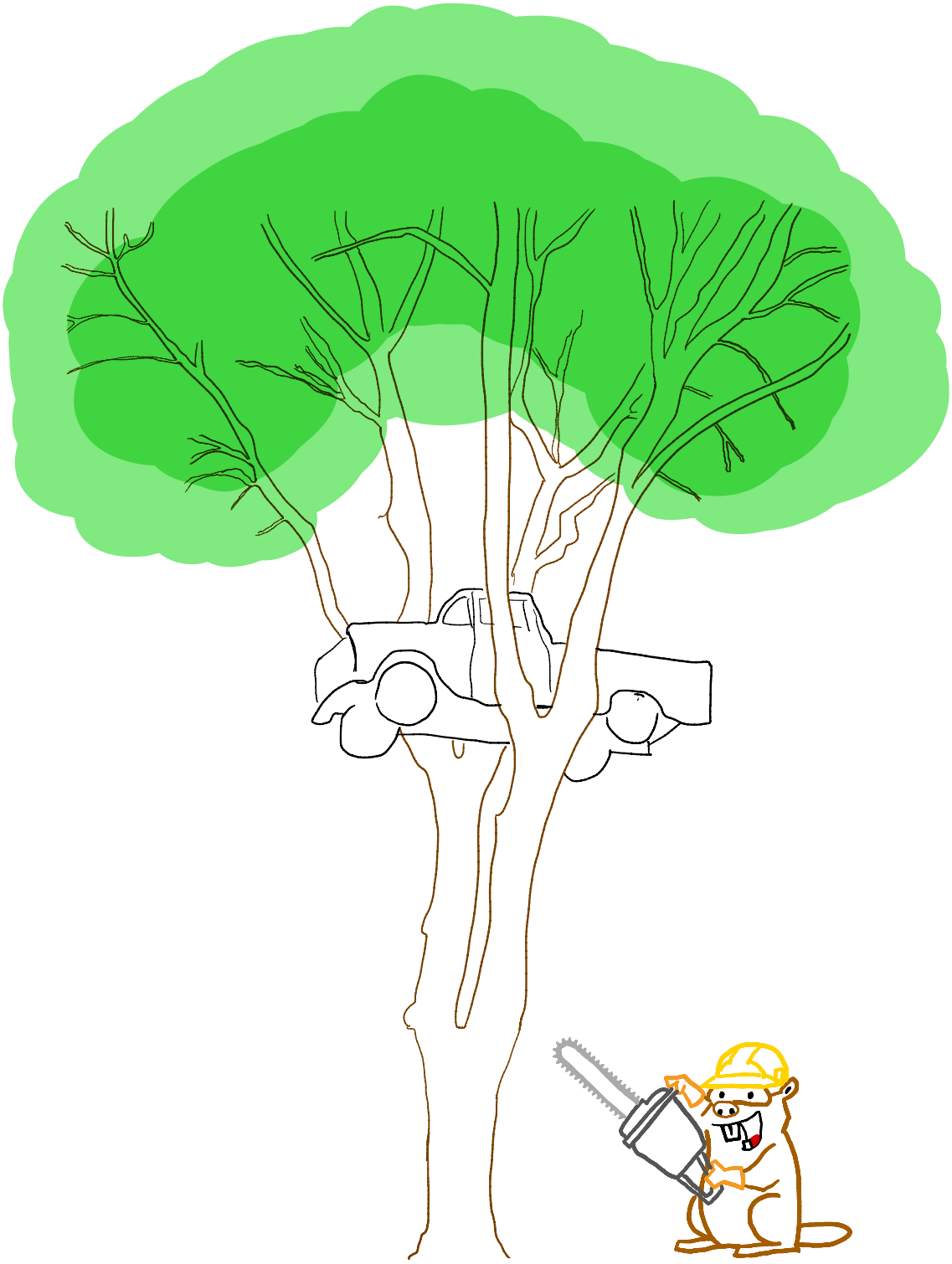
Title: Parking on random trees

Keywords: Random trees, parking, percolation, algorithms

Abstract: This thesis deals with the study of parking models on random graphs and trees in a broad sense. First we investigate two algorithms which enable us to find a large independent set of a graph, that is a subset of the vertices of the graph where no pair of vertices are connected to each other. The first one uses a greedy procedure to construct an independent set which is maximal for the inclusion order. In the generic case, this subset has a positive density and we give example of large (random) graphs for which we can explicitly compute the law of the size of this greedy independent set. The second algorithm is Karp–Sipser algorithm which is optimal in the sense that there exists an independent set with the maximal possible size which contains the subset of vertices produced by Karp–Sipser algorithm. However, this algorithm stops when the unexplored subgraph contains no more leaves and we call this subgraph the Karp–Sipser core. We give the precise localization of the phase transition for the existence of a giant Karp–Sipser core for a configuration model with vertices of degree 1, 2 and 3, and we precisely analyse its size at criticality.

Then, we examine the (dynamical) parking model introduced by Konheim and Weis on the line. In this version, we consider a rooted tree where each vertex represents a park spot and the edges are oriented towards the root. Cars arrive on the vertices of the tree, each car tries to park on its arrival node and if the spot is already occupied, it drives towards the root and parks as soon as possible. We expect to observe a phase transition. Indeed, if few cars arrive, most of them can park on the tree whereas if the "density" of cars is too large, then a positive proportion of them will not manage to park and exit the tree. We first formalize the existence of this phase transition for a sequence of trees which converges locally under slight hypotheses. We then give the localisation of this transition for critical Bienaymé–Galton–Watson trees using again the local limit, and for the infinite binary trees via a combinatorial decomposition. On critical trees, we also show that the phase transition is sharp. Moreover, for a good choice of trees and car arrivals, a coupling between the parking model and the Erdős–Rényi random graph model enables us to study the critical window of the phase transition and gives information about the geometry of the clusters of parked cars. Lastly, we establish an unexpected link between the parking model and planar maps by using a "last car" decomposition.

Parking on (random) trees



Alice Contat

Table des matières

Chapitre 1 :

Introduction

Cette thèse s'articule essentiellement en deux parties. La première sera centrée sur un modèle de parking statique qui permet d'obtenir des grands ensembles indépendants d'un graphe (aléatoire ou non). Elle contient les articles [65] et [52]. La deuxième représente le cœur de ce manuscrit et traite du modèle de parking dynamique sur des arbres. Les travaux [64, 67, 12, 63] y figurent.

L'introduction de cette thèse est divisée en quatre sections. Dans un premier temps (Section 1.1), nous présentons le contexte général des arbres et graphes aléatoires, notamment la notion de limite locale ainsi que le principe d'exploration markovienne qui seront des outils cruciaux dans nos travaux. Les parties suivantes donnent un aperçu de nos contributions principales. En particulier, la Section 1.2 rassemble nos résultats issus de [65, 52] liés à la construction de grands ensembles indépendants. La troisième section (Section 1.3) présente nos contributions propres au modèle de parking dynamique sur les arbres [64, 67, 12] tandis que la dernière (Section 1.4) montre le lien inattendu entre ce modèle et celui des cartes planaires [63].

Contents

1.1 Arbres et graphes aléatoires

Nous introduisons dans cette partie les différents modèles d'arbres et de graphes qui forment le support de cette thèse.

1.1.1 Arbres et graphes

Graphes. Commençons par rappeler qu'un graphe $\mathbf{g}$ est la donnée d'un ensemble de sommets V et d'un (multi-)ensemble E de paires de sommets appelées arêtes. Selon le contexte, on considérera des arêtes *orientées* (si $(x,y) \in E$, alors il y a une arête allant du sommet x vers le sommet y), ou non (si $\{x,y\} \in E$, il y a une arête reliant les sommets x et y). Également, on s'autorisera parfois des *arêtes multiples* (plusieurs arêtes reliant la même paire de sommets) et des *boucles* (arêtes dont les deux extrémités sont le même sommet), voir Figure 1.1. En l'absence d'arête multiple et de boucle, on parle de graphe *simple*.

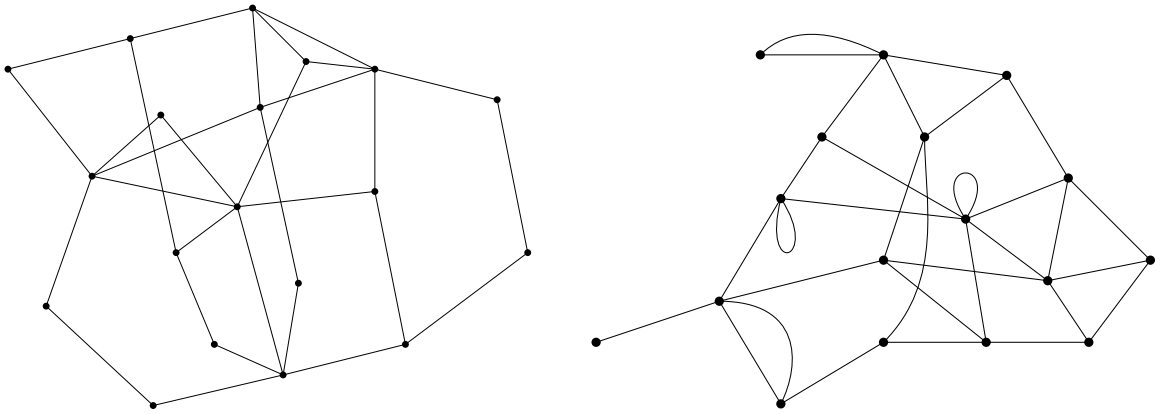


Figure 1.1 : À gauche, un graphe simple avec 19 sommets et 30 arêtes. Sur la droite, un (multi-)graphe, avec 17 sommets et 32 arêtes dont 2 boucles et 4 arêtes multiples. Dans les deux cas, les arêtes ne sont pas orientées, et les deux graphes sont connexes.

Tous les graphes que nous considérons ont un nombre fini ou au plus dénombrable de sommets et d'arêtes, et ils sont *localement finis*, c'est-à-dire que le *degré* (nombre d'arêtes incidentes, les boucles étant comptées deux fois) de chaque sommet est fini. On s'intéressera souvent à la *distance* (de graphe) entre deux sommets $x,y \in V$, qui est le plus petit nombre d'arêtes d'un chemin non orienté reliant x et y. Si pour toute paire de sommets, il existe un chemin reliant les deux sommets, on dira que le graphe est *connexe*. Pour un graphe fixé $\mathbf{g}$, on notera $|\mathbf{g}|$ son nombre de sommets ou *taille*, sauf mention contraire.

Arbres. Parmi les graphes, une catégorie nous intéressera particulièrement : *les arbres*. Un arbre est un graphe (simple) connexe et acyclique. En particulier, on peut remarquer qu'un arbre fini possède $n-1$ arêtes s'il est composé de $n \geqslant 1$ sommets. Quand un sommet particulier, que l'on appelle *racine*,

est distingué, on munit naturellement l'arbre d'une structure généalogique dont l'ancêtre commun est la racine. Plus précisément, étant donné un sommet x à distance $r \geqslant 1$ de la racine, il y a un unique sommet y à distance $r-1$ de la racine situé sur l'unique chemin reliant x à la racine. On dit alors que x est un *enfant* de y, et que y est donc le *parent* (unique) de x, voir Figure 1.2. En général, on représente la racine en bas de l'arbre et on la note souvent $\varnothing$ lorsqu'il n'y a pas d'ambiguïté[1].

Enfin, on appelle *arbre plan* un arbre enraciné muni d'un ordre cyclique autour de chaque sommet : pour chaque sommet x de l'arbre, les enfants de x sont numérotés, et l'on représente le premier enfant de x à gauche dans le plan, le deuxième enfant directement à droite du premier et ainsi de suite. Pour une définition plus formelle d'un arbre plan comme un sous-ensemble de l'arbre d'Ulam, voir par exemple [144] ou [73]. On peut montrer qu'il y a $\frac{1}{n}\binom{2n-2}{n-1}$ arbres plans avec n sommets.

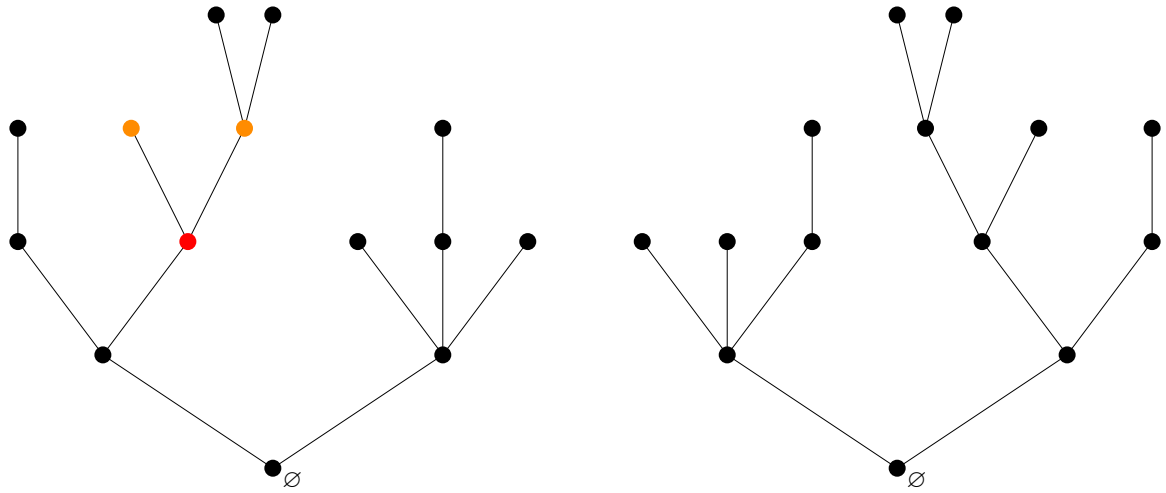


Figure 1.2 : Deux exemples d'arbres enracinés en $\varnothing$. À gauche, les deux sommets oranges sont les enfants du sommet rouge. Ce sont les mêmes arbres enracinés mais ce ne sont pas les même arbres plans.

Arbres de Bienaymé–Galton–Watson. Un des modèles les plus étudiés d'arbres aléatoires est le modèle de Bienaymé–Galton–Watson. Il a été introduit par Bienaymé puis indépendamment par Galton et Watson pour étudier l'extinction des patronymes des familles nobles en Angleterre. Soit ν une mesure de probabilité sur $\{0,1,2,\dots\}$. Un arbre de Bienaymé–Galton–Watson de loi de reproduction ν est un arbre *plan* où l'on part d'un individu racine à la génération 0, et où chaque individu a un nombre aléatoire d'enfants (numérotés) suivant la loi ν, indépendamment pour tous les individus. Autrement dit, la racine de l'arbre a X enfants où X a loi ν et les sous-arbres issus de chacun de ces enfants sont des arbres de Bienaymé–Galton–Watson de loi de reproduction ν indépendants, et indépendant de X. Voir [144] pour une définition formelle.

[1] Cela provient de la notation de Neveu pour les arbres plans, voir [144].

Une des forces du modèle de Bienaymé–Galton–Watson est qu'il englobe, pour de bons choix de loi de reproduction, de nombreux modèles combinatoires. Par exemple, si on veut obtenir un arbre non-plan étiqueté (aussi appelé *arbre de Cayley*) de taille n uniformément au hasard, il suffit de tirer un arbre de Bienaymé–Galton–Watson avec loi de reproduction $\mathsf{Poisson}(1)$ conditionné à avoir taille n, et conditionnellement à cet arbre, assigner des étiquettes aux sommets uniformément au hasard et oublier l'orientation autour de chaque sommet. De même, un arbre de Bienaymé–Galton–Watson de loi de reproduction géométrique (critique) conditionné à avoir n sommets a la même loi qu'un arbre plan (non étiqueté) de taille n uniforme.

Graphe d'Erdős-Rényi. Le modèle de graphe aléatoire le plus "populaire" en probabilités est celui d'Erdős–Rényi. C'est un graphe simple non orienté noté $G(n,p)$ pour $n \geqslant 1$ et $p \in [0,1]$, ayant pour ensemble de sommets $\{1,\dots,n\}$ et pour chaque paire (i,j), une arête est présente entre les sommets i et j avec probabilité p, et ce, indépendamment pour chaque arête possible. En fait, le modèle sous cette forme a été introduit par Gilbert [98] en 1959, et la même année, Erdős et Rényi [83, 84] ont introduit un modèle assez similaire mais avec un nombre d'arêtes m fixé. Nous utiliserons plutôt cette version, et même, on s'autorisera une version avec des arêtes multiples et des boucles. Plus précisément :

Définition 1.1. *Soit $n,m \in \{0,1,2,\dots\}$. Le (multi-)graphe d'Erdős–Rényi $G(n,m)$ est un graphe aléatoire avec n sommets numérotés $\{1,\dots,n\}$ et m arêtes (non ordonnées) indépendantes tel que pour tout $1 \leqslant i \leqslant m$, les deux extrémités de la i-ième arête sont deux sommets indépendants et uniformes.*

De cette définition, il est aisé de voir qu'on peut coupler les graphes à nombre de sommets n fixé pour définir un processus $(G(n,m) : m \geqslant 0)$ tel pour que pour tout $m \geqslant 0$, on passe de $G(n,m)$ à $G(n,m+1)$ en ajoutant une arête dont les extrémités sont deux sommets indépendants choisis uniformément au hasard indépendamment de $G(n,m)$.

Version gelée. Nous nous intéresserons également à une version modifiée du graphe d'Erdős–Rényi, introduite dans [67], que nous appelons "gelée" car nous allons supprimer certaines arêtes, et donc ralentir, ou "geler" partiellement la croissance de certaines composantes connexes (celles qui ne sont pas des arbres). Plus précisément, nous définissons un processus de graphes[2] $(F(n,m) : m \geqslant 0)$ dont les n sommets étiquetés $\{1,\dots,n\}$ peuvent être soit blancs, soit bleus. On se donne une famille de variables aléatoires représentant les extrémités des potentielles arêtes $((X_m,Y_m) : m \geqslant 1)$ indépendantes et uniformes sur $\{1,\dots,n\}^2$. Initialement, le graphe $F(n,0)$ est composé des n sommets étiquetés $\{1,\dots,n\}$ coloriés en blanc et ne contient pas d'arête. Ensuite, on ajoute les arêtes une par une selon la règle suivante, voir Figure 1.3 : pour tout $m \geqslant 1$,

- Si X_m et Y_m sont deux sommets blancs dans $F(n,m-1)$, alors on ajoute une arête entre X_m et Y_m à $F(n,m-1)$ pour obtenir $F(n,m)$. De plus, si cette arête crée un cycle, alors on colorie tous les sommets de la composante connexe de X_m et Y_m en bleu.

[2] F pour *frozen* en anglais.

- Si X_m est un sommet bleu, alors on laisse $F(n,m) = F(n,m-1)$.
- Enfin si X_m est un sommet blanc et Y_m est un sommet bleu, on ajoute une arête entre X_m et Y_m à $F(n,m-1)$ et on colorie tous les sommets de la composante connexe de X_m (et Y_m) pour obtenir $F(n,m)$.

On dit que les composantes bleues sont les composantes *gelées* du graphe et l'ensemble de ces composantes gelées forme le *congélateur*.

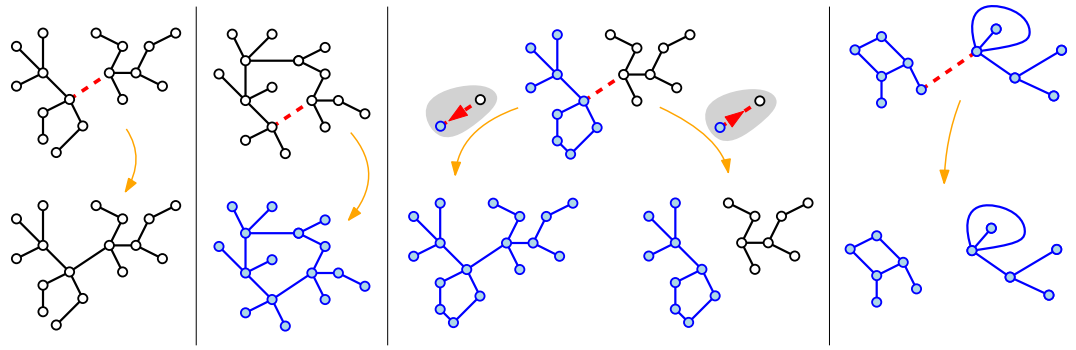


Figure 1.3 : Illustration des règles de transition pour le modèle de graphe d'Erdős–Rényi gelé. La nouvelle arête potentielle est en pointillé rouge. Si elle est entre deux sommets blancs (deux premières figures à gauche), on la garde, et on colorie sa composante en bleue si elle crée un cycle (deuxième figure). Si elle est entre un sommet bleu et blanc, on ne la garde que si elle va du sommet blanc vers le bleu et toute la nouvelle composante est déclarée gelée et bleue (troisième figure). On supprime l'arête si elle arrive entre deux sommets bleus (quatrième figure).

Notons qu'on peut construire la version classique et la version gelée du graphe d'Erdős–Rényi à partir de la même suite d'arêtes $((X_m, Y_m) : m \geqslant 1)$.

Modèle de configuration. Un autre modèle bien étudié dans la littérature probabiliste est le *modèle de configuration*. À l'inverse du graphe d'Erdős–Rényi où les degrés des sommets sont aléatoires, ce modèle permet d'obtenir un graphe aléatoire où les degrés des sommets sont prescrits.

Définition 1.2. *Soit* $\mathbf{d} = (d_i)_{i\geqslant 1}$ *une suite d'entiers telle que* $\sum_{i\geqslant 1} id_i = 2m$ *est pair. Notons* $n = \sum_{i\geqslant 1} d_i$. *Le modèle de configuration sur* $\mathbf{d}$, *noté* $CM(\mathbf{d})$, *est le (multi-)graphe aléatoire obtenu à partir de* n *sommets, dont* d_i *sommets de degré* i *pour tout* $i \geqslant 1$, *en appariant les* $2m$ *demi-arêtes ou "pattes" des sommets uniformément au hasard.*

La condition de parité sur la somme des id_i garantit qu'on puisse apparier toutes les demi-arêtes. Notons que ici, on s'autorise les arêtes multiples et les boucles. On s'est également restreint au cas où il n'y a pas de sommet de degré 0 puisque ces sommets ne jouent pas un rôle important dans le graphe. Pour une définition plus rigoureuse, il faudrait en fait étiqueter les demi-arêtes que l'on apparie, voir par exemple [41, Section 2.4].

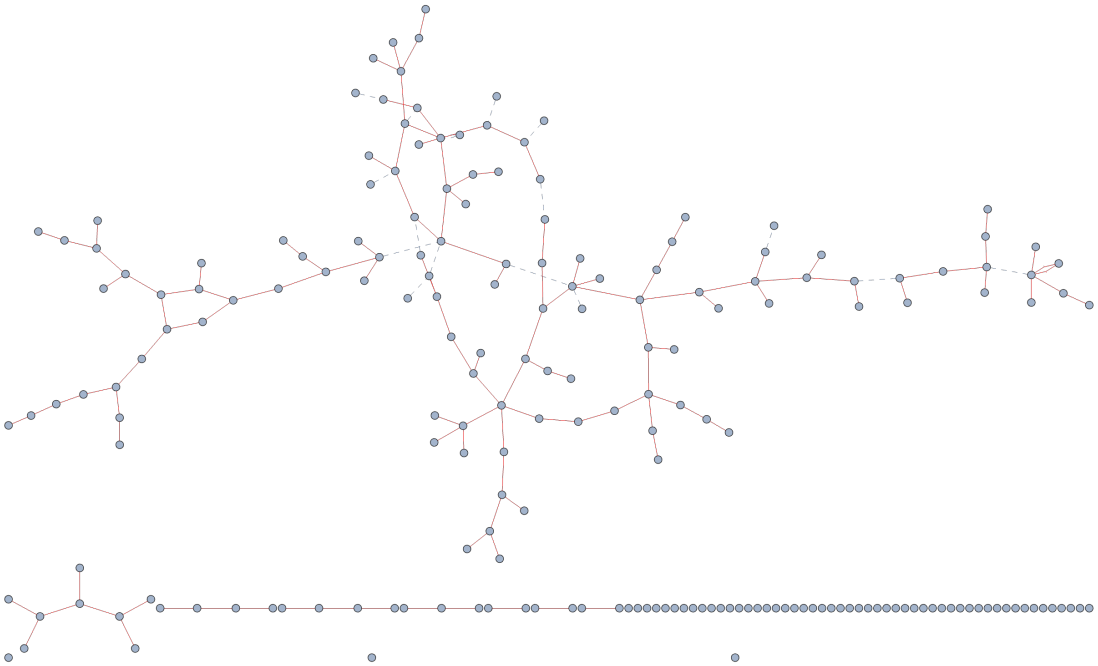


Figure 1.4 : Une réalisation de $F(200,130)$ et de $G(200,130)$ où les deux graphes sont construits en utilisant les mêmes arêtes. Les arêtes rouges sont les arêtes de $G(200,130)$ qui sont gardées dans $F(200,130)$ et celles en pointillé bleu sont les arêtes supprimées. Notons qu'on ne supprime jamais d'arêtes entre deux composantes qui n'ont pas de cycle.

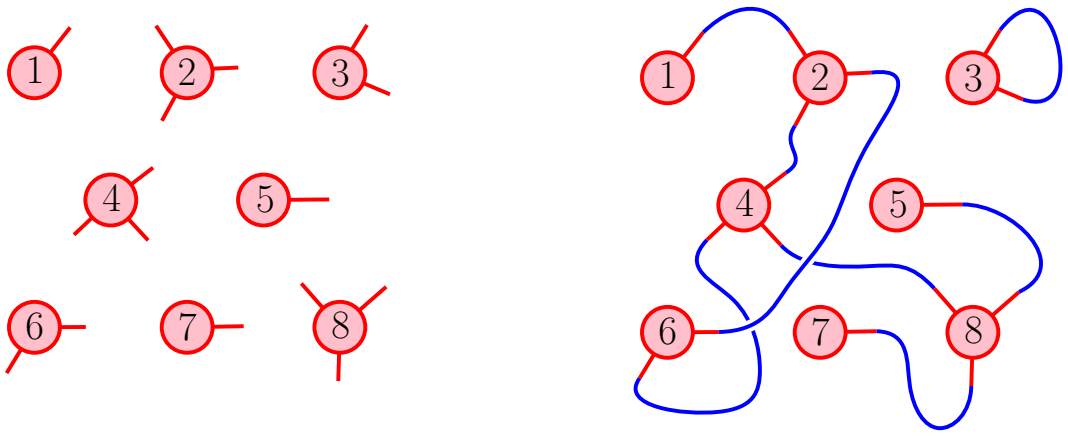


Figure 1.5 : Un exemple d'appariement d'arêtes pour le modèle de configuration.

1.1.2 Limite locale et méthode objective.

Il y a plusieurs manières d'étudier un "grand" graphe : on peut soit regarder de manière globale le graphe, "vu de loin", soit s'intéresser aux propriétés locales du graphe, autour d'un point "typique".

C'est ce second point de vue que nous allons décrire et adopter dans cette section. Rappelons que les graphes que l'on considère (au moins dans cette section) ne sont pas plans (ce ne sont pas des cartes), et sont localement finis.

Graphe enraciné. La notion de convergence *locale* a été introduite par Benjamini et Schramm [26]. Heuristiquement, la topologie locale rend compte du "paysage" local autour d'un point du graphe. Pour ce faire, il faut pouvoir distinguer un sommet du graphe et on va donc considérer des graphes enracinés (avec un sommet distingué) et connexes. Si $(\mathbf{g}_1, \rho_1)$ et $(\mathbf{g}_2, \rho_2)$ sont deux graphes connexes enracinés en ρ_1 et ρ_2 respectivement, alors on définit la distance locale

$$\mathrm{d}_{\mathrm{loc}}((\mathbf{g}_1, \rho_1), (\mathbf{g}_2, \rho_2)) = (1 + \sup\{r \geqslant 0 : B_r(\mathbf{g}_1, \rho_1) = B_r(\mathbf{g}_2, \rho_2)\})^{-1}$$

où $B_r(\mathbf{g}, \rho)$ désigne la boule de rayon r autour de la racine ρ de $\mathbf{g}$, c'est-à-dire le sous-graphe induit par tous les sommets situés à distance (de graphe) au plus r de ρ. Notons que l'égalité entre les deux boules est vue à isométrie de graphes préservant la racine près. En effet, nous voulons que cette distance nous permette de comparer la géométrie locale de deux graphes enracinés, quelque soit l'étiquetage de leurs sommets. On notera $\mathcal{G}^\bullet$ l'ensemble des graphes localement finis, connexes et enracinés en un sommet, vus à isométrie (préservant la racine) près. En particulier, deux graphes enracinés sont à une distance (locale) plus petite que $1/(r+1)$ si les voisinages autour de leur racine respective coïncident au moins jusqu'à une distance r. La topologie induite par cette distance sur $\mathcal{G}^\bullet$ est appelée *topologie locale*, et $(\mathcal{G}^\bullet, \mathrm{d}_{\mathrm{loc}})$ est un espace métrique séparable et complet.

Une suite déterministe $(\mathbf{g}_n, \rho_n)_{n\geqslant 1}$ de graphes connexes enracinés converge pour la topologie locale si pour tout $r \geqslant 0$, la boule de rayon r autour de la racine ρ_n de $\mathbf{g}_n$ converge (donc est constante à partir d'un certain rang). Par extension, si les graphes $\mathbf{g}_n$ ne sont pas nécessairement connexes, on dit que $(\mathbf{g}_n, \rho_n)_{n\geqslant 1}$ converge pour la topologie locale si la composante connexe de ρ_n dans $\mathbf{g}_n$ converge pour la topologie locale. On peut également s'intéresser à la convergence en loi dans l'espace métrique $(\mathcal{G}^\bullet, \mathrm{d}_{\mathrm{loc}})$. Une manière de formuler cette convergence que nous utiliserons dans la suite, est de dire qu'une suite de graphes enracinés éventuellement aléatoires $(G_n, \rho_n)_{n\geqslant 1}$ converge en loi pour la topologie locale vers un graphe (G_∞, ρ_∞), lui aussi localement fini, si pour tout $r \geqslant 0$, la boule $B_r(G_n, \rho_n)$ de rayon r autour de ρ_n dans G_n converge en loi vers $B_r(G_\infty, \rho_\infty)$, autrement dit si pour toute fonction f continue bornée telle que $f(\mathbf{g}, \rho)$ ne dépend que de la boule $B_r(\mathbf{g}, \rho)$ pour un $r \geqslant 0$, alors

$$\mathbb{E}\left[f(G_n, \rho_n)\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[f(G_\infty, \rho_\infty)\right].$$

Graphe non-enraciné. Si on veut donner une définition de convergence locale pour des graphes non enracinés mais finis, on peut également se donner un choix (déterministe ou aléatoire) de racine. Le point de vue adopté par la convergence au sens de Benjamini et Schramm est de choisir un sommet racine uniformément au hasard. Introduisons donc $\mathcal{G}_{\mathrm{fini}}$ l'ensemble des graphes finis, connexes et vus à isométrie près. Plus précisément, si on a $(G_n)_{n\geqslant 1}$ une suite de graphes (aléatoires) de $\mathcal{G}_{\mathrm{fini}}$, pour chaque $n \geqslant 1$, conditionnellement à G_n, on note X_n un sommet uniforme de G_n.

On dit alors que $(G_n)_{n\geqslant 1}$ converge vers $(G_\infty, \rho_\infty) \in \mathcal{G}^\bullet$ au sens de Benjamini-Schramm si (la composante connexe de X_n dans) (G_n, X_n) converge vers (G_∞, ρ_∞) pour la topologie locale, c'est-à-dire si pour tout $r \geqslant 0$ et toute fonction f continue bornée telle que $f(\mathbf{g}, \rho)$ ne dépend que de la boule $B_r(\mathbf{g}, \rho)$, alors

$$\mathbb{E}\left[f(G_n, X_n)\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[f(G_\infty, \rho_\infty)\right].$$

On peut se demander quelles sont les limites possibles pour le graphe enraciné aléatoire (G_∞, ρ_∞). On dira que f est une *fonction de transport* si f est une fonction qui prend en entrée un graphe et deux points (ordonnés) de ce graphe et renvoie un réel positif, telle s'il existe une isométrie de graphes qui envoie un graphe $\mathbf{g}_1$ sur un graphe $\mathbf{g}_2$ et une paire de points (x_1, y_1) de $\mathbf{g_1}$ sur une paire de points (x_2, y_2) de $\mathbf{g_2}$, alors $f((\mathbf{g}_1, x_1, y_1)) = f((\mathbf{g}_2, x_2, y_2))$ (f est invariante par isométrie de graphes qui préserve deux points). Une propriété importante [74, Théorème 7] est que si $(G_n)_{n\geqslant 1}$ converge au sens de Benjamini–Schramm vers un graphe (infini ou non) (G_∞, ρ_∞), alors (G_∞, ρ_∞) est *unimodulaire*, c'est-à-dire que pour toute fonction de transport f,

$$\mathbb{E}\left[\sum_{x\in G_\infty} f(G_\infty, \rho_\infty, x)\right] = \mathbb{E}\left[\sum_{x\in G_\infty} f(G_\infty, x, \rho_\infty)\right].$$

On dit aussi que (G_∞, ρ_∞) vérifie le principe de *transport de masse*. De plus, les graphes unimodulaires ne peuvent avoir que $0, 1, 2$ ou une infinité de bouts. Ce nombre peut éventuellement être aléatoire, mais si on se restreint au cas des arbres, des simplifications apparaissent. La notion de bouts coïncide pour les arbres avec celle d'épines dorsales, c'est-à-dire de suite de sommets $(v_i)_{i\geqslant 1}$ telle que v_i et v_{i+1} sont reliés, vue à égalité (éventuellement décalée) à partir d'un certain rang, près, voir par exemple [74, Théorème 14]. Si on se donne une suite d'arbres qui a une limite au sens de Benjamini–Schramm, alors la limite est un arbre avec presque sûrement 1 ou 2 épines dorsales, voir par exemple [74, Théorème 13].

Benjamini-Schramm quenched. Enfin, dans certains cas, nous aurons besoin d'une version encore un peu plus forte de convergence locale, qui introduit une forme d'indépendance. Soit $(G_n)_{n\geqslant 1}$ une suite de graphes aléatoires finis, et pour chaque $n \geqslant 1$, conditionnellement à G_n, on note X_n et Y_n deux sommets indépendant choisis uniformément au hasard, et (G_∞, ρ_∞) un graphe enraciné. On dit que $(G_n)_{n\geqslant 1}$ converge vers (G_∞, ρ_∞) au sens de *Benjamini–Schramm quenched* si les suites $(G_n, X_n)_{n\geqslant 1}$ et $(G_n, Y_n)_{n\geqslant 1}$ convergent localement vers deux copies indépendantes de même loi que (G_∞, ρ_∞). En particulier, la suite $(G_n)_{n\geqslant 1}$ converge vers (G_∞, ρ_∞) au sens de Benjamini–Schramm.

Autrement dit, la suite $(G_n)_{n\geqslant 1}$ converge vers (G_∞, ρ_∞) au sens Benjamini–Schramm quenched si pour tout $r \geqslant 0$ pour toutes fonctions f et g continues bornées telles que $f(\mathbf{g}, \rho)$ et $g(\mathbf{g}, \rho)$ ne dépendent que de la boule $B_r(\mathbf{g}, \rho)$,

$$\mathbb{E}\left[f(G_n, X_n) g(G_n, Y_n)\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[f(G_\infty, \rho_\infty)\right] \mathbb{E}\left[g(G_\infty, \rho_\infty)\right].$$

Cette notion de convergence est similaire à la notion de convergence presque sûre au sens de Benjamini–

Schramm [97, 162]. Une autre façon de formuler cette convergence est de regarder la mesure empirique

$$\mu(G_n) = \frac{1}{|V(G_n)|} \sum_{v \in V(G_n)} \delta_{(G_n, v)}.$$

Alors $(G_n)_{n \geqslant 1}$ converge vers (G_∞, ρ_∞) au sens de Benjamini–Schramm quenched si la mesure $\mu(G_n)$ converge en probabilité vers $\delta_{(G_\infty, \rho_\infty)}$.

La plupart des graphes aléatoires issus de modèles "classiques" ont une limite au sens de Benjamini–Schramm, voire au sens de Benjamini–Schramm quenched. Nous donnons deux exemples ci-dessous.

Premier exemple : Arbres de Bienaymé–Galton–Watson critiques. Intéressons nous au cas des arbres de Bienaymé–Galton–Watson. On se fixe une loi de reproduction ν et on note $\mathcal{T}$ un arbre de Bienaymé–Galton–Watson de loi de reproduction $\nu = \sum_{k \geqslant 0} \nu_k \delta_k$. On suppose que ν a pour espérance 1, et que ν est apériodique, c'est-à-dire pour tout n assez grand $\mathbb{P}(|\mathcal{T}| = n) > 0$, où on rappelle que $|\mathcal{T}|$ désigne le nombre de sommets de $\mathcal{T}$. On note $\mathcal{T}_n$ la version de $\mathcal{T}$ conditionnée à avoir taille n (lorsque cela est possible), et ρ_n la racine de l'arbre $\mathcal{T}_n$.

Introduisons les deux arbres infinis qui vont apparaître lorsque l'on s'intéresse aux limites locales de ces arbres. D'une part, on a l'*arbre de Kesten* [117] que l'on peut décrire de la manière suivante : c'est un arbre plan dont les sommets de l'arbre peuvent être soit "normaux", soit "mutants". On part d'un sommet racine de type mutant. Tous les sommets ont un nombre aléatoire d'enfants indépendamment les uns des autres. Pour les sommets normaux, ce nombre d'enfants a pour loi ν et tous les enfants sont normaux. Les sommets mutants se reproduisent selon la loi biaisée par la taille $\overline{\nu} = \sum_{k \geqslant 0} k \nu_k \delta_k$. Notons que $\overline{\nu}$ est bien une mesure de probabilité puisque l'on a supposé que $\sum_{k \geqslant 1} k \nu_k = 1$, et que cette loi est supportée par les entiers strictement positifs. Parmi les enfants d'un sommet mutant, l'un d'eux est choisi uniformément au hasard (indépendamment de toutes les générations précédentes) et est déclaré mutant, et les autres sont normaux. Il y a donc exactement un sommet mutant à chaque génération, voir Figure 1.6.

L'autre arbre qui apparaît est l'arbre d'Aldous[3] [7] construit de la manière suivante : on part d'un arbre de Bienaymé–Galton–Watson de loi de reproduction ν, enraciné sur le sommet initial. On ajoute à la racine un enfant vers le bas, qui est le point de départ (mutant) d'un arbre de Kesten. Dans le cas de l'arbre d'Aldous, on représente souvent l'arbre de Kesten sous-jacent vers le bas, voir Figure 1.6. La relation parents/enfants est alors inversée le long de la lignée de mutants. On peut remarquer que si ν est la loi de Poisson de paramètre 1, alors l'arbre de Kesten et l'arbre d'Aldous ont même loi puisque dans ce cas, si X a loi $\nu = \mathsf{Poisson}(1)$, alors $X + 1$ a loi $\overline{\nu}$.

Munis de ces deux définitions, nous pouvons maintenant énoncer les résultats de convergence locale. D'abord, la suite d'arbres enracinés $(\mathcal{T}_n, \rho_n)_{n \geqslant 1}$ converge localement vers l'arbre de Kesten, voir par exemple [2]. En revanche la suite d'arbres $(\mathcal{T}_n)_{n \geqslant 1}$ converge au sens de Benjamini–Schramm (donc les arbres sont enracinés uniformément au hasard) vers l'arbre d'Aldous, voir par exemple [163].

[3] en anglais, *Aldous' sin-tree*, où le mot *sin* fait référence au fait qu'il y a un seul chemin infini (*single infinite path)*, comme pour l'arbre de Kesten

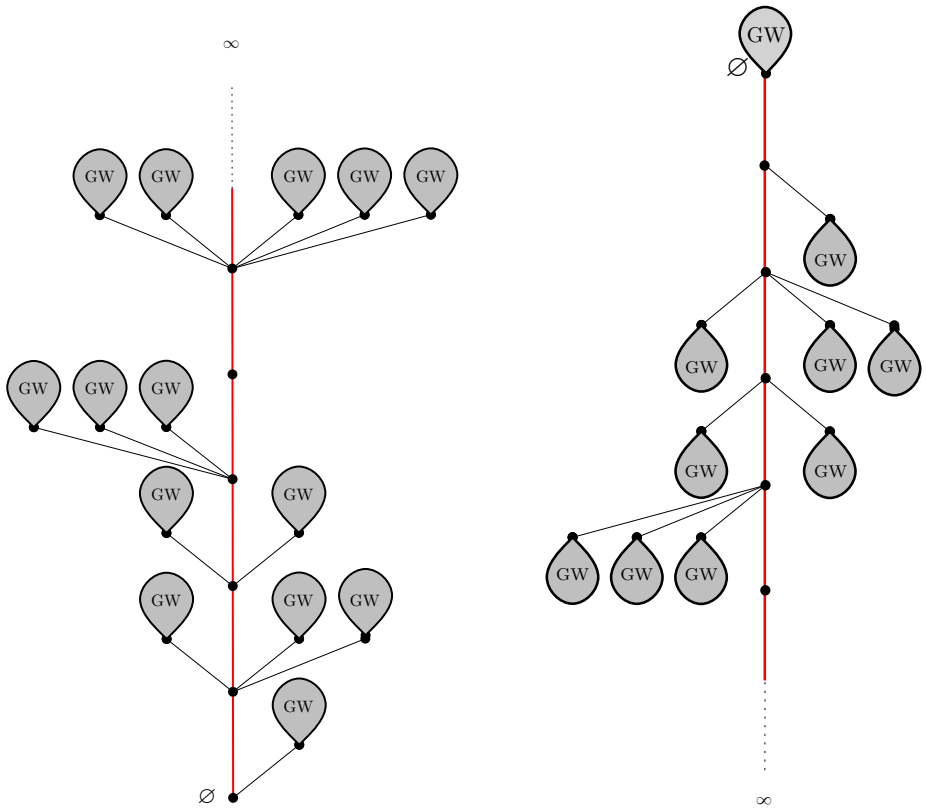


Figure 1.6 : à gauche, l'arbre de Kesten. Le sommet racine est mutant et le sommet rouge représente le chemin infini partant de la racine qui relie tous les sommets mutants. À droite, l'arbre d'Aldous où on a dessiné l'arbre de Kesten sous-jacent vers le bas.

Ces convergences sont d'ailleurs toujours vraies au sens plan, en conservant l'orientation cyclique naturelle autour de chaque sommet.

Deuxième exemple : graphe d'Erdős–Rényi. On peut aussi s'intéresser à la limite locale du graphe d'Erdős–Rényi $G(n,p)$. Comme les sommets ne jouent pas de rôle particulier, le graphe $G(n,p)$ enraciné en le sommet numéroté 1 a la même loi que le graphe $G(n,p)$ enraciné uniformément au hasard. Il suffit donc de s'intéresser à la limite de $G(n,p)$ au sens de Benjamini–Schramm. Le régime "intéressant" est lorsque le degré moyen d'un sommet est fixé, c'est-à-dire lorsque p est d'ordre c/n pour un $c>0$ fixé lorsque n est grand. Le résultat est alors le suivant : pour tout $c>0$, le graphe $G\left(n,\lfloor c/n\rfloor\right)$ converge au sens de Benjamini–Schramm vers un arbre de Bienaymé–Galton–Watson de loi de reproduction $\mathsf{Poisson}(c)$. Bien qu'il puisse y avoir des cycles dans $G\left(n,\lfloor c/n\rfloor\right)$, cela ne se voit pas dans la limite locale. Ce résultat est souvent utilisé pour montrer la transition de phase pour l'existence d'une composante géante à $c=1$.

1.1.3 Exploration markovienne

La plupart des modèles que nous définirons et étudierons dans la suite possèdent la structure suivante : d'abord on choisit un (grand) arbre ou graphe aléatoire, puis on ajoute sur ce graphe un modèle ou algorithme déterministe ou aléatoire. Une des principales techniques que nous utiliserons est de ne pas révéler l'arbre ou le graphe aléatoire entièrement dès le début, mais en fait de se servir de notre modèle ou algorithme pour le révéler pas-à-pas, au fur et à mesure de nos besoins.

Le modèle de configuration est l'exemple typique de graphes aléatoires sur lequel on peut appliquer cette technique. En effet, plutôt que d'apparier toutes les demi-arêtes à la fois pour obtenir un appariement uniforme, on peut en fait les choisir deux par deux uniformément et ainsi révéler pas-à-pas le graphe. Si on choisit deux demi-arêtes uniformément au hasard indépendantes et qu'on les apparie, puis qu'on apparie toutes les demi-arêtes restantes uniformément et indépendamment, alors l'appariement total est uniforme. Nous verrons une application de cette technique dans le chapitre 3. Cette technique d'exploration markovienne est aussi très utilisée pour les *cartes* ou graphes planaires. Elle est appelée dans ce cas *processus d'épluchage* ou *peeling process* en anglais, et nous l'expliciterons dans la section 1.4.1. Nous introduirons également une nouvelle exploration markovienne des arbres de Cayley qui sera un outil crucial pour les chapitres 2 et 6.

1.2 Parking statique sur des graphes

Nous motivons et présentons dans cette section nos travaux [52, 65] qui concernent une version statique d'un problème de parking.

1.2.1 Ensemble indépendant de sommets et appariement d'arêtes

Commençons par présenter les deux problèmes qui vont nous intéresser dans cette section, très classiques en théorie des graphes : trouver un *couplage* ou *appariement*, et trouver un *ensemble indépendant de sommets*. Un *couplage* d'un graphe est un sous-ensemble de arêtes qui ne contient pas de paire d'arêtes adjacentes. D'un point de vue "parking", on peut imaginer que des voitures se garent sur les arêtes et débordent sur les sommets adjacents, de sorte que d'autres voitures ne peuvent pas se garer sur les arêtes adjacentes aux arêtes "occupées". De manière similaire, un *ensemble indépendant* d'un graphe est un sous-ensemble de sommets qui ne contient pas de paire de sommets voisins. L'interprétation parking est similaire : on peut imaginer des voitures se garant sur les sommets et empiétant sur les arêtes adjacentes, si bien que les sommets voisins des sommets "garés" ne peuvent pas contenir de nouvelles voitures. On donne un exemple en Figure 1.7.

Étant donné un graphe, trouver un couplage ou trouver un ensemble indépendant de *taille* maximale peuvent sembler deux problèmes très similaires à première vue. Pourtant, le problème qui concerne les sommets est bien plus difficile que celui qui concerne les arêtes ! En effet, d'un point de vue algorithmique, trouver un couplage de taille maximale se fait en temps polynomial en la taille du graphe. Le premier à proposer un tel algorithme est Edmonds en 1985 (voir [82] et [151, Théorème

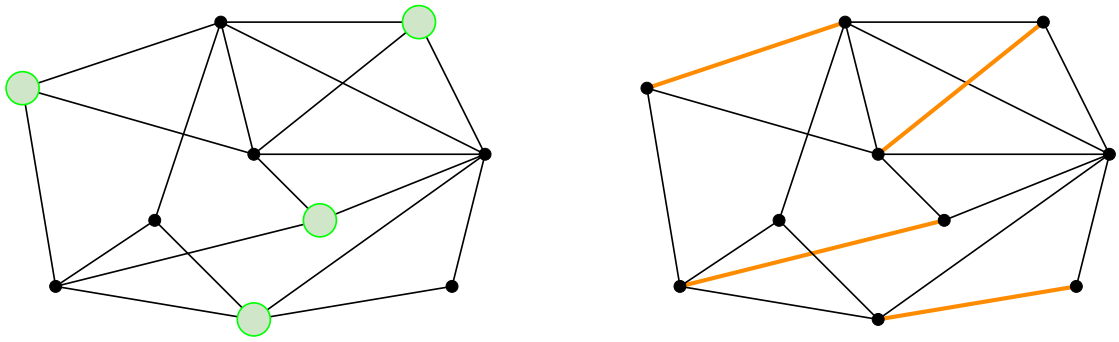


Figure 1.7 : À gauche, les sommets verts forment un ensemble indépendant du graphe. À droite, les arêtes oranges forment un couplage du graphe. Dans les deux cas, le couplage ou l'ensemble indépendant est maximal au sens où on ne peut pas ajouter d'arête ou de sommet, mais ils ne sont pas de taille maximale.

9.1.8]). Depuis, différentes améliorations ont été proposées, voir [151, Tableau 9.1.1] ou [38, 96, 171]. En revanche, construire un ensemble indépendant de taille maximale est un problème NP-complet [95]. Il est possible de faire un peu mieux que la méthode naïve consistant à examiner tous les sous-ensembles de sommets possibles et vérifier s'ils forment un ensemble indépendant, ce qui prendrait un temps $O(2^n \cdot n^2)$. De nombreux travaux ont consisté à trouver des algorithmes de plus en plus performants [175, Table 1], mais le temps nécessaire pour déterminer un ensemble indépendant de taille maximale reste exponentiel en le nombre de sommets du graphe, et ce, même si on se restreint à des classes de graphes ayant des degrés bornés (voir par exemple [176]). En fait, la présence d'un sommet dans un ensemble indépendant de taille maximale a une influence importante sur la présence d'autres sommets, même très lointains, ce qui n'est pas le cas pour les couplages.

1.2.2 Stratégie gloutonne

Puisque déterminer un ensemble indépendant de taille maximale est un problème difficile, commençons par étudier une stratégie dite "gloutonne"[4] qui permet de générer un ensemble indépendant relativement grand.

Algorithme glouton. Comme on vient de le dire, il est très difficile de déterminer un ensemble indépendant de taille maximale. En revanche, on peut facilement en construire un "grand", maximal pour l'inclusion, en ajoutant des sommets (compatibles) un par un, tant que c'est possible. On examine les sommets dans un certain ordre (déterministe ou aléatoire) et lorsqu'on ajoute un sommet, on "bloque" ses voisins puisqu'ils ne pourront pas faire partie de l'ensemble indépendant final. Plus précisément, on divise les sommets en trois catégories : les sommets indéterminés, les sommets actifs et les sommets bloqués. Initialement, tous les sommets sont indéterminés. À chaque étape, on choisit

[4] ou *greedy* en anglais

un sommet indéterminé, change son statut en "actif" et tous les voisins indéterminés de ce sommet deviennent bloqués. On s'arrête lorsque qu'il n'y a plus de sommet indéterminé. Notons qu'à chaque étape, l'ensemble des sommets actifs forme un ensemble indépendant de sommets et à la fin de l'algorithme, l'ensemble des sommets actifs est maximal pour l'inclusion (s'il on ajoute un autre sommet, l'ensemble ne sera plus indépendant). Dans la suite, on appellera *stratégie gloutonne* cette stratégie lorsque conditionnellement au graphe de départ, les sommets sont examinés dans un ordre aléatoire uniforme.

Constante d'encombrement. Bien sûr, l'ensemble obtenu ne sera pas *maximum*, au sens où il n'a pas la plus grande taille possible. Cependant, pour des graphes peu denses, l'ensemble indépendant que l'on obtient est déjà relativement grand, et il a en général une densité positive lorsque la taille du graphe est grande. Génériquement, lorsqu'on prend une suite de graphes (peu denses) de plus en plus grands, la densité (proportion de sommets) de l'ensemble indépendant obtenu par la stratégie gloutonne converge en probabilité vers une constante que l'on appelle alors la *constante d'encombrement* ou en anglais, *jamming constant*. L'existence d'une constante d'encombrement a été montrée par McDiarmid [140] pour des graphes d'Erdős–Rényi ou par Wormald dans le cas de graphes réguliers [173, 174], ce dernier utilisant la *méthode de l'équation différentielle*. On peut également mentionner [27, 46] qui ont étudié le cas du modèle de configuration. Tous ces résulats sont englobés par un résultat de Krivelevich, Mészáros, Michaeli et Shikelman [125], qui montrent dans un cadre assez général en utilisant la *méthode objective* d'Aldous [14], l'existence d'une constante d'encombrement. Pour énoncer leur résultat, nous avons besoin d'introduire pour un graphe $\mathbf{g}$ fini et x un sommet de $\mathbf{g}$, le nombre $\mathcal{N}(\mathbf{g}, x, r)$ de chemins de longueur au plus r partant de x. Etant donné une suite de graphes finis $(G_n)_{n\geqslant 1}$ (éventuellement aléatoires), on dira que $(G_n)_{n\geqslant 1}$ a une croissance de chemins sous-polynomiale si

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{\mathbb{E}\left[\mathcal{N}(G_n, X_n, r) \wedge M\right]}{r!} \xrightarrow[r\to\infty]{} 0,$$

où $(X_n)_{n\geqslant 1}$ est une suite de points indépendants choisis uniformément au hasard dans G_n. On peut alors résumer leur résultat de la manière suivante.

Théorème 1 ([125])**.** *Soit $(G_n)_{n\geqslant 1}$ une suite de graphes finis. Supposons que $(G_n)_{n\geqslant 1}$ converge au sens de Benjamini–Schramm quenched vers (G_∞, ρ_∞) et est à croissance de chemins sous-polynomiale. Notons I_n la taille (aléatoire) de l'ensemble indépendant obtenu par la stratégie gloutonne sur G_n. Alors,*

$$\frac{I_n}{|G_n|} \xrightarrow[n\to\infty]{(\mathbb{P})} \alpha,$$

où α est une constante qui ne dépend que de (G_∞, ρ_∞).

Ce théorème nous donne donc l'existence d'une constante d'encombrement avec des hypothèses peu restrictives sur la suite de graphes (G_n). Bien souvent, le défi est de déterminer cette constante d'encombrement.

Sur la ligne. Un exemple de graphe très simple auquel on peut penser pour appliquer cet algorithme est la ligne de taille n, c'est-à-dire le graphe avec pour sommets $(1, 2, \cdots, n)$ et arêtes $(i, i+1)$ pour $1 \leqslant i < n$ entre 1. Bien sûr, il est facile d'obtenir un ensemble indépendant de taille maximale sur la ligne : il suffit de prendre un sommet sur deux alternativement, voir Figure 1.8. La densité asymptotique d'un ensemble indépendant de taille maximale est de $1/2$.

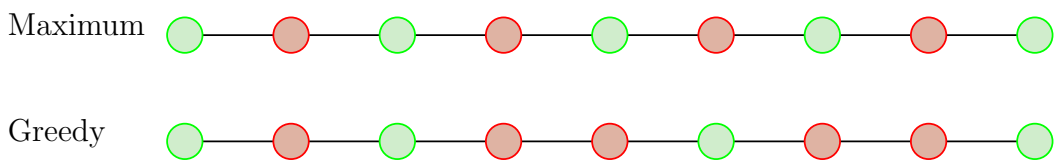


Figure 1.8 : En haut, les sommets verts forment un ensemble indépendant de taille maximale sur une ligne de longueur 9. En bas, les sommets verts forment un ensemble indépendant sur cette même ligne obtenu pour une réalisation de la stratégie gloutonne.

Cependant, on peut se demander quelle est la taille d'un ensemble indépendant obtenu par la stratégie gloutonne. La densité asymptotique d'un tel ensemble converge dans ce cas vers une constante appelée *constante de Rényi discrète*. De manière indépendante, Flory [93] et Page [145] ont montré que cette constante vaut $\frac{1-e^{-2}}{2}$. Page obtient de plus les moments asymptotiques de la taille de l'ensemble indépendant obtenu par la stratégie gloutonne. Notons que dans le cas de la ligne discrète, il est facile de passer d'ensemble indépendant à couplage en échangeant le rôle des sommets et des arêtes.

Il existe une variante continue très similaire à ce problème : on se place sur l'intervalle $[0, x)$ pour $x \in \mathbb{R}_+$ et on essaie de garer des voitures de longueur 1 jusqu'à saturation : à chaque étape, on tire un réel U_k uniformément au hasard et indépendamment du passé entre 0 et $x-1$ et on gare une voiture sur l'intervalle $[U_k, U_k+1)$ si elle n'intersecte aucune voiture déjà garée. On s'arrête lorsqu'il n'y a plus aucun intervalle de longueur 1 disponible pouvant accueillir une nouvelle voiture. On peut alors regarder la densité des voitures dans l'intervalle $[0, x)$. La densité moyenne de voitures quand x tend vers l'infini converge en probabilité vers une limite

$$m := \int_0^\infty \mathrm{d}x \ \exp\left(-2\int_0^x \mathrm{d}y \ \frac{1-\mathrm{e}^{-y}}{y}\right),$$

appelée *constante de Rényi* [155].

Sur la grille. Plus généralement, on peut s'intéresser à la grille de dimension d. Encore une fois, il est facile de voir qu'on peut obtenir un ensemble indépendant de taille maximale en prenant un sommet sur deux alternativement dans chaque direction. En revanche, déterminer la valeur de constante d'encombrement de la grille (qui existe par le Théorème 1), même en dimension 2 reste à ce jour un problème ouvert. Une conjecture de 1960 de Palásti [146] suggère qu'en dimension 2, la constante

d'encombrement serait simplement le carré de la constante de Rényi (en dimension 1), pour le cas discret comme pour la variante continue. Cependant, dans les deux cas, les simulations numériques récentes suggèrent que cette conjecture serait erronée, bien que ce n'ait pas encore été prouvé ou infirmé. On peut se référer à [90, Section 5.3] pour un aperçu des travaux sur ce sujet.

Sur des arbres de Cayley. Quand le graphe que l'on considère est un arbre, il y a également un algorithme simple permettant de déterminer un ensemble indépendant de taille maximale. Nous l'étudierons d'ailleurs dans la sous-section suivante. Néanmoins, il est tout de même pertinent d'étudier la stratégie gloutonne. Intéressons-nous au cas des arbres de Cayley à n sommets, c'est-à dire des arbres non planaires dont les sommets sont étiquetés entre 1 et n.

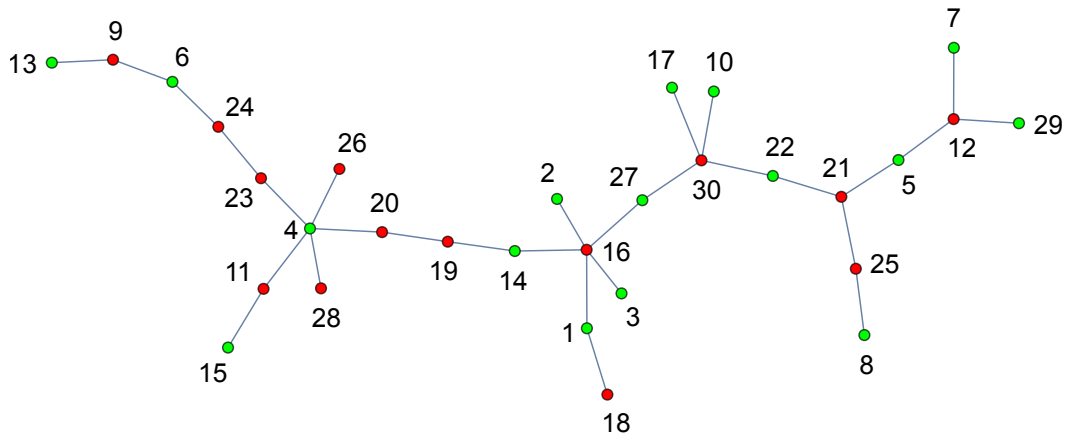


Figure 1.9 : Un arbre de Cayley avec 30 sommets. Les sommets verts forment un ensemble indépendant, obtenu en appliquant la stratégie gloutonne en utilisant les étiquettes des sommets pour déterminer l'ordre dans lequel on ajoute les sommets.

Soit $\mathcal{T}_n$ un arbre de Cayley choisi uniformément au hasard parmi les n^{n-2} arbres de Cayley à n sommets possibles, et notons $\mathcal{T}_n^\bullet$ l'arbre enraciné obtenu à partir de $\mathcal{T}_n$ en distinguant un sommet uniformément au hasard. Meir et Moon [141] montrent que la taille maximale d'un ensemble indépendant de sommets sur $\mathcal{T}_n$ se concentre autour de βn avec $\beta \approx 0.5671$ l'unique solution de $xe^x = 1$. En ce qui concerne la stratégie gloutonne, l'existence d'une constante d'encombrement pour $\mathcal{T}_n$ est donnée par le Théorème 1, et même, la preuve met en évidence le "fait intriguant" que cette constante vaut $1/2$. Dans [65], nous donnons une explication probabiliste et combinatoire à cette constante $1/2$ en donnant la loi de la taille I_n de l'ensemble indépendant obtenu par la stratégie gloutonne sur $\mathcal{T}_n$ et en montrant de I_n a presque la même loi que $n - I_n$ la taille de son complémentaire.

Théorème 1.1 (*[65]*)

- *La taille de l'ensemble indépendant glouton sur $\mathcal{T}_n$ a la même loi que le nombre de sommets à hauteur paire dans $\mathcal{T}_n^\bullet$.*
- *Pour $1 \leqslant k \leqslant n-1$,*
$$\mathbb{P}(I_n = k) = \binom{n}{k} \frac{k^{n-k}(n-k)^{k-1}}{n^{n-1}}.$$
- *Il existe une variable aléatoire $\mathcal{E}_n$ à valeurs dans $\{0,1\}$ telle que*
$$I_n \overset{(d)}{=} (n - I_n) + \mathcal{E}_n.$$
De plus, $\mathbb{P}(\mathcal{E}_n = 1) \to 1/4$ quand n tend vers $+\infty$.

La preuve du premier item de ce théorème repose essentiellement sur le fait qu'un arbre de Cayley uniforme de taille n est uniforme par réenracinement uniforme au hasard, voir par exemple [129, fin de la Section 1.5]. En utilisant cette invariance par réenracinement, on remarque que la taille de l'ensemble indépendant glouton et le nombre de sommets à hauteur paire dans $\mathcal{T}_n^\bullet$ satisfont la même équation récursive qui caractérise leur loi. Le calcul de la loi explicite repose sur des résultats de Féray et Kortchemski [89] sur les arbres de Bienaymé–Galton–Watson bi-types alternants. On établira d'ailleurs un théorème central limit local pour I_n : pour tout $A \geqslant 0$,

$$\mathbb{P}\left(I_n = \left\lfloor \frac{n}{2} + x\sqrt{n} \right\rfloor\right) \underset{n\to\infty}{\sim} \frac{1}{\sqrt{n}} \frac{1}{\sqrt{2\pi(1/4)^2}} \exp\left(-\frac{x^2}{2\cdot(1/4)^2}\right)$$

uniformément pour $x \in [-A, A]$.

Cette égalité en loi suggère d'ailleurs une quasi-symétrie de la loi de I_n autour de $n/2$. Mais, il y a en fait une petite dissymétrie dans la loi du nombre de sommet à hauteur paire (et donc dans la loi de I_n) puisque par exemple, la racine de $\mathcal{T}_n^\bullet$ est toujours à hauteur paire. C'est ce que met d'ailleurs en évidence la troisième partie du théorème. La technique utilisée pour montrer cette partie est très différente, même si on pourrait probablement déduire ce résultat de la loi exacte de I_n. La preuve repose sur une construction markovienne de l'ensemble indépendant glouton ainsi qu'une nouvelle exploration markovienne des arbres de Cayley. L'idée principale est de considérer l'arbre comme initialement inconnu, et de le découvrir au fur et à mesure qu'on applique la stratégie gloutonne.

Couplage glouton sur des arbres de Cayley. Similairement, on peut s'intéresser au couplage obtenu par la stratégie gloutonne. De manière totalement analogue, on peut montrer l'existence d'une constante d'encombrement pour les couplages gloutons en utilisant la limite locale. Grâce à notre exploration markovienne des arbres de Cayley, on peut montrer que cette constante d'encombrement de couplage est égale à $3/8$ pour les arbres de Cayley, et obtenir un théorème central limite local (avec variance $1/96$). Ce résultat a été obtenu par une méthode différente par Dyer, Frieze et Pittel [81].

Couplage glouton sur des arbres plans. D'un point de vue combinatoire, il paraît également naturel d'étudier le couplage obtenu par la stratégie gloutonne pour des arbres invariants en loi par réenracinement sur une arête (orientée) uniforme. C'est le cas des arbres plans (non étiquetés) uniformes à n sommets, qui peuvent aussi être obtenus comme arbres de Bienaymé–Galton–Watson de loi de reproduction $\mathsf{Geom}(1/2)$ conditionnés à avoir n sommets (et enracinés sur l'arête allant du sommet racine vers son premier fils). De manière semblable, on montre que dans le cas d'arbres plans uniformes à n sommets, la constante d'encombrement de couplage vaut $1/3$, et on peut obtenir exactement la loi de la taille du couplage glouton ainsi qu'un théorème central limite local.

Théorème 1.2 (*[65]*)

Soit $\widetilde{M}_n$ la taille d'un couplage obtenu par la stratégie gloutonne sur un arbre plan uniforme à n sommets. Alors la loi de $\widetilde{M}_n$ est donnée par

$$\mathbb{P}(\widetilde{M}_n = k) = \frac{n}{k}\frac{\binom{n+k-2}{2k-1}\binom{n-k-1}{k-1}}{\binom{2n-2}{n-1}} = \frac{n!(n-1)!(n+k-2)!}{k!(2k-1)!(n-2k)!(2n-2)!},$$

pour tout $1 \leqslant k \leqslant \lfloor n/2 \rfloor$. En conséquence, pour tout $A \geqslant 0$,

$$\mathbb{P}\left(\widetilde{M}_n = \left\lfloor \frac{n}{3} + x\sqrt{n} \right\rfloor\right) \underset{n\to+\infty}{\sim} \frac{1}{\sqrt{n}} \frac{1}{\sqrt{2\pi(2/9)^2}} \exp\left(-\frac{x^2}{2\cdot(2/9)^2}\right)$$

uniformément pour $x \in [-A, A]$.

1.2.3 Stratégie optimale : l'algorithme de Karp–Sipser

En fait, pour construire un ensemble indépendant (ou un couplage), il est souvent possible de faire mieux que la stratégie gloutonne en partant d'une observation simple : s'il existe une feuille dans le graphe, alors elle fait partie d'un ensemble indépendant de taille maximale. En effet, si un ensemble indépendant ne contient pas cette feuille, alors on peut l'ajouter à cet ensemble et enlever l'unique voisin de cette feuille (s'il y était). Le nouvel ensemble sera toujours indépendant et sa taille ne peut qu'augmenter. C'est cette stratégie qu'exploite l'algorithme de *Karp–Sipser* [115].

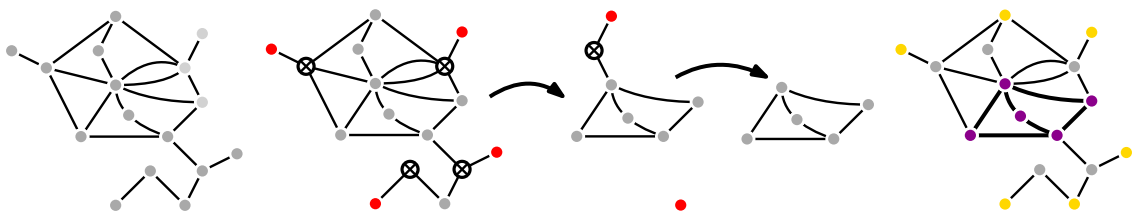


Figure 1.10 : Illustration de la phase 1 de l'algorithme de Karp–Sipser. Sur la figure de droite, les sommets en jaune appartiennent à un des ensemble indépendant de taille maximale. Le sous-graphe induit par les sommets violets forme le cœur de Karp–Sipser du graphe.

Phase 1. Fixons nous un graphe $\mathbf{g} = (V, E)$. La première phase de l'algorithme est la suivante : en reprenant les notations de la section précédente, on démarre avec tous les sommets inexplorés $\mathcal{U}_0 = V$, et aucun sommet actif ou bloqué $\mathcal{A}_0 = \mathcal{B}_0 = \emptyset$. À chaque étape $k \geqslant 1$, on regarde le graphe $\mathbf{g}_{k-1}$ induit par $\mathbf{g}$ sur l'ensemble de sommets inexplorés $\mathcal{U}_{k-1}$ (on garde toutes les arêtes de $\mathbf{g}$ entre les sommets de $\mathcal{U}_{k-1}$). Si $\mathbf{g}_{k-1}$ contient (au moins) une feuille, alors on ajoute une des feuilles a_k de $\mathbf{g}_{k-1}$ aux sommets actifs, et on ajoute son unique voisin b_k aux sommets bloqués pour obtenir $\mathcal{B}_k$. On ajoute également à $\mathcal{A}_{k-1}$ toutes les autres feuilles dont l'unique voisins est aussi b_k pour obtenir $\mathcal{A}_k$. On continue ainsi jusqu'à ce que le graphe $\mathbf{g}_k$ ne contienne plus de feuille et on note $\tau = \inf\{k \geqslant 0 : \mathbf{g}_k \text{ ne contient pas de feuille}\}$ l'instant auquel l'algorithme s'arrête. Le sous-graphe $\mathbf{g}_\tau$ que l'on obtient à la fin de cette phase de l'algorithme est appelé *cœur de Karp–Sipser* de $\mathbf{g}$ et on le notera $\mathsf{KSCore}(\mathbf{g})$. Cette phase est optimale au sens où il existe un ensemble indépendant de sommets de taille maximale qui contient A_τ. Il est également optimal pour obtenir un couplage de taille maximale, puisqu'il en existe un qui contient les arêtes $(\{a_k, b_k\} : 1 \leqslant k \leqslant \tau)$.

Phase 2. Il n'y a pas de stratégie "simple" pour trouver un couplage ou un ensemble indépendant de taille maximale à partir du cœur de Karp–Sipser d'un graphe. Cependant, il existe une deuxième phase de l'algorithme qui permet de donner une bonne approximation d'un couplage de taille maximale, mais elle ne fonctionne pas pour trouver un ensemble indépendant de taille maximale. En effet, du fait des potentielles corrélations à longue portée pour l'appartenance d'un sommet à un ensemble indépendant de taille maximale, il est très difficile d'obtenir même une approximation d'un tel ensemble.

Pour approcher un couplage de taille maximale, la stratégie de la phase 2 de l'algorithme de Karp–Sipser consiste à choisir une arête uniformément au hasard et de supprimer (ou bloquer) les sommets correspondants (et donc ignorer les arêtes qui leur sont incidentes). Si on a créé au moins une feuille, on applique à nouveau la phase 1 de l'algorithme, jusqu'à ce qu'il n'y ait plus de feuille dans la partie inexplorée du graphe. Puis on recommence (à choisir une arête au hasard puis réappliquer la phase 1) jusqu'à ce qu'il n'y ait plus de sommet inexploré.

Transition de phase. Puisque la phase 2 de l'algorithme fait des erreurs importantes pour obtenir un ensemble indépendant de taille maximale, il est important de comprendre la taille du cœur de Karp–Sipser que nous fournit la première phase de l'algorithme. Pour des modèles "classiques" de graphes aléatoires, on peut observer une transition de phase en fonction de la densité des arêtes : tant qu'il y a peu d'arêtes, la taille du cœur de Karp–Sipser sera petite devant la taille du graphe initial, alors que si la densité d'arêtes devient trop grande, la taille du cœur de Karp–Sipser devient linéaire en la taille du graphe initial. Par exemple, dans le cas du graphe d'Erdős–Rényi $G(n, \frac{c}{n})$, les performances de l'algorithme de l'algorithme ont été analysées par Karp et Sipser [115] eux-mêmes puis raffinées par Aronson, Frieze et Pittel [19]. On peut résumer leur résultat ainsi. Lorsque $c < e$, le cœur de Karp–Sipser est vide avec grande probabilité lorsque n tend vers l'infini, et donc la phase 1 de l'algorithme fournit un ensemble indépendant et un couplage de taille maximale. Lorsque $c > e$, la taille du cœur de Karp–Sipser est linéaire en n lorsque n est grand, et la phase 2 de l'algorithme

fournit un couplage, dont la taille diffère au plus de $O(n^{1/5}\log(n)^{12})$ de la taille maximale. Enfin, lorsque $c = e$, la taille du cœur de Karp–Sipser est petite devant n lorsque $n \to \infty$, mais ils ne donnent pas de précisions sur son ordre de grandeur.

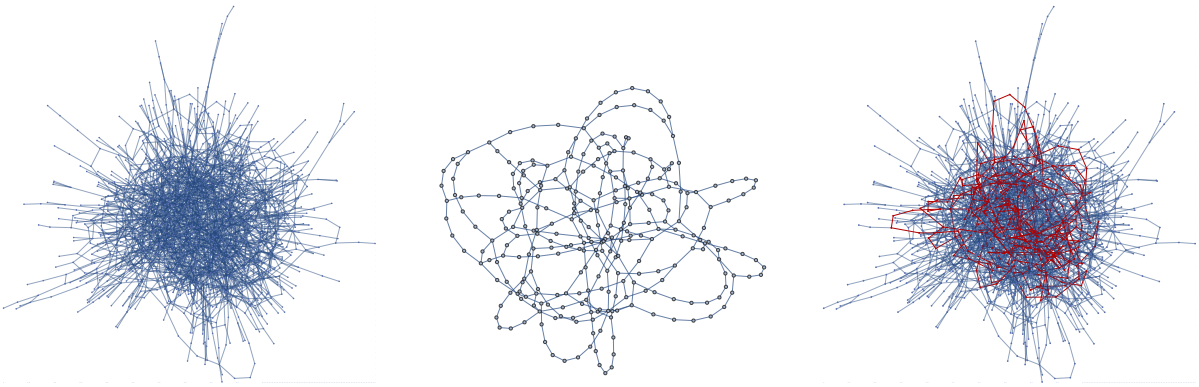


Figure 1.11 : À gauche, la composante géante d'un graphe $G(n, \frac{e}{n})$ pour $n = 2000$ et au milieu, son cœur de Karp–Sipser. À droite, le cœur de Karp–Sipser est dessiné en rouge à l'intérieur du graphe initial.

Ce résultat de transition de phase pour la taille du cœur de Karp–Sipser a ensuite été étendu au cas du modèle de configuration [39, 110]. Dans un travail en collaboration avec Thomas Budzinski et Nicolas Curien [52], nous avons étudié plus précisément la taille et la géométrie du cœur de Karp–Sipser au point critique dans le cas d'un modèle de configuration. Dans un souci de cohérence avec les notations de l'article [52] et du chapitre 3 correspondant, nous noterons jusqu'à la fin de cette section $|\mathbf{g}|$ le nombre de demi-arêtes du graphe $\mathbf{g}$ et non plus son nombre de sommets, et il sera souvent appelé n. Nous espérons que le lecteur nous pardonnera.

On se fixe donc une suite $\mathbf{d}^n = (d_1^n, d_2^n, d_3^n)_{n\geqslant 1}$ telle que

$$n = d_1^n + 2d_2^n + 3d_3^n \text{ est pair.}$$

On imagine que $\mathbf{d}^n$ représente le nombre de sommets de degré 1, 2 et 3 et on s'intéresse $\mathrm{CM}(\mathbf{d}^n)$ un modèle de configuration obtenu en appariant les demi-arêtes uniformément au hasard. On suppose de plus que

$$\frac{d_1^n}{n} \xrightarrow[n\to\infty]{} p_1, \quad \frac{2d_2^n}{n} \xrightarrow[n\to\infty]{} p_2, \quad \text{et} \quad \frac{3d_3^n}{n} \xrightarrow[n\to\infty]{} p_3, \tag{1.1}$$

de sorte à ce que la proportion asymptotique de demi-arêtes incidentes à un sommet de degré i est p_i pour $1 \leqslant i \leqslant 3$. Si (u_n) est une suite de réels positifs et (X_n) une suite de variables aléatoires, on notera $X_n = O_{\mathbb{P}}(u_n)$ si la suite $\left(u_n^{-1}X_n\right)$ est tendue, et on notera $X_n = o_{\mathbb{P}}(u_n)$ si $u_n^{-1}X_n$ converge vers 0 en probabilité.

Nous montrons d'abord un critère explicite pour la transition de phase.

Théorème 1.3 (*[52]*)

Sous l'hypothèse (1.1), *on définit*

$$\Theta = (p_3 - p_1)^2 - 4p_1.$$

- ***Cas sous-critique.*** *Si* $\Theta < 0$, *alors, quand* $n \to \infty$, *on a*

$$|\mathrm{KSCore}(\mathrm{CM}(\mathbf{d}^n))| = O_{\mathbb{P}}(\log^2 n).$$

- ***Cas surcritique.*** *Si* $\Theta > 0$, *alors*

$$n^{-1} \cdot |\mathrm{KSCore}(\mathrm{CM}(\mathbf{d}^n))| \xrightarrow[n\to\infty]{(\mathbb{P})} \frac{4\Theta}{3+\Theta}.$$

- ***Cas critique.*** *Enfin, si* $\Theta = 0$, *alors* $|\mathrm{KSCore}(\mathrm{CM}(\mathbf{d}^n))| = o_{\mathbb{P}}(n)$.

L'idée principale pour montrer ce théorème est de construire le graphe en même temps qu'on explore l'ensemble indépendant. Ensuite la méthode de l'équation différentielle [173] nous permet d'approcher le nombre renormalisé de sommets de degrés 1, 2 et 3 à chaque étape par la solution d'un système d'équations différentielles. L'analyse de ce système déterministe nous permet de prouver le critère ci-dessus.

Pour analyser plus précisément le cas critique, nous avons besoin d'hypothèses plus restrictives sur la suite des degrés, et surtout un contrôle plus précis des fluctuations initiales. On se fixe une suite de degrés telle que

$$d^n_{1,c} = n(1 - \frac{\sqrt{3}}{2}) + O(1), \quad 2d^n_{2,c} = 0, \quad \text{et} \quad 3d^n_{3,c} = n\frac{\sqrt{3}}{2} + O(1), \tag{1.2}$$

et $d^n_1 + 3d^n_3 = n$ est pair pour que l'on puisse apparier les arêtes. En particulier, on a bien $\Theta = (\sqrt{3}-1)^2 - 4(1 - \frac{\sqrt{3}}{2}) = 0$ et on est bien dans le cas critique présenté dans le Théorème 1.3. Nous obtenons alors le résultat suivant.

Théorème 1.4 (*[52]*)

Soit $D_2(n)$ *(resp.* $D_3(n)$*) le nombre de demi-arêtes attachées à des sommets de degré* 2 *(resp.* 3*) dans* $\mathrm{KSCore}(\mathrm{CM}(\mathbf{d}^n_{\mathrm{crit}}))$. *Alors, on a*

$$\begin{pmatrix} n^{-3/5} \cdot D_2(n) \\ n^{-2/5} \cdot D_3(n) \end{pmatrix} \xrightarrow[n\to\infty]{(d)} \cdot \begin{pmatrix} 3^{-3/5}2^{14/5} \cdot \vartheta^{-2} \\ 3^{-2/5}2^{16/5} \cdot \vartheta^{-3} \end{pmatrix},$$

où $\vartheta = \inf\{t \geqslant 0 : B_t = t^{-2}\}$, *pour un mouvement brownien linéaire standard* $(B_t : t \geqslant 0)$ *démarré de* 0. *De plus, conditionnellement à* $(D_2(n), D_3(n))$, *le graphe* $\mathrm{KSCore}(\mathrm{CM}(\mathbf{d}^n_{\mathrm{crit}}))$ *est un modèle de configuration* $\mathrm{CM}((0, D_2(n), D_3(n)))$.

La preuve de ce théorème utilise la même exploration (markovienne) que celle utilisée pour montrer la transition de phase. Cependant, pour étudier le cas critique, nous avons besoin d'avoir un contrôle beaucoup plus fin de la chaîne de Markov juste avant son extinction (quand le nombre de sommets de degré 1 atteint 0). Plus précisément, nous montrons que εn étapes avant la fin, les nombres de sommets (inexplorés) restants de degré 1, 2 et 3 sont d'ordres respectivement $\varepsilon^2 n$, εn et $\varepsilon^{3/2} n$ "en espérance". D'un autre côté, nous montrons que les fluctuations du nombre de sommets de degré 1 sont d'ordre $\varepsilon^{3/4}\sqrt{n}$. Ainsi le nombre de 1 atteint 0 lorsque ses fluctuations sont du même ordre de grandeur que son espérance, c'est-à-dire pour $\varepsilon \approx n^{-2/5}$. Pour montrer que cette heuristique est valide, nous établissons un contrôle précis de la chaine de Markov échelle par échelle.

1.2.4 Perspectives

Plusieurs questions émergent naturellement de cette présentation. D'abord, nous étudions précisément ce qui se passe au point critique pour le cœur de Karp–Sipser, mais on pourrait vouloir étendre notre analyse à l'ensemble de la fenêtre critique. De plus, notre résultat pour la taille du cœur de Karp–Sipser est pour l'instant restreint au cas du modèle de configuration avec des sommets de degrés 1, 2 et 3, mais il semble tout à fait naturel de l'élargir au cas de graphe d'Erdős–Rényi. D'ailleurs, l'idée d'une exploration markovienne pas-à-pas qui est la base de notre preuve a déjà été exploitée pour montrer cette transition de phase et analyser la seconde phase de l'algorithme dans le cas du graphe d'Erdős–Rényi [19]. En se basant sur les résultats de cet article, il nous semble tout à fait possible d'adapter notre analyse fine de la chaîne de Markov et de ses fluctuations dans le cas critique pour montrer que le cœur de Karp–Sipser a également une taille d'ordre $n^{3/5}$ pour le graphe d'Erdős–Rényi. D'ailleurs, nous remarquons que l'exposant qui apparait à la fin de la phase 2 de l'algorithme (dans le cas surcritique) est le même que celui concernant le nombre d'erreurs commises dans la phase 2. Nous pensons que ces deux quantités sont en fait liées.

1.3 Parking dynamique sur des arbres enracinés

Dans cette partie, nous présentons un autre modèle de parking pour lequel les voitures se déplacent sur un arbre enraciné à la recherche d'une place libre.

1.3.1 Règle de parking et motivations

Commençons par rappeler le modèle de parking (dynamique) sur un arbre. On se donne $\mathbf{t}$ un arbre *enraciné*, et une configuration d'arrivées de voitures, c'est-à-dire un étiquetage des sommets $(a_u : u \in \mathbf{t})$ par des entiers positifs ou nuls. Cet étiquetage représente le nombre de voitures arrivant sur chaque sommet, et chaque sommet de $\mathbf{t}$ peut accueillir (au plus) une voiture garée. La règle pour garer les voitures est la suivante : lorsqu'une voiture arrive sur un sommet libre, elle s'y gare. Sinon, elle se déplace en direction de la racine et occupe la première place disponible. Si elle ne trouve aucune

place disponible le long de son trajet, alors, elle sort de l'arbre par la racine et contribue au *flux* de voitures sortantes, voire figure 1.12.

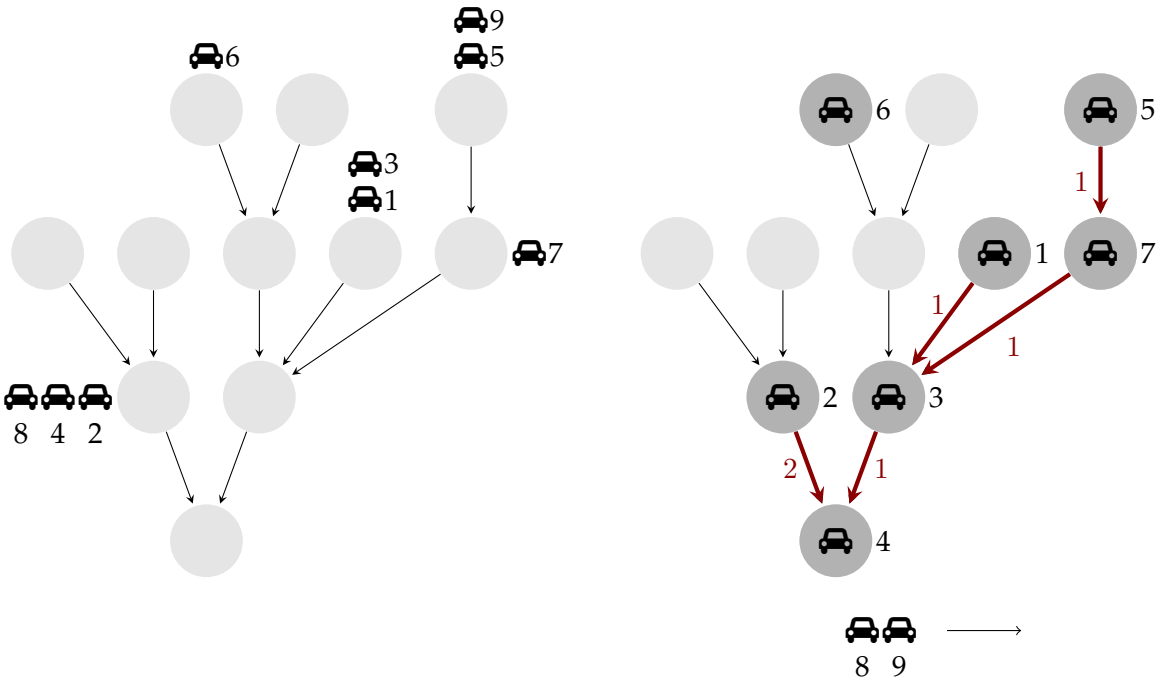


Figure 1.12 : À gauche, un arbre enraciné avec 11 sommets où les arêtes sont dirigés vers le sommet racine, ainsi que 9 voitures arrivant sur ses sommets. À droite, la configuration finale une fois que les 9 voitures sont garées, ainsi que le flux de voitures traversant chaque arête. Remarquons que deux voitures n'ont pas trouvé de place disponible et sortent de l'arbre.

Une propriété importante de ce modèle est sa propriété *abélienne* : quelque soit l'ordre dans lequel les voitures se garent, la configuration finale de places occupées et le flux de voitures sortantes seront les mêmes. On peut d'ailleurs retrouver la configuration initiale d'arrivées de voitures à partir de la configuration finale de places occupées, du flux de voitures traversant chaque arrête et du flux sortant.

Ce modèle a d'abord été introduit dans le cas où l'arbre $\mathbf{t}$ est simplement une ligne (orientée vers la racine, disons à gauche) par Konheim et Weiss [122] dans les années 60 pour des motivations informatiques. Depuis, de nombreux travaux ont exploré ce modèle, montrant notamment son lien avec le coalescent additif [57, 35, 29]. L'étude de ce modèle sur des arbres est bien plus récente, initiée par Lackner et Panholzer en 2016 [128]. Ils étudient le processus de parking dans le cas où l'arbre sous-jacent est aléatoire, et nous décrirons plus précisément leur modèle dans la section suivante.

Une des motivations de l'introduction du modèle de parking sur des arbres vient de considérations hydrologiques. On peut notamment essayer de modéliser de l'eau qui ruisselle dans une pente. Une première simplification est de découper notre pente en petites cellules, chaque cellule pouvant absorber une "unité" d'eau. Imaginons qu'il pleuve et qu'une quantité aléatoire d'eau arrive sur chaque cellule,

de manière indépendante et uniforme sur les cellules. Autrement dit, la quantité d'eau qui arrive est i.i.d. sur chaque cellule. Si la quantité d'eau qui arrive est inférieure à une "unité", elle s'infiltre. Sinon l'eau va ruisseler vers le bas. Selon notre choix de découpage en cellules de la pente, on peut envisager différentes dynamiques pour le ruissellement. Par exemple, sur le réseau en diamant (voir figure 1.13) il est raisonnable d'imaginer que de chaque cellule, l'eau ruisselle sur la case soit en bas à gauche soit en bas à droite avec même probabilité $(1/2)$ indépendamment sur chaque cellule. Sur un réseau carré, on peut imaginer que l'eau ruisselle soit sur la case juste en dessous (avec proba $1-2p$), ou va en bas à gauche ou à droite, chacun avec proba p, et conditionne la configuration finale au fait qu'il n'y ait pas de croisement. Vu d'un point en bas de la pente, l'ensemble des points depuis lesquels l'eau peut ruisseler jusqu'à ce point, forme un arbre. Jones [111] appelle l'arbre obtenu un *arbre de drainage*. Malheureusement, l'étude du modèle de parking sur ce type d'arbre est difficile du fait de sa géométrie, notamment des fortes dépendances entre les différentes branches. Même s'il existe des résultats sur les arbres de drainages, notamment leurs lien avec le chateau brownien [55, 161], le modèle de parking sur ces arbres ne fait pas (encore !) partie des modèles étudiés dans la littérature.

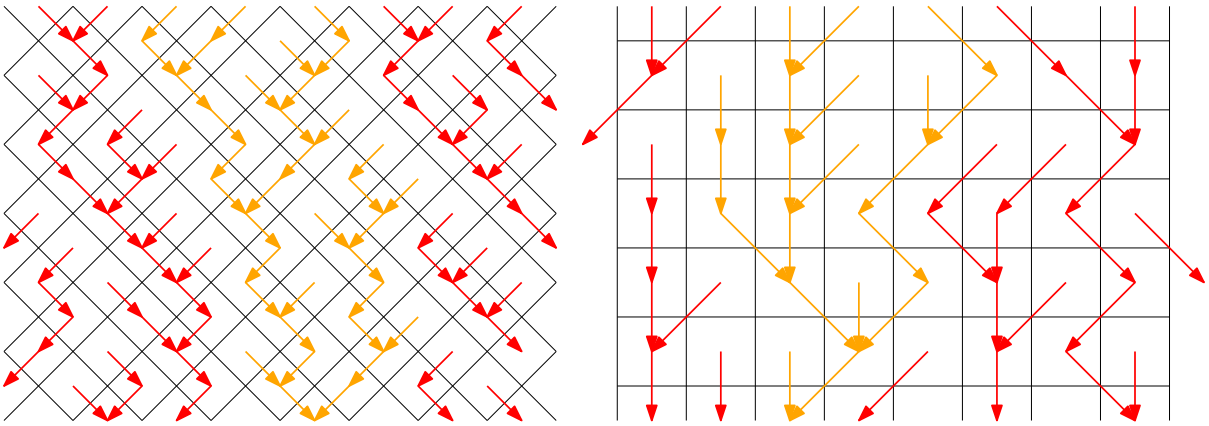


Figure 1.13 : À gauche, un exemple de ruissellement sur un réseau en diamant et à droite, un exemple sur le réseau carré. En rouge et orange, est indiqué sur chaque cellule la direction dans laquelle ruisselle l'eau si elle ne peut pas être absorbée. Vu d'une cellule du bas, les cellules pouvant ruisseler jusqu'à ce point forment naturellement un arbre, par exemple l'arbre orange.

Dans la suite, on s'intéressera à plusieurs caractéristiques observables du modèle, par exemple le flux sortant de voitures ou les "composantes" de voitures garées. On peut aussi considérer la probabilité d'obtenir une configuration où toutes les voitures sont garées à la fin (le flux sortant de voitures est nul), ou encore la probabilité que la racine de l'arbre soit une place libre.

1.3.2 Transition de phase (universelle via la limite locale)

Les premiers résultats sur ce modèle mettent en évidence une transition de phase pour le flux sortant de voitures. En effet, on peut imaginer que si très peu de voitures arrivent sur notre arbre, alors

elles vont quasiment toutes trouver une place disponible (et ce, proche de leur sommet d'arrivées). En revanche, lorsque la densité des arrivées de voitures est trop importante, alors une proportion positive d'entre elles ne va pas arriver à se garer et contribuera au flux sortant. Nous commençons par rendre rigoureuse cette heuristique dans un cadre assez général.

Transition de phase via la limite locale. Soit $(\mathcal{T}_n)_{n\geqslant 1}$ une suite d'arbres enracinés avec n sommets. Pour tout entier $k \geqslant 0$, on considère une famille de lois de probabilité $(\mu_{(k),\alpha} : \alpha \in \mathbb{R}_+)$ stochastiquement croissante en α. Pour tout $\alpha \in \mathbb{R}_+$ fixé, on considère le modèle α de parking sur $\mathcal{T}_n$, c'est-à-dire que conditionnellement à $\mathcal{T}_n$, on se donne des arrivées de voitures aléatoires $(A_x : x \in \mathcal{T}_n)$ indépendantes sur chaque sommet et de loi $\mu_{(k),\alpha}$ sur chaque sommet de degré $k \geqslant 0$, et on applique la procédure de parking décrite ci-dessus. On note respectivement $m_{(k),\alpha}$ et $\sigma^2_{(k),\alpha}$ l'espérance et la variance de $\mu_{(k),\alpha}$. Enfin on note $\varphi_\alpha(\mathcal{T}_n)$ le flux de voitures sortantes pour tout $\alpha \in \mathbb{R}_+$, et on appelle *composantes de voitures garées* les composantes connexes de la forêt obtenue en ne gardant que les arêtes de l'arbre initial dont les extrémités contiennent toutes les deux une voiture garée dans la configuration finale.

Théorème 1.5 (*Chapitre 4*)

On suppose que $\mathcal{T}_n$ converge au sens de Benjamini–Schramm quenched vers un arbre infini enraciné $(\mathcal{T}_\infty, \rho)$ qui n'a qu'une seule épine dorsale presque sûrement. Pour s'assurer une continuité des arrivées de voitures $(\mu_{(k),\alpha} : k \geqslant 0, \alpha \in \mathbb{R}_+)$ en α, on suppose que

$$\sum_{k\geqslant 0} \nu_k m_{(k),\alpha} = \alpha,$$

où $\nu_k = \mathbb{P}(\deg(\rho) = k \text{ dans } \mathcal{T}_\infty)$, et on suppose que pour tout $\alpha \in \mathbb{R}_+$, il existe une constante K_α telle que $m_{(k),\alpha} < K_\alpha$ pour tout $k \geqslant 0$. Soit

$$\alpha_c := \inf\{\alpha \in \mathbb{R}_+ : \ \mathbb{P}(\text{il existe une composante infinie de voitures garées dans } \mathcal{T}_\infty) > 0\}.$$

Alors, on a

$$\frac{\varphi_\alpha(\mathcal{T}_n)}{n} \xrightarrow[n\to\infty]{(\mathbb{P})} \begin{cases} 0 & si \quad \alpha < \alpha_c, \\ C_\alpha & si \quad \alpha > \alpha_c, \end{cases}$$

où $C_\alpha > 0$ est une constante strictement positive si $\alpha > \alpha_c$.

Notons qu'ici, même si l'arbre $\mathcal{T}_\infty$ est infini, il n'y a pas d'ambiguité pour définir la dynamique de parking sur $\mathcal{T}_\infty$ puisque pour chaque sommet de $\mathcal{T}_\infty$, le sous-arbre au dessus de ce sommet est fini, et le statut final du sommet ne dépend que de ce sous-arbre. Nous verrons comment contourner la difficulté qui apparaît si ça n'était pas le cas dans le chapitre 7.

La preuve de ce théorème en chapitre 4 consiste essentiellement à utiliser le caractère unimodulaire et ergodique de la limite au sens de Benjamini–Schramm. Nous nous sommes volontairement restreint au cas des arbres dont la limite au sens de Benjamini–Schramm n'a qu'une seule épine dorsale.

Comme le Théorème 1 de la section précédente, le théorème ci-dessus nous montre un résultat universel pour une large classe de graphes en utilisant leur limite locale, en l'occurrence l'existence d'une transition de phase pour le modèle de parking sur de nombreux arbres. Mais le challenge en général est de déterminer les constantes qui entrent en jeu dans le modèle, ici la valeur de α_c, qui nous permet de localiser précisément la transition.

Cas particuliers. D'un point de vue chronologique, cette transition de phase a d'abord été montrée pour des choix particuliers de suite d'arbres $(\mathcal{T}_n)$. Lackner et Panholzer sont les premiers à s'intéresser au modèle de parking sur des arbres. Ils étudient d'un point de vue combinatoire le modèle de parking sur un arbre de Cayley de taille n avec m arrivées ordonnées de voitures sur les sommets. Rappelons qu'un arbre de Cayley de taille n est un arbre non-plan avec n sommets étiquetés de 1 à n. Quitte à rendre le nombre total d'arrivées de voitures aléatoire, leur modèle rentre en fait dans le cadre de notre Théorème 1.5. Le point de vue adopté par Lackner et Panholzer est très différent : ils énumèrent et obtiennent une formule explicite pour le nombre de configurations initiales (arbre et arrivées de voitures) pour lesquelles toutes les voitures sont garées dans la configuration finale [128, Théorème 4.4 et 4.5]. Évidemment, il faut pour cela que le nombre de voitures m soit inférieur ou égal au nombre de places disponibles, c'est-à-dire de sommets dans l'arbre n. En se rappelant qu'il y a n^{n-1} arbres de Cayley enracinés de taille n différents et n^m choix possibles pour les arrivées de m voitures, on peut calculer la probabilité $p_{n,m}$ d'obtenir une configuration finale où toutes les voitures sont garées, pour une configuration intiale choisie uniformément au hasard. Le régime pour lequel on observe une transition de phase est lorsque le nombre de voitures par sommet est d'ordre constant quand n tend vers l'infini, c'est-à-dire lorsque $m = \lfloor \alpha n \rfloor$ pour un $\alpha \in [0,1]$ fixé. Dans ce régime, Lackner et Panholzer remarquent [128, Corollaire 4.7] qu'on a les équivalents asymptotiques suivants lorsque n tend vers l'infini :

$$p_{n,\lfloor \alpha n \rfloor} \underset{n\to\infty}{\sim} \begin{cases} c_\alpha & \text{si} \quad \alpha < 1/2, \\ c_{1/2} n^{-1/6} & \text{si} \quad \alpha = 1/2, \\ \frac{c_\alpha}{n} d_\alpha^{-n} & \text{si} \quad \alpha > 1/2, \end{cases}$$

où $c_\alpha > 0$ et $d_\alpha > 0$ sont des constantes explicites qui ne dépendent que de α. Leur preuve utilise une décomposition combinatoire en fonction de la dernière voiture, idée que nous exploiterons à nouveau dans la section 1.4.2 et le chapitre 8. Ceci est le premier indice de l'apparition d'une transition de phase à $\alpha = 1/2$. Ce modèle a ensuite été étudié d'un point de vue probabiliste par Goldschmidt et Przykucki [100]. En effet, il est très proche du modèle de parking où l'arbre sous-jacent est un arbre de Bienaymé–Galton–Watson conditionné à avoir taille n avec loi de reproduction $\mathsf{Poisson}(1)$, et où les arrivées de voitures sont i.i.d. sur les sommets et ont pour loi $\mathsf{Poisson}(\alpha)$. Un argument de poissonisation et dépoissonisation permet de passer du modèle discret (nombre fixé de voitures) au modèle aléatoire (arrivées de voitures aléatoires) et réciproquement. Goldschmidt et Przykucki donnent une explication probabiliste à cette transition de phase et retrouvent le résultat de Lackner et Panholzer en utilisant la méthode objective [14], c'est-à-dire un argument de limite locale. Ils s'intéressent de plus au flux de voitures sortantes et montrent que lorsque $\alpha \leqslant 1/2$, l'espérance du

flux de voitures est bornée uniformément en n (par 1) alors que cette espérance tend vers l'infini quand $n \to \infty$ pour $\alpha > 1/2$.

Deux autres cas particuliers ont été étudiés dans la littérature. Chen et Goldschmidt [60] ont regardé un modèle où l'arbre sous-jacent est un arbre plan uniforme de taille n, ou arbre de Bienaymé–Galton–Watson conditionné à avoir taille n avec loi de reproduction géométrique de paramètre $1/2$, et avec toujours des arrivées de voitures de loi de Poisson de paramètre α sur les sommets. Ils montrent une transition de phase semblable à $\alpha_c = \sqrt{2} - 1$. Enfin, motivé par la modélisation de réseaux hydrologiques, Owen Dafydd Jones étudie le cas d'arbre de Bienaymé–Galton–Watson avec loi de reproduction $\nu = \beta(\delta_0 + \delta_2) + (1 - 2\beta)\delta_1$ pour $\beta \in (0, 1/2)$, et des arrivées de voitures de loi $\mu = (1 - \alpha)\delta_0 + \alpha\delta_2$. Il donne la localisation précise α_c où a lieu la transition (en fonction de β). On peut aussi mentionner Panholzer qui s'est intéressé à de nombreux modèles combinatoires [148].

Universalité via la limite locale des GW. Les premiers à obtenir la localisation de la transition de phase du modèle de parking pour une large classe d'arbres sont Curien et Hénard en 2019 [75]. Notamment, ils donnent un critère explicite pour déterminer la phase (sous-critique, critique ou surcritique) du modèle de parking lorsque l'arbre sous-jacent est un arbre de Bienaymé–Galton–Watson critique (l'espérance de la loi de reproduction est 1) et pour des arrivées de voitures i.i.d. sur les sommets (conditionnellement à l'arbre). Le plus frappant dans leur résultat est que la localisation de la transition de phase dépend de manière relativement simple des paramètres du modèle. Plaçons nous dans un cadre encore un peu plus général que dans [75], qui est celui étudié dans [64]. D'abord, étant donné un arbre enraciné $\mathbf{t}$, on suppose que les arrivées de voitures sur chaque sommet de $\mathbf{t}$ sont des variables aléatoires indépendantes et que leur loi ne dépend que du nombre d'enfants du sommet dans $\mathbf{t}$. Pour $k \geqslant 0$, on note la loi d'arrivées des voitures sur un sommet ayant k enfants

$$\mu_{(k)} \text{ de moyenne } m_{(k)} \geqslant 0 \text{ et de variance finie } \sigma^2_{(k)}.$$

Pour ces arrivées de voitures, on notera $\varphi(\mathbf{t})$ le flux de voitures sortant de l'arbre $\mathbf{t}$. En ce qui concerne le choix de l'arbre, nous optons pour différentes versions d'un arbre de Bienaymé–Galton–Watson critique de loi de reproduction

$$\nu \neq \delta_1 \text{ de moyenne 1 et de variance finie } \Sigma^2,$$

la version "classique" (non conditionnée) $\mathcal{T}$, la version conditionnée à avoir n sommets $\mathcal{T}_n$[5]. On suppose qu'il existe $k \geqslant 1$ tel que $\nu_k > 0$ et pour lequel la loi d'arrivées de voitures $\mu_{(k)}$ est différente de δ_1. Pour caractériser la localisation de la transition de phase, nous avons besoin d'introduire la loi de reproduction biaisée par la taille $\overline{\nu} = \sum_{k \geqslant 1} k\nu_k \delta_k$ et les quantités

$$\mathbb{E}_{\overline{\nu}}[m] := \sum_{k=0}^{\infty} k\nu_k\, m_{(k)}, \quad \mathbb{E}_{\nu}[m] := \sum_{k=0}^{\infty} \nu_k\, m_{(k)} \quad \text{et} \quad \mathbb{E}_{\nu}[\sigma^2 + m^2 - m] := \sum_{k=0}^{\infty} \nu_k \left(\sigma^2_{(k)} + m^2_{(k)} - m_{(k)}\right).$$

Notons en particulier que si les lois d'arrivées de voitures ne dépendent pas du nombre d'enfants du sommet et ont pour moyenne commune m et variance σ^2, alors

[5] On se restreint aux valeurs de n pour lesquelles $\mathbb{P}(|\mathcal{T}| = n) > 0$.

$$\mathbb{E}_{\overline{\nu}}[m] = \mathbb{E}_{\nu}[m] = m \quad \text{et} \quad \mathbb{E}_{\nu}[\sigma^2 + m^2 - m] = \sigma^2 + m^2 - m.$$

Nous avons établi le théorème suivant.

Théorème 1.6 (*[64]*)

On suppose que $\mathbb{E}_{\overline{\nu}}[m] \leqslant 1$. *Le processus de parking sur un arbre de Bienaymé–Galton–Watson subit une transition de phase qui ne dépend que du signe de la quantité*

$$\Theta := (1 - \mathbb{E}_{\overline{\nu}}[m])^2 - \Sigma^2 \mathbb{E}_{\nu}[\sigma^2 + m^2 - m]. \tag{1.3}$$

Plus précisément, nous avons trois régimes différents présentant les caractéristiques suivantes

	sous-critique $\Theta > 0$	*critique* $\Theta = 0$	*surcritique* $\Theta < 0$
$\varphi(\mathcal{T}_n)$ *quand* $n \to \infty$	*converge en loi*	$\xrightarrow[n\to\infty]{(\mathbb{P})} \infty$ *mais est* $o(n)$	$\sim cn$ *avec* $c > 0$
$\Sigma^2 \mathbb{E}[\varphi(\mathcal{T})] + \mathbb{E}_{\overline{\nu}}[m] - 1$	$-\sqrt{\Theta}$	0	∞
$\mathbb{P}$ (*⌀ contient une voiture dans* $\mathcal{T}$)	$\mathbb{E}_{\nu}[m]$	$\mathbb{E}_{\nu}[m]$	$\mathbb{E}_{\nu}[m] - c$

La preuve de ce résultat repose essentiellement sur l'utilisation de la limite locale de $\mathcal{T}_n$ quand n tend vers l'infini, qui est l'arbre de Kesten $\mathcal{T}_\infty$ présenté plus haut. Grace à l'utilisation d'une formule "tous-pour-un", ainsi que l'introduction de temps aléatoires pour les arrivées des voitures, on peut écrire une équation différentielle ordinaire pour l'espérance du flux sortant de voitures au temps t. Cette équation se résolvant facilement, on peut à partir de cette équation déterminer si l'espérance du flux de voitures sortantes est finie ou infinie, une fois toutes les voitures arrivées et garées.

1.3.3 Transition continue et abrupte

Mieux qu'une simple transition, nous montrons que pour le modèle très général précédent, la transition de phase est non seulement continue, mais également abrupte, au sens où les caractéristiques observables du modèle, comme le flux de voitures sortantes ou la taille des composantes de voitures garées, décroissent exponentiellement lorsque l'on s'écarte du point critique. Pour ce faire, nous avons besoin d'une hypothèse supplémentaire indiquant que la loi du nombre d'arrivées de voitures sur un sommet "typique" de l'arbre a une queue exponentielle, c'est-à-dire que la série

$$F(z) = \sum_{k \geqslant 0} \nu_k \sum_{i \geqslant 0} \mu_{(k)}(i) z^i \tag{$H_{\exp}$}$$

a un rayon de convergence strictement plus grand que 1. Dans ce régime, nous montrons le théorème suivant.

Théorème 1.7 (*[64]*)

On suppose ($H_{\exp}$). *Soit* $\varepsilon > 0$. *Dans le régime surcritique, c'est-à-dire quand* $\Theta < 0$, *il existe* $\delta > 0$, *et* $n_0 \geqslant 0$ *tel que pour tout* $n \geqslant n_0$,

$$\mathbb{P}\left(|\varphi(\mathcal{T}_n) - cn| \geqslant \varepsilon n\right) \leqslant \mathrm{e}^{-\delta n},$$

où $c > 0$ *est donné par le Théorème 1.6. Dans le régime sous-critique, c'est-à-dire quand* $\Theta > 0$, *il existe* $\delta > 0$, *et* $n_0 \geqslant 0$ *tel que pour tout* $n \geqslant n_0$,

$$\mathbb{P}\left(|\varphi(\mathcal{T})| \geqslant \varepsilon n\right) \leqslant \mathrm{e}^{-\delta n}.$$

La preuve de ce théorème utilise des techniques très différentes dans le cas surcritique et le cas sous-critique. Le cas surcritique utilise une borne exponentielle pour la distribution des arbres "pendants" ou franges, que nous établissons. Le cas sous-critique utilise un point de vue bien plus analytique. Nous utilisons la définition récursive des arbres de Bienaymé–Galton–Watson (décomposition à la racine) pour écrire une équation récursive sur $z \mapsto W(z)$ la fonction génératrice du flux de voitures sortant de l'arbre. Bien que l'équation que nous obtenons ait une singularité en $z = 1$, nous résolvons cette singularité grâce au développement de Newton–Puiseux de la série pour montrer que son rayon de convergence est strictement plus grand que 1.

Une application importante de ce théorème est qu'il permet d'estimer la taille des composantes de voitures garées dans les phases surcritiques et sous-critiques. Rappelons qu'on appelle composantes de voitures garées les composantes connexes de la forêt obtenue en ne gardant que les arêtes reliant deux sommets adjacents dans l'arbre et contenant tous les deux une voiture garée dans la configuration finale. Notons $|\mathsf{C}_{\max}(n)|$ la taille de la plus grande composante de voitures garées dans $\mathcal{T}_n$, et $|\mathsf{C}_2(n)|$ la taille de la deuxième plus grande. Le résultat suivant nous donne des informations sur la taille asymptotique de ces deux composantes.

Corollaire 1.1 (*[64]*)

$$\textit{(surcritique } \Theta < 0) \qquad \frac{|\mathsf{C}_{\max}(n)|}{n} \xrightarrow[n\to\infty]{(\mathbb{P})} C \quad \textit{et} \quad \mathbb{P}(|\mathsf{C}_2(n)| \geqslant A\ln(n)) \underset{n\to\infty}{\to} 0,$$

$$\textit{(sous-critique } \Theta > 0) \qquad \mathbb{P}(|\mathsf{C}_{\max}(n)| \geqslant A\ln(n)) \underset{n\to\infty}{\to} 0,$$

où $C \in (0,1)$ *et* $A > 0$ *sont des constantes dépendant des lois* ν *et* $\mu_{(k)}$ *pour* $k \geqslant 0$.

On peut remarquer que ces tailles de composantes sont très similaires à ce que l'on observe pour le graphe d'Erdős–Rényi. Ce n'est d'ailleurs pas une simple coïncidence, ce que nous mettons en évidence dans la section suivante.

1.3.4 Étude de la fenêtre critique

Nous avons montré dans la section précédente l'existence d'une transition de phase pour le modèle de parking sur des arbres "critiques". Nous avons aussi vu que les régimes sous-critiques et surcritiques sont plutôt bien compris, au contraire du régime critique pour lequel nous pouvons seulement dire que le flux de voitures sortantes tend vers l'infini mais est petit devant la taille de l'arbre n quand $n \to \infty$. On peut donc se demander plus précisément quel est son ordre de grandeur, et quelle est la taille des composantes de voitures garées, leur géométrie...

C'est l'objet de notre travail [67] écrit en collaboration avec Nicolas Curien. Bien que nous pensons que les résultats qui vont être présentés dans cette section valent pour un ensemble de modèles très généraux, nous restreignons le cadre d'étude au même modèle que celui étudié par Lackner et Panholzer dans [128] : arbre de Cayley $\mathcal{T}_n$ de taille n choisi uniformément au hasard et m voitures arrivant uniformément au hasard et indépendamment sur les sommets de $\mathcal{T}_n$. Nous décrivons l'évolution de la taille des composantes dans la fenêtre critique, et introduisons pour cela, pour tout $\lambda \in \mathbb{R}$,

$$m_n(\lambda) := \left\lfloor \frac{n}{2} + \frac{\lambda}{2} n^{2/3} \right\rfloor \wedge 0.$$

Nous introduisons également une notion de composantes de voitures garées un peu différente de celle utilisée précédemment. On appelle composantes *proches* de voitures garées les composantes connexes de la forêt obtenue en gardant les arêtes entre les places de parking occupées et leur parent dans l'arbre sous-jacent.

Pour tout $\lambda \in \mathbb{R}$, notons $C_n^*(\lambda)$ la taille de la composante proche de voitures garées de la racine (éventuellement vide), $(C_{n,i}(\lambda) : i \geqslant 1)$ les tailles des autres composantes ordonnées par ordre décroissant, et $D_n(\lambda)$ le flux de voitures sortantes dans $\mathcal{T}_n$ après l'arrivée de $m_n(\lambda)$ voitures sur $\mathcal{T}_n$.

Théorème 1.8 (*[67]*)

Pour la topologie de Skorokhod sur $\mathbb{D}\mathrm{adlag}(\mathbb{R}, \ell^2 \times \mathbb{R}_+ \times \mathbb{R}_+)$, *nous avons la convergence en loi suivante*

$$\begin{pmatrix} n^{-2/3} \cdot C_{n,i}(\lambda), & i \geqslant 1 \\ n^{-2/3} \cdot C_{n,*}(\lambda) & \\ n^{-1/3} \cdot D_n(\lambda) & \end{pmatrix}_{\lambda \in \mathbb{R}} \xrightarrow[n\to\infty]{(d)} \begin{pmatrix} \mathscr{C}_i(\lambda), & i \geqslant 1 \\ \mathscr{C}_*(\lambda) & \\ \mathscr{D}(\lambda) & \end{pmatrix}_{\lambda \in \mathbb{R}.}$$

où les processus $\mathscr{C}_i, \mathscr{C}_*$ *et* $\mathscr{D}$ *sont construits explicitement.*

La preuve de ce théorème repose sur un couplage discret entre le processus de parking sur $\mathcal{T}_n$ où l'on ajoute les voitures une par une uniformément au hasard, et la version gelée du processus de graphe d'Erdős–Rényi $(F(n,m) : m \geqslant 0)$ défini plus haut. En particulier, nous explicitons une correspondance entre les composantes gelées de $F(n,m)$ et la composante (proche) de voitures garées la racine dans $\mathcal{T}_n$ avec m arrivées de voitures, et entre les composantes non gelées de $F(n,m)$ et les autres composantes proches dans $\mathcal{T}_n$. Notamment, les arêtes du graphe d'Erdős–Rényi "classique" $G(n,m)$ que l'on supprime pour obtenir sa version gelée $F(n,m)$ correspondent aux voitures qui ne

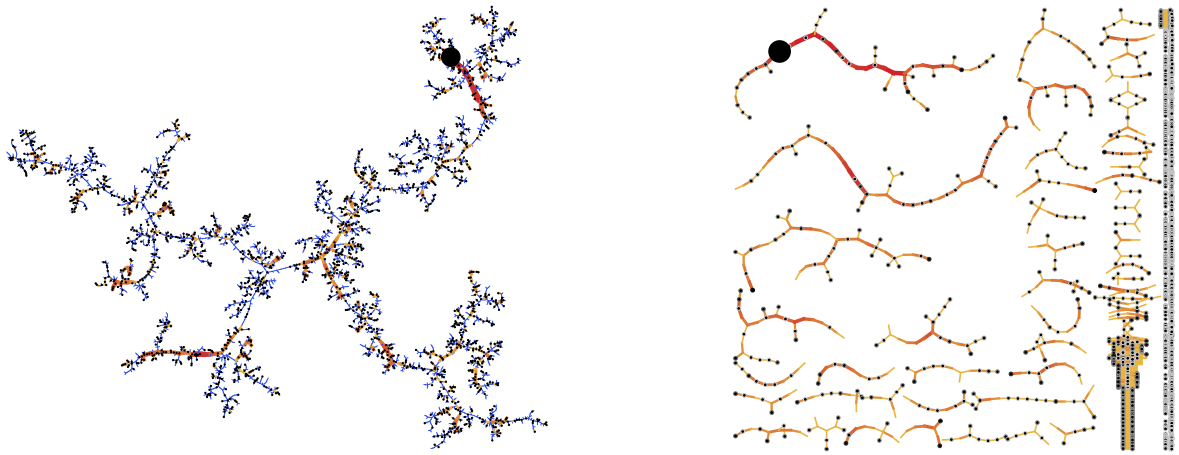


Figure 1.14 : À gauche, une simulation d'un arbre de parking critique $\mathcal{T}_{5000}$ avec 2500 voitures. Les couleurs et épaisseurs des arêtes indiquent le flux de voitures les traversant. La racine de l'arbre est représentée par un cercle noir. À droite, la décomposition du même arbre en composantes de voitures garées.

trouvent pas de place disponible. Nous faisons ensuite une analyse détaillée de la limite des tailles de composantes dans $(F(n,m) : m \geqslant 0)$, en s'appuyant sur ses similarités avec sa version classique $(G(n,m) : m \geqslant 0)$. En effet, pour la version classique, la limite des tailles des composantes a été très étudiée dans la littérature et s'appelle dans la fenêtre critique, le *coalescent multiplicatif*. Nous verrons dans le chapitre 6 comment définir précisément la limite pour la version gelée, que l'on appellera le *coalescent multiplicatif gelé*.

En fait, notre couplage donne bien plus d'informations que simplement la taille des composantes. On donne par exemple la probabilité que toutes les voitures arrivent à se garer.

Proposition 1.1 (*[67]*)

Pour $n \geqslant 1$ et $m \geqslant 0$, nous avons

$$\mathbb{P}(m \text{ voitures i.i.d. uniformes se garent toutes sur } \mathcal{T}_n)\left(1-\frac{m}{n}\right) = \mathbb{P}(G(n,m) \text{ est acyclique}).$$

Notre couplage nous permet également de décrire la géométrie des composantes proches de voitures garées. Il est aisé de voir que ces composantes sont des *"nearly parked trees"*, c'est-à-dire des arbres de Cayley avec $N \geqslant 1$ sommets et $N-1$ arrivées de voitures numérotées, conditionnés à ce que toutes les voitures se garent et que la racine reste vide dans la configuration finale. On définit donc $\mathcal{P}_N$ un nearly parked tree uniforme à N sommets pour $N \geqslant 1$ et on note ρ sa racine.

Proposition 1.2 (*[67]*)

L'espérance de la hauteur d'un point typique dans nearly parked tree à N sommets est

$$\frac{1}{N}\mathbb{E}\left[\sum_{x\in P_N} \mathrm{d}(\rho,x)\right] = \sum_{h=1}^{N-1}\binom{N}{h+1}\left(\frac{h+1}{N}\right)^2 N^{1-h}\left(\frac{1}{2}\right)_h \underset{N\to\infty}{\sim} \frac{\Gamma(3/4)}{2^{1/4}\sqrt{\pi}}\cdot N^{3/4},$$

où d *désigne la distance de graphe (dans* P_N*) et* $(x)_a = x(x+1)\cdots(x+a-1)$ *est le symbole de Pochhammer.*

1.3.5 Et sur d'autres arbres ?

Tout ce que nous avons décrit jusque là concerne le modèle de parking sur des arbres de Bienaymé–Galton–Watson critique, quand les sommets ont en moyenne 1 enfant chacun. Mais on peut se demander ce qui se passe sur d'autres arbres.

Arbres sous-critiques. D'abord, intéressons-nous aux arbres de Bienaymé–Galton–Watson *sous-critiques*. Premièrement, si on regarde l'arbre non-conditionné $\mathcal{T}$, alors l'espérance de sa taille (son nombre de sommets) est finie donc le nombre de voitures arrivant sur l'arbre est également d'espérance finie. L'espérance du flux seront donc aussi toujours finie, et ceci est d'ailleurs valable même quand il y a plus d'une voiture par sommets en moyenne. Si on s'intéresse à la limite locale d'un arbre de Bienaymé–Galton–Watson $\mathcal{T}_n$ avec une loi de reproduction ν sous-critique conditionné à avoir taille n, alors deux phénomènes différents [2, 1, 105, 112] se produisent en fonction de ν :

- Si ν est "générique", alors l'arbre $\mathcal{T}_n$ a une limite locale qui est un arbre de Kesten pour une certaine loi $\widetilde{\nu}$ critique. Dans ce cas, la transition de phase pour $\mathcal{T}_n$ est la même que pour un arbre de Bienaymé–Galton–Watson avec loi de reproduction $\widetilde{\nu}$. Comme $\widetilde{\nu}$ dépend un peu plus subtilement de ν, la transition de phase ne dépend pas uniquement des deux premiers moments de ν, mais elle ne dépend bien que des deux premiers moments de la loi des arrivées de voitures.
- Si ν est "non-générique", alors la "limite locale" $\mathcal{T}_n$ contient presque sûrement un sommet x de degré infini à hauteur finie. Ainsi, une proportion positive des enfants de ce sommet recevra plus de 2 voitures arrivées (pour une loi d'arrivées de voitures non triviale). Le flux de voitures sortant de x est donc infini, et celui de la racine aussi.

Arbres surcritiques. Le problème dans le cas d'arbres surcritiques (plus d'un enfant par sommet en moyenne) est un peu plus difficile. L'existence d'une transition de phase a été montrée [22] par Bahl, Barnet et Junge. Cependant, il n'existe à notre connaissance que des bornes numériques sur la localisation de la transition dans des cas particuliers, hormis notre travail [12] que nous présentons ci-dessous.

Arbre binaire infini. Dans un travail avec David Aldous, Nicolas Curien et Olivier Hénard, nous nous sommes intéressés au cas de l'arbre binaire infini et avons déterminé la localisation de la transition. Plus précisément, notre modèle est le suivant. On se place sur l'arbre binaire infini $\mathbb{B} = \cup_{n\geqslant 0}\{0,1\}^n$ avec $\{0,1\}^0 = \emptyset$ la racine de l'arbre, et des arêtes entre u et $u0$, et entre u et $u1$ pour tout mot $u \in \mathbb{B}$. On fait arriver des voitures $(A_u : u \in \mathbb{B})$ sur chaque sommet de l'arbre $\mathbb{B}$. On suppose que les arrivées de voitures $(A_u : u \in \mathbb{B})$ sont i.i.d. avec une loi fixée $\mu = (\mu_k : k \geqslant 0)$ dont le support est $\{0,1,2,\dots\}$.

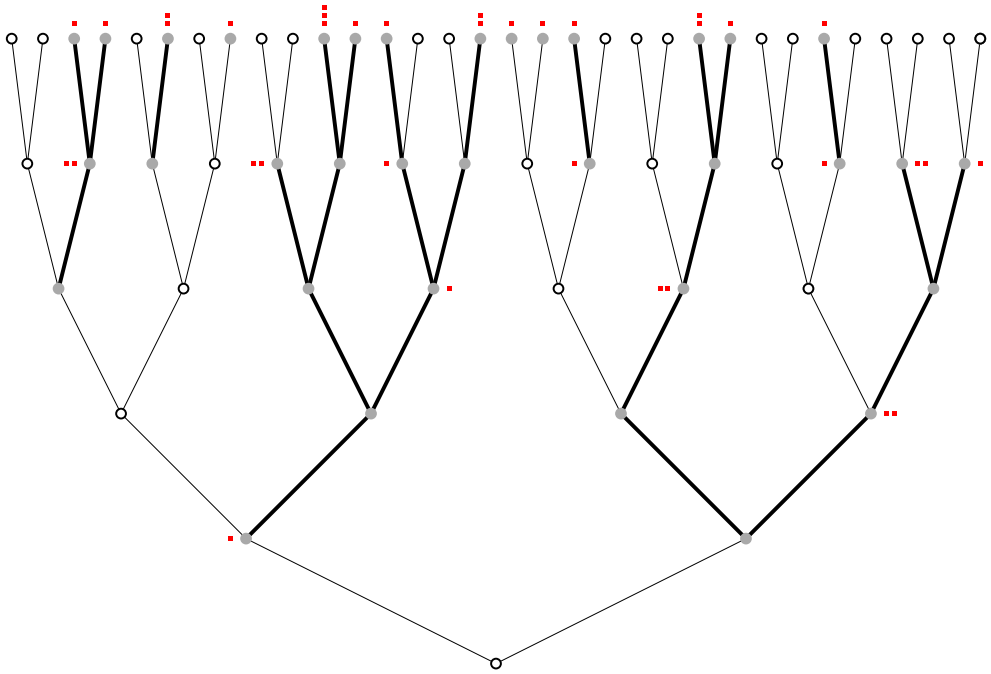


Figure 1.15 : Illustration du processus de parking sur les 5 premiers niveaux de l'arbre binaire infini. Les voitures, incluant celles venant des niveaux supérieurs, sont représentées par les carrés rouges. Les sommets occupés dans la configuration finale sont représentés en gris tandis que les sommets libres sont en blanc. Les composantes connexes de voitures garées sont représentées avec des arêtes plus épaisses.

Pour éviter les cas triviaux, on suppose que $\mu_0 + \mu_1 < 1$, sinon les voitures se garent toujours sur leur noeud d'arrivée et le flux de voitures est toujours nul. Notons

$$G(x) = \sum_{k \geqslant 0} \mu_k x^k$$

la fonction génératrice de la loi μ. On introduit également la variable aléatoire

$$X := \text{nombre de voitures qui passent par la racine,}$$

de sorte à ce que le flux de voitures sortantes est $(X-1)_+ = \max(X-1, 0)$. Dans cette section, nous revenons à une notion de composantes connexes plus classiques, c'est-à-dire que comme dans la section 1.3.3, nous appelons composantes de voitures garées les composantes connexes de la forêt obtenue en ne gardant les arêtes entre deux sommets reliés dans l'arbre que s'ils contiennent tous les deux une voiture dans la configuration finale. Nous nous intéressons également aux composantes de places vides, c'est-à-dire les composantes connexes de la forêt obtenue en ne gardant les arêtes entre deux sommets reliés dans l'arbre que si aucun des deux ne contient une voiture dans la configuration finale. Alors nous montrons la dichotomie suivante en fonction du choix de la loi μ :

- soit le nombre X de voitures qui visitent la racine $\emptyset$ a une espérance finie, et toutes les composantes de voitures garées sont finies presque sûrement. On appelle cela la phase *sous-critique*.

- soit $X = \infty$ presque sûrement et en fait, tous les sommets de $\mathbb{B}$ sont occupés dans la configuration finale, et on appelle cette phase le régime *surcritique*.

Nous distinguons de plus le régime *critique*, quand il n'est pas possible d'augmenter stochastiquement μ en restant dans le régime sous-critique. Comme dans le cas des arbres de Bienaymé–Galton–Watson critiques, le processus de parking est dans le régime surcritique dès que $\mathbb{E}[A_\emptyset] > 1$, puisqu'il y a plus de voitures que de places de parking en moyenne. Bien que le modèle semble a priori plus simple que le modèle précédent avec des arbres de Bienaymé–Galton–Watson critiques, la transition de phase dépend de manière beaucoup plus subtile de la fonction génératrice G de la loi d'arrivée de voitures μ, comme le montre le théorème suivant.

Théorème 1.9 (*[12]*)

On suppose qu'il existe $t_c \in (0, \infty)$ tel que

$$t_c = \min\{t \geqslant 0 : \ 2(G(t) - tG'(t))^2 = t^2 G(t) G''(t)\}. \tag{$\star$}$$

Alors, le processus de parking est sous-critique si et seulement si

$$(t_c - 2)G(t_c) \geqslant t_c(t_c - 1)G'(t_c). \tag{1.4}$$

La condition ($\star$) sur l'existence d'un t_c est en fait légère et peu restrictive. Elle est par exemple vérifiée dès lors que les fonctions génératrices ont un rayon de convergence infini. Nous donnerons d'ailleurs dans le chapitre 7 un moyen de déterminer si le processus de parking est surcritique ou sous-critique si cette condition ($\star$) n'est pas vérifiée.

Nous obtenons des informations supplémentaires dans le régime critique : supposons donc que ($\star$) est vérifiée et que μ est critique. Nous verrons que cela correspond en fait au cas où (7.1) est une égalité. Rappelons que X désigne le nombre de voitures qui passent par la racine et on introduit $p_k = \mathbb{P}(X = k)$, et $p_\circ = p_0$ et $p_\bullet = p_1$ pour les probabilités d'avoir la racine vide, ou occupée mais avec un flux sortant nul. Dans le régime critique, nous montrons le résultat suivant.

Théorème 1.10 (*[12]*)

Supposons que la condition ($\star$) est vérifiée et que (7.1) est une égalité. Alors, la racine est soit vide, soit appartient à une composante finie de voitures garées presque sûrement, et nous obtenons

$$p_\circ = \frac{t_c^2}{4(t_c - 1)G(t_c)} \quad \text{and} \quad p_\bullet = \sqrt{\frac{p_\circ}{\mu_0}} - p_\circ.$$

Mentionnons deux conséquences de ce théorème.

- $\mathbb{E}[X] < \infty$, même dans le régime critique.
- Un porisme de ce théorème est que les composantes de places vides sont en fait des arbres de Bienaymé–Galton–Watson dont la loi de reproduction ξ est donnée par

$$\mathbb{P}(\xi = 0) = \frac{p_\bullet^2}{(p_\circ + p_\bullet)^2}, \quad \mathbb{P}(\xi = 1) = \frac{2p_\bullet p_\circ}{(p_\circ + p_\bullet)^2}, \quad \mathbb{P}(\xi = 2) = \frac{p_\circ^2}{(p_\circ + p_\bullet)^2}.$$

Nous montrons en autres que dans tout le régime sous-critique (et critique), on a $p_\circ > \frac{1}{2}$, et ces arbres sont donc surcritiques. Ainsi, même dans le régime critique pour le modèle de parking, il y a des composantes infinies de places disponibles (et même un nombre infini). Au contraire, les composantes de voitures garées sont toujours finies, même dans le régime critique.

La preuve de ces deux résultats utilise essentiellement la même décomposition combinatoire en composantes de voitures garées, et leur énumération. Pour les compter, nous nous appuyons sur une autre décomposition combinatoire, cette fois-ci directement à la racine, "à la Tutte".

1.3.6 Perspectives

Arbres surcritiques généraux. Une extension naturelle du modèle, qui n'est pas contenue dans cette thèse, serait de localiser la transition de phase pour des arrivées de voitures i.i.d. sur un arbre de Bienaymé–Galton–Watson surcritique général. Comme mentionné plus haut, l'existence d'une transition de phase a déjà été montrée [22] mais il n'y a à notre connaissance pas de caractérisation de sa localisation dans le cas général. La méthode que nous développons dans le cas de l'arbre binaire infini nous semble en fait assez universelle. Le point crucial est de parvenir à énumérer les composantes (pondérées) de voitures garées ou *fully parked trees*. Ceci a d'ailleurs déjà été fait par Linxiao Chen [59] dans le cas d'arbres plans, pour des arbres de Bienaymé–Galton–Watson surcritique avec loi de reproduction géométrique de paramètre p pour $p \in (1/2, 1)$. Nous espérons obtenir une caractérisation de la localisation de la transition de phase pour des arbres surcritiques généraux grâce aux techniques utilisées dans [12].

Universalité de la phase critique ? Nous pensons que cette transition de phase revêt un caractère universel, c'est-à-dire que de nombreuses caractéristiques ont le même comportement dans les différentes phases pour un grand nombre de choix d'arrivées de voitures et d'arbres sous-jacents. Par exemple, nous pensons que pour un arbre de Bienaymé–Galton–Watson critiques où la loi de reproduction a une variance finie, et pour des arrivées de voitures dont les lois sont à queues légères, la taille des composantes de voitures garées est aussi d'ordre $n^{2/3}$, comme pour les arbre de Cayley avec arrivées de voitures uniformes. Ce résultat devrait découler d'un couplage similaire à celui présenté dans le chapitre 6 entre le modèle de parking et la version gelée du processus d'Erdős–Rényi, mais il faut cette fois-ci définir une version gelée d'un modèle de configuration [66]. Cependant, différentes classes d'universalité devraient apparaître lorsque les lois d'arrivées de voitures sont à queues lourdes. Ces différentes classes d'universalité sont d'ailleurs déjà présentes dans [59] et [12] : une phase générique, une phase dense et une phase diluée, selon la queue de distribution de la loi d'arrivées de voitures.

Quand $\alpha_c = 0$... Lorsque $\alpha_c = 0$, dès qu'on met des arrivées de voiture qui sont fixes avec n la taille de l'arbre, le modèle de parking sera toujours dans la phase surcritique. Mais on peut se demander quel est l'ordre de grandeur précis du nombre de voitures qu'on peut faire arriver sur l'arbre pour observer une transition entre phase sous-critique et phase surcritique. C'est le cas notamment des

arbres qui ont aussi de grands degrés à la limite, par exemple les arbres de Bienaymé–Galton–Watson critiques dont la loi de reproduction a une variance infinie (stable, à queue lourde), ou encore des arbres récursifs (uniforme ou à attachement préférentiel). Plus précisément, dans ce cas, dès que la loi des arrivées de voitures (i.i.d. sur les sommets) est fixe avec la taille de l'arbre, alors le modèle de parking sera nécessairement dans la phase surcritique. Par exemple, si on regarde un arbre récursif uniforme à n sommets, alors le degré de la racine est d'ordre $\log(n)$. Il faut donc que l'espérance du nombre de voitures par sommets soit au plus d'ordre $1/\log(n)$ pour que le modèle de parking puisse être sous-critique. Pour toutes ces familles d'arbres, il nous paraît donc pertinent de déterminer l'ordre de grandeur du nombre de voitures par sommet (pour les arrivées) qui nous permet d'observer une transition de phase pour le flux sortant.

Un modèle un peu différent. Enfin, on peut mentionner une extension naturelle du modèle de parking. Revenons au cas le plus simple possible d'arbre : la ligne (de taille n). Konheim et Weiss montrent que pour des arrivées de $m = \lfloor \alpha n \rfloor$ voitures uniformes sur la ligne avec $\alpha \in [0,1]$, la transition de phase pour le flux sortant a lieu pour $\alpha_c = 1$. Mais que se passerait-t-il si les voitures (garées ou sorties) avaient le droit de quitter leur place, de choisir un nouveau sommet d'arrivée et d'essayer de se garer ? Si on met n voitures, le flux de voitures sortant serait certainement toujours d'ordre $\sqrt{n}$. Plus précisément, on peut se demander quelle est la loi de répartition de n voitures sur une ligne de taille n dans la configuration finale qui est invariante par l'opération de choisir une voiture uniformément au hasard (garée ou sortie), l'enlever de sa place, lui choisir un nouveau sommet d'arrivée et essayer de la garer.

1.4 Parking et cartes planaires

Un dernier aspect de cette thèse est de montrer les nombreuses similarités entre le modèle de parking et le modèle des cartes planaires.

1.4.1 Définition de cartes et arbres de peeling

Commençons par une introduction au modèle des cartes planaires et aux propriétés qui nous seront utiles.

Cartes. D'abord, rappelons qu'une carte est un graphe (non orienté) fini et connexe (non nécessairement simple), plongé sur une surface et vu à homéomorphisme préservant l'orientation près. On ne s'intéressera ici qu'aux cartes *planaires*, c'est-à-dire que la surface sur laquelle les graphes sont plongés est la sphère de dimension 2, mais plus généralement on parlera de cartes *de genre* $g \geqslant 0$ lorsque la surface sous-jacente est un tore de genre g. Pour éviter les problèmes de symétrie, les cartes que nous considérons sont *enracinées* sur une arête orientée distinguée, voir Figure 1.16. Le nombre d'arêtes incidentes à une face est appelé *degré de la face* (ou parfois son *périmètre* ou sa *longueur*). Nous allons nous intéresser en particulier à deux types de cartes : les *quadrangulations*, où toutes les

faces ont degrés 4, et les triangulations, où toutes les faces ont degré 3. Parfois, nous considérerons des cartes *avec une frontière*, c'est-à-dire avec une face distinguée qui peut avoir un degré différent (mais ce degré devra être pair dans le cas des quadrangulations). Dans ce cas, nous imposerons que l'arête racine se trouve sur cette frontière, avec la face distinguée à la droite de l'arête racine (en allant dans le sens de l'orientation de la flèche). On appellera cette face la *face externe*. Notons que pour une carte planaire, on a la formule d'Euler

$$\#\text{sommets} + \#\text{faces} - \#\text{arêtes} = 2.$$

Pour les quadrangulations ou triangulations, on peut également également écrire une relation entre le nombre de façes et le nombre d'arêtes, si bien qu'il n'y a plus qu'un degré de liberté pour la taille d'une quadrangulation ou triangulation planaire (sans bord).

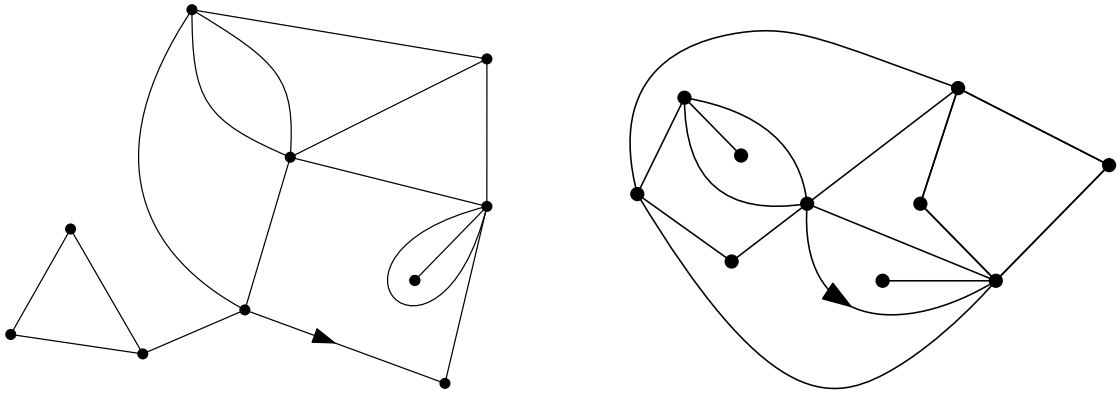


Figure 1.16 : À gauche, une carte planaire enracinée avec 10 sommets. À droite, une quadrangulation planaire à 10 sommets, 8 faces et 16 arêtes.

Algorithme et arbre de peeling. Une manière possible de décrire et d'énumérer les triangulations ou quadrangulations est de le faire arêtes par arêtes, récursivement, inspiré par une décomposition de Tutte [168]. La difficulté vient du fait que, lorsqu'on enlève une arête à une quadrangulation, elle ne reste pas une quadrangulation car l'une des faces a un degré différent. On se retrouve alors avec une quadrangulation avec une frontière (et il faut décider d'une règle pour ré-enraciner la nouvelle quadrangulation). Plus précisément, lorsque qu'on enlève l'arête racine à une quadrangulation avec une frontière de degré $2p$ pour $p \geqslant 1$, alors l'un des deux évènements suivant se produit, voir Figure 1.17 :

- soit la quadrangulation reste connexe. Cela signifie qu'on a découvert une nouvelle face de la quadrangulation, qui est donc un quadrangle. On se retrouve donc avec une quadrangulation avec une frontière de degré $2(p+1)$ que l'on réenracine sur l'arête "la plus à gauche" de la face que l'on vient de découvrir.

- soit la quadrangulation n'est plus connexe et on se retrouve avec deux quadrangulations de bord de degré $2p_1$ et $2p_2$ avec $p_1 + p_2 + 1 = p$, que l'on réenracine en utilisant les deux extrémités de l'arêtes que l'on vient d'enlever.

Notons qu'on donne ici un choix de réenracinement, ou *algorithme de peeling* possible mais on pourrait s'en donner un autre, tant qu'il ne dépend que du degré du bord et des étapes précédentes, et non de la quandrangulation tout entière. Dans le cas où $p = 0$, la carte est juste une *carte-sommet*, composée d'un seul sommet sans arête.

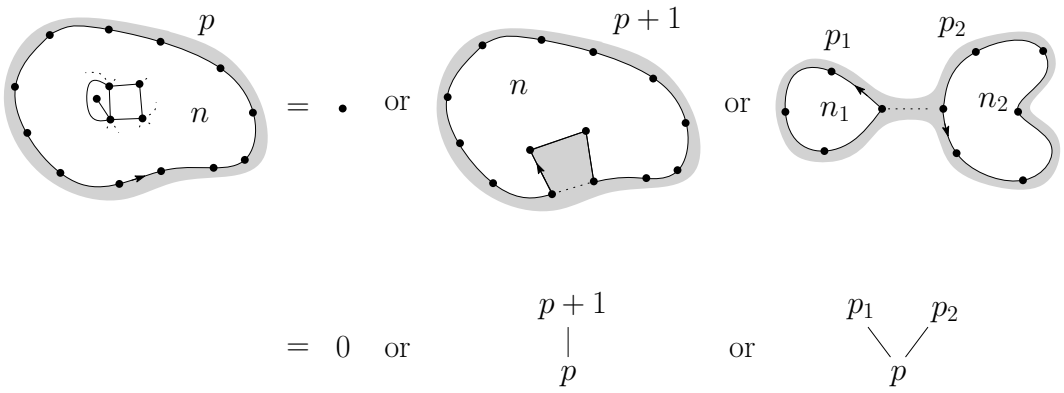


Figure 1.17 : Illustration de la décomposition récursive de Tutte dans le cas des quadrangulations, et en dessous, la correspondance dans l'arbre de peeling.

Il est possible d'itérer cette exploration ou *épluchage*, arête par arête, jusqu'à ne se retrouver qu'avec des cartes-sommet. Cette technique d'épluchage a d'abord été introduite par le physicien Watabiki [172], puis reprise par Omer Angel sous une forme un peu différente et pour des triangulations dans [15], et sous une forme très semblable par Timothy Budd [51]. On peut encoder cette exploration dans un arbre appelé *arbre de peeling* en se souvenant à chaque instant de la longueur du bord de la carte explorée, voir figure 1.18.

D'autres moyens d'énumérer les cartes. Depuis, de nombreuses méthodes ont permis l'énumération de cartes. On peut mentionner par exemple les intégrales de matrices [45, 164], les bijections avec d'autres arbres "à la Schaeffer" [69, 159, 160] ou encore les correspondances avec la hiérarchie KP [56, 134]. Nous ne détaillerons pas ces méthodes ici, puisque notre travail utilise essentiellement une nouvelle décomposition des arbres de peeling.

1.4.2 Décomposition selon la dernière voiture

Nous présentons dans cette section notre travail [63], qui introduit de nouvelles équations pour l'énumération des quandrangulations et triangulations planaires, sans bord ou avec un ou deux bords.

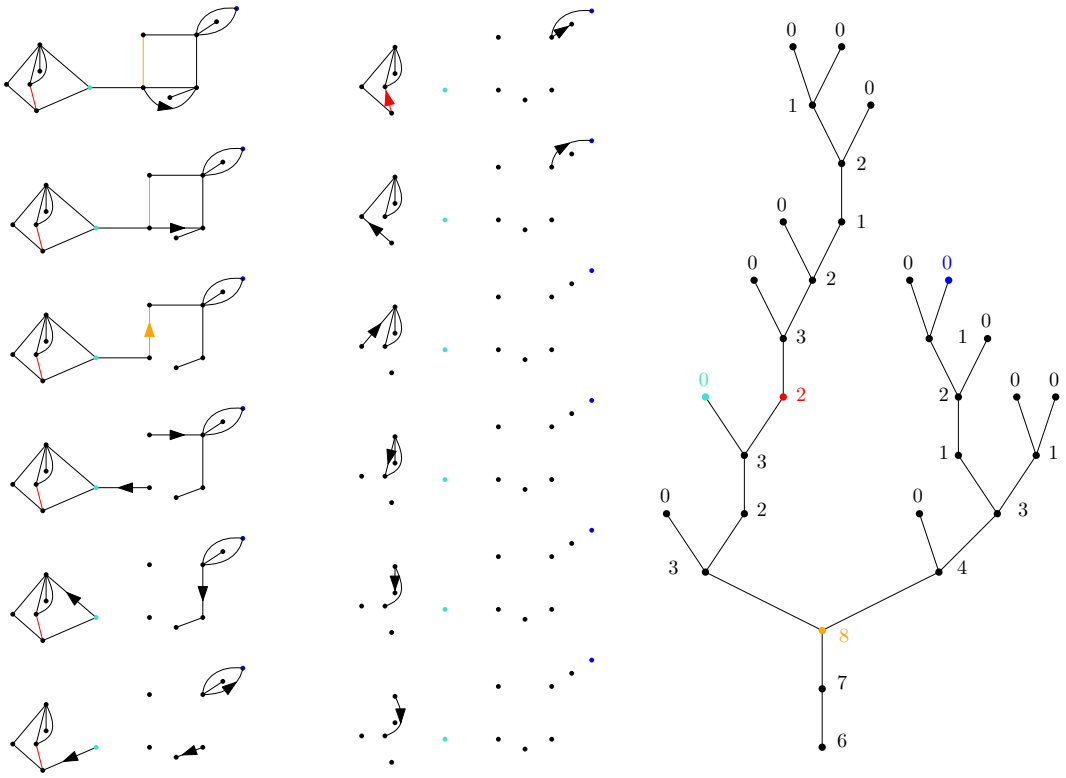


Figure 1.18 : Exemple pas-à-pas du processus d'exploration d'une quadrangulation planaire avec $n = 12$ sommets et une frontière de degrée $2p$ où $p = 6$, et son arbre de peeling correspondant. En bleu, deux sommets de la quadrangulation et leurs feuilles étiquetées 0 correspondantes dans l'arbre de peeling. En rouge et orange, deux arêtes de la quadrangulation et leurs sommets internes correspondants dans l'arbre.

Nous utilisons une nouvelle méthode qui s'inspire d'une décomposition issue du modèle de parking : une décomposition par rapport à la dernière voiture introduite par Lackner et Panholzer [128].

Décomposition des arbres fortement garés. Commençons par décrire cette décomposition pour un certain type d'arbres de parking, les arbres *fortement garés* qui sont des arbres de Cayley enracinés sur l'un des n sommets et avec $n + p$ arrivées de voitures étiquetées sur les sommets, tel que dans la configuration finale, toutes les places sont occupées (donc p voitures contribuent aux flux de voitures sortantes) et il y a un flux strictement positif de voitures traversant chaque arête de l'arbre, voir figure 1.19.

Prenons un arbre fortement garé et imaginons qu'on enlève de cet arbre la dernière voiture arrivée. Si le flux de voitures sortantes est positif, cette voiture était nécessairement une voiture sortante (sinon on aurait eu une arête de l'arbre avec un flux de voitures nul en dessous de la place où aurait été garée cette voiture). Si le flux de voitures est nul, alors pour la même raison, cette dernière voiture voiture s'était garée sur la racine.

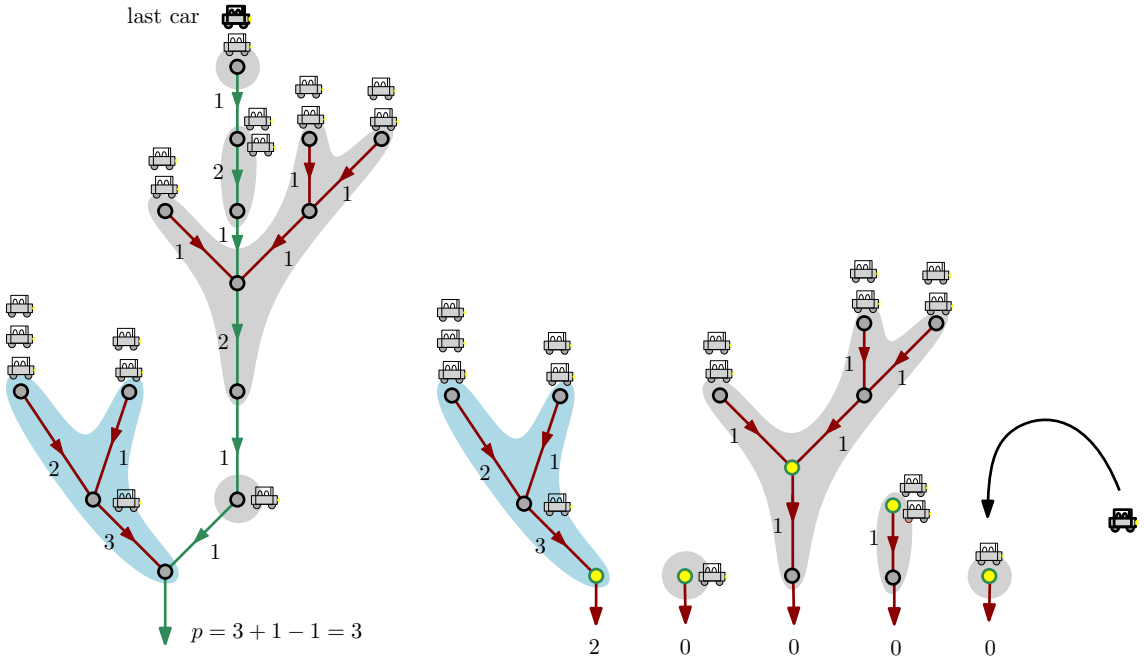


Figure 1.19 : Illustration de la décomposition en dernière voiture d'un arbre fortement garé (à gauche). En retirant les arêtes dont le flux devient nul quand on enlève la dernière voiture, on se retrouve avec une suite d'arbres fortement garés avec un point distingué, dont le premier peut avoir un flux positif de voitures sortantes.

En enlevant cette voiture, on peut créer des arêtes où le flux de voitures est nul. On se retrouve donc avec une suite d'arbres fortement garés, chacun avec un sommet distingué pour se souvenir de la façon dont il faudra recoller les arbres pour retrouver l'arbre initial. Plus précisément, introduisons la fonction génératrice bivariée des arbres fortement garés

$$\mathbf{S}(x,y) = \sum_{n\geqslant 1, p\geqslant 0} \frac{\mathrm{SP}(n,p)}{n!(n+p)!} x^n y^p$$

où $\mathrm{SP}(n,p)$[6] désigne le nombre d'arbres fortement garés avec n sommets et $n+p$ voitures. Alors, la décomposition en dernière voiture donne l'équation suivante sur $\mathbf{S}$

$$y\partial_y \mathbf{S}(x,y) + x\partial_x \mathbf{S}(x,y) - \mathbf{S}^\bullet(x,0) = \frac{xy\partial_x \mathbf{S}(x,y)}{1-\mathbf{S}^\bullet(x,0)},$$

où $\mathbf{S}^\bullet(x,y) = x\partial_x \mathbf{S}(x,y)$ est la fonction génératrice des arbres fortement garés avec un sommet distingué. On pourrait se dire que résoudre cette équation nécessite de connaitre a priori $\mathbf{S}(x,0)$, mais en fait elle détermine entièrement $\mathbf{S}(x,y)$. Si on applique la décomposition par rapport à la dernière voiture directement aux arbres fortement garés sans voitures sortantes ($p=0$), on trouve d'ailleurs une équation très simple

$$\mathbf{S}^\bullet(x,0) = \frac{x}{1-\mathbf{S}^\bullet(x,0)},$$

[6] pour *strongly parked trees* en anglais.

qu'il est aisé de résoudre pour trouver $\mathrm{SP}(n,n) = (2n-2)!$. Ce résultat avait déjà été montré par King et Yan [119].

Décomposition "dernière voiture" des arbres de peeling. L'idée de notre travail [63] est d'appliquer cette décomposition aux arbres de peeling des cartes pour obtenir de nouvelles équations. Notre décomposition permet par exemple d'énumérer les quadrangulations. On note Q_n le nombre de quadrangulations enracinées (sans bord) avec n sommets, et on introduit $\mathfrak{Q}$ la série génératrice correspondante

$$\mathfrak{Q}(x) := \sum_{n\geqslant 2} Q_n x^n = x^2 + 2x^3 + 9x^4 + 54x^5 + 378x^6 + \cdots .$$

Notons que par convention, la carte avec deux sommets reliés par une arête est considérée comme une quadrangulation, ce qui explique le terme x^2.

Théorème 1.11 (*[63]*)

En notant $\mathfrak{Q}^\bullet = x\mathfrak{Q}'(x)$, la "décomposition en dernière voiture" des arbres de peeling des quadrangulations donne l'équation suivante

$$\mathfrak{Q}^\bullet = 2x^2 + 6x\left(\frac{\mathfrak{Q}^\bullet - \mathfrak{Q}}{1-\mathfrak{Q}^\bullet/x}\right).$$

Cette équation caractérise $\mathfrak{Q}$ et est équivalente à l'équation récursive suivante : $Q_2 = 1$ et pour $n \geqslant 3$,

$$nQ_n = \sum_{k=2}^{n-1} k(n+1-k)Q_k Q_{n+1-k} + (4n-10)Q_{n-1}.$$

La forme de l'équation qu'on obtient est très proche d'une équation récursive similaire obtenue en montrant des correspondances avec une équation issue de la physique, appelée *équation KP* (pour Kadomtsev–Petviashvili), comme montré par exemple dans [134, Corollary 2] ou [56]. Cependant, ces deux équations n'ont pas l'air équivalentes, l'équation venant de la hiérarchie KP s'écrivant avec nos notations

$$\mathfrak{Q}^\bullet - \mathfrak{Q} = 4x(2\mathfrak{Q}^\bullet - 3\mathfrak{Q}) + 3(2\mathfrak{Q}^\bullet - 3\mathfrak{Q})^2 + x^2.$$

On peut également déduire de notre décomposition en dernière voiture des équations sur les quadrangulations avec un bord. On note donc $Q_n^{(p)}$ le nombre de quadrangulations enracinées avec n sommets et un bord de degré $2p$, et $\mathbf{Q}$ la série génératrice bivariée correspondante

$$\mathbf{Q}(x,y) := \sum_{n\geqslant 1}\sum_{p=0}^{n} Q_n^{(p)} x^n y^p = x + y\left(x^2 + 2x^3 + 9x^4 + \cdots\right) + y^2\left(2x^3 + 9x^4 + 54x^5 + \cdots\right) + \cdots .$$

Notons que par convention, la carte-sommet est une quadrangulation avec un bord de longueur 0, ce qui explique le terme x. Notre décomposition en dernière voiture donne d'équation suivante

$$\mathbf{Q}^\bullet = x + 6y\mathbf{Q}^\bullet\left(\frac{\mathfrak{Q}^\bullet - \mathfrak{Q}}{1-\mathfrak{Q}^\bullet/x}\right) + 2xy\left(3\mathbf{Q}^\bullet - 2\mathbf{Q} - y\partial_y\mathbf{Q}\right),$$

où $\mathbf{Q}^\bullet = x\partial_x \mathbf{Q}$.

On obtient également des équations similaires dans le cas de triangulations. Pour cela, introduisons T_n le nombre de triangulations enracinées (sans bord) avec n sommets et la série génératrice correspondante

$$\mathfrak{T}(x) = \sum_{n\geqslant 2} T_n x^n = x^2 + 4x^3 + 32x^4 + 336x^5 + \dots$$

Théorème 1.12 (*[63]*)

La décomposition en dernière voiture des arbres de peeling des triangulations donne l'équation suivante

$$3\mathfrak{T}^\bullet - 4\mathfrak{T} = 2\left(\frac{\mathfrak{T}^\bullet - \mathfrak{T}}{1 - \mathfrak{T}^\bullet/x}\right),$$

où $\mathfrak{T}^\bullet = x\mathfrak{T}'(x)$. En posant $T_2 = 1$, cette équation caractérise $\mathfrak{T}$ et est équivalente à l'équation récursive suivante

$$T_n = \frac{1}{n-2}\sum_{k=2}^{n-1}(3k-4)(n+1-k)T_k T_{n+1-k}.$$

Comme pour les quadrangulations, cette équation semble très proche de l'équation provenant de la hiérarchie KP :

$$\mathfrak{T}^\bullet - \mathfrak{T} = (6\mathfrak{T}^\bullet - 8\mathfrak{T} + x)^2,$$

voir [101, Théorème 5.4 et Équation 45]. Cependant, il ne nous semble pas possible de déduire une équation de l'autre par le calcul. On définit également $T_n^{(p)}$ le nombre de triangulations avec une frontière de degré p et n sommets au total, et $\mathbf{T}$ la série génératrice bivariée correspondante.

$$\mathbf{T}(x,y) = \sum_{n\geqslant 1, p\geqslant 0} T_n^{(p)} x^n y^p = x + y(x^2 + 4x^3 + 32x^4 + \dots) + \dots.$$

Notre décomposition en dernière voiture montre que $\mathbf{T}$ satisfait l'équation différentielle suivante

$$6\mathbf{T}^\bullet - 2y\partial_y \mathbf{T} - 6\mathbf{T} + y\partial_y(y\mathbf{T}) = \frac{4y}{x}\mathbf{T}^\bullet\left(\frac{\mathfrak{T}^\bullet - \mathfrak{T}}{1 - \mathfrak{T}^\bullet/x}\right) + y\left(4\mathbf{T}^\bullet - 3\mathbf{T} - y\partial_y \mathbf{T}\right).$$

où $\mathbf{T}^\bullet = x\partial_x \mathbf{T}$.

Nous verrons qu'on peut également obtenir des équations pour énumérer les triangulations et quadrangulations avec deux bords, voir la fin du Chapitre 8.

1.4.3 Perspectives

Notre décomposition en dernière voiture est complètement explicite sur les arbres de peeling. Malheureusement, son action sur la carte de départ elle-même (celle qui correspond à l'arbre de peeling) n'est pour l'instant pas bien comprise. En fait, cette décomposition dépend fortement de notre choix d'algorithme de peeling au départ et devrait donc donner de nombreuses bijections en fonction de l'algorithme de peeling choisi. Nous nous demandons s'il existe un algorithme de peeling pour lequel notre transformation nous permettrait de construire une nouvelle bijection "simple" entre, par

exemple, une quadrangulation enracinée avec un sommet distingué et la suite de cartes correspondant à la décomposition en dernières voitures de son arbre de peeling.

Partie I

Parking statique sur des graphes

Chapitre 2 :

Surprising identities for the greedy independent set on Cayley trees

Les résultats de ce chapitre sont issus de l'article [65], publié dans Journal of Applied Probability.

We prove a surprising symmetry between the law of the size G_n of the greedy independent set on a uniform Cayley tree $\mathcal{T}_n$ of size n and that of its complement. We show that G_n has the same law as the number of vertices at even height in $\mathcal{T}_n$ rooted at a uniform vertex. This enables us to compute the exact law of G_n. We also give a Markovian construction of the greedy independent set, which highlights the symmetry of G_n and whose proof uses a new Markovian exploration of *rooted* Cayley trees which is of independent interest.

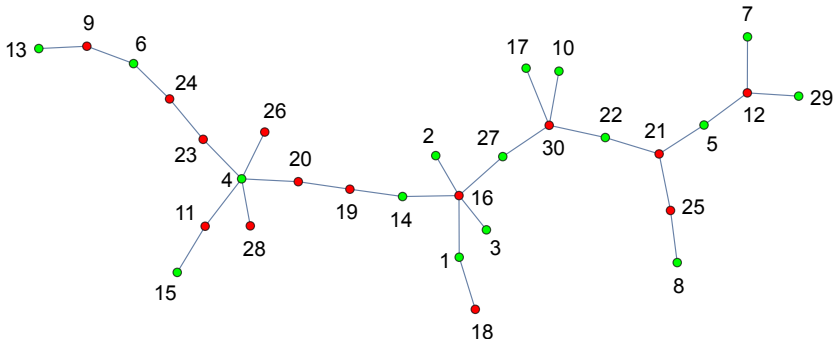


Figure 2.1: Example of the greedy independent set obtained on a tree of size 30. The labels represent the order in which vertices are inspected in the construction of the greedy independent set. The green vertices are the active vertices whereas the red vertices are the blocked vertices.

Contents

2.1 Introduction

An *independent* set of a graph $\mathcal{G} = (V, E)$ is a subset of vertices where no pair of vertices are connected to each other. Finding an independent set of maximal size is a notoriously difficult problem in general [95]. However, using a greedy procedure, we can construct a *maximal* (for the inclusion order) independent set by inspecting the vertices of the graph one by one in a random order, adding the current vertex and blocking its neighbours if it is not connected to any previously added vertex. More precisely, the vertices are divided in three possible statuses: the undetermined vertices $\mathcal{U}_k$, the active vertices $\mathcal{A}_k$ and the blocked vertices $\mathcal{B}_k$. Initially, we start with $\mathcal{U}_0 = V$ and $\mathcal{A}_0 = \mathcal{B}_0 = \emptyset$. At step $k \geqslant 1$, we choose an undetermined vertex v_k uniformly at random, change its status to active and change the status of all its undetermined neighbours to blocked. We stop at $\tau = \min\{k \geqslant 0,\, \mathcal{U}_k = \emptyset\}$. Note that at each step k, no vertices of $\mathcal{A}_k$ are neighbours and $\mathcal{A}_\tau$ is a maximal independent set, which we call the (random) *greedy* independent set, see Figure 2.1.

Of course the independent set obtained by the greedy algorithm is usually not *maximum* in the sense that it does not have the maximal possible size. In the case of trees, finding an independent set of maximal size is much simpler than in general. However, from a probabilistic or combinatorial point of view the greedy independent set is still worth investigation even on (random) trees. Greedy independent sets on (random) graphs have been studied extensively with a particular focus on the proportion of vertices of the graph in the greedy independent set called the *greedy independence ratio* or *jamming constant*. Recently, Krivelevich, Mészáros, Michaeli and Shikelman [125] used Aldous' objective method [14] to prove under mild assumptions that if a sequence of random finite graphs with a root vertex chosen uniformly at random converges locally, then the sequence of their greedy independence ratios also converges in probability.

Recall that a *Cayley tree* of size n is an unrooted and unordered tree over the n labeled vertices $\{1, \dots, n\}$ and we let $\mathcal{T}_n$ be a random Cayley tree sampled uniformly at random among the n^{n-2} Cayley trees of size n. We shall denote by $\mathcal{T}_n^\bullet$ the rooted tree obtained from $\mathcal{T}_n$ by distinguishing a vertex uniformly at random. Using the local limit of $\mathcal{T}_n^\bullet$ given by Kesten's infinite tree, Krivelevich,

Mészáros, Michaeli and Shikelman [125, Section 6.3] proved the "intriguing fact" that the asymptotic greedy independence ratio of uniform Cayley trees is $1/2$. Meir and Moon proved in [141] that the size of a maximum independent set of a uniform Cayley tree concentrates around ρn where $\rho \approx 0.5671$ is the unique solution of $xe^x = 1$.

In this note we prove a much stronger, and perhaps surprising statement concerning the size of the greedy independent set on a uniform Cayley tree showing that it has (almost) the same law as that of its complement! We denote by G_n the size of the greedy independent set $|\mathcal{A}_\tau|$ on a uniform Cayley tree $\mathcal{T}_n$ and H_n the number of vertices at even height in $\mathcal{T}_n^\bullet$. Our first observation is that G_n has the same law as H_n, which enables us to compute the exact law of G_n.

Theorem 2.1

The size G_n of the greedy independent set on $\mathcal{T}_n$ has the same law as the number H_n of vertices at even height in $\mathcal{T}_n^\bullet$. As a consequence, for $1 \leqslant k \leqslant n-1$,

$$\mathbb{P}(G_n = k) = \mathbb{P}(H_n = k) = \binom{n}{k} \frac{k^{n-k}(n-k)^{k-1}}{n^{n-1}}. \tag{2.1}$$

The proof of Theorem 2.1 relies on the invariance of Cayley trees under rerooting at a uniform vertex. The exact computation of the law of H_n is a consequence of a result of [89] on bi-type alternating Galton–Watson trees. This equality in distribution of G_n and H_n suggests that their common law is almost symmetric with respect to $n/2$. But H_n (as well as G_n) has a little drift caused by the root vertex of $\mathcal{T}_n^\bullet$ which is always at even height. Indeed, it follows from Theorem 2.1 that G_n/n converges in probability to $1/2$ (thus, recovering [125, Section 6.3]) and we also have a local central limit theorem for G_n: for all $A \geqslant 0$,

$$\mathbb{P}\left(G_n = \left\lfloor \frac{n}{2} + x\sqrt{n} \right\rfloor\right) \underset{n\to\infty}{\sim} \frac{1}{\sqrt{n}} \frac{1}{\sqrt{2\pi(1/4)^2}} \exp\left(-\frac{x^2}{2\cdot(1/4)^2}\right) \tag{2.2}$$

uniformly for $x \in [-A, A]$. We also give a "Markovian" construction of the greedy independent set which brings to light this symmetry of G_n.

Theorem 2.2

There exists a random variable $\mathcal{E}_n$ with values in $\{0,1\}$ such that we have

$$G_n \overset{(d)}{=} (n - G_n) + \mathcal{E}_n.$$

Moreover $\mathbb{P}(\mathcal{E}_n = 1) \to 1/4$ as n goes to ∞.

This symmetry between G_n and $n - G_n$ is striking because the geometry of a greedy independent set and that of its complement are totally different (see Figure 2.1). The main idea for the proof of Theorem 2.2 is to consider the underlying tree as "unknown" and to discover it as we build the greedy independent set. See [27, 46, 110] for similar applications of this technique for other random graph models. In our case, we develop in Section 2.3 a new type of Markovian explorations of uniform Cayley trees which is inspired by Pitman's famous algorithm [149] but is more *flexible* in the sense

that we can choose which vertex to explore at each step of the procedure. By choosing the next vertex to explore as the next vertex inspected by the greedy algorithm, we connect these Markovian explorations with the construction of the greedy independent set. We expect these explorations to be applicable to a wider range of contexts, e.g. we will shed new light on Aldous-Broder algorithm [6, 48] and Pitman's construction [149] of $\mathcal{T}_n$ using particular cases of our Markovian explorations (see Section 2.3). Similar explorations are used in a forthcoming work [67] on the parking process on Cayley trees.

Independent sets of edges. Instead of an independent set of vertices, we could have considered an independent set of edges or a *matching*, that is a set of edges in which no pair of elements have a vertex in common. As for the vertices, we can construct a maximal independent set of edges greedily by inspecting the edges one by one in a uniform random order, keep it if our edge set stays independent and stop once we have inspected all the edges. Denoting by M_n the number of edges kept after applying this algorithm on a uniform Cayley tree $\mathcal{T}_n$ of size n, our Markovian exploration allows us to show that M_n/n concentrates around $3/8$ and to obtain a Central Limit Theorem for M_n. We do not include the details here since this result has already been shown by Dyer, Frieze and Pittel in [81, Theorem 2] by other means.
But in the case of plane trees i.e. on rooted and ordered unlabeled trees, the invariance under rerooting at a uniform edge plays the same role as the invariance under rerooting at a uniform vertex above and enables us to compute the exact law of the size of the greedy independent set of edges. More precisely, we let $\widetilde{\mathcal{T}}_n$ be a uniform plane tree i.e. rooted and ordered (unlabeled) tree of size n (i.e. with n vertices) and let $\widetilde{M}_n$ be the size of the greedy independent set of edges obtained on $\widetilde{\mathcal{T}}_n$.

Theorem 2.3

The size $\widetilde{M}_n$ of the greedy independent set of edges on $\widetilde{\mathcal{T}}_n$ has law given by

$$\mathbb{P}(\widetilde{M}_n = k) = \frac{n}{k}\frac{\binom{n+k-2}{2k-1}\binom{n-k-1}{k-1}}{\binom{2n-2}{n-1}} = \frac{n!(n-1)!(n+k-2)!}{k!(2k-1)!(n-2k)!(2n-2)!},$$

for $1 \leqslant k \leqslant \lfloor n/2 \rfloor$. As a consequence, for all $A \geqslant 0$,

$$\mathbb{P}\left(\widetilde{M}_n = \left\lfloor \frac{n}{3} + x\sqrt{n} \right\rfloor\right) \underset{n\to+\infty}{\sim} \frac{1}{\sqrt{n}} \frac{1}{\sqrt{2\pi(2/9)^2}} \exp\left(-\frac{x^2}{2\cdot(2/9)^2}\right)$$

uniformly for $x \in [-A, A]$.

We prove this theorem at the end of Section 2.2.

Peleg Michaeli has recently informed us that the formula 2.1 was also independently proved by Alois Panholzer in 2020 in [147].

Acknowledgments. The author is very grateful to Mireille Bousquet-Mélou, Vincent Delecroix and Philippe Duchon for enlightening discussions during the 2021 Aléa days. In particular, we are indebted to Mireille Bousquet-Mélou for first computing the law of G_n in Theorem 2.1. We thank Igor

Kortchemski for a discussion about bicolored trees. We also thank Matthieu Jonckheere for stimulating discussions and Nicolas Curien for his constant guidance. I also thank the two anonymous referees for their careful reading and their valuable suggestions that helped clarifying the paper, as well as Peleg Michaeli for pointing [147] and catching shortcomings in the first version of the paper.

2.2 Greedy independent sets and bicolored trees

In this section, we prove Theorems 2.1 and 2.3 by relating the construction of greedy independent sets to bicolored Galton–Watson trees with alternating colors. Our main tool will be invariance of the underlying random tree with respect to rerooting at a uniform vertex or edge.

2.2.1 Independent set of vertices for Cayley trees.

The following lemma is well known (see for instance [129, End of Section 1.5]) and implies in particular the invariance of a uniform Cayley tree under independent uniform relabeling of the vertices.

Lemma 2.1. *If $\mathcal{T}_n$ is a Galton–Watson plane tree with* Poisson(1) *offspring distribution conditioned to have n vertices, then the tree $\mathcal{T}_n^\bullet$ obtained by labeling its vertices by $\{1,\dots,n\}$ uniformly at random, forgetting the planar ordering but keeping the root vertex, is a uniform rooted Cayley tree.*

In particular, if σ_n is a uniform permutation of $\{1,2,\dots,n\}$ independent of $\mathcal{T}_n$, then the tree $\mathcal{T}_n^{\sigma_n}$ obtained by relabeling the vertex i of $\mathcal{T}_n$ by $\sigma_n(i)$ for $1 \leqslant i \leqslant n$ is still a uniform Cayley tree. In the rest of this section, we shall always consider that $\mathcal{T}_n$ and $\mathcal{T}_n^\bullet$ are built from $\mathcal{T}_n$ as above.

Remark. By invariance under rerooting, the tree $\mathcal{T}_n^\bullet$ has the same law, seen as unlabeled rooted tree, as the tree $\mathcal{T}_n$ rooted at any deterministic vertex $i \in \{1,\dots,n\}$. In the next section, we shall always suppose that our Cayley trees are rooted at the vertex with label n, but in this section, it is better to think of them as rooted on a uniform vertex.

We decompose the proof of Theorem 2.1 in two parts. We first prove that G_n has the same law as H_n and then compute explicitly the law of H_n using bi-type alternating Galton-Watson trees.

Proof of the first half of Theorem 2.1. The proof of this lemma relies on Lemma 2.1. We will show that G_n and H_n obey the characteristic same recursive distributional equation.
Let us start with H_n. Recall that $\mathcal{T}_n^\bullet$ is built from the conditioned Galton–Watson plane tree $\mathcal{T}_n$ by assigning uniform labels, keeping the root vertex and forgetting about the plane ordering. If we denote by $K \geqslant 0$ the number of vertices of height 2 in $\mathcal{T}_n$, by $T_1,\dots,T_K$ the plane trees (ordered from left to right) attached to theses vertices in $\mathcal{T}_n$, by $N_1,\dots,N_K$ their respective sizes (see Figure 2.2), then by the Markov branching property of the Galton–Watson measure [102], conditionally on $(K,N_1,\dots,N_K)$, the plane rooted trees $(T_1,\dots,T_K)$ are independent Galton–Watson plane trees with Poisson(1) offspring distribution conditioned to have sizes $(N_1,\dots,N_K)$. Since the number of vertices at even height in $\mathcal{T}_n$ is just 1 (for the root vertex) plus the sum of the number of vertices at

even height in every tree T_i for $1 \leqslant i \leqslant K$, it follows that

$$H_n \overset{(d)}{=} 1 + \sum_{i=1}^{K} H_{N_i}^{(i)}, \tag{2.3}$$

where $(H_j^{(i)} : 1 \leqslant i, j)$ are independent variables also independent of $(K, N_1, \ldots, N_K)$ and $H_j^{(i)}$ has law H_j for every $i, j \geqslant 1$.

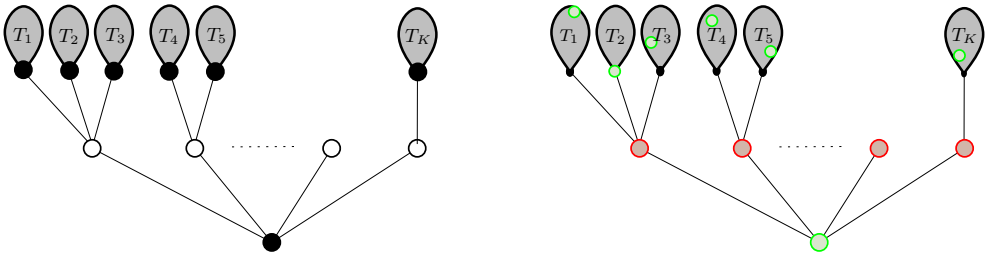


Figure 2.2: On the left, an illustration of the recursive equation in law for the number H_n of vertices at even height on $\mathscr{T}_n$. In black at the bottom, the root vertex, in white its neighbours, and the next vertices to include are the black roots in each tree T_i. On the right, an illustration of the first step of the greedy algorithm. With our coupling with $\mathscr{T}_n$, the first vertex that we add in the greedy independent set is the root vertex of $\mathscr{T}_n$ (in green at the bottom and in red, its neighbours). The next vertices (in green) to inspect in each T_i are then not necessarily the root vertices of T_i.

Let us now move on to the size of the greedy independent set. By construction, it is built first by including a uniform vertex V with label in $\{1, \ldots, n\}$ and blocking its neighbours (which are the vertices at distance 1 from the vertex V). We can assume that this vertex is actually the root vertex of $\mathcal{T}_n^\bullet$ (or equivalently of $\mathscr{T}_n$) since it is a uniform vertex of $\{1, \ldots, n\}$. Using the same notation as above (see Figure 2.2), the crucial observation is that the greedy independent set is obtained by joining the existing root vertex of $\mathscr{T}_n$ together with the independent sets obtained by applying the greedy algorithm on the trees $T_1, \ldots, T_K$ independently. The difference with the case of H_n above is that, in each tree T_i, the next vertex to inspect is not necessarily the root of T_i (induced by $\mathscr{T}_n$) but a "new" uniform vertex of T_i (see Figure 2.2). But by invariance of uniform Cayley trees under uniform rerooting, we still have

$$G_n \overset{(d)}{=} 1 + \sum_{i=1}^{K} G_{N_i}^{(i)} \tag{2.4}$$

where $(G_j^{(i)} : 1 \leqslant i, j)$ are independent variables also independent of $(K, N_1, \ldots, N_K)$ and $G_j^{(i)}$ has law G_j for $i, j \geqslant 1$.
Moreover, the equations (2.3) and (2.4) characterize the law of H_n and G_n since $N_1 + \cdots + N_K \leqslant n - 1$ almost surely. Hence G_n and H_n have the same law and we get the desired result.

□

Proof of the second half of Theorem 2.1. By Lemma 2.1, the variable H_n has the same law as the number of vertices at even height in a Galton–Watson tree with Poisson(1) offspring distribution conditioned to have n vertices. To compute it, we artificially introduce two types of vertices (white vertices and black vertices) and let $\mathcal{T}_b$ be a two-type alternating Galton–Watson tree with a black root and with Poisson(1) offspring distribution (for both types of vertices), so that all vertices at even height are black and have white children, and all vertices at odd height are white and have black children. We denote by N_b (resp. N_w) the number of black (resp. white) vertices in $\mathcal{T}_b$. Using the result of [58] and more precisely [89, Corollary 3.4] and denoting by S_j the sum of j i.i.d. random variables with law Poisson(1) for $j \geqslant 1$, so that S_j has law Poisson(j), we obtain, for all $k \geqslant 1$,

$$\begin{aligned}\mathbb{P}(H_n = k) &= \mathbb{P}(N_b = k, N_w = n-k | N_b + N_w = n)\\ &= \frac{n}{k}\frac{\mathbb{P}(S_k = n-k)\mathbb{P}(S_{n-k} = k-1)}{\mathbb{P}(S_n = n-1)}\\ &= \binom{n}{k}\frac{k^{n-k}(n-k)^{k-1}}{n^{n-1}}.\end{aligned}$$

□

A straightforward application of Stirling's formula gives Equation 2.2. We then easily deduce the local central limit theorem and the law of large numbers.

2.2.2 Independent set of edges for plane trees.

As mentioned in the introduction, this construction of the greedy independent set of vertices on uniform Cayley trees, which are invariant under rerooting at a uniform vertex, suggests a similar construction for the greedy independent set of edges on uniform plane trees which are invariant under rerooting at a uniform edge. Recall that we denote by $\widetilde{\mathcal{T}}_n$ a uniform plane tree of size n (i.e. with n vertices). It can be seen as a graph which is properly embedded in the plane, has only one face of degree $2n-2$, and is rooted at the oriented edge going from the root vertex to its leftmost child (see Figure 2.3). The crucial observation is that if, conditionally on $\widetilde{\mathcal{T}}_n$, we let $\vec{e}$ be a uniform oriented edge of $\widetilde{\mathcal{T}}_n$, then the tree $\widetilde{\mathcal{T}}_n^{\vec{e}}$ obtained be rerooting the tree $\widetilde{\mathcal{T}}_n$ at the edge $\vec{e}$ is still a uniform plane tree.

Moreover, as uniform Cayley trees, uniform plane trees of size n can be seen as conditioned Galton–Watson trees with the appropriate offspring distribution (see for instance [9]).

Lemma 2.2. *A Galton–Watson plane tree with* Geom(1/2) *offspring distribution conditioned to have n vertices is a uniform (rooted) plane tree of size n, where a random variable X has law* Geom(1/2) *if for all $k \geqslant 0$, we have $\mathbb{P}(X = k) = 2^{-k-1}$.*

Using this lemma and the invariance under rerooting at a uniform oriented edge, we now prove Theorem 2.3.

Proof of Theorem 2.3. By invariance under rerooting at a uniform oriented edge, we can assume that our greedy algorithm on $\widetilde{\mathcal{T}}_n$ first includes the root edge, blocks its "neighbouring edges" i.e. the edges

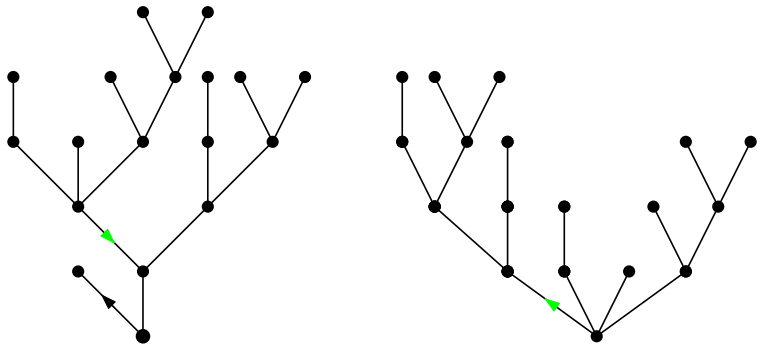


Figure 2.3: On the left, a plane tree rooted at the black oriented edge. On the right, the usual representation of this plane tree rerooted at the green oriented edge.

adjacent to one of its endpoints (see Figure 2.4, left). We denote by K_2 the number of children of the root vertex minus 1 (or number of brothers of the root edge), by K_1 the number of children of the first child of the root, and by $T_1^1, \dots, T_{K_1}^1$ and $T_1^2, \dots T_{K_2}^2$ the trees attached to theses vertices (ordered from left to right, see Figure 2.4), which have respective sizes $N_1^1, \dots, N_{K_1}^1$ and $N_1^2, \dots, N_{K_2}^2$. Then, conditionally on $(K_1, K_2, N_1^1, \dots, N_{K_1}^1, N_1^2, \dots, N_{K_2}^2)$, the trees $(T_1^1, \dots T_{K_1}^1, T_1^2, \dots, T_{K_2}^2)$ are independent Galton–Watson trees with $\mathsf{Geom}(1/2)$ offspring distribution conditioned to have respective sizes $(N_1^1, \dots, N_{K_1}^1, N_1^2, \dots, N_{K_2}^2)$. By the same argument as in the proof of Theorem 2.1, we deduce that $\widetilde{M}_n$ satisfies the following recursive distributional equation

$$\widetilde{M}_n \overset{(d)}{=} 1 + \sum_{i=1}^{K_1} \widetilde{M}_{N_i^1}^{(i)} + \sum_{i=1}^{K_2} \widetilde{M}_{N_i^2}^{(K_1+i)}, \tag{2.5}$$

where $(\widetilde{M}_j^{(i)} : 1 \leqslant i, j)$ are independent variables also independent of

$$(K_1, K_2, N_1^1, \dots, N_{K_1}^1, N_1^2, \dots, N_{K_2}^2)$$

and $\widetilde{M}_j^{(i)}$ has law $\widetilde{M}_j$ for $1 \leqslant i, j$.

As in the proof of Theorem 2.1 we interpret the last recursive distributional equation using two-type Galton–Watson trees. Specifically, let $\mathbb{T}$ be a two-type alternating Galton–Watson tree, with green and red vertices, with a green root and where the red vertices have Bernoulli of parameter $1/2$ offspring distribution and the offspring distribution of the green vertices is the law of the sum of two independent variables with law $\mathsf{Geom}(1/2)$. We denote by N^g (resp. N^r) the number of green (resp. red) vertices in $\mathbb{T}$. Writing N_n^g for the number of green vertices in the tree $\mathbb{T}$ conditioned to have $n-1$ vertices in total, then N_n^g obeys the same recursive equation (2.5) as $\widetilde{M}_n$ (see Figure 2.4).

Using again [89, Corollary 3.4] and denoting by $\widetilde{S}_j^g$ the sum of j i.i.d. random variables with law $\mathsf{Geom}(1/2)$ and $\widetilde{S}_j^r$ the sum of j i.i.d. random variables with Bernoulli law of parameter $1/2$ for

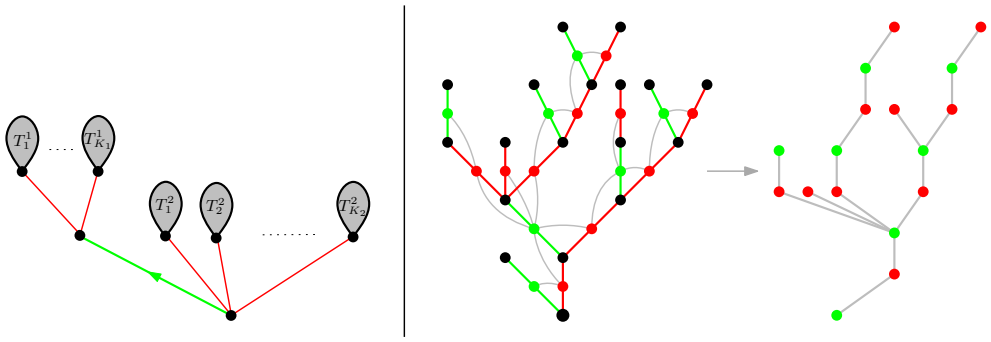


Figure 2.4: On the left, an illustration of the first step of the greedy algorithm for the independent set of edges. The first edge that we add in the independent set is the green root edge, and we block its neighbouring edges (red). The next edge to inspect is a uniform edge in each tree T_i^j, but can be taken as the root edge by invariance under rerooting. On the right, a plane tree $\mathbf{t}$ with black vertices and with edges colored as follows: we color the root edge in green, its neighbouring edges in red and reapply the procedure in each tree by taking first the root edge (in green). Its corresponding bi-type alternating plane tree $\mathbf{t}_g$ is obtained by considering the edges of $\mathbf{t}$ as the vertices of $\mathbf{t}_g$: the children of a green vertex in $\mathbf{t}_g$ correspond to its children followed by its brothers in $\mathbf{t}$, and a red vertex has a green child in $\mathbf{t}_g$ if the red corresponding edge has (at least) a child in $\mathbf{t}$.

$j \geqslant 1$, so that $\widetilde{S}_j^r$ has binomial distribution of parameters j and $1/2$, we obtain, for all $k \geqslant 1$,

$$\begin{aligned}\mathbb{P}(\widetilde{M}_n = k) &= \mathbb{P}(N^g = k, N^r = n-1-k | N^g + N^r = n-1)\\ &= \frac{1}{k}\frac{\mathbb{P}(\widetilde{S}^g_{2k} = n-1-k)\mathbb{P}(\widetilde{S}^r_{n-1-k} = k-1)}{\mathbb{P}(\widetilde{S}^g_n = n-1)}\\ &= \frac{n}{k}\frac{\binom{n+k-2}{2k-1}\binom{n-k-1}{k-1}}{\binom{2n-2}{n-1}}.\end{aligned}$$

□

2.3 Markovian explorations of a rooted tree

In this section, we introduce the Markovian explorations of Cayley trees which we will use to prove the symmetry of the law of G_n (Theorem 2.2) and which we shall call "peeling explorations" by analogy with the peeling process in the theory of random planar maps, see [72]. Given a Cayley tree $\mathbf{t}$, an *exploration* of $\mathbf{t}$ will be a sequence of forests $(\mathbf{f}_0, \ldots, \mathbf{f}_{n-1})$ starting from the forest $\mathbf{f}_0$ made of n isolated vertices and ending at $\mathbf{f}_{n-1} = \mathbf{t}$, such that we pass from $\mathbf{f}_i$ to $\mathbf{f}_{i+1}$ by adding one edge. In our setup, the edge to add at each step will be the edge linking the vertex *peeled* at step i to its parent. To do so, we thus need to root our tree $\mathbf{t}$ by distinguishing a vertex and orienting all edges towards it. That is why in the rest of the paper (and contrary to the previous section) we will always see a

Cayley tree $\mathbf{t}$ as rooted at the vertex of index n.
Specifically, for $0 \leqslant k \leqslant n-1$, we let $\mathcal{F}^*(n,k)$ be the set of (unordered) rooted forests on labeled vertices $\{1,\dots,n\}$ with exactly k edges where one of the trees is distinguished: in the following, the distinguished tree and its vertices are seen as being *blue* whereas the other trees and vertices are *white*, see Figure 2.5. We denote by $\mathcal{F}_n^* = \cup_{k=0}^{n-1}\mathcal{F}^*(n,k)$ the set of such rooted forests on $\{1,\dots,n\}$. Below, the blue component will correspond to the component of the "real" root of the tree in our peeling exploration of rooted Cayley tree. Given a forest $\mathbf{f} \in \mathcal{F}^*(n,k)$, we say that a rooted Cayley tree $\mathbf{t}$ *contains* $\mathbf{f}$ if $\mathbf{t}$ can be obtained from $\mathbf{f}$ by adding edges between each white root of a tree and another compatible vertex i.e. a vertex contained in a different tree of $\mathbf{f}$. Moreover, if v_1 is the root of a white tree and v_2 is a vertex of another tree, we denote by $\mathbf{f}_{v_1\to v_2}$ the forest obtained from $\mathbf{f}$ by adding an edge from v_1 to v_2 and coloring the resulting component with the color of v_2. Assume now that we have a function $\mathfrak{a}$ called the *peeling algorithm* which associates $\mathbf{f} \in \mathcal{F}_n^* \setminus \mathcal{F}^*(n,n-1)$ with a white root of a tree of the forest $\mathbf{f}$.

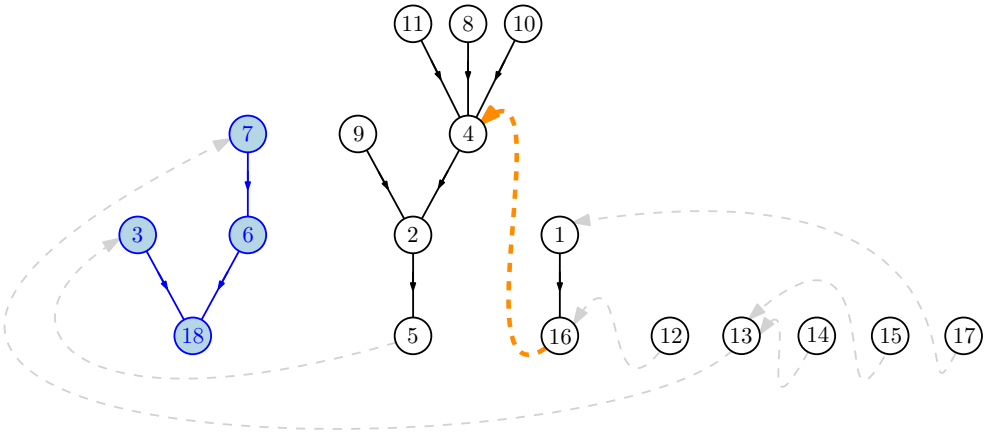


Figure 2.5: Example of a forest $\mathbf{F}_{10}^{\mathfrak{a}}$ which can be obtained after 10 steps of exploration of a given tree $\mathbf{t}$ of size 18. The edges which are still unknown are in dashed. If the vertex labeled $\mathfrak{a}(\mathbf{F}_{10}^{\mathfrak{a}}) = 16$ is the next peeled vertex, then the next edge to be added will be the dashed orange edge.

Definition 2.1. *Let $\mathbf{t}$ be a Cayley tree with n vertices rooted at the vertex labeled n. The peeling exploration of $\mathbf{t}$ with algorithm $\mathfrak{a}$ is the sequence*

$$\mathbf{F}_0^{\mathfrak{a}} \longrightarrow \mathbf{F}_1^{\mathfrak{a}} \longrightarrow \dots \longrightarrow \mathbf{F}_{n-2}^{\mathfrak{a}} \longrightarrow \mathbf{F}_{n-1}^{\mathfrak{a}}$$

obtained by starting from $\mathbf{F}_0^{\mathfrak{a}}$ the unique element of $\mathcal{F}^*(n,0)$ with isolated vertices, which are all white except the blue vertex n, and at step $i \geqslant 1$, the forest $\mathbf{F}_i^{\mathfrak{a}}$ is obtained from $\mathbf{F}_{i-1}^{\mathfrak{a}}$ and $\mathfrak{a}(\mathbf{F}_{i-1}^{\mathfrak{a}})$ as follows: if v is the parent of $\mathfrak{a}(\mathbf{F}_{i-1}^{\mathfrak{a}})$ in $\mathbf{t}$, then we add the edge from $\mathfrak{a}(\mathbf{F}_{i-1}^{\mathfrak{a}})$ to v to $\mathbf{F}_{i-1}^{\mathfrak{a}}$ and color the vertices of the resulting component with the color of v to obtain $\mathbf{F}_i^{\mathfrak{a}} = \left(\mathbf{F}_{i-1}^{\mathfrak{a}}\right)_{\mathfrak{a}(\mathbf{F}_{i-1}^{\mathfrak{a}})\to v}$.

Notice that $\mathbf{F}_{n-1}^{\mathfrak{a}}$ is the tree $\mathbf{t}$ with blue vertices, regardless of the choice of $\mathfrak{a}$. Moreover, when $\mathbf{t} = \mathcal{T}_n$ is a uniform Cayley tree rooted at n, then the sequence $(\mathbf{F}_i^{\mathfrak{a}})_{0\leqslant i\leqslant n-1}$ is a Markov chain:

Proposition 2.1. *Fix a peeling algorithm $\mathfrak{a}$. If $\mathcal{T}_n$ is a uniform Cayley tree of size n rooted at n, then the exploration $(\mathbf{F}_i^{\mathfrak{a}})_{0\leqslant i\leqslant n-1}$ of $\mathcal{T}_n$ with algorithm $\mathfrak{a}$ is a Markov chain whose probability transitions are described as follows: for $i \geqslant 0$, conditionally on $\mathbf{F}_i^{\mathfrak{a}}$ and on $\mathfrak{a}(\mathbf{F}_i^{\mathfrak{a}})$, we denote by $m \geqslant 1$ the size of the connected component of $\mathfrak{a}(\mathbf{F}_i^{\mathfrak{a}})$ and by $\ell \geqslant 1$ the number of blue vertices in $\mathbf{F}_i^{\mathfrak{a}}$.*

- *For every blue vertex v, the parent of $\mathfrak{a}(\mathbf{F}_i^{\mathfrak{a}})$ is v with probability $(\ell+m)/(\ell n)$ and in that case, we add an edge from $\mathfrak{a}(\mathbf{F}_i^{\mathfrak{a}})$ to v and color the resulting component blue to obtain $\mathbf{F}_{i+1}^{\mathfrak{a}}$.*
- *For every white vertex v which does not belong to the component of $\mathfrak{a}(\mathbf{F}_i^{\mathfrak{a}})$, the parent of $\mathfrak{a}(\mathbf{F}_i^{\mathfrak{a}})$ is v with probability $1/n$ and in that case we add an edge from $\mathfrak{a}(\mathbf{F}_i^{\mathfrak{a}})$ to v to obtain $\mathbf{F}_{i+1}^{\mathfrak{a}}$.*

This proposition is a direct corollary of the following lemma.

Lemma 2.3. *If $\mathbf{f} \in \mathcal{F}^*(n,k)$ is a forest on $\{1,\dots,n\}$ with k edges and $\ell \geqslant 1$ blue vertices (hence $n-k-1$ white trees and a blue tree), then the number $N(\mathbf{f})$ of rooted Cayley trees containing $\mathbf{f}$ is equal to ℓn^{n-k-2}.*

Proof. The proof is similar to that of Pitman in [149, Lemma 1]. We introduce the number $N^*(\mathbf{f})$ of *refining* sequences of $\mathbf{f}$, that is the number of sequences $(\mathbf{f}_j,\dots,\mathbf{f}_1)$ where $\mathbf{f}_j = \mathbf{f}$ and for $i \leqslant j$, the forest $\mathbf{f}_i$ has a blue tree and $i-1$ white trees, and $\mathbf{f}_{i-1}$ can be obtained from $\mathbf{f}_i$ by adding an edge as above. Note that $j = n-k$ since $\mathbf{f}$ has $n-k-1$ white trees. Given a fixed target $\mathbf{f}_1$, any forest $\mathbf{f}_j$ that contains a $\mathbf{f}_1$ has $j-1$ fewer edges than $\mathbf{f}_1$. Hence there are $(j-1)!$ refining sequences from $\mathbf{f}_j$ to $\mathbf{f}_1$ and $N^*(\mathbf{f}_j) = (j-1)!N(\mathbf{f}_j)$. Now we prove the result on $N(\mathbf{f})$ by induction over $j-1$, the number of white trees of $\mathbf{f}$. When $j=1$, then $\mathbf{f} \in \mathcal{F}^*(n,n-1)$ is simply a blue rooted tree and the result is clear.

Suppose the result holds for some $j-1 \geqslant 0$. Let $\mathbf{f} \in \mathcal{F}^*(n,n-j-1)$ be a forest on $\{1,\dots,n\}$ with $n_1 \geqslant 1$ blue vertices and j white trees $(\mathbf{t}_2,\dots,\mathbf{t}_{j+1})$ of size $(n_2,\dots,n_{j+1})$. Then, if r_i is the root of the white tree $\mathbf{t}_i$,

- for each blue vertex v_1, by the induction hypothesis, there are $N^*(\mathbf{f}_{r_i\to v_1}) = (n_1+n_i)n^{j-2}(j-2)!$ refinements of $\mathbf{f}_{r_i\to v_1}$ and there are n_1 such vertices v_1,
- and for each vertex v_j in a white tree $\mathbf{t}_j$ with $i \neq j$, by the induction hypothesis, there are $N^*(\mathbf{f}_{r_i\to v_j}) = n_1 n^{j-2}(j-2)!$ refinements of $\mathbf{f}_{r_i\to v_j}$ and there are $n-(n_i+n_1)$ such vertices v_j.

In total, the number of refinements of $\mathbf{f}$ is

$$N^*(\mathbf{f}) = \sum_{i=2}^{j+1} n_1(n_1+n_i)n^{j-2}(j-2)! + (n-(n_i+n_1))n_1 n^{j-2}(j-2)! = (j-1)!n_1 n^{j-1},$$

and we get the desired result.

□

Now we can prove Proposition 6.5. **Proof.** [Proof of Proposition 6.5] It suffices to notice that since $\mathcal{T}_n$ is uniform, for all $i \geqslant 0$, conditionally on $\mathbf{F}_i^{\mathfrak{a}}$, the tree $\mathcal{T}_n$ is a uniform tree among those which contain $\mathbf{F}_i^{\mathfrak{a}}$. Hence, for every (compatible) vertex v,

$$\mathbb{P}(\mathbf{F}_{i+1}^{\mathfrak{a}} = (\mathbf{F}_i^{\mathfrak{a}})_{\mathfrak{a}(\mathbf{F}_i^{\mathfrak{a}})\to v} | \mathbf{F}_i^{\mathfrak{a}}, \mathfrak{a}(\mathbf{F}_i^{\mathfrak{a}})) = \frac{|\{\mathbf{t} \text{ s.t } \mathbf{t} \text{ contains } (\mathbf{F}_i^{\mathfrak{a}})_{\mathfrak{a}(\mathbf{F}_i^{\mathfrak{a}})\to v}\}|}{|\{\mathbf{t} \text{ s.t } \mathbf{t} \text{ contains } \mathbf{F}_i^{\mathfrak{a}}\}|}.$$

Using Lemma 6.1, we recognize the transition probabilities given in Proposition 6.5 and obtain the desired result.

□

The strength of Proposition 6.5 is that different peeling algorithms yield different explorations (hence different types of information) of the same underlying tree. Let us illustrate this with several examples of peeling algorithms.

Pitman's algorithm and $\mathfrak{a}_{\text{unif}}$. A natural choice of algorithm $\mathfrak{a}$ is, given a forest $\mathbf{f}$, to choose a root of a white tree of $\mathbf{f}$ uniformly at random for $\mathfrak{a}_{\text{unif}}(\mathbf{f})$. This does not seem to enter our setup since $\mathfrak{a}_{\text{unif}}$ is not a deterministic function of $\mathbf{f}$. However we can imagine that we first condition on the randomness involved in $\mathfrak{a}_{\text{unif}}$ making it deterministic. Once $\mathfrak{a}_{\text{unif}}$ is fixed, we apply it to a random Cayley tree $\mathcal{T}_n$, thus independent of the choice of $\mathfrak{a}_{\text{unif}}$. The Markov chain obtained in Proposition 6.5 to this peeling is reminiscent of (but not identical to) Pitman's construction [149] of uniform rooted Cayley trees, which we quickly recall: Start from the forest made of the n isolated vertices $\{1, \dots, n\}$ and at step $1 \leqslant k \leqslant n-1$, pick a vertex V_k uniformly at random and then pick a root R_k uniformly at random among the $n-k$ trees which do not contain V_k, and add the directed edge from R_k to V_k. Note that in Pitman's construction, we first pick a uniform vertex then a uniform compatible root and the "real" root of the Cayley tree is found out only at the end of the exploration whereas in our peeling exploration with algorithm $\mathfrak{a}_{\text{unif}}$, we first fix the root as the vertex labeled n and during the exploration we choose uniformly a root of a white tree whose parent is chosen almost uniformly among the compatible vertices.

Building branches and Aldous-Broder algorithm. Let us mention another choice of algorithm which sheds new light on Aldous-Broder's construction [6, 48] of $\mathcal{T}_n$ by using a random walk on the complete graph *with loops*. We assume in this paragraph that $n \geqslant 2$. Given a forest $\mathbf{f} \in \mathcal{F}_n^*$, let $\mathfrak{a}_{AB}(\mathbf{f})$ be the root of the tree which contains the white vertex with the smallest label. For example, starting at $\mathbf{F}_0^{\mathfrak{a}_{AB}}$, then $\mathfrak{a}_{AB}(\mathbf{F}_i^{\mathfrak{a}_{AB}})$ is the root of the component of 1 until the exploration hits the blue root n. Let $T_1^{\mathfrak{a}_{AB}} = \min\{i \geqslant 1 \text{ s.t. } 1 \text{ is a blue vertex in } F_{i-1}^{\mathfrak{a}_{AB}}\}$ be the length of the first "branch" built in this exploration. Then, for $k \geqslant 2$,

$$\mathbb{P}(T_1^{\mathfrak{a}_{AB}} = k) = \frac{k}{n} \prod_{i=2}^{k-1} \left(1 - \frac{i}{n}\right).$$

Recall now Aldous-Broder's algorithm [6, 48] on the complete graph on n vertices, that is the graph with vertex set $\{1, \dots, n\}$, and all possible edges including the loops (i, i): Consider a random walk

$(X_i : i \geqslant 0)$ starting at $X_0 := n$ and where the X_i's are i.i.d. uniform random variables on $\{1, \dots n\}$ for $i \geqslant 1$. We denote by $\tau_k := \inf\{i : X_i = k\} < \infty$ for any $k \in \{1, \dots, n\}$. The walk induces the tree $\mathcal{T}^{AB}$ which consists of the tree on $\{1, \dots, n\}$ with the edges $(X_{\tau_k - 1}, X_{\tau_k})$ for $1 \leqslant k \leqslant n-1$ (we only take the edges which lead to newly discovered vertices). Consider $T_1^{AB} = \min\{i \geqslant 1 \text{ s.t. } X_i \in \{X_0, \dots, X_{i-1}\}\}$ the time of the first repetition of the walk $(X_i : i \geqslant 0)$. Then by construction, $\mathbb{P}(T_1^{AB} = 1) = 1/n$ and for $k \geqslant 2$,

$$\mathbb{P}(T_1^{AB} = k) = \frac{k}{n} \prod_{i=1}^{k-1} \left(1 - \frac{i}{n}\right) = \frac{n-1}{n} \mathbb{P}(T_1^{\mathfrak{a}_{AB}} = k) = \mathbb{P}(T_1^{AB} \neq 1) \cdot \mathbb{P}(T_1^{\mathfrak{a}_{AB}} = k).$$

We deduce that conditionally on $X_1 \neq n$, the law of the length T_1^{AB} of the first branch is identical to that of the length of the branch linking n and 1 in $\mathcal{T}_n$, see [8]. A similar result holds for the next branches in both constructions. This has been already observed by Camarri and Pitman in the more general context of p-trees in [54, Corollary 3].

Greedy construction. In the next section, we use a peeling algorithm (with additional decorations) tailored to the construction of the greedy independent set on $\mathcal{T}_n$.

2.4 Markovian construction of the greedy independent set

This section is devoted to the proof of Theorem 2.2 which use our Markovian exploration of Cayley trees.

2.4.1 Markovian construction and its transitions

Recall from the introduction the greedy algorithm: given a Cayley tree, we inspect its vertices sequentially in a uniform random order and at each step, if the considered vertex is undetermined, we change its status to active and change the status of all its undetermined neighbours to blocked. By Lemma 2.1, we have an invariance property of $\mathcal{T}_n$ under independent uniform relabeling. Hence we can and will directly use the labels of $\mathcal{T}_n$ to define the order of exploration of the greedy independent set on $\mathcal{T}_n$.

Recall that we rooted our tree at the vertex n. The idea here is to link the greedy construction to a peeling exploration of $\mathcal{T}_n$: we explore and inspect at each step a vertex and update its status and possibly *that of its parent in the tree but not these of its children.* In particular, the root n of the tree is possibly the last vertex to be considered (when it is not previously blocked).

More precisely, given the Cayley tree $\mathcal{T}_n$ rooted at n, we divide its vertices into three statuses (undetermined, active or blocked) as before and set

$$\widetilde{\mathcal{U}}_0^n = \{1, \dots, n\}, \qquad \widetilde{\mathcal{A}}_0^n = \widetilde{\mathcal{B}}_0^n = \emptyset.$$

Then inductively at step i, we inspect v_i the *undetermined* vertex with the smallest label and we let w_i be its parent in $\mathcal{T}_n$ (unless $v_i = n$). Then we update their statuses as follows:

- If $v_i = n$, then the vertex n becomes active i.e. $\widetilde{\mathcal{U}}_{i+1}^n = \emptyset$ and $\widetilde{\mathcal{A}}_{i+1}^n = \widetilde{\mathcal{A}}_i^n \cup \{v_i\}$
- If w_i is undetermined, then v_i becomes active (it is taken in the greedy independent set) and w_i becomes blocked i.e. $\widetilde{\mathcal{U}}_{i+1}^n = \widetilde{\mathcal{U}}_i^n \setminus \{v_i, w_i\}$ and $\widetilde{\mathcal{A}}_{i+1}^n = \widetilde{\mathcal{A}}_i^n \cup \{v_i\}$ and $\widetilde{\mathcal{B}}_{i+1}^n = \widetilde{\mathcal{B}}_i^n \cup \{w_i\}$.
- If w_i is blocked (resp. active), then v_i becomes active (resp. blocked), that is $\widetilde{\mathcal{U}}_{i+1}^n = \widetilde{\mathcal{U}}_i^n \setminus \{v_i\}$ and $\widetilde{\mathcal{A}}_{i+1}^n = \widetilde{\mathcal{A}}_i^n \cup \{v_i\}$ (resp. $\widetilde{\mathcal{B}}_{i+1}^n = \widetilde{\mathcal{B}}_i^n \cup \{v_i\}$).

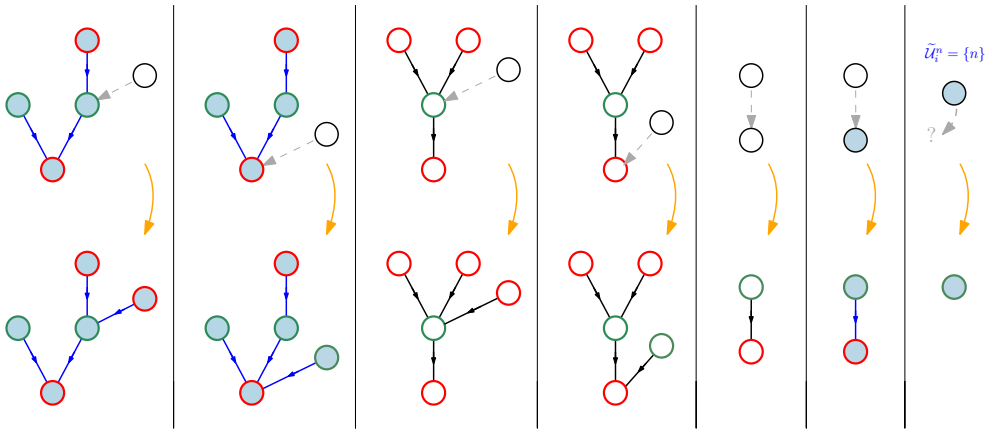


Figure 2.6: Illustration of the possible transitions in the peeling exploration with $\mathfrak{a}_{\text{Greedy}}$ together with the updating of the status of the vertices. The interior color represents the color of the vertices (blue or white) in the peeling exploration and the boundary color represents the status of the vertices in the greedy construction of the independent set: black for the undetermined vertices, green for the active vertices and red for the blocked vertices. The new edge which we explore (between v_i and w_i) is in dotted gray. When $v_i = n$, there is no peeling step but v_i is the last undetermined vertex and it becomes active.

As before, at each step i, the set $\widetilde{\mathcal{A}}_i^n$ is an independent set and we stop the process when all vertices are active or blocked i.e. at time $\theta_n = \inf\{i \geqslant 0,\ \widetilde{\mathcal{U}}_i^n = \emptyset\}$. It is then easy to check that the random subset $\widetilde{\mathcal{A}}_{\theta_n}^n$ is equal to the greedy independent set $\mathcal{A}_\tau$ defined in the introduction (if we inspect vertices according to their labels in the tree). The advantage of the above construction over the one of the introduction where we update the status of all neighbours of the inspected vertex is that it can be seen as a Markovian exploration of $\mathcal{T}_n$ in the sense of the preceding section (where we additionally keep track of the status of the vertices). Indeed, it is equivalent to the peeling of $\mathcal{T}_n$ started at the forest $\mathbf{F}_0^{\mathfrak{a}_{\text{Greedy}}}$ made of an undetermined blue vertex with label n and $n-1$ isolated white undetermined vertices and where at step $i+1$, we peel the vertex

$$v_i = \min \widetilde{\mathcal{U}}_i^n := \mathfrak{a}_{\text{Greedy}}(\mathbf{F}_i^{\mathfrak{a}_{\text{Greedy}}})$$

the undetermined vertex with the smallest label and update the status of the vertices as above, see Figure 6.3. Recall the two possible colors for the vertices during a peeling exploration: the vertices

which are in the connected component of the vertex labeled n which is the "real root" of the tree are *blue* whereas the other vertices are *white*. We make the following remarks which are easily checked by induction: with the possible exception of the vertex n, the undetermined vertices are white whereas the active or blocked vertices can be white or blue. Moreover at each step $i < \theta_n$, the forest $\mathbf{F}_i^{\mathfrak{a}_{\text{Greedy}}}$ is made of white isolated undetermined vertices, white trees of size at least 2 rooted at a blocked vertex, and a blue tree, see Figure 2.7.

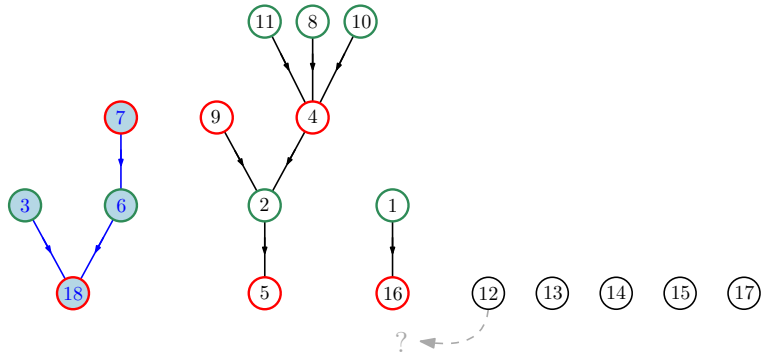


Figure 2.7: Illustration of a possible value of $\mathbf{F}_{10}^{\mathfrak{a}_{\text{Greedy}}}$. As in Figure 6.3, the interior color represents the color of the vertices in the peeling exploration and the boundary color represents the status of the vertices in the greedy construction of the independent set: black for the undetermined vertices, green for the active vertices and red for the blocked vertices. The new vertex to peel has label 12 here.

A small caveat is that a priori the peeling exploration with algorithm $\mathfrak{a}_{\text{Greedy}}$ is not defined after time θ_n: Generically, at that time, although the status in the greedy independent set of all vertices is known, the whole geometric structure of $\mathcal{T}_n$ is not completely revealed since there are many white vertices in $\mathbf{F}_{\theta_n}^{\mathfrak{a}_{\text{Greedy}}}$ (we never peel the blocked roots of white trees). For the sake of completness we may continue the peeling exploration (for example by peeling the white (blocked) root of a tree with the smallest label), but we shall not use it.

We denote by $(U_i^n, A_i^n, B_i^n)_{i\geqslant 0}$ the number of undetermined, active and blocked vertices in the process on a uniform Cayley tree $\mathcal{T}_n$ with the greedy peeling algorithm $\mathfrak{a}_{\text{Greedy}}$ i.e. the size of $\widetilde{\mathcal{U}}_i^n$, $\widetilde{\mathcal{A}}_i^n$ and $\widetilde{\mathcal{B}}_i^n$. For the active and blocked vertices, we distinguish the number of white vertices $A_i^{n,w}$ and $B_i^{n,w}$ from the number of blue vertices $A_i^{n,b}$ and $B_i^{n,b}$. Then with the notation of Theorem 2.2, we have $G_n = A_{\theta_n}^n$ and using a decorated version of Proposition 6.5, we see that

$$(U_i^n, A_i^{n,w}, B_i^{n,w}, A_i^{n,b}, B_i^{n,b} : 0 \leqslant i \leqslant \theta_n)$$

is a Markov chain. To describe its probability transitions, we introduce $\Delta X_i := X_{i+1} - X_i$ the increment of a random process $(X_j : j \geqslant 0)$ at time $i \geqslant 0$. Suppose that $\widetilde{\mathcal{U}}_i^n$ is not empty. On the one hand, when $n \in \widetilde{\mathcal{U}}_i^n$, then n is the only blue vertex in $\mathbf{F}_i^{\mathfrak{a}_{\text{Greedy}}}$, hence $A_i^{n,b} = B_i^{n,b} = 0$. On the other hand, when $n \notin \widetilde{\mathcal{U}}_i^n$, then $A_i^{n,b} \geqslant 1$ and $B_i^{n,b} \geqslant 1$. Note that as long as $\widetilde{\mathcal{U}}_i^n$ is neither $\{n\}$ nor the emptyset $\emptyset$, the vertex $\mathfrak{a}_{\text{Greedy}}(\mathbf{F}_i^{\mathfrak{a}_{\text{Greedy}}})$ is a white isolated vertex in $\mathbf{F}_i^{\mathfrak{a}_{\text{Greedy}}}$. Therefore, conditionally

on $(U_i^n, A_i^{n,w}, B_i^{n,w}, A_i^{n,b}, B_i^{n,b})$, the transitions of the increments of $((U_i^n, A_i^{n,w}, B_i^{n,w}, A_i^{n,b}, B_i^{n,b}) : i \geqslant 0)$ are given by:

If $A_i^{n,b} = B_i^{n,b} = 0$ and $U_i^n = 1$ i.e. if $\widetilde{\mathcal{U}}_i^n = \{n\}$ then almost surely,

$$\Delta(U_i^n, A_i^{n,w}, B_i^{n,w}, A_i^{n,b}, B_i^{n,b}) = (-1, 0, 0, +1, 0).$$

If $A_i^{n,b} = B_i^{n,b} = 0$ and $U_i^n \geqslant 2$,

with prob.	$\frac{U_i^n - 2}{n}$	$\frac{A_i^{n,w}}{n}$	$\frac{B_i^{n,w}}{n}$	$\frac{2}{n}$
ΔU_i^n	-2	-1	-1	-2
$\Delta A_i^{n,w}$	$+1$	0	$+1$	0
$\Delta B_i^{n,w}$	$+1$	$+1$	0	0
$\Delta A_i^{n,b}$	0	0	0	$+1$
$\Delta B_i^{n,b}$	0	0	0	$+1$

If $A_i^{n,b} \geqslant 1, B_i^{n,b} \geqslant 1$ and $U_i^n \geqslant 1$,

with prob.	$\frac{U_i^n - 1}{n}$	$\frac{A_i^{n,w}}{n}$	$\frac{B_i^{n,w}}{n}$	$\frac{A_i^{n,b}(A_i^{n,b}+B_i^{n,b}+1)}{(A_i^{n,b}+B_i^{n,b})n}$	$\frac{B_i^{n,b}(A_i^{n,b}+B_i^{n,b}+1)}{(A_i^{n,b}+B_i^{n,b})n}$
ΔU_i^n	-2	-1	-1	-1	-1
$\Delta A_i^{n,w}$	$+1$	0	$+1$	0	0
$\Delta B_i^{n,w}$	$+1$	$+1$	0	0	0
$\Delta A_i^{n,b}$	0	0	0	0	$+1$
$\Delta B_i^{n,b}$	0	0	0	$+1$	0

(2.6)

2.4.2 Symmetry of the greedy independent set

We now prove Theorem 2.2 which states the symmetry between the law of $A_{\theta_n}^n$ the size of the maximal independent set obtained by the greedy algorithm and $B_{\theta_n}^n$ that of its complement. The main observation is the following lemma.

Lemma 2.4. *For $j \geqslant 0$, the laws of $(A_i^{n,w}, A_i^{n,b})_{0 \leqslant i \leqslant j+1}$ and $(B_i^{n,w}, B_i^{n,b})_{0 \leqslant i \leqslant j+1}$ are the same on the event $\widetilde{\mathcal{U}}_i^n \neq \{n\}$ for all $i \leqslant j$.*

Proof. Notice then that the role of $(A_i^{n,w}, A_i^{n,b})$ and that of $(B_i^{n,w}, B_i^{n,b})$ are exchangeable in the probability transitions (2.6) as long as $\widetilde{\mathcal{U}}_i^n \neq \{n\}$. The lemma follows. □

Thanks to this lemma, we notice that conditionally on $\widetilde{\mathcal{U}}_i^n \neq \{n\}$ for all $0 \leqslant i \leqslant \theta_n$, then the size $A_{\theta_n}^n$ of the greedy independent set has the same law as that its complement $B_{\theta_n}^n$. Otherwise, if there exists $i \geqslant 0$ such that $\widetilde{\mathcal{U}}_i^n = \{n\}$, then this i corresponds to $\theta_n - 1$ and in that case, $A_{\theta_n - 1}^n$ and $B_{\theta_n - 1}^n$ have the same law and $A_{\theta_n}^n = G_n = A_{\theta_n - 1}^n + 1$ and $B_{\theta_n}^n = B_{\theta_n - 1}^n$. Hence, the random variable $\mathcal{E}_n$ in Theorem 2.2 is the indicator function of the event $\{\exists i \geqslant 0,\ \widetilde{\mathcal{U}}_i^n = \{n\}\}$ and to prove Theorem 2.2, it only remains to show that

$$\mathbb{P}(\mathcal{E}_n = 1) = \mathbb{P}(\exists i,\ \widetilde{\mathcal{U}}_i^n = \{n\}) \xrightarrow[n \to \infty]{} 1/4.$$

To compute the above probability and also obtain the fluctuations of G_n, we study the *fluid limit* of our system (see [85, Chapter 11]). The transition probabilities (2.6) in both cases point to the idea of gathering blue and white vertices and focusing only on their status. Indeed, the Markov chain $(U_i^n, A_i^n, B_i^n : i \geqslant 0)$ has bounded increments and for $i < \theta_n$,

$$\left| \mathbb{E}\left[\begin{pmatrix} \Delta U_i^n \\ \Delta A_i^n \\ \Delta B_i^n \end{pmatrix} \middle| \begin{pmatrix} U_i^n \\ A_i^n \\ B_i^n \end{pmatrix} \right] - F\left(\frac{U_i^n}{n}, \frac{A_i^n}{n}, \frac{B_i^n}{n} \right) \right| \leqslant \frac{3}{n}, \quad \text{where } F(x,y,z) = \begin{pmatrix} -2x - y - z \\ x + z \\ x + y \end{pmatrix}. \tag{2.7}$$

This suggests to study the deterministic system which is solution of the following equations:

$$\begin{cases} u'(t) = -2u(t) - a(t) - b(t), & u(0) = 1, \\ a'(t) = u(t) + b(t), & a(0) = 0, \\ b'(t) = u(t) + a(t), & b(0) = 0. \end{cases}$$

The solution is given by

$$\begin{cases} u(t) = 2e^{-t} - 1, \\ a(t) = b(t) = 1 - e^{-t}. \end{cases}$$

In particular $t^* = \min\{t \geqslant 0,\ u(t) = 0\} = \ln(2)$ and $a(t^*) = 1/2$. We can now apply [85, Theorem 2.1 p456] to the process $X_n(t) = 1/n \cdot (U^n_{\lfloor tn \rfloor}, A^n_{\lfloor tn \rfloor}, B^n_{\lfloor tn \rfloor})$ which starts at $X_n(0) = (1,0,0)$ with, using the notation of [85], the function F given by (2.7) and

$$\begin{cases} \beta_{(-2,1,1)}(x,y,z) = x, \\ \beta_{(-1,0,1)}(x,y,z) = y, \\ \beta_{(-1,1,0)}(x,y,z) = z. \end{cases}$$

We obtain that at least for all $t < t^*$, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{0 \leqslant s \leqslant t} \left| \left(\frac{U^n_{\lfloor sn \rfloor}}{n}, \frac{A^n_{\lfloor sn \rfloor}}{n}, \frac{B^n_{\lfloor sn \rfloor}}{n} \right) - (u(s), a(s), b(s)) \right| > \varepsilon \right) \xrightarrow[n \to \infty]{} 0.$$

Since $u'(t^*) < 0$, we can apply [85, Theorem 4.1 p464] which states that the stopping time of our Markov chain also concentrates around its expected value and has Gaussian fluctuations around it:

$$\sqrt{n}\left(\frac{\theta_n}{n} - t^* \right) \xrightarrow[n \to \infty]{(d)} \mathcal{N}(0, 3/4). \tag{2.8}$$

Lastly, notice that $\mathcal{E}_n = 1$ when the vertex n stays undetermined at each step $1 \leqslant i \leqslant \theta_n - 1$, that is, with probability $1 - 2/n$ at each step, and by the the previous equation $(1 - 2/n)^{\theta_n}$ converges in probability towards e^{-2t^*} and is bounded by 1. Hence we obtain

$$\mathbb{P}(\mathcal{E}_n = 1) = \mathbb{E}\left[\left(1 - \frac{2}{n} \right)^{\theta_n - 1} \right] \xrightarrow[n \to \infty]{} \mathrm{e}^{-2t^*} = \frac{1}{4},$$

which concludes the proof of Theorem 2.2.

Chapitre 3 :

The critical Karp–Sipser core of random graphs

Les résultats de ce chapitre ont été obtenus en collaboration avec Thomas Budzinski et Nicolas Curien et ont été soumis pour publication [52].

We study the Karp–Sipser core of a random graph made of a configuration model with vertices of degree $1, 2$ and 3. This core is obtained by recursively removing the leaves as well as their unique neighbors in the graph. We settle a conjecture of Bauer & Golinelli [24] and prove that at criticality, the Karp–Sipser core has size $\approx \mathsf{Cst} \cdot \vartheta^{-2} \cdot n^{3/5}$ where ϑ is the hitting time of the curve $t \mapsto \frac{1}{t^2}$ by a linear Brownian motion started at 0. Our proof relies on a detailed multi-scale analysis of the Markov chain associated to the Karp-Sipser leaf-removal algorithm close to its extinction time.

3.1 Introduction

The Karp–Sipser algorithm. Let $\mathfrak{g}$ be a finite graph. The Karp–Sipser algorithm [116] consists in removing recursively the vertices of degree 1 in $\mathfrak{g}$ as well as their unique neighbors, see Figure 3.2. The initial motivation of Karp & Sipser for considering this algorithm is that the leaves[1] and isolated vertices removed during this process form an independent set of $\mathfrak{g}$ which has very high density. We recall that an independent set in $\mathfrak{g}$ is a subset of vertices, no two of which are adjacent. The problem of finding an independent set of maximal size is in general a NP-hard problem, and the Karp–Sipser algorithm provides a fair lower bound (it is furthermore "optimal" as long as there are remaining leaves in the graph).

[1] Here and in the rest of the paper, the concept of leaf is a dynamical concept, as a vertex in the initial graph which is not a leaf may become one later.

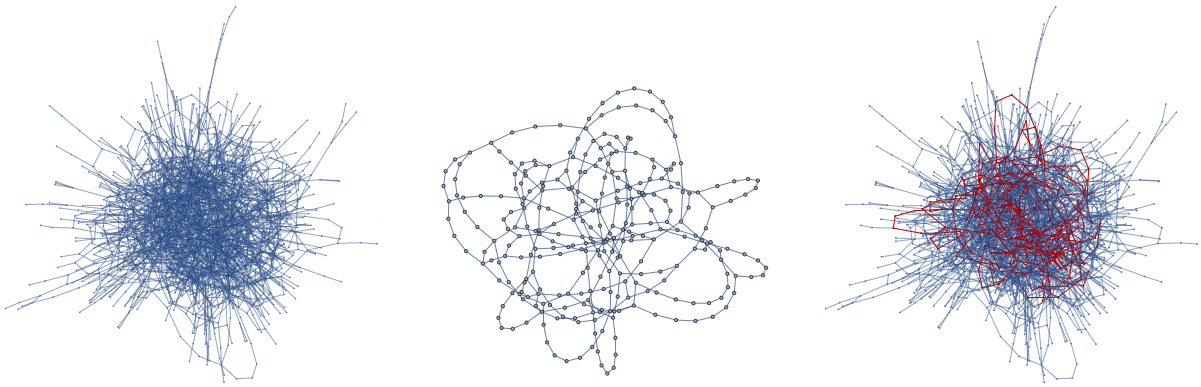


Figure 3.1: (Left). The giant component of an Erdős–Rényi random graph $\mathrm{G}(n, \frac{\mathrm{e}}{n})$ with $n = 2000$ on the left and **(Middle)** its Karp–Sipser core. **(Right).** The Karp–Sipser core in red inside the original graph.

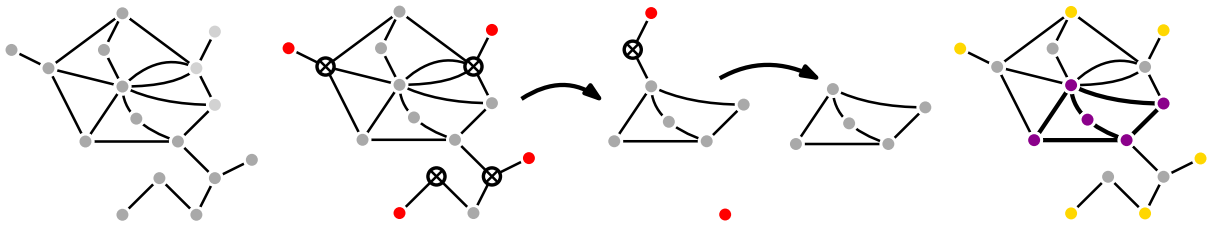


Figure 3.2: Illustration of the Karp–Sipser algorithm. The first 4 figures show the initial graph, as well as the recursive deletion process of the leaves (in red) together with their unique neighbor (crosses), until no leaf is left: we then obtain the Karp–Sipser core (fourth figure). On the right, the initial graph is represented together with the Karp–Sipser core in thick lines and the independent set formed by the removed "leaves" in yellow.

The Karp–Sipser core of random graphs. A striking property of the leaf-removal process is its Abelian property: whatever the order in which we decide to recursively remove the leaves and their neighbors, we always obtain the same subgraph of $\mathfrak{g}$ (with no leaves) which we will call the *Karp–Sipser core* of $\mathfrak{g}$ and denote by $\mathsf{KSCore}(\mathfrak{g})$, see [24, Appendix] or [124, Section 1.6.1]. Beware that the above notion differs from the usual k-core of a graph[2], see Section 3.5. By the above remark, the Karp–Sipser algorithm creates an independent set (the leaves removed during the algorithm) whose size is within at most $|\mathsf{KSCore}(\mathfrak{g})|$ from the maximal size of an independent set in $\mathfrak{g}$.

The performance of the Karp–Sipser algorithm on the Erdős–Rényi random graph $\mathrm{G}(n, \frac{c}{n})$ has been analyzed in the pioneer work [116] and later refined in the breakthrough [19] which established a phase transition as $n \to \infty$ depending on the value of c:

- if $c < \mathrm{e}$, then as $n \to \infty$, the size $|\mathsf{KSCore}(\mathrm{G}(n, \frac{c}{n}))|$ is of order $O(1)$;

[2]The k-core of $\mathfrak{g}$ is the largest subset V of its vertices such that for any $v \in V$, the induced degree of v within V is at least k.

- if $c > \mathbf{e}$, then as $n \to \infty$, the size $|\mathsf{KSCore}(\mathrm{G}(n, \frac{c}{n}))|$ is of order n.

Those works have later been extended to the configuration model [39, 110]. However, the careful analysis of the critical case $c = \mathbf{e}$ was open as of today to the best of our knowledge. In [24], based on numerical simulations, the physicists Bauer & Golinelli predicted that $|\mathsf{KSCore}(\mathrm{G}(n, \frac{\mathbf{e}}{n}))|$ should be of order $n^{3/5}$. The main result of this work (Theorem 3.2) is to settle this conjecture in the case of a random graph with degrees $1, 2$ and 3.

Model and results. In this paper we shall consider a random graph model closely related to $\mathrm{G}(n, \frac{c}{n})$ but for which the analysis of the Karp–Sipser algorithm is simpler. Namely, we fix a sequence of numbers $\mathbf{d}^n = (d_1^n, d_2^n, d_3^n)_{n \geqslant 1}$ such that

$$n = d_1^n + 2d_2^n + 3d_3^n \text{ is even.}$$

We imagine $\mathbf{d}^n$ as the number of vertices of degree $1, 2$ and 3 and consider a random multi-graph $\mathrm{CM}(\mathbf{d}^n)$ sampled by pairing the edges emanating from the $d_1^n + d_2^n + d_3^n$ vertices uniformly at random. This is a special instance of the so-called configuration model introduced by Bollobas [40], see [170] for background. In the rest of the paper we shall further assume that

$$\frac{d_1^n}{n} \xrightarrow[n\to\infty]{} p_1, \quad \frac{2d_2^n}{n} \xrightarrow[n\to\infty]{} p_2, \quad \text{and} \quad \frac{3d_3^n}{n} \xrightarrow[n\to\infty]{} p_3, \tag{3.1}$$

so that the proportion of half-edges which are incident to a vertex of degree i is p_i. Our goal will be to analyze $\mathsf{KSCore}(\mathrm{CM}(\mathbf{d}^n))$. A phase transition has been observed in [110] for the size of the Karp–Sipser core but its location depending on (p_1, p_2, p_3) was not explicit. Our first contribution is to make this threshold precise. For a graph $\mathfrak{g}$, we will write $|\mathfrak{g}|$ for twice the number of edges of $\mathfrak{g}$, and call this quantity the *size* of $\mathfrak{g}$. If (u_n) is a sequence of positive numbers and (X_n) a sequence of random variables, we will write $X_n = O_{\mathbb{P}}(u_n)$ if $\left(u_n^{-1} X_n\right)$ is tight, and we will write $X_n = o_{\mathbb{P}}(u_n)$ if $u_n^{-1} X_n$ converges to 0 in probability.

Theorem 3.1 (*Explicit phase transition*)

Under the assumptions (3.1), let

$$\Theta = (p_3 - p_1)^2 - 4p_1. \tag{3.2}$$

- ***Subcritical phase.*** *If $\Theta < 0$, then as $n \to \infty$ we have*

$$|\mathsf{KSCore}(\mathrm{CM}(\mathbf{d}^n))| = O_{\mathbb{P}}(\log^2 n).$$

- ***Supercritical phase.*** *If $\Theta > 0$, then*

$$n^{-1} \cdot |\mathsf{KSCore}(\mathrm{CM}(\mathbf{d}^n))| \xrightarrow[n\to\infty]{(\mathbb{P})} \frac{4\Theta}{3 + \Theta}.$$

- ***Critical phase.*** *If $\Theta = 0$, then $|\mathsf{KSCore}(\mathrm{CM}(\mathbf{d}^n))| = o_{\mathbb{P}}(n)$.*

Sketch of proof of the phase transition. The proof of this theorem uses classical techniques. We shall reveal the random graph $\mathrm{CM}(\mathbf{d}^n)$ by pairing its half-edges two-by-two as we perform the Karp–Sipser leaf removal algorithm (a.k.a. peeling algorithm). More precisely, when we remove a leaf, we reveal its neighbor in the graph and remove it as well, which decreases the degrees of some other vertices. During this process, the number of remaining vertices of degree $1, 2$ and 3 evolves as an $(\mathbb{Z}_{\geqslant 0})^3$-valued Markov chain with explicit probability transitions. This is, of course, a recurrent idea in random graph theory and has already been used many times for the Karp–Sipser algorithm itself [116, 19]. More precisely, we shall erase leaves uniformly at random one-by-one (in contrast with [110], where all possible leaves are erased at each round) and use the *differential equation method* [173] to prove that the renormalized number of vertices of degree $1, 2$ and 3 is well approximated by a differential equation on $\mathbb{R}^3$ for which we are able to find explicit solutions. In a sense, this returns to the roots of this method since it was Karp & Sipser [116] who first introduced it in the context of random graphs following earlier works of Kurtz [126] in population models.

Remark (A spectral parallel to the Karp–Sipser phase transition). The *nullity* of a graph is the multiplicity of 0 in the spectrum of its adjacency matrix. It is easy to see that the leaf-removal process on a graph $\mathfrak{g}$ leaves its *nullity* invariant and so the Karp–Sipser algorithm can also be used to study the later, see [25, 158]. The phase transition for the emergence of a Karp–Sipser core of positive proportion in $\mathrm{G}(n, \frac{\mathrm{e}}{n})$ has a parallel phase transition[3] for the emergence of extended states (an absolutely continuous part) at zero in $\mathrm{G}(n, \frac{\mathrm{e}}{n})$, see [25, 70]. We wonder whether a similar result holds true for the configuration models we study.

We now turn to the detailed analysis of the *critical case* which is the main goal of our work. For this we fix a particular degree sequence $\mathbf{d}^n_{\mathrm{crit}} = (d^n_{1,c}, d^n_{2,c}, d^n_{3,c})$ such that $d^n_{1,c} + 3d^n_{3,c} = n$ is even (to be able to perform the configuration model) and

$$d^n_{1,c} = n(1 - \frac{\sqrt{3}}{2}) + O(1), \quad 2d^n_{2,c} = 0, \quad \text{and } 3d^n_{3,c} = n\frac{\sqrt{3}}{2} + O(1). \tag{3.3}$$

In particular we have $\Theta = (\sqrt{3}-1)^2 - 4(1 - \frac{\sqrt{3}}{2}) = 0$ so we are indeed in the critical case of Theorem 3.1. By definition, the core $\mathrm{KSCore}(\mathrm{CM}(\mathbf{d}^n_{\mathrm{crit}}))$ has only vertices of degrees 2 or 3. Our main result is then the following:

Theorem 3.2 (*Geometry of the critical Karp–Sipser core*)

Let $D_2(n)$ (resp. $D_3(n)$) be the total number of half-edges attached to a vertex of degree 2 (resp. 3) in $\mathrm{KSCore}(\mathrm{CM}(\mathbf{d}^n_{\mathrm{crit}}))$. Then we have

$$\begin{pmatrix} n^{-3/5} \cdot D_2(n) \\ n^{-2/5} \cdot D_3(n) \end{pmatrix} \xrightarrow[n\to\infty]{(d)} \cdot \begin{pmatrix} 3^{-3/5}2^{14/5} \cdot \vartheta^{-2} \\ 3^{-2/5}2^{16/5} \cdot \vartheta^{-3} \end{pmatrix},$$

where $\vartheta = \inf\{t \geqslant 0 : B_t = t^{-2}\}$, for a standard linear Brownian motion $(B_t : t \geqslant 0)$ started from 0. Moreover, conditionally on $(D_2(n), D_3(n))$, the graph $\mathrm{KSCore}(\mathrm{CM}(\mathbf{d}^n_{\mathrm{crit}}))$ is a configuration model.

[3] Unfortunately, this does not seem to be an easy corollary of the "geometric" phase transition.

Remark (Bauer & Golinelli's prediction). The above theorem confirms a long-standing prediction of Bauer & Golinelli [24] stated in the case of the Erdős-Rényi random graph: based on Monte-Carlo simulations they proposed a few possible sets of critical exponents [24, Table 1] and our theorem confirms their prediction. See also [99, 124] for later developments.

Note that our assumptions on the initial degree sequence are much stronger than for Theorem 3.1 since the size of the critical core is quite sensitive to initial conditions. Our proof still works if the error $O(1)$ is replaced by $O(n^{1/2})$, and the result should remain true as long as the initial error is $o(n^{3/5})$, see Section 3.5 for a discussion on the near-critical regime. Although our main result only considers the graph $\mathrm{CM}(\mathbf{d}^n_{\mathrm{crit}})$, we believe that the above limiting result holds for a large variety of random graphs which are critical for the Karp–Sipser algorithm. In particular, we expect a similar result for configuration models with bounded degrees and for the Erdős–Rényi graph $G(n, \frac{e}{n})$, but the number of vertices of degree $2 \leqslant k \leqslant 5$ in the core should be of order $n^{(5-k)/5}$. In particular, we conjecture that there are no vertices of degree 6 or more in $\mathsf{KSCore}(G(n, \frac{e}{n}))$.

Ideas of proof. The proof of Theorem 3.2 uses the same Markov chain as the one used to study the phase transition. The difference is that we need to study the behaviour of this chain right before its extinction, at a scale much finer than n. More precisely, we can expect from the differential equation approximation that εn steps before extinction, the number of vertices of unmatched degrees 1, 2 and 3 are respectively of order $\varepsilon^2 n$, εn and $\varepsilon^{3/2} n$. On the other hand, a variance computation shows that the fluctuations of the number of vertices of degree 1 are of order $\varepsilon^{3/4}\sqrt{n}$. Finally, the time at which we can expect the Markov chain to terminate is the time where the fluctuations exceed the expected value, that is at $\varepsilon = n^{-2/5}$. However, checking that the differential equation approximation remains good until that scale requires some careful control of the Markov chain accross scales. In particular, the reason why the fluctuations become much smaller than $\sqrt{n}$ in the end of the process is that the drift of our Markov chain induces a "self-correcting" effect.

Acknowledgments. The last two authors were supported by ERC 740943 GeoBrown and by ANR RanTanPlan. The first author is grateful to the Laboratoire de Mathématiques d'Orsay, where most of this work was done, for its hospitality. We warmly thank Matthieu Jonckheere for a stimulation discussion about [110] and Justin Salez for enlightening explanations about maximal matchings and independent sets in random graphs.

3.2 Karp–Sipser exploration of the configuration model

As we mentioned in the introduction the main idea (already present in [116, 19, 110, 39, 124]) is to explore the random configuration model $\mathrm{CM}(\mathbf{d}^n)$ at the same time as we run the Karp–Sipser algorithm to discover its core. Let us explain this in details. Fix a degree sequence $\mathbf{d}^n = (d_1^n, d_2^n, d_3^n)$ such that $n = d_1^n + 2d_2^n + 3d_3^n$ is even. We shall expose the $\frac{n}{2}$ edges of $\mathrm{CM}(\mathbf{d}^n)$ one by one and create a process

$$(X_k^n, Y_k^n, Z_k^n : k \geqslant 0)$$

where X^n, Y^n, Z^n represent respectively the number of unmatched half-edges linked to vertices of *unmatched degree*[4] $1, 2, 3$. The process of the sum is denoted by $S^n = X^n + Y^n + Z^n$. In particular, we always have $(X_0^n, Y_0^n, Z_0^n) = (d_1^n, 2d_2^n, 3d_3^n)$ and $S_0^n = n$ with our conventions.

As long as $X_k^n > 0$, the process evolves as follows. Since $X_k^n > 0$, there are still vertices of unmatched degree 1. We pick ℓ (for leaf) one of these vertices uniformly at random and reveal its neighbor v in the graph. Now, in the Karp–Sipser algorithm this vertex is "destroyed" so we shall erase v from the configuration *as well as the connections is has with other vertices of the graph.* More precisely, we reveal the neighbors of v in $\mathrm{CM}(\mathbf{d}^n)$ and erase all the connections we create when doing so. In particular, if v is connected to a vertex $w \neq \ell$ of unmatched degree d via i edges, then after the operation w becomes a vertex of unmatched degree $d - i$. After that, the vertices of unmatched degree 0 are simply removed. We listed all 13 combinatorial possibilities (recall that our vertices have degree $1, 2$ or 3) in Figure 3.3. The stopping time of the algorithm is

$$\theta^n := \inf\{k \geqslant 0 : X_k^n = 0\}.$$

Finally, we extend the process (X^n, Y^n, Z^n) to any k by setting $(X_k^n, Y_k^n, Z_k^n) = (X_{\theta^n}^n, Y_{\theta^n}^n, Z_{\theta^n}^n)$ for $k \geqslant \theta^n$. We denote by $(\mathcal{F}_k)_{k\geqslant 0}$ the natural filtration generated by this exploration. The starting point of our investigations is the following.

Proposition 3.1. *The process $(X_k^n, Y_k^n, Z_k^n)_{0\leqslant k\leqslant \theta^n}$ is a Markov process whose probability transitions are described in Figure 3.3. Furthermore, for any stopping time τ, the remaining pairing of the unmatched edges conditionally on $\mathcal{F}_\tau$ is uniform.*

Proof. This is standard: the above exploration procedure of $\mathrm{CM}(\mathbf{d}^n)$ is Markovian and preserves the fact that the remaining pairing of edges is uniform.

□

In particular, notice that at the stopping time θ^n, the graph made by pairing the remaining unmatched edges is precisely the Karp–Sipser core of $\mathrm{CM}(\mathbf{d}^n)$ and so the second part of Theorem 3.2 is already proved.

3.3 Phase transition *via* fluid limit of the Markov chain

In this section, we prove Theorem 3.1. The main ingredient is a deterministic fluid limit result for the Markov chain (X^n, Y^n, Z^n).

3.3.1 Fluid limit for the Markov chain

For a process indexed by discrete time $(\mathfrak{H}_k : k \geqslant 0)$ we use the notation $\Delta\mathfrak{H}_k = \mathfrak{H}_{k+1} - \mathfrak{H}_k$ for $k \geqslant 0$. Given the transitions of the Markov chain (X^n, Y^n, Z^n) the following should come as no surprise.

[4] The unmatched degree of a vertex at time k is the number of half-edges attached to this vertex which are still unmatched at time k.

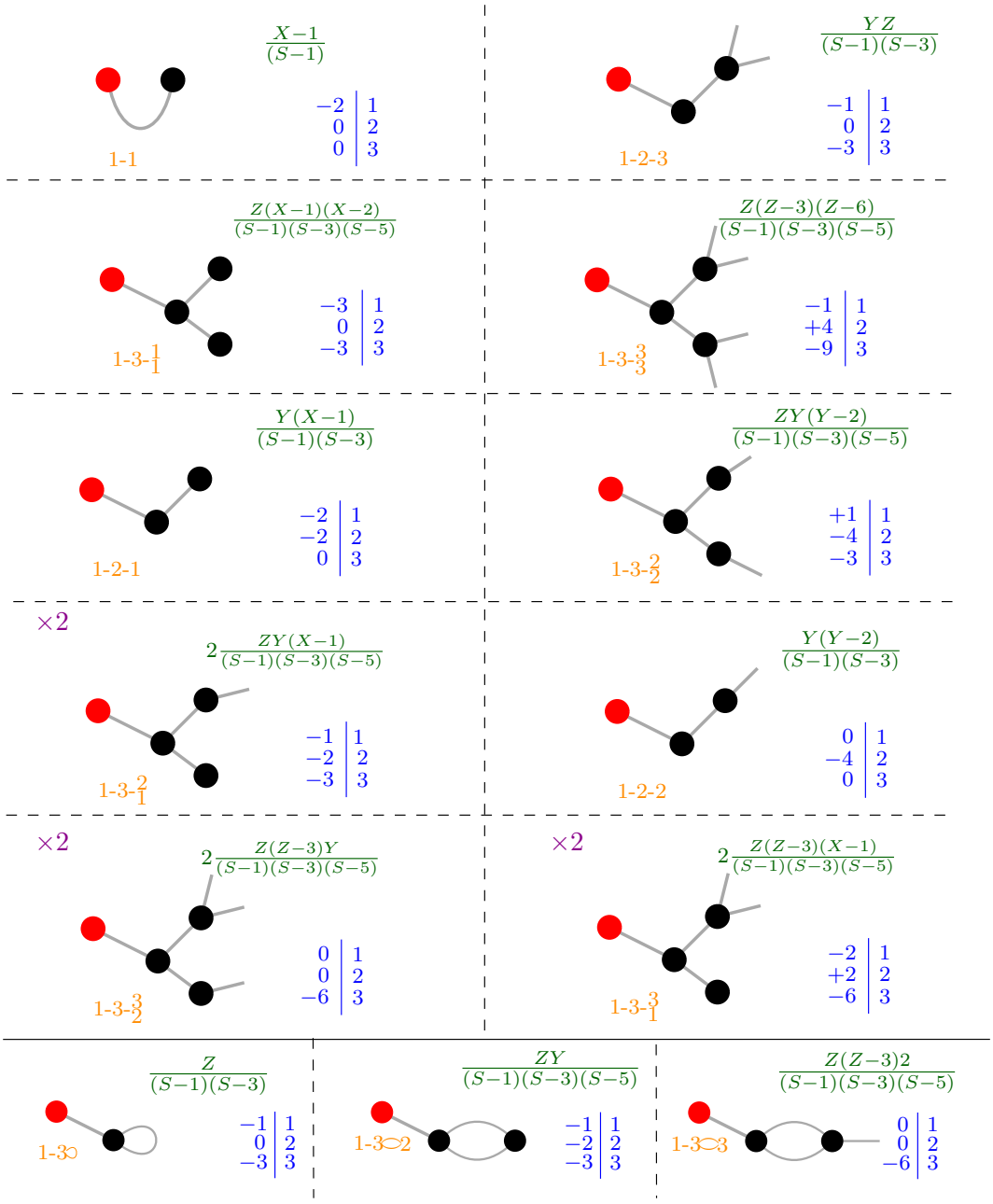


Figure 3.3: Transitions probabilities of the Markov chain (X^n, Y^n, Z^n): as long as $X^n > 0$, a vertex ℓ of degree 1 (in red above) is picked and its neighbor v is revealed. The vertices ℓ, v are then removed from the configuration model as well as the connections they created. The probability of each event is indicated in green in the upper right corner. The variation of X, Y, Z are displayed in blue. A symmetry factor is indicated when needed in purple in the upper left corner. Notice in particular that the last three cases on the bottom have probabilities of smaller order $O(1/S)$, so they will not participate to the large scale limit.

Proposition 3.2 (Fluid limit). *Suppose that* $\mathbf{d}^n = (d_1^n, d_2^n, d_3^n)$ *satisfies* (3.1). *Then we have the following convergence in probability for the uniform norm:*

$$\left(\frac{X^n_{\lfloor tn\rfloor}}{n}, \frac{Y^n_{\lfloor tn\rfloor}}{n}, \frac{Z^n_{\lfloor tn\rfloor}}{n}\right)_{0\leqslant t\leqslant \theta^n/n} \xrightarrow[n\to\infty]{(\mathbb{P})} (\mathscr{X}(t), \mathscr{Y}(t), \mathscr{Z}(t))_{0\leqslant t\leqslant t_{\mathrm{ext}}}, \tag{3.4}$$

where $(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$ *is the unique solution*[5] *to the differential equation* $(\mathscr{X}', \mathscr{Y}', \mathscr{Z}') = \phi(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$ *with* ϕ *defined below* (3.5) *with initial conditions* (p_1, p_2, p_3) *and where* t_{ext} *is the first hitting time of* 0 *by the continuous process* $\mathscr{X}$. *Moreover,* $\theta^n/n \to t_{\mathrm{ext}}$ *in probability as* $n \to \infty$.

Proof. It is a standard application of the differential equation method. Indeed, the increments of the Markov chain (X^n, Y^n, Z^n) are bounded and using the exact transitions (Figure 3.3), the conditional expected drifts

$$\mathbb{E}\left[\Delta X_k^n, \Delta Y_k^n, \Delta Z_k^n \mid \mathcal{F}_k\right]$$

converge for large values of n towards $\phi\left(\frac{X_k^n}{n}, \frac{Y_k^n}{n}, \frac{Z_k^n}{n}\right)$ where the function ϕ is defined by

$$\phi\begin{pmatrix}\mathscr{X}\\ \mathscr{Y}\\ \mathscr{Z}\end{pmatrix} = \begin{pmatrix} -2\mathbf{x} - \mathbf{yz} - 3\mathbf{x}^2\mathbf{z} - 2\mathbf{yx} + \mathbf{zy}^2 - 2\mathbf{zxy} - \mathbf{z}^3 - 4\mathbf{z}^2\mathbf{x} \\ 4\mathbf{z}^3 - 2\mathbf{xy} - 4\mathbf{zy}^2 - 4\mathbf{xyz} - 4\mathbf{y}^2 + 4\mathbf{z}^2\mathbf{x} \\ -3\mathbf{yz} - 3\mathbf{zy}^2 - 12\mathbf{z}^2\mathbf{y} - 3\mathbf{zx}^2 - 6\mathbf{xyz} - 12\mathbf{z}^2\mathbf{x} - 9\mathbf{z}^3 \end{pmatrix}, \tag{3.5}$$

$$\text{with } \mathscr{S} := \mathscr{X} + \mathscr{Y} + \mathscr{Z} \text{ and where } \begin{pmatrix}\mathbf{x}\\ \mathbf{y}\\ \mathbf{z}\end{pmatrix} := \frac{1}{\mathscr{S}}\begin{pmatrix}\mathscr{X}\\ \mathscr{Y}\\ \mathscr{Z}\end{pmatrix} \text{ is the proportion vector.} \tag{3.6}$$

For any $\delta > 0$, the convergence of the conditional expected drifts to ϕ is uniform on $\{n^{-1}\cdot S^n \geqslant \delta\}$ and $(x,y,z) \mapsto \phi(x,y,z)$ is Lipschitz on $\{(x,y,z) \in \mathbb{R}_+^3 : \delta^{-1} \geqslant x+y+z \geqslant \delta\}$ as $\nabla\phi(x,y,z)$ is of the form $\frac{P(x,y,z)}{(x+y+z)^4}$, where P is a polynomial. Therefore, by [173, Theorem 1], the equation $(\mathscr{X}', \mathscr{Y}', \mathscr{Z}') = \phi(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$ with initial condition (p_1, p_2, p_3) has a unique solution until the time t_{ext}^δ where $\mathscr{X}$ first hits δ, and the convergence (3.4) holds for $0 \leqslant t \leqslant t_{\mathrm{ext}}^\delta$. Moreover, let $t_{\mathrm{ext}} = \lim_{\delta\to 0} t_{\mathrm{ext}}^\delta$. Since ϕ is bounded by an absolute constant, the solution $(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$ is Lipschitz on $[0, t_{\mathrm{ext}})$, so we can extend it uniquely in a continuous way to $[0, t_{\mathrm{ext}}]$, and by continuity t_{ext} is indeed the first time where $\mathscr{X}$ hits 0. We know that (3.4) holds on every compact subset of $[0, t_{\mathrm{ext}})$. Moreover, the increments of (X^n, Y^n, Z^n) are bounded by an absolute constant, so the functions $n^{-1}\cdot(X^n, Y^n, Z^n)$ are uniformly Lipschitz and the previous convergence extends to a uniform convergence on $[0, t_{\mathrm{ext}}]$.

We now only need to check that $\frac{\theta^n}{n}$ converges in probability to t_{ext}. We notice that deterministically, if $k < \theta^n$, then $S_{k+1}^n \leqslant S_k^n - 2$, which implies $\theta^n \leqslant n$, so up to extraction we may assume that $\frac{\theta^n}{n}$ converges to some random variable $\widetilde{t}_{\mathrm{ext}}$. By convergence of the process and the definition of t_{ext}, it is immediate that $\widetilde{t}_{\mathrm{ext}} \geqslant t_{\mathrm{ext}}$. For the other direction, we treat two cases separately:

[5] More precisely, by *solution*, we mean that $(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$ is a continuous function from $[0, t_{\mathrm{ext}}]$ to $\mathbb{R}^3$ such that $\mathscr{X}$ first hits 0 at time t_{ext} and $(\mathscr{X}'(t), \mathscr{Y}'(t), \mathscr{Z}'(t)) = \phi(\mathscr{X}(t), \mathscr{Y}(t), \mathscr{Z}(t))$ for all $0 \leqslant t < t_{\mathrm{ext}}$.

- if $\mathscr{S}(t_{\text{ext}}) = 0$, then let $\varepsilon > 0$, and let $\delta > 0$ be such that $\mathscr{S}(t_{\text{ext}} - \delta) < \varepsilon$. With probability $1 - o(1)$ as $n \to +\infty$, we have $S^n_{\lfloor (t_{\text{ext}}-\delta)n \rfloor} < 2\varepsilon n$. Since S^n decreases by at least two at each step, this implies $\theta^n \leqslant (t_{\text{ext}} - \delta)n + \varepsilon n$, so $\widetilde{t}_{\text{ext}} \leqslant t_{\text{ext}}$.
- if $\mathscr{S}(t_{\text{ext}}) > 0$, we first argue that the first component of $\phi(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$ remains bounded from above by a negative constant along the whole trajectory. Indeed, since $\mathscr{S}$ is bounded from below, we have $\mathscr{Z}' \geqslant -c\mathscr{Z}$ for some constant c along the trajectory. Hence $\mathscr{Z}$ is bounded from below by a positive constant on $[0, t_{\text{ext}}]$, so $\mathbf{y}$ is bounded away from 1. Since the first component of $\phi(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$ is at most $-\mathbf{yz} + \mathbf{y}^2\mathbf{z} = -\mathbf{yz}(1 - \mathbf{y})$, this proves our claim. Therefore, with high probability, the conditional expected drift $\mathbb{E}\left[\Delta X^n_k \mid X^n_k\right]$ is also bounded from above by a negative constant $-c$ along the trajectory. Since the increments are bounded, by the weak law of large numbers this ensures $\widetilde{t}_{\text{ext}} \leqslant t^{\varepsilon}_{\text{ext}} + \frac{1}{c}\varepsilon$ for all $\varepsilon > 0$, so $\widetilde{t}_{\text{ext}} = t_{\text{ext}}$.

□

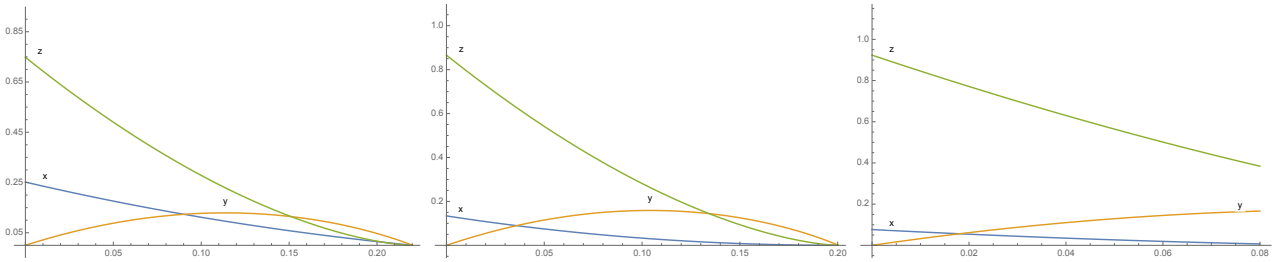


Figure 3.4: Illustration of the differential system $(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$ in terms of "number of legs" in the subcritical (left), critical (center) and supercritical (right) cases.

3.3.2 Solving the differential equation

In this section, our goal will be to gather information about the solutions to (3.5), which will give Theorem 3.1 and be an important tool in the proof of Theorem 3.2. As indicated by the system (3.5), we will see that the solutions are easier to express in terms of proportions. We refer to Figures 3.4 and 3.5 for a visualization of the trajectories of these solutions.

Proposition 3.3. *We fix $p_1 > 0$ and $p_2, p_3 \geqslant 0$ with $p_1 + p_2 + p_3 = 1$. Let $(\mathscr{X}(t), \mathscr{Y}(t), \mathscr{Z}(t))_{0 \leqslant t \leqslant t_{\text{ext}}}$ be the solution to (3.5) with initial condition (p_1, p_2, p_3). Recall from (3.2) the definition*

$$\Theta = (p_3 - p_1)^2 - 4p_1 \in [-3, 1].$$

- *If $\Theta < 0$ (subcritical case), then $\mathscr{X}(t_{\text{ext}}) = \mathscr{Y}(t_{\text{ext}}) = \mathscr{Z}(t_{\text{ext}}) = 0$. Moreover, for $t < t_{\text{ext}}$ sufficiently close to t_{ext}, we have $\mathscr{Z}(t) < \mathscr{X}(t)$.*
- *If $\Theta > 0$ (supercritical case), then*

$$\mathscr{X}(t_{\text{ext}}) = 0, \quad \mathscr{Y}(t_{\text{ext}}) = \frac{4\Theta}{3 + \Theta}\left(1 - \sqrt{\Theta}\right) > 0, \quad \text{and} \quad \mathscr{Z}(t_{\text{ext}}) = \frac{4\Theta^{3/2}}{3 + \Theta} > 0. \tag{3.7}$$

- *If* $\Theta = 0$ *(critical case), then* $\mathscr{X}(t_{\text{ext}}) = \mathscr{Y}(t_{\text{ext}}) = \mathscr{Z}(t_{\text{ext}}) = 0$*, and more precisely as* $\varepsilon \to 0$*:*

$$\begin{cases} \mathscr{X}(t_{\text{ext}} - \varepsilon) & \sim & 3\varepsilon^2, \\ \mathscr{Y}(t_{\text{ext}} - \varepsilon) & \sim & 4\varepsilon, \\ \mathscr{Z}(t_{\text{ext}} - \varepsilon) & \sim & 4\sqrt{3}\varepsilon^{3/2}. \end{cases} \tag{3.8}$$

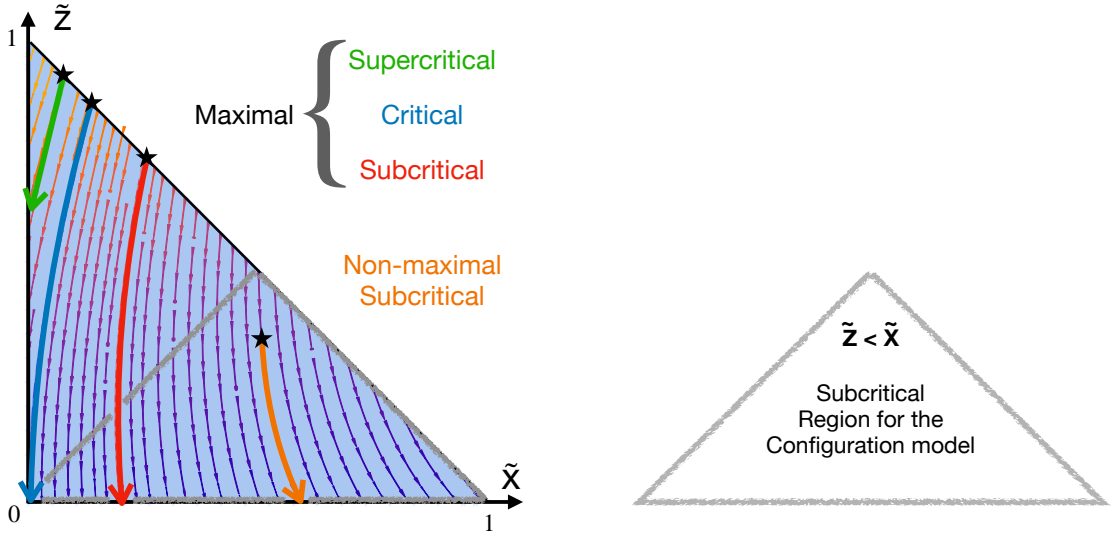


Figure 3.5: Illustration of the differential system $\tilde{\mathbf{x}}, \tilde{\mathbf{z}}$ with the vector field. The maximal solutions start from $\tilde{\mathbf{x}}(0) + \tilde{\mathbf{z}}(0) = 1$. A maximal supercritical (resp. critical, resp. subcritical) solution is shown in green (resp. blue, resp. red). A non-maximal subcritical solution is displayed in orange. Note that any subcritical solution terminates in the gray region which is subcritical for the configuration model itself.

Proof. We will first obtain an explicit (up to time-change) solution to (3.5). We recall that $\mathscr{S} = \mathscr{X} + \mathscr{Y} + \mathscr{Z}$ is the fluid limit of the sum process and that $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are the proportions whose sum is constant and equal to 1.

Using $\mathbf{y} = 1 - \mathbf{x} - \mathbf{z}$, the system (3.5) translates into the following system on $\mathbf{x}, \mathbf{z}$ and $\mathscr{S}$:

$$\begin{cases} \mathbf{x}' & = & \frac{1}{\mathscr{S}}(\mathbf{x} - \mathbf{z})\mathbf{z}, \\ \mathbf{z}' & = & \frac{1}{\mathscr{S}}(-2 + \mathbf{x} - \mathbf{z})\mathbf{z}, \\ \mathscr{S}' & = & 2(-2 + \mathbf{x} - \mathbf{z}), \end{cases} \tag{3.9}$$

where again $\mathscr{S}(0) = 1$ and $\mathbf{x}(0), \mathbf{z}(0) \geqslant 0$ satisfy $\mathbf{x}(0) + \mathbf{z}(0) \leqslant 1$.

In order to get rid of $\mathscr{S}$ in this system, we perform a time change: for $t \in [0, t_{\text{ext}}]$, we write

$$\gamma(t) = \int_0^t \frac{\mathrm{d}s}{\mathscr{S}(s)} \in [0, +\infty].$$

We also define the functions $\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}}$ on $[0, u_{\text{ext}}]$, with $u_{\text{ext}} = \int_0^{t_{\text{ext}}} \frac{\mathrm{d}s}{\mathscr{S}(s)}$, by $\tilde{\mathbf{x}}(u) = \mathbf{x}\left(\gamma^{-1}(u)\right)$ and $\tilde{\mathbf{z}}(u) = \mathbf{z}\left(\gamma^{-1}(u)\right)$. We obtain the system

$$\begin{cases} \tilde{\mathbf{x}}' &= (\tilde{\mathbf{x}} - \tilde{\mathbf{z}})\tilde{\mathbf{z}}, \\ \tilde{\mathbf{z}}' &= (-2 + \tilde{\mathbf{x}} - \tilde{\mathbf{z}})\tilde{\mathbf{z}}. \end{cases}$$

We find solutions to this system as follows: by subtracting the second line to the first one, we have $\tilde{\mathbf{x}}' - \tilde{\mathbf{z}}' = 2\tilde{\mathbf{z}}$ and the second line implies that $\tilde{\mathbf{x}} - \tilde{\mathbf{z}} = \left(\frac{\tilde{\mathbf{z}}'}{\tilde{\mathbf{z}}} + 2\right)$. Deriving the second identity and comparing, we deduce the following second-order non-linear one-dimensional differential equation:

$$2\left(\tilde{\mathbf{z}}\right)^3 = \tilde{\mathbf{z}}''\tilde{\mathbf{z}} - \left(\tilde{\mathbf{z}}'\right)^2.$$

A complete family of solutions is given by

$$\begin{cases} \tilde{\mathbf{z}}(u) &= \dfrac{b^2}{\sinh(b(u+u_0))^2}, \\ \tilde{\mathbf{x}}(u) &= \left(\dfrac{b}{\tanh(b(u+u_0))} - 1\right)^2 + 1 - b^2, \\ \tilde{\mathbf{y}}(u) &= \dfrac{-2b^2}{\tanh^2(b(u+u_0))} + \dfrac{2b}{\tanh(b(u+u_0))} + 2b^2 - 1, \end{cases} \tag{3.10}$$

where $b, u_0 \in \mathbb{R}$. We notice that along these solutions, the quantity $(\tilde{\mathbf{z}} - \tilde{\mathbf{x}})^2 - 4\tilde{\mathbf{x}}$ is constant, and is equal to $4(b^2 - 1)$, this quantity is equal to the Θ defined by (3.2):

$$(\tilde{\mathbf{z}} - \tilde{\mathbf{x}})^2 - 4\tilde{\mathbf{x}} \equiv 4(b^2 - 1) = (p_3 - p_1)^2 - 4p_1 = \Theta. \tag{3.11}$$

We also notice that $\tilde{\mathbf{y}}$ is always increasing and that $\tilde{\mathbf{y}} < 0$ for u small enough, which has no meaning in our context. Therefore, every solution is contained in a *maximal* solution, i.e. a solution where the initial condition (p_1, p_2, p_3) satisfies $p_2 = 0$. Since we are only interested in the behavior near extinction and since the right-hand side of the formulas (3.7) depends only on b, we may restrict ourselves to maximal solutions, i.e. assume $p_2 = 0$. Other solutions can be deduced from this by a time shift in $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}})$, which translates into a time shift in $(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$. From $\tilde{\mathbf{y}}(0) = 0$, we get

$$u_0 = \frac{1}{2b} \log\left(1 + 2b + \frac{2b\sqrt{(4b^2 - 1)}}{2b - 1}\right) > 0,$$

so $p_1 = 1 - \frac{1}{2}\sqrt{4b^2 - 1}$ and $p_3 = \frac{1}{2}\sqrt{4b^2 - 1}$.

We now come back to the true solutions $(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$ in each of the three cases of Proposition 3.3. For this, we need to study the time change $\gamma : [0, t_{\text{ext}}] \to [0, u_{\text{ext}}]$. By definition of γ and the third line of (3.9), for all $t \in [0, t_{\text{ext}})$, we have

$$\begin{cases} \frac{1}{\mathscr{S}(t)} &= \gamma'(t), \\ \mathscr{S}'(t) &= 2(-2 + \tilde{\mathbf{x}}(\gamma(t)) - \tilde{\mathbf{z}}(\gamma(t)). \end{cases}$$

Multiplying those lines and integrating both sides using the exact expressions of $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{z}}$, we find $\frac{\mathrm{d}}{\mathrm{d}t}\log \mathscr{S}(t) = -4\frac{\mathrm{d}}{\mathrm{d}t}\log\left(\sinh(b \cdot (\gamma(t) + u_0))\right)$ so the following quantity is constant:

$$\mathscr{S}(t)\sinh^4\left(b \cdot (\gamma(t) + u_0)\right) = \mathscr{S}(t)\left(\frac{b^2}{\mathbf{z}(t)}\right)^2 = \frac{b^4}{\tilde{\mathbf{z}}(0)^2} = \frac{4b^4}{4b^2 - 1}. \tag{3.12}$$

Note that this last equation, combined with the expression of $\mathscr{S}'(t)$, provides a differential equation satisfied by $\mathscr{S}$, from which we could express $\mathscr{S}$ as the inverse bijection of an explicit function. However, this will not be needed in the proof. Given those findings, the rest of the proof is made of easy calculations. Let us proceed. We refer to Figures 3.4 and 3.5 for visualization of the system in terms of proportions or in "number of legs".

Subcritical regime. For $\Theta < 0$, we have $\frac{1}{2} < b < 1$. In this case, we observe that $\tilde{\mathbf{x}}(u) \geqslant 1 - b^2$ is bounded away from 0, so the same is true for $\mathbf{x}(t)$ on $[0, t_{\text{ext}}]$. It follows that $\mathscr{S}(t_{\text{ext}}) = \frac{\mathscr{X}(t_{\text{ext}})}{\mathbf{x}(t_{\text{ext}})} = 0$. Therefore, by (3.12), we have $\mathbf{z}(t_{\text{ext}}) = \sqrt{\left(\frac{4b^2-1}{4}\right)\mathscr{S}(t_{\text{ext}})} = 0$. In particular, for t sufficiently close to t_{ext}, we have $\mathbf{z}(t) < \mathbf{x}(t)$, so $\mathscr{Z}(t) < \mathscr{X}(t)$. Note that this also implies $u_{\text{ext}} = +\infty$.

Supercritical regime. For $\Theta > 0$, we have $1 < b < \frac{\sqrt{5}}{2}$. In this case, the function $\tilde{\mathbf{x}}$ first hits 0 at time

$$\hat{u}_{\text{ext}} = -u_0 + \frac{1}{b}\operatorname{Arccoth}\frac{1+\sqrt{b^2-1}}{b}.$$

This implies that $u_{\text{ext}} \leqslant \hat{u}_{\text{ext}}$. We claim that we have equality. Indeed, if this is not the case, we have $\mathbf{x}(t_{\text{ext}}) = \tilde{\mathbf{x}}(u_{\text{ext}}) > 0$, so $\mathscr{S}(t_{\text{ext}}) = 0$, so (3.12) implies $\tilde{\mathbf{z}}(u_{\text{ext}}) = 0$ with $u_{\text{ext}} < +\infty$, which is not possible given the explicit expression of $\tilde{\mathbf{z}}$. Therefore, we have $\tilde{\mathbf{x}}(u_{\text{ext}}) = 0$. Using (3.11) we can compute

$$\mathbf{z}(t_{\text{ext}}) = \tilde{\mathbf{z}}(u_{\text{ext}}) = 2\sqrt{b^2-1} \quad \text{and} \quad \mathbf{y}(t_{\text{ext}}) = 1 - 2\sqrt{b^2-1}$$

and finally, using (3.12):

$$\mathscr{S}(t_{\text{ext}}) = \frac{4}{4b^2-1}\tilde{\mathbf{z}}(u_{\text{ext}}) = \frac{16(b^2-1)}{4b^2-1},$$

which, once translated in terms of Θ, gives (3.7).

Critical regime. For $\Theta = 0$, the maximal solution starts from $p_3 = \frac{\sqrt{3}}{2}$ and $p_1 = 1 - \frac{\sqrt{3}}{2}$, and we have $b = 1$. In particular, using $u_0 \geqslant 0$, we have $\tilde{\mathbf{x}}(u) > 0$ for all $u \geqslant 0$. By the same argument as in the supercritical regime, this implies $u_{\text{ext}} = +\infty$. Therefore, by the exact expression of $\tilde{\mathbf{x}}, \tilde{\mathbf{y}}, \tilde{\mathbf{z}}$, as $t \to t_{\text{ext}}$, we have $\mathbf{x}(t), \mathbf{z}(t) \to 0$ and $\mathbf{y}(t) \to 1$. Therefore, by (3.12) at $t = t_{\text{ext}}$, we have $\mathscr{S}(t_{\text{ext}}) = 0$, so $\mathscr{Y}(t_{\text{ext}}) = \mathscr{Z}(t_{\text{ext}}) = 0$.

Hence, letting $t \to t_{\text{ext}}$ in the third equation of (3.9), we have $\mathscr{S}'(t) \to -4$ as $t \to t_{\text{ext}}$, so $\mathscr{S}(t_{\text{ext}} - \varepsilon) \sim 4\varepsilon$ as $\varepsilon \to 0$. Injecting this in (3.12), we find $\mathbf{z}(t_{\text{ext}} - \varepsilon) \sim \sqrt{3\varepsilon}$, so $\mathscr{Z}(t_{\text{ext}} - \varepsilon) \sim 4\sqrt{3}\varepsilon^{3/2}$. Finally, we know from (3.11) that $(\mathbf{z} - \mathbf{x})^2 - 4\mathbf{x}$ is constant equal to 0, so $\mathbf{x}(t_{\text{ext}} - \varepsilon) \sim \frac{1}{4}\mathbf{z}(t_{\text{ext}} - \varepsilon)^2 \sim \frac{3}{4}\varepsilon$, which gives the asymptotics for $\mathscr{X}$.

□

3.3.3 Phase transition: proof of Theorem 3.1

Subcritical regime. We assume that (p_1, p_2, p_3) is subcritical, and consider the associated solution $(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$ to the differential equation. By Proposition 3.3, let $t_1 < t_{\text{ext}}$ be such that $\mathscr{Z}(t_1) < \mathscr{X}(t_1)$. By Proposition 3.2, we have

$$\frac{1}{n}\left(X^n_{\lfloor t_1 n\rfloor}, Y^n_{\lfloor t_1 n\rfloor}, Z^n_{\lfloor t_1 n\rfloor}\right) \xrightarrow[n\to+\infty]{(\mathbb{P})} (\mathscr{X}(t_1), \mathscr{Y}(t_1), \mathscr{Z}(t_1))\,.$$

Moreover, by Proposition 3.1, conditinally on $\mathcal{F}_{\lfloor t_1 n\rfloor}$, the remaining graph after $\lfloor t_1 n\rfloor$ steps of the Karp–Sipser algorithm is a configuration model with respectively $X^n_{\lfloor t_1 n\rfloor}$, $Y^n_{\lfloor t_1 n\rfloor}$ and $Z^n_{\lfloor t_1 n\rfloor}$ half-edges belonging to vertices of degree 1, 2 and 3. Since $n^{-1}Z^n_{\lfloor t_1 n\rfloor} \approx \mathscr{Z}(t_1) < \mathscr{X}(t_1) \approx n^{-1}X^n_{\lfloor t_1 n\rfloor}$ this is a *subcritical* configuration model (do not confuse with subcriticality in terms of the Karp–Sipser core). In particular, by [142, Theorem 1.b] there is a constant $c = c(p_1, p_2, p_3)$ such that with high probability the remaining subgraph after $\lfloor t_1 n\rfloor$ steps has fewer than $c\log(n)$ cycles and all of its connected components have size at most $c\log(n)$. On the other hand, by construction, the Karp–Sipser core is included in the union of all the cycles of $G^n_{\lfloor t_1 n\rfloor}$, so it has size $O_{\mathbb{P}}(\log^2 n)$.

Remark (True size of the subcritical KS-core). The above bound $O_{\mathbb{P}}(\log^2 n)$ for the size of the subcritical Karp–Sipser core is very crude towards the end of the proof. We expect the actual order of magnitude of the KS-core to be $O_{\mathbb{P}}(1)$ as in the Erdős–Rényi case [19].

Critical and supercritical regime. In this case, combining Proposition 3.2 and our explicit computations of the solutions, we obtain that $(X^n/S^n, Y^n/S^n, Z^n/S^n, n^{-1}\cdot S^n)(\theta^n)$ converges to

$$(\mathbf{x}(t_{\text{ext}}), \mathbf{y}(t_{\text{ext}}), \mathbf{z}(t_{\text{ext}}), \mathscr{S}(t_{\text{ext}})) = \left(0, 1-2\sqrt{b^2-1}, 2\sqrt{b^2-1}, \frac{16(b^2-1)}{4b^2-1}\right).$$

In particular the number of half-edges of the Karp-Sipser core is equal to $S^n_{\theta^n} = Y^n_{\theta^n} + Z^n_{\theta^n}$, so it is asymptotically $o_{\mathbb{P}}(n)$ if $b = 1$ (critical case). If $b > 1$, it is linear in n, which concludes the proof of Theorem 3.1 after a quick computation.

3.4 Analysis of the critical case

In this section, we shall prove our main result Theorem 3.2. In the rest of the paper, we shall thus suppose that the initial conditions (3.3) are in force. Let us first explain the heuristics to help the reader follow the proof. We refer to Figure 3.6 for an illustration.

We have seen above that in the critical regime, the asymptotic size of the Karp-Sipser core is $o_{\mathbb{P}}(n)$ and that almost all vertices have degree 2 (i.e. with density 1 since $\mathbf{y}(t_{\text{ext}}) = 1$). Recall that the process stops at time

$$\theta^n = \inf\{k \geqslant 0 : X^n_k = 0\},$$

which by Proposition 3.2 is $\approx t_{\text{ext}}\cdot n$. To analyse this stopping time and understand the size of the KS-core, we need to be more precise in the analysis of the fluctuations of the process (X^n, Y^n, Z^n) around

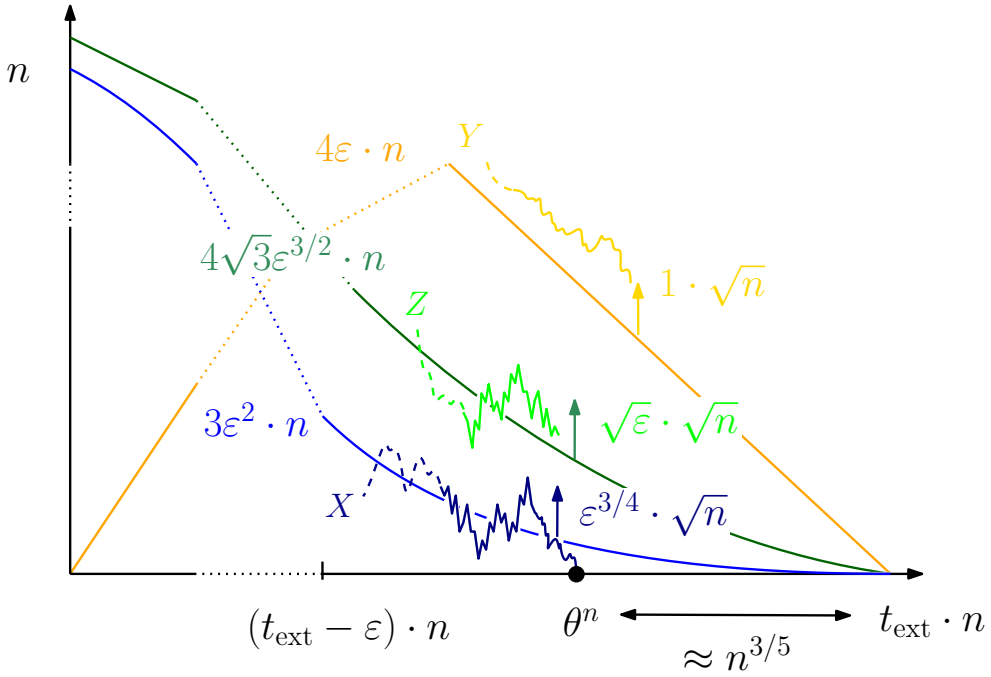


Figure 3.6: Heuristics for the proof of Theorem 3.2. The variations of the processes (X, Y, Z) around its deterministic fluid limit when $k = (t_{\text{ext}} - \varepsilon_k)n$ are displayed above. In particular, in the case of X, the number of degree 1 vertices, those variations may cause X to touch 0 when $\varepsilon_k \approx n^{-2/5}$ so that there are $\varepsilon_k n \approx n^{3/5}$ vertices of degree 2 and $\varepsilon_k^{3/2} n \approx n^{2/5}$ vertices of degree 3 remaining in the graph.

its fluid limit $n \cdot (\mathscr{X}, \mathscr{Y}, \mathscr{Z})$. To this end, we define the fluctuations processes $(A_k^n, B_k^n, C_k^n)_{0 \leqslant k \leqslant \theta^n}$ by

$$\begin{cases} X_k^n &= n\mathscr{X}\left(\frac{k}{n}\right) + A_k^n \\ Y_k^n &= n\mathscr{Y}\left(\frac{k}{n}\right) + B_k^n \\ Z_k^n &= n\mathscr{Z}\left(\frac{k}{n}\right) + C_k^n \end{cases}$$

To simplify notation, the n in the exponent will be implicit for the rest of the paper when there is no ambiguity, even if we will often look at the asymptotic as n goes to infinity.

When we are sufficiently far from the end of the process, i.e. when $k \approx tn$ for $0 \leqslant t < t_{\text{ext}}$ we know from Proposition 3.2 that (X, Y, Z) is well approximated by $n \cdot (\mathscr{X}, \mathscr{Y}, \mathscr{Z})$ and classical results (see Lemma 3.2) will show that the fluctuations A, B and C renormalized by a factor $1/\sqrt{n}$ converge to Gaussian variables whose variances depend on t. To analyse the algorithm towards the end we will use the notation, for $0 \leqslant k \leqslant (t_{\text{ext}} n) \wedge \theta$,

$$\boxed{\varepsilon_k := t_{\text{ext}} - \frac{k}{n} \geqslant 0 \qquad \text{so that} \qquad k = (t_{\text{ext}} - \varepsilon_k)n.} \tag{3.13}$$

Notice the bold font for $\boldsymbol{\varepsilon}$ to avoid confusion. Recall from Equation (3.8) that $\mathscr{X}(k/n), \mathscr{Y}(k/n)$ and $\mathscr{Z}(k/n)$ are of order respectively $\varepsilon_k^2, \varepsilon_k$ and $\varepsilon_k^{3/2}$. We will see below that the order of magnitude of

$n^{-1/2} \cdot A_k, n^{-1/2} \cdot B_k$ and $n^{-1/2} \cdot C_k$ are respectively $\varepsilon_k^{3/4}, 1$ and $\varepsilon_k^{1/2}$. In particular, the fluctuations A of X become of the same order of magnitude as its deterministic approximation $n\mathscr{X}$ when

$$n\varepsilon_k^2 \approx n\mathscr{X}(t_{\text{ext}} - \varepsilon_k) \approx A_k \approx \sqrt{n} \cdot \varepsilon_k^{3/4} \quad \text{i.e. when} \quad \varepsilon_k \approx n^{-2/5} \iff nt_{\text{ext}} - k \approx n^{3/5},$$

and this explains heuristically why $\theta_n = t_{\text{ext}} n + O(n^{3/5})$ and why the size of the Karp-Sipser core is given essentially by $Y_{\theta_n} \approx n^{3/5}$. The rest of this section makes those heuristic rigorous and proves our main result Theorem 3.2.

We first provide estimations of the conditional expected drifts and variances of the increments of the fluctuation processes (A, B, C) in Propositions 3.4 and 3.5. These propositions support the above heuristics and lead us to introduce the renormalized fluctuations processes

$$\widetilde{A}_k = \frac{A_k}{\varepsilon_k^{3/4}\sqrt{n}}, \qquad \widetilde{B}_k = \frac{B_k}{\sqrt{n}}, \quad \text{and} \quad \widetilde{C}_k = \frac{C_k}{\varepsilon_k^{1/2}\sqrt{n}},$$

which, at least heuristically, should be tight in k. After that, our proof consists in two main steps. First we will show that with high probability as $n \to \infty$, we can bound –with some log's– the process $(\widetilde{A}, \widetilde{B}, \widetilde{C})$ up to time $O(n^{3/5})$ before $t_{\text{ext}} n$, see Proposition 3.6. To do so we will extensively use the fact that for $\widetilde{C}$ and $\widetilde{A}$, the conditional expected drifts tend "to pull them back to 0" so that the processes remain small over all scales. Finally, in a second step, we will show that when $k = nt_{\text{ext}} - tn^{3/5}$ for $x \in \mathbb{R}$ the fluctuation process $\widetilde{A}$ is well approximated by a stochastic differential equation, see Proposition 3.8. The fluctuations $\widetilde{B}$ and $\widetilde{C}$ are, at this scale, still negligible in front of their differential method approximation.

3.4.1 Drift and variance estimates

In this section we compute the conditional expected drift and variance of the fluctuations processes A, B, C. Recall the very important notation ε_k introduced in (3.13). As explained above, it will turn out that $\theta \equiv \theta^n$ is located around $t_{\text{ext}} n - O(n^{3/5})$ and in the forthcoming Propositions 3.4 and 3.5 we shall allow a little room and only look at times $k < \theta$ such that $\varepsilon_k \geqslant n^{-2/5-1/100}$ (and indeed the fraction $1/100$ is somehow arbitrary). We thus put

$$\tilde{\theta} = \theta \wedge \left(t_{\text{ext}} n - n^{3/5-1/100} \right). \tag{3.14}$$

Recall from above the notation

$$\widetilde{A}_k := \frac{X_k - n\mathscr{X}\left(\frac{k}{n}\right)}{\varepsilon_k^{3/4}\sqrt{n}} \quad \widetilde{B}_k := \frac{Y_k - n\mathscr{Y}\left(\frac{k}{n}\right)}{\sqrt{n}} \quad \text{and} \quad \widetilde{C}_k := \frac{Z_k - n\mathscr{Z}\left(\frac{k}{n}\right)}{\varepsilon_k^{1/2}\sqrt{n}}.$$

Recall also that $\mathcal{F}_k$ is the σ-algebra generated by $(X_i, Y_i, Z_i)_{0 \leqslant i \leqslant k}$. We have chosen the normalization so that the processes $\widetilde{A}_k, \widetilde{B}_k$ and $\widetilde{C}_k$ are of order 1 and fluctuate at the time-scale $\varepsilon_k n$, which is why the conditional expected drift and variances are all of order $\frac{1}{\varepsilon_k n}$.

Proposition 3.4 (Drift estimates). *There exists a constant $K > 0$ such that for all $\delta > 0$, there is $\eta \equiv \eta(\delta) > 0$ such that the following holds for n large enough. For any $(t_{\text{ext}} - \eta)n \leqslant k < \tilde{\theta}$, if we have $|\widetilde{A}_k|, |\widetilde{B}_k|, |\widetilde{C}_k| < 1000 \log n$ then:*

$$\left| \mathbb{E}\left[\Delta \widetilde{A}_k | \mathcal{F}_k\right] + \frac{1}{\varepsilon_k n}\frac{1}{4}\widetilde{A}_k \right| \leqslant \frac{\delta}{\varepsilon_k n}|\widetilde{A}_k| + \frac{K\varepsilon_k^{1/4}}{\varepsilon_k n}\max\left(|\widetilde{B}_k|, |\widetilde{C}_k|\right) + \frac{K}{\varepsilon_k n}n^{-1/30}, \tag{3.15}$$

$$\left| \mathbb{E}\left[\Delta \widetilde{B}_k | \mathcal{F}_k\right] \right| \leqslant \frac{K}{\varepsilon_k n}\sqrt{\varepsilon_k}\max\left(|\widetilde{A}_k|, |\widetilde{B}_k|, |\widetilde{C}_k|\right) + \frac{K}{\varepsilon_k n}n^{-1/30}, \tag{3.16}$$

$$\left| \mathbb{E}\left[\Delta \widetilde{C}_k | \mathcal{F}_k\right] - \frac{1}{\varepsilon_k n}\left(\frac{3\sqrt{3}}{2}\widetilde{B}_k - \widetilde{C}_k\right) \right| \leqslant \frac{\delta}{\varepsilon_k n}\max\left(|\widetilde{B}_k|, |\widetilde{C}_k|\right) + \frac{K}{\varepsilon_k n}\varepsilon_k^{3/4}|\widetilde{A}_k| + \frac{K}{\varepsilon_k n}n^{-1/30}. \tag{3.17}$$

Proposition 3.5 (Variance estimates). *There exists a constant K such that for all $\delta > 0$, there is $\eta \equiv \eta(\delta) > 0$ such that the following holds for n large enough. For any $(t_{\text{ext}} - \eta)n \leqslant k < \tilde{\theta}$, if we have $|\widetilde{A}_k|, |\widetilde{B}_k|, |\widetilde{C}_k| < 1000 \log n$ then:*

$$\left| \mathrm{Var}\left(\Delta \widetilde{A}_k | \mathcal{F}_k\right) - \frac{2\sqrt{3}}{\varepsilon_k n} \right| \leqslant \frac{\delta}{\varepsilon_k n} + \frac{K}{\varepsilon_k n}n^{-1/30} + \frac{K\varepsilon_k^{1/2}}{\varepsilon_k n}\widetilde{A}_k^2 + \frac{K}{n}\max\left(\widetilde{B}_k^2, \widetilde{C}_k^2\right), \tag{3.18}$$

$$\mathrm{Var}\left(\Delta \widetilde{A}_k | \mathcal{F}_k\right) \leqslant \frac{2\sqrt{3} + \delta}{\varepsilon_k n} + \frac{K}{\varepsilon_k n}n^{-1/30}, \tag{3.19}$$

$$\mathrm{Var}\left(\Delta \widetilde{B}_k | \mathcal{F}_k\right) \leqslant \frac{K\varepsilon_k}{\varepsilon_k n} \tag{3.20}$$

$$\mathrm{Var}\left(\Delta \widetilde{C}_k | \mathcal{F}_k\right) \leqslant \frac{K\varepsilon_k^{1/2}}{\varepsilon_k n}. \tag{3.21}$$

The proofs of the above two propositions follow by examining precisely the probability transitions of the Markov chain (X, Y, Z) given by Figure 3.3 and basic (though important) analysis of the behavior of the function ϕ (defined by (3.5)) and its gradient $\nabla\phi$ near t_{ext}. Let us start with a deterministic lemma based on (3.8) controlling X, Y, Z from the processes $\widetilde{A}, \widetilde{B}, \widetilde{C}$:

Lemma 3.1. *There are absolute constants $C, c > 0$ such that if $|\widetilde{A}_k|, |\widetilde{B}_k|, |\widetilde{C}_k| < 1000 \log n$ and $X_k > 0$ and $\varepsilon_k \in [n^{-2/5-1/100}, \eta]$, for n large enough we have*

$$X_k \leqslant C\varepsilon_k^2 n \times n^{1/100}, \quad Y_k \leqslant C\varepsilon_k n, \quad Z_k \leqslant C\varepsilon_k^{3/2} n$$

and

$$S_k \geqslant Y_k \geqslant c\varepsilon_k n.$$

Proof. Recall the asymptotics (3.8). We simply write

$$X_k \leqslant n\mathscr{X}\left(\frac{k}{n}\right) + \varepsilon_k^{3/4}\sqrt{n}\widetilde{A}_k \underset{(3.8)}{\leqslant} C'\varepsilon_k^2 n + 1000\varepsilon_k^{3/4}\sqrt{n}\log n.$$

The assumption $k \leqslant t_{\text{ext}}n - n^{3/5-1/100}$, i.e. $\varepsilon_k \geqslant n^{-2/5-1/100}$, implies that the second term is $O\left(\varepsilon_k^2 n \times n^{1/100}\right)$. The other two upper bounds can be proved in the same way. Finally, we have

$$S_k \geqslant Y_k = n\mathscr{Y}(k/n) + \sqrt{n}\widetilde{B}_k \geqslant c'\varepsilon_k n - 1000\sqrt{n}\log n,$$

which is enough to prove the lower bound on S_k since $\varepsilon_k n \geqslant n^{3/5-1/100}$ is much larger than $\sqrt{n}\log n$ if n is large enough.

□

Proof of Proposition 3.4. Recall the definition of ϕ in (3.5) given in terms of proportions, so that using the notation $s = x+y+z$ we have

$$\begin{aligned}
\phi_X(x,y,z) &= -2\frac{x}{s} - \frac{yz}{s^2} - 3\frac{x^2z}{s^3} - 2\frac{xy}{s^2} + \frac{y^2z}{s^3} - 2\frac{xyz}{s^3} - \frac{z^3}{s^3} - 4\frac{xz^2}{s^3},\\
\phi_Y(x,y,z) &= 2\left(2\frac{z^3}{s^3} - \frac{xy}{s^2} - 2\frac{y^2z}{s^3} - 2\frac{xyz}{s^3} - 2\frac{y^2}{s^2} + 2\frac{xz^2}{s^3}\right),\\
\phi_Z(x,y,z) &= 3\left(-\frac{yz}{s^2} - \frac{y^2z}{s^3} - 4\frac{yz^2}{s^3} - \frac{x^2z}{s^3} - 2\frac{xyz}{s^3} - 4\frac{xz^2}{s^3} - 3\frac{z^3}{s^3}\right),
\end{aligned}$$

and the fluid limit equation is $\mathscr{X}' = \phi_X(\mathscr{X},\mathscr{Y},\mathscr{Z})$, and similarly for the two other coordinates.

We start with the estimate (3.15) on $\widetilde{A}$. We first decompose the conditional expected drift as follows:

$$\begin{aligned}
\mathbb{E}\left[\Delta\widetilde{A}_k|\mathcal{F}_k\right] &= \frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}}\mathbb{E}\left[\Delta A_k|\mathcal{F}_k\right] + \left(\frac{\varepsilon_k^{3/4}}{\varepsilon_{k+1}^{3/4}} - 1\right)\widetilde{A}_k\\
&= \frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}}\left(\mathbb{E}\left[\Delta X_k|\mathcal{F}_k\right] - n\left(\mathscr{X}\left(\frac{k+1}{n}\right) - \mathscr{X}\left(\frac{k}{n}\right)\right)\right) + \left(\frac{\varepsilon_k^{3/4}}{\varepsilon_{k+1}^{3/4}} - 1\right)\widetilde{A}_k
\end{aligned}$$

Therefore, by decomposing $1/4 = 1 - 3/4$, we can decompose the left-hand side of (3.15) as follows:

$$\begin{aligned}
&\left|\mathbb{E}\left[\Delta\widetilde{A}_k|\mathcal{F}_k\right] + \frac{1}{\varepsilon_k n}\frac{1}{4}\widetilde{A}_k\right|\\
&\leqslant \left|\left(\frac{\varepsilon_k^{3/4}}{\varepsilon_{k+1}^{3/4}} - 1\right)\widetilde{A}_k - \frac{3}{4}\frac{1}{\varepsilon_k n}\widetilde{A}_k\right| && (3.22)\\
&+ \frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}}\left|\mathbb{E}\left[\Delta X_k|\mathcal{F}_k\right] - \phi_X\left(\frac{X_k}{n},\frac{Y_k}{n},\frac{Z_k}{n}\right)\right| && (3.23)\\
&+ \frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}}\left|\phi_X\left(\frac{X_k}{n},\frac{Y_k}{n},\frac{Z_k}{n}\right) - \phi_X\left((\mathscr{X},\mathscr{Y},\mathscr{Z})\left(\frac{k}{n}\right)\right) - \left(\frac{A_k}{n},\frac{B_k}{n},\frac{C_k}{n}\right)\cdot\nabla\phi_X\left((\mathscr{X},\mathscr{Y},\mathscr{Z})\left(\frac{k}{n}\right)\right)\right| && (3.24)\\
&+ \frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}}\left|\left(\frac{A_k}{n},\frac{B_k}{n},\frac{C_k}{n}\right)\cdot\nabla\phi_X\left((\mathscr{X},\mathscr{Y},\mathscr{Z})\left(\frac{k}{n}\right)\right) + \frac{1}{\varepsilon_k}\frac{A_k}{n}\right| && (3.25)\\
&+ \frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}}\left|\phi_X\left((\mathscr{X},\mathscr{Y},\mathscr{Z})\left(\frac{k}{n}\right)\right) - n\left(\mathscr{X}\left(\frac{k+1}{n}\right) - \mathscr{X}\left(\frac{k}{n}\right)\right)\right|. && (3.26)
\end{aligned}$$

We will bound each of these five error terms one by one. More precisely, we will prove that the terms (3.22), (3.23), (3.24) and (3.26) are all $O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$, whereas the other terms in (3.15) come from (3.25). We start with (3.22), which is easy. We simply write $\varepsilon_k = t_{\text{ext}} - \frac{k}{n}$ and $\varepsilon_{k+1} = t_{\text{ext}} - \frac{k+1}{n}$.

This implies $\frac{\varepsilon_{k+1}}{\varepsilon_k} = 1 - \frac{1}{\varepsilon_k n}$, so

$$\frac{\varepsilon_k^{3/4}}{\varepsilon_{k+1}^{3/4}} - 1 = \frac{3}{4}\frac{1}{\varepsilon_k n} + O\left(\frac{1}{(\varepsilon_k n)^2}\right),$$

where the constant is absolute. Finally, using $\varepsilon_k n \geqslant n^{3/5-1/100}$ we have

$$\frac{|\widetilde{A}_k|}{(\varepsilon_k n)^2} \leqslant \frac{100 n^{-3/5+1/100}\log n}{\varepsilon_k n},$$

so we can bound (3.22) by $\frac{K}{\varepsilon_k n} n^{-1/30}$.

We now move on to (3.23). The drift $\mathbb{E}[\Delta X_k | \mathcal{F}_k]$ can be expressed as the sum over all the cases of Figure 6.3 of the probability of each case multiplied by the variation of X in this case. For example, the probability for the first case is $\frac{X_k - 1}{S_k - 1}$. Approximating $\mathbb{E}[\Delta X_k|\mathcal{F}_k]$ by $\phi_X\left(\frac{X_k}{n}, \frac{Y_k}{n}, \frac{Z_k}{n}\right)$ is then equivalent to approximating $\frac{X_k-1}{S_k-1}$ by $\frac{X_k/n}{S_k/n}$, and similarly for all the other terms. But we have

$$\frac{X_k - 1}{S_k - 1} = \frac{X_k/n}{S_k/n} \times \frac{1 - \frac{1}{X_k}}{1 - \frac{1}{S_k}} = \frac{X_k/n}{S_k/n}\left(1 - O\left(\frac{1}{X_k}\right) + O\left(\frac{1}{S_k}\right)\right) = \frac{X_k/n}{S_k/n} + O\left(\frac{1}{S_k}\right),$$

since $S_k \geqslant X_k$. When we do the same computation for all the cases of Figure 6.3, we also get an error $O(\frac{1}{S_k})$. Note that for the last three cases on the bottom right of Figure 6.3, the probability is already $O(\frac{1}{S_k})$, so these cases do not contribute to $\phi_X(x,y,z)$. So we can bound (3.23) by

$$\frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}} O\left(\frac{1}{S_k}\right) \underset{\text{Lem.3.1}}{=} O\left(\frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}} \times \frac{1}{\varepsilon_k n}\right) \underset{\varepsilon_k \geqslant n^{-\frac{2}{5}-\frac{1}{100}}}{=} O\left(\frac{1}{\varepsilon_k n} \times \frac{1}{(n^{-\frac{2}{5}-\frac{1}{100}})^{3/4}\sqrt{n}}\right) = O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right).$$

We move on to (3.24). We want to estimate the error when we do a linear approximation of ϕ_X near $(\mathscr{X}, \mathscr{Y}, \mathscr{Z})\left(\frac{k}{n}\right)$, so we will need to bound the second derivatives of ϕ_X near this point. More precisely, we write $(v_1, v_2, v_3) = \left(\frac{A_k}{n}, \frac{B_k}{n}, \frac{C_k}{n}\right)$. By the Taylor-Lagrange formula we can bound (3.24) by

$$\frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}} \sum_{1 \leqslant i,j \leqslant 3} |v_i| \times |v_j| \times \max_{\substack{|u_1 - \mathscr{X}(k/n)| \leqslant |v_1| \\ |u_2 - \mathscr{Y}(k/n)| \leqslant |v_2| \\ |u_3 - \mathscr{Z}(k/n)| \leqslant |v_3|}} \left|\frac{\partial^2 \phi_X}{\partial x_i \partial x_j}(u_1, u_2, u_3)\right|. \tag{3.27}$$

By the assumptions of the proposition, we have the bounds:

$$|v_1| \leqslant 1000\,\varepsilon_k^{3/4}\frac{\log n}{\sqrt{n}}, \quad |v_2| \leqslant 1000\frac{\log n}{\sqrt{n}}, \quad |v_3| \leqslant 1000\,\varepsilon_k^{1/2}\frac{\log n}{\sqrt{n}}. \tag{3.28}$$

On the other hand, we can compute the second order derivatives of ϕ_X, which are of the form $\frac{P(x,y,z)}{(x+y+z)^4}$ for some polynomial P. By Lemma 3.1, we know that u_1, u_2 and u_3 are respectively $O\left(\varepsilon_k^2 n^{1/100}\right)$, $O\left(\varepsilon_k n^{1/100}\right)$ and $O\left(\varepsilon_k^{3/2} n^{1/100}\right)$, and the sum $u_1 + u_2 + u_3$ is of order ε_k. Hence, we can consider the term with the highest order in the numerator. For example, we find

$$\frac{\partial^2 \phi_X}{\partial x_1^2} = \frac{12u_2^2 + 28u_2u_3 + 10u_3^2}{(u_1 + u_2 + u_3)^4},$$

and the highest order term in the numerator is $u_2^2 = O\left(\varepsilon_k^2 n^{1/50}\right)$. On the other hand, the denominator is of order ε_k^4, so we get

$$\frac{\partial^2 \phi_X}{\partial x_1^2}(u_1, u_2, u_3) = O\left(\varepsilon_k^{-2} n^{1/50}\right).$$

The bounds on $\frac{\partial^2 \phi_X}{\partial x_i \partial x_j}(u_1, u_2, u_3)$ that we obtain for all second-order partial derivatives are summarized in the following table:

$i \backslash j$	1	2	3
1	$O\left(\varepsilon_k^{-2} n^{1/50}\right)$	$O\left(\varepsilon_k^{-3/2} n^{1/50}\right)$	$O\left(\varepsilon_k^{-2} n^{1/50}\right)$
2	$O\left(\varepsilon_k^{-3/2} n^{1/50}\right)$	$O\left(\varepsilon_k^{-1} n^{1/50}\right)$	$O\left(\varepsilon_k^{-3/2} n^{1/50}\right)$
3	$O\left(\varepsilon_k^{-2} n^{1/50}\right)$	$O\left(\varepsilon_k^{-3/2} n^{1/50}\right)$	$O\left(\varepsilon_k^{-2} n^{1/50}\right)$

Combining this with (3.28), we find that each term of (3.27) is

$$O\left(\frac{\varepsilon_k^{-1} n^{1/50} \log^2 n}{\varepsilon_{k+1}^{3/4} n \sqrt{n}}\right) = O\left(\frac{1}{\varepsilon_k n} \times \frac{n^{1/50} \log^2 n}{\varepsilon_{k+1}^{3/4} \sqrt{n}}\right) \underset{2\varepsilon_{k+1} \geqslant n^{-2/5-1/100}}{=} O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right),$$

which bounds (3.24). Note that it was necessary to handle one by one the terms of (3.27) and not to bound everything crudely by $\|v\|^2 \times \|D^2 \phi_X\|$ (we would have obtained an additional factor ε_k^{-1}, which is too large).

Let us now bound (3.25). We first compute the gradient of $\nabla \phi_X$:

$$\nabla \phi_X(x, y, z) = \frac{1}{(x+y+z)^3}\left(-4y^2 - 9yz + xz - 3z^2, 4xy + 6xz + 2z^2, -x^2 - 2yz + 3xy + 3xz\right). \tag{3.29}$$

On the other hand, by (3.8), when $\varepsilon \to 0$, we have

$$\mathscr{X}(t_{\text{ext}} - \varepsilon) \sim 3\varepsilon^2, \quad \mathscr{Y}(t_{\text{ext}} - \varepsilon) \sim 4\varepsilon \quad , \mathscr{Z}(t_{\text{ext}} - \varepsilon) \sim 4\sqrt{3}\varepsilon^{3/2}.$$

Therefore, we can replace (x, y, z) in (3.29) by $(\mathscr{X}(t), \mathscr{Y}(t), \mathscr{Z}(t))$ and let $t \to t_{\text{ext}}$. We find that there are constants $K, \eta > 0$ such that, for any $0 < \varepsilon < \eta$, we have:

$$\left|\frac{\partial \phi_X}{\partial x}\left((\mathscr{X}, \mathscr{Y}, \mathscr{Z})(t_{\text{ext}} - \varepsilon)\right) - \frac{1}{\varepsilon}\right| \leqslant \frac{\delta}{\varepsilon},$$
$$\left|\frac{\partial \phi_X}{\partial y}\left((\mathscr{X}, \mathscr{Y}, \mathscr{Z})(t_{\text{ext}} - \varepsilon)\right)\right| \leqslant K,$$
$$\left|\frac{\partial \phi_X}{\partial y}\left((\mathscr{X}, \mathscr{Y}, \mathscr{Z})(t_{\text{ext}} - \varepsilon)\right)\right| \leqslant \frac{K}{\varepsilon^{1/2}}.$$

This is the value of η that we take in Proposition 3.4. We can now replace ε by $\varepsilon_k \in (0, \eta)$ and we obtain the following bound on (3.25):

$$\begin{aligned}
&\frac{\delta}{\varepsilon_k n} \times \frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}}|A_k| + \frac{K}{n} \times \frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}}|B_k| + \frac{K}{\varepsilon_k^{1/2} n} \times \frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}}|C_k| \\
&= \frac{\delta}{\varepsilon_k n}|\widetilde{A}_k| + \frac{K}{\varepsilon_k n}\varepsilon_k^{1/4}|\widetilde{B}_k| + \frac{K}{\varepsilon_k n}\varepsilon_k^{1/4}|\widetilde{C}_k| \\
&\leqslant \frac{\delta}{\varepsilon_k n}|\widetilde{A}_k| + \frac{1000K}{\varepsilon_k n}\varepsilon_k^{1/4}\log n.
\end{aligned}$$

We finally treat the term (3.26). We recall that $\mathscr{X}$ solves the equation $\mathscr{X}' = \phi_X(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$, so this is just a linear approximation, so we will need to bound the second derivative $\mathscr{X}''$. More precisely, (3.26) is bounded by

$$\frac{n}{\varepsilon_{k+1}^{3/4}\sqrt{n}} \times \left(\frac{1}{n}\right)^2 \times \max_{\left[\frac{k}{n}, \frac{k+1}{n}\right]} \left|\mathscr{X}''\right|. \tag{3.30}$$

Moreover, by differentiating $\mathscr{X}' = \phi_X(\mathscr{X}, \mathscr{Y}, \mathscr{Z})$, we have

$$\begin{aligned}\mathscr{X}''(t) &= \left(\phi_X \frac{\partial \phi_X}{\partial x} + \phi_Y \frac{\partial \phi_X}{\partial y} + \phi_Z \frac{\partial \phi_X}{\partial z}\right)(\mathscr{X}(t), \mathscr{Y}(t), \mathscr{Z}(t)) \\ &= \frac{\mathscr{Z}(t)}{\mathscr{S}(t)^4}\left(\mathscr{X}(t)^2 - 2\mathscr{X}(t)\mathscr{Y}(t) + 8\mathscr{Y}(t)\mathscr{Z}(t) + 11\mathscr{Z}(t)^2\right).\end{aligned}$$

This is a continuous function of t on $[0, t_{\text{ext}})$. Moreover, by (3.8), we have

$$\mathscr{X}(t_{\text{ext}} - \varepsilon) \sim_{\varepsilon \to 0} \frac{4\sqrt{3}\varepsilon^{3/2}}{(4\varepsilon)^4} \times 8 \times 4\varepsilon \times 4\sqrt{3}\varepsilon^{3/2} = 6,$$

so $\mathscr{X}''$ is bounded by a constant K. Plugging this into (3.30), we can bound (3.26) by $\frac{K}{\varepsilon_{k+1}^{3/4} n^{3/2}} = O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$.

We now move on to the estimates (3.16) and (3.17). Since the proof is similar, we will not do it in full details and only stress the differences with the proof of (3.15). The decomposition of the error into five terms is the same with the following modifications:

- the first term (3.22) becomes $\left|\left(\frac{\varepsilon_k^{1/2}}{\varepsilon_{k+1}^{1/2}} - 1\right)\widetilde{C}_k - \frac{1}{2}\frac{1}{\varepsilon_k n}\widetilde{C}_k\right|$ for $\widetilde{C}$, and disappears completely for $\widetilde{B}$;
- in the terms (3.23), (3.24), (3.25) and (3.26), the factors $\frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}}$ become $\frac{1}{\sqrt{n}}$ for $\widetilde{B}$ and $\frac{1}{\varepsilon_{k+1}^{1/2}\sqrt{n}}$ for $\widetilde{C}$;
- in the fourth term (3.25), the drift $\frac{1}{\varepsilon_k n}\widetilde{A}_k$ becomes 0 for $\widetilde{B}$ and $\frac{3\sqrt{3}}{2}\widetilde{B}_k - \frac{3}{2}\widetilde{C}_k$ of $\widetilde{C}$.

The first and second term can then be bounded by $O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$ in the exact same way as for $\widetilde{A}$ (this bound actually becomes cruder for (3.23), since now the factor $\varepsilon_{k+1}^{3/4}$ in the denominator disappears or become larger).

The bound on the fifth term (3.26) is also very similar: we now have

$$\mathscr{Y}''(t) = -2\frac{\mathscr{Z}}{\mathscr{S}^4}\left(\mathscr{X}\mathscr{Y} + 4\mathscr{Y}^2 + 8\mathscr{X}\mathscr{Z} + 21\mathscr{Y}\mathscr{Z} + 20\mathscr{Z}^2\right)(t) = O\left((t_{\text{ext}} - t)^{-1/2}\right)$$

$$\mathscr{Z}''(t) = 3\frac{\mathscr{Z}}{\mathscr{S}^4}\left(\mathscr{X}^2 + 4\mathscr{X}\mathscr{Y} + 4\mathscr{Y}^2 + 8\mathscr{X}\mathscr{Z} + 14\mathscr{Y}\mathscr{Z} + 11\mathscr{Z}^2\right)(t) = O\left((t_{\text{ext}} - t)^{-1/2}\right).$$

Therefore, the analog of (3.26) for $\widetilde{B}$ (resp. $\widetilde{C}$) is $O\left(\frac{1}{\sqrt{n}} \times n \times \left(\frac{1}{n}\right)^2 \times \varepsilon_k^{-1/2}\right) = O\left(\frac{\varepsilon_k^{-1/2}}{n^{3/2}}\right)$ (resp. $O\left(\frac{\varepsilon_k^{-1}}{n^{3/2}}\right)$). In both cases, this is $O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$.

The analog of the third term (3.24) is still very similar, but requires to be more careful. Indeed (3.27) becomes respectively

$$\frac{1}{\varepsilon_{k+1}^{1/2}\sqrt{n}} \sum_{1\leqslant i,j\leqslant 3} |v_i| \times |v_j| \times \max_{\substack{|u_1-\mathscr{X}(k/n)|\leqslant|v_1|\\ |u_2-\mathscr{Y}(k/n)|\leqslant|v_2|\\ |u_3-\mathscr{Z}(k/n)|\leqslant|v_3|}} \left|\frac{\partial^2\phi_Y}{\partial x_i\partial x_j}(u_1,u_2,u_3)\right|. \tag{3.31}$$

and

$$\frac{1}{\sqrt{n}} \sum_{1\leqslant i,j\leqslant 3} |v_i| \times |v_j| \times \max_{\substack{|u_1-\mathscr{X}(k/n)|\leqslant|v_1|\\ |u_2-\mathscr{Y}(k/n)|\leqslant|v_2|\\ |u_3-\mathscr{Z}(k/n)|\leqslant|v_3|}} \left|\frac{\partial^2\phi_Z}{\partial x_i\partial x_j}(u_1,u_2,u_3)\right|. \tag{3.32}$$

for $\widetilde{B}$ and $\widetilde{C}$. Moreover, when we compute the second order partial derivatives $\frac{\partial^2\phi_Y}{\partial x_i\partial x_j}$ and $\frac{\partial^2\phi_Y}{\partial x_i\partial x_j}$, we get respectively the following tables:

$i\backslash j$	1	2	3
1	$O\left(\varepsilon_k^{-2}n^{1/50}\right)$	$O\left(\varepsilon_k^{-2}n^{1/50}\right)$	$O\left(\varepsilon_k^{-2}n^{1/50}\right)$
2	$O\left(\varepsilon_k^{-2}n^{1/50}\right)$	$O\left(\varepsilon_k^{-3/2}n^{1/50}\right)$	$O\left(\varepsilon_k^{-2}n^{1/50}\right)$
3	$O\left(\varepsilon_k^{-2}n^{1/50}\right)$	$O\left(\varepsilon_k^{-2}n^{1/50}\right)$	$O\left(\varepsilon_k^{-3/2}n^{1/50}\right)$

$i\backslash j$	1	2	3
1	$O\left(\varepsilon_k^{-3/2}n^{1/50}\right)$	$O\left(\varepsilon_k^{-3/2}n^{1/50}\right)$	$O\left(\varepsilon_k^{-2}n^{1/50}\right)$
2	$O\left(\varepsilon_k^{-3/2}n^{1/50}\right)$	$O\left(\varepsilon_k^{-3/2}n^{1/50}\right)$	$O\left(\varepsilon_k^{-2}n^{1/50}\right)$
3	$O\left(\varepsilon_k^{-2}n^{1/50}\right)$	$O\left(\varepsilon_k^{-2}n^{1/50}\right)$	$O\left(\varepsilon_k^{-2}n^{1/50}\right)$

In both cases, using (3.28), we find that each term of (3.31) or (3.32) is

$$O\left(\frac{\varepsilon_k^{-3/2}n^{1/50}\log^2 n}{\varepsilon_{k+1}^{1/2}n^{3/2}}\right) = O\left(\frac{1}{\varepsilon_k n}\times\frac{n^{1/50}\log^2 n}{\varepsilon_k\sqrt{n}}\right) \underset{\varepsilon_k\geqslant n^{-\frac{2}{5}-\frac{1}{100}}}{=} O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right).$$

Finally, to handle the analog of the fourth term (3.25), we just need to compute the gradients of ϕ_Y and ϕ_Z:

$$\nabla\phi_Y(x,y,z) = \frac{1}{(x+y+z)^3}\left(2xy+6y^2+6yz-8z^2, -2x^2-6xy-6xz+4yz+12z^2, 8xz+4y^2+12yz\right),$$

$$\nabla\phi_Z(x,y,z) = \frac{1}{(x+y+z)^3}\left(3xz+9yz+15z^2, 6yz+12z^2, -3x^2-9xy-15xz-6y^2-12yz\right).$$

As in the first case, we can now replace (x,y,z) by $(\mathscr{X}(t),\mathscr{Y}(t),\mathscr{Z}(t))$, use (3.8) and identify the

highest order terms in $t_{\text{ext}} - t$. We find that there is a constant K such that for all $0 \leqslant t < t_{\text{ext}}$:

$$\begin{aligned}
\left|\frac{\partial \phi_Y}{\partial x}(\mathscr{X}(t),\mathscr{Y}(t),\mathscr{Z}(t))\right| &\leqslant \frac{K}{t_{\text{ext}}-t},\\
\left|\frac{\partial \phi_Y}{\partial y}(\mathscr{X}(t),\mathscr{Y}(t),\mathscr{Z}(t))\right| &\leqslant \frac{K}{(t_{\text{ext}}-t)^{1/2}},\\
\left|\frac{\partial \phi_Y}{\partial z}(\mathscr{X}(t),\mathscr{Y}(t),\mathscr{Z}(t))\right| &\leqslant \frac{K}{t_{\text{ext}}-t},\\
\left|\frac{\partial \phi_Z}{\partial x}(\mathscr{X}(t),\mathscr{Y}(t),\mathscr{Z}(t))\right| &\leqslant \frac{K}{(t_{\text{ext}}-t)^{1/2}}.
\end{aligned}$$

Moreover, there is $\eta > 0$ (depending on δ) such that, if $t_{\text{ext}} - \eta \leqslant t < t_{\text{ext}}$, then

$$\begin{aligned}
\left|\frac{\partial \phi_Z}{\partial y}(\mathscr{X}(t),\mathscr{Y}(t),\mathscr{Z}(t)) - \frac{3\sqrt{3}}{2}\frac{1}{(t_{\text{ext}}-t)^{1/2}}\right| &\leqslant \frac{\delta}{(t_{\text{ext}}-t)^{1/2}},\\
\left|\frac{\partial \phi_Z}{\partial y}(\mathscr{X}(t),\mathscr{Y}(t),\mathscr{Z}(t)) + \frac{3}{2}\frac{1}{t_{\text{ext}}-t}\right| &\leqslant \frac{\delta}{t_{\text{ext}}-t}.
\end{aligned}$$

From here, taking $t = \frac{k}{n}$ and replacing (A_k, B_k, C_k) by $\left(\varepsilon_k^{3/4}\sqrt{n}\widetilde{A}_k, \sqrt{n}\widetilde{B}_k, \varepsilon_k^{1/2}\sqrt{n}\widetilde{C}_k\right)$, we easily obtain the claimed bound on (3.25).

□

Proof of Proposition 3.5. Just like in the proof of Proposition 3.4, we first introduce the following functions (again with the notation $s = x + y + z$):

$$\begin{aligned}
\psi_X(x,y,z) &= 4\frac{x}{s} + 4\frac{xy}{s^2} + \frac{yz}{s^2} + \frac{y^2z}{s^3} + 9\frac{x^2z}{s^3} + 2\frac{xyz}{s^3} + 2\frac{xz^2}{s^3} + \frac{z^3}{s^3},\\
\psi_Y(x,y,z) &= \frac{xy}{s^2} + 4\frac{y^2}{s^2} + 4\frac{y^2z}{s^3} + 2\frac{xyz}{s^3} + 2\frac{xz^2}{s^3} + 4\frac{z^3}{s^3},\\
\psi_Z(x,y,z) &= \frac{yz}{s^2} + \frac{y^2z}{s^3} + 8\frac{yz^2}{s^3} + \frac{x^2z}{s^3} + 2\frac{xyz}{s^3} + 8\frac{xz^2}{s^3} + 9\frac{z^3}{s^3}.
\end{aligned}$$

These functions are respectively the fluid limit approximations of $\mathbb{E}[(\Delta X_k)^2 \mid \mathcal{F}_k]$, $\mathbb{E}[(\Delta Y_k)^2 \mid \mathcal{F}_k]$ and $\mathbb{E}[(\Delta Z_k)^2 \mid \mathcal{F}_k]$ and can be computed from the transitions given in Figure 6.3 as before.

Variance of $\widetilde{A}$. Let us start by establishing (3.18). We first note that, since adding a function of A_k does not change the conditional variance on $\mathcal{F}_k$, we have

$$\operatorname{Var}\left(\Delta\widetilde{A}_k|\mathcal{F}_k\right) = \operatorname{Var}\left(\Delta\widetilde{A}_k + \left(\frac{1}{\varepsilon_k^{3/4}\sqrt{n}} - \frac{1}{\varepsilon_{k+1}^{3/4}\sqrt{n}}\right)A_k|\mathcal{F}_k\right) = \frac{1}{\varepsilon_{k+1}^{3/2}n}\operatorname{Var}\left(\Delta A_k|\mathcal{F}_k\right) = \frac{1}{\varepsilon_{k+1}^{3/2}n}\operatorname{Var}\left(\Delta X_k|\mathcal{F}_k\right).$$

Therefore, we can write

$$\operatorname{Var}\left(\Delta\widetilde{A}_k|\mathcal{F}_k\right) = \frac{1}{\varepsilon_{k+1}^{3/2}n}\mathbb{E}\left[(\Delta X_k)^2|\mathcal{F}_k\right] - \frac{1}{\varepsilon_{k+1}^{3/2}n}\mathbb{E}\left[\Delta X_k|\mathcal{F}_k\right]^2,$$

so

$$\begin{aligned}\left|\mathrm{Var}\left(\Delta\widetilde{A}_k|\mathcal{F}_k\right)-\frac{2\sqrt{3}}{\varepsilon_k n}\right| &\leqslant \frac{1}{\varepsilon_{k+1}^{3/2}n}\mathbb{E}\left[\Delta X_k|\mathcal{F}_k\right]^2 \\ &+\frac{1}{\varepsilon_{k+1}^{3/2}n}\left|\mathbb{E}\left[(\Delta X_k)^2|\mathcal{F}_k\right]-\psi_X\left(\frac{X_k}{n},\frac{Y_k}{n},\frac{Z_k}{n}\right)\right| \\ &+\frac{1}{\varepsilon_{k+1}^{3/2}n}\left|\psi_X\left(\frac{X_k}{n},\frac{Y_k}{n},\frac{Z_k}{n}\right)-\psi_X\left(\mathscr{X}\left(\frac{k}{n}\right),\mathscr{Y}\left(\frac{k}{n}\right),\mathscr{Z}\left(\frac{k}{n}\right)\right)\right| \\ &+\frac{1}{\varepsilon_{k+1}^{3/2}n}\left|\psi_X\left(\mathscr{X}\left(\frac{k}{n}\right),\mathscr{Y}\left(\frac{k}{n}\right),\mathscr{Z}\left(\frac{k}{n}\right)\right)-\left(2\sqrt{3}\sqrt{\varepsilon_k}\right)\right| \\ &+\frac{1}{\varepsilon_{k+1}^{3/2}n}\left|\left(2\sqrt{3}\sqrt{\varepsilon_k}\right)-\left(2\sqrt{3}\sqrt{\varepsilon_{k+1}}\right)\right|\end{aligned} \tag{3.33}$$

The first term can be bounded by using Proposition 3.4 and $\varepsilon_k \geqslant n^{-2/5-1/100}$. Moreover, by the exact same argument as for term (3.23) in the proof of Proposition 3.4, the second term is

$$O\left(\frac{1}{\varepsilon_{k+1}^{3/2}n}\times\frac{1}{S_k}\right)\underset{\text{Lem.3.1}}{=}O\left(\frac{1}{\varepsilon_k^{5/2}n^2}\right)\underset{\varepsilon_k\geqslant n^{-\frac{2}{5}-\frac{1}{100}}}{=}O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right).$$

We now bound the third term of (3.33). If we write $(v_1,v_2,v_3)=\left(\frac{A_k}{n},\frac{B_k}{n},\frac{C_k}{n}\right)$, this is bounded by

$$\frac{1}{\varepsilon_{k+1}^{3/2}\sqrt{n}}\sum_{i=1}^{3}|v_i|\times\max_{\substack{|x-\mathscr{X}(k/n)|\leqslant|v_1|\\|y-\mathscr{Y}(k/n)|\leqslant|v_2|\\|z-\mathscr{Z}(k/n)|\leqslant|v_3|}}\left|\frac{\partial\psi_X}{\partial x}(x,y,z)\right|. \tag{3.34}$$

Just like for ϕ_X in the proof of Proposition 3.4, we can compute the gradient of ψ_X: the partial derivatives are of the form $\frac{P(x,y,z)}{(x+y+z)^4}$, where P is a homogeneous polynomial of degree 3. By using Lemma 3.1, just like in (3.27), we have that x, y and z are respectively $O\left(\varepsilon_k^2 n^{1/100}\right)$, $O\left(\varepsilon_k n^{1/100}\right)$ and $O\left(\varepsilon_k^{3/2}n^{1/100}\right)$ and that the sum $x+y+z$ is of order ε_k. Therefore, by considering the higher order terms in the polynomial $P(x,y,z)$, we obtain the following estimates:

$$\frac{\partial\psi_X}{\partial x}(x,y,z)=O\left(\varepsilon_k^{-1}n^{3/100}\right),\quad \frac{\partial\psi_X}{\partial y}(x,y,z)=O\left(\varepsilon_k^{-1/2}n^{3/100}\right),\quad \frac{\partial\psi_X}{\partial z}(x,y,z)=O\left(\varepsilon_k^{-1}n^{3/100}\right).$$

Combining this with (3.34), we get that the third term of (3.33) is

$$O\left(\frac{n^{3/100}\log n}{\varepsilon_k^2 n^{3/2}}\right).$$

using $\varepsilon_k \geqslant n^{-2/5-1/100}$, this is $O\left(\frac{n^{-1/30}}{\varepsilon_k n}\right)$.

We now bound the fourth term of (3.33). For this, we use again (3.8). In particular, when we write down $\psi_X\left(\mathscr{X},\mathscr{Y},\mathscr{Z}\right)(t_{\text{ext}}-\varepsilon)$, the highest order terms in ε are $\frac{\mathscr{Y}\mathscr{Z}}{\mathscr{S}^2}\sim\sqrt{3}\sqrt{\varepsilon}$ and $\frac{\mathscr{Y}^2\mathscr{Z}}{\mathscr{S}^3}\sim\sqrt{3}\sqrt{\varepsilon}$, so we have

$$\psi_X\left(\mathscr{X},\mathscr{Y},\mathscr{Z}\right)(t_{\text{ext}}-\varepsilon)\sim_{\varepsilon\to0}2\sqrt{3}\sqrt{\varepsilon}.$$

In particular, taking $\varepsilon = \varepsilon_k$, if we choose η small enough the fourth term of (3.33) is bounded by $\frac{\delta}{\varepsilon_k n}$. Finally the fifth term is also smaller than $\frac{\delta}{\varepsilon_k n}$ if ε_k is small enough. Gathering-up the pieces we have established (3.18).

The bound (3.19) follows from the same proof by noticing that the only term of (3.33) which makes the errors $\frac{\widetilde{A}^2}{\varepsilon_k^{1/2} n}, \frac{\widetilde{B}^2}{n}, \frac{\widetilde{C}^2}{n}$ appear is the $-\mathbb{E}\left[\Delta X_k | \mathcal{F}_k\right]^2$, which is negative.

Variance of $\widetilde{B}$. The bound (3.20) is immediate: for the same reason as with $\widetilde{A}$, we have

$$\mathrm{Var}\left(\Delta \widetilde{B}_k | \mathcal{F}_k\right) = \frac{1}{n} \mathrm{Var}\left(\Delta Y_k | \mathcal{F}_k\right) \leqslant \frac{9}{n},$$

since $|\Delta Y_k|$ is bounded by 3.

Variance of $\widetilde{C}$. Finally, we prove (3.21): as before, we can write

$$\begin{aligned}\mathrm{Var}\left(\Delta \widetilde{C}_k | \mathcal{F}_k\right) &= \frac{1}{\varepsilon_{k+1} n} \mathbb{E}\left[(\Delta Z_k)^2 | \mathcal{F}_k\right] - \mathbb{E}\left[\Delta Z_k | \mathcal{F}_k\right]^2 \\ &\leqslant \frac{1}{\varepsilon_{k+1} n} \left| \mathbb{E}\left[(\Delta Z_k)^2 | \mathcal{F}_k\right] - \Psi_Z\left(\frac{X_k}{n}, \frac{Y_k}{n}, \frac{Z_k}{n}\right) \right| + \frac{1}{\varepsilon_{k+1} n} \Psi_Z\left(\frac{X_k}{n}, \frac{Y_k}{n}, \frac{Z_k}{n}\right).\end{aligned}$$

By the exact same argument as for $\widetilde{A}$, the first term is

$$O\left(\frac{1}{\varepsilon_{k+1} n} \frac{1}{S_k}\right) = O\left(\frac{1}{\varepsilon_k^2 n^2}\right) = O\left(\frac{\varepsilon_k^{1/2}}{\varepsilon_k n}\right),$$

where the first equality comes from Lemma (3.1) and the second from $\varepsilon_k \geqslant n^{-2/5-1/100}$. On the other hand, noticing that every term in $\psi_Z(x,y,z)$ has a factor $\frac{z}{s}$, we can write

$$\psi_Z\left(\frac{X_k}{n}, \frac{Y_k}{n}, \frac{Z_k}{n}\right) = O\left(\frac{Z_k}{S_k}\right) = O\left(\frac{\varepsilon_k^{3/2} n}{\varepsilon_k n}\right) = O\left(\varepsilon_k^{1/2}\right),$$

where the second inequality comes from Lemma 3.1. This proves (3.21).

□

3.4.2 Rough behaviour of $\widetilde{A}, \widetilde{B}$ and $\widetilde{C}$

In this section we will use our drift and variance estimates to control $\widetilde{A}, \widetilde{B}, \widetilde{C}$. Recall notation from (3.14) and (3.13). We shall get a rather rough control on $\widetilde{A}, \widetilde{B}$ and $\widetilde{C}$ (Proposition 3.6) and later refine the one on $\widetilde{A}$. In the rest of this subsection, on top of the constant $K > 0$ given by Propositions 3.4 and 3.5, we fix

$$\delta = \frac{1}{100}$$

for definiteness and let $0 < \eta \equiv \eta(\delta) < 1/2$ so that we can apply the above propositions. In particular, the value of η does not depend on n, nor on the coming $\epsilon > 0$ and *its value may be decreased for convenience* by keeping the same δ. In the coming pages $\mathsf{Cst} > 0$ is a constant (which may depend on the constant K or the now-fixed $\delta = \frac{1}{100}$) and that may increase from line to line, but

whose value does not depend upon n (provided it is large enough), nor η, nor on the forthcoming ϵ. On the contrary K_ϵ is a constant that depends upon ϵ but also upon η in an implicit way.

The value η shall give our "starting scale" $k_0 = \lfloor (t_{\text{ext}} - \eta)n \rfloor$ which is such that $\varepsilon_{k_0} = \eta$ and we shall then look at times $k_0 \leqslant k \leqslant \tilde{\theta}$. We start by controlling the fluctuations at k_0.

Lemma 3.2 (Fluctuations in the bulk). *For all $\epsilon > 0$ there exists $K_\epsilon > 0$ so that for all n large enough, with probability at least $1 - \epsilon$ we have*

$$\max(|\widetilde{A}_{k_0}|, |\widetilde{B}_{k_0}|, |\widetilde{C}_{k_0}|) < K_\epsilon \text{ and } \widetilde{\theta} > k_0. \tag{3.35}$$

Proof. Classical results entail that on top of the law of large numbers for the process $n^{-1} \cdot (X^n, Y^n, Z^n)$ given in Proposition 3.2, we have a functional central limit theorem for their fluctuations, as long as we stay in the bulk. More precisely, for $0 \leqslant t \leqslant (t_{\text{ext}} - \eta)$, the solution given by the differential equation (3.5) is bounded away from 0, i.e.

$$\inf\{\min(\mathscr{X}(t), \mathscr{Y}(t), \mathscr{Z}(t)) : 0 \leqslant t \leqslant (t_{\text{ext}} - \eta)\} > 0, \tag{3.36}$$

and thanks to our hypothesis (3.3), the initial fluctuations A_0, B_0 and C_0 are bounded so that $(A_0, B_0, C_0)/\sqrt{n}$ converges to $(0,0,0)$[6]. Therefore, we can apply [85, Theorem 2.3 p 458], which implies that

$$\left(\left(\frac{A_{\lfloor tn \rfloor}}{\sqrt{n}}, \frac{B_{\lfloor tn \rfloor}}{\sqrt{n}}, \frac{C_{\lfloor tn \rfloor}}{\sqrt{n}} \right) : 0 \leqslant t \leqslant t_{\text{ext}} - \eta \right)$$

converges as n goes to infinity weakly to a continuous random processes driven by a nice stochastic differential equation. Furthermore [85, Theorem 2.3 p 458] entails that the terminal value

$$\left(\frac{A_{\lfloor (t_{\text{ext}} - \eta)n \rfloor}}{\sqrt{n}}, \frac{B_{\lfloor (t_{\text{ext}} - \eta)n \rfloor}}{\sqrt{n}}, \frac{C_{\lfloor (t_{\text{ext}} - \eta)n \rfloor}}{\sqrt{n}} \right)$$

converges towards a Gaussian law whose covariance depends on η only. Given (3.36), this implies that w.h.p. we have $X_k > 0$ for all $0 \leqslant k \leqslant (t_{\text{ext}} - \eta)n$ (in other words $\theta > (t_{\text{ext}} - \eta)n$) and that $|\widetilde{A}_{\lfloor (t_{\text{ext}} - \eta)n \rfloor}|, |\widetilde{B}_{\lfloor (t_{\text{ext}} - \eta)n \rfloor}|, |\widetilde{C}_{\lfloor (t_{\text{ext}} - \eta)n \rfloor}|$ are tight. The statement of the lemma follows.

□

After this initial control, we shall provide a rough upper bound on the fluctuation processes.

Proposition 3.6 (Rough upper bounds). *For all $\epsilon > 0$, there exists a constant $K_\epsilon > 0$ such that for n large enough, with probability at least $1 - \epsilon$ we have*

$$\max_{k_0 \leqslant k < \tilde{\theta}} \left\{ \frac{|\widetilde{A}_k|}{|\log(\varepsilon_k)|^{3/4}}, |\widetilde{B}_k|, |\widetilde{C}_k| \right\} \leqslant K_\epsilon. \tag{3.37}$$

Remark (The truth). The proof of the proposition shows that we can replace the $3/4$ exponent by $1/2 + \delta$ for all $\delta > 0$. We anyway expect an "iterated logarithm" behavior for $\widetilde{A}$ so that we could

[6] We could have allowed $o(\sqrt{n})$ fluctuations, but not $o(n)$ as in Theorem 3.1.

replace $|\log(\varepsilon_k)|$ by $|\log\log(\varepsilon_k)|$. In the same vein, a little more effort would yield that $\widetilde{B}$ and $\widetilde{C}$ "converge"[7] but our estimates will be largely sufficient for our purposes.

Proof. In light of the form of the drift of $\widetilde{C}$ obtained in Equation (3.17), we will rather consider the process $\widetilde{E}_k = \widetilde{C}_k - \frac{3\sqrt{3}}{2}\widetilde{B}_k$ instead of $\widetilde{C}$, but notice we can control $|\widetilde{C}_k| \leqslant \frac{3\sqrt{3}}{2}|\widetilde{B}_k| + |\widetilde{E}_k|$ using the processes $\widetilde{B}$ and $\widetilde{E}$ so that it is sufficient to prove the proposition after replacing $\widetilde{C}$ by $\widetilde{E}$. Introduce L the first time at which one of the those three processes becomes large, i.e.

$$L = \tilde{\theta} \wedge \min\left\{k \geqslant k_0 : \max\left(\frac{|\widetilde{A_k}|}{|\log \varepsilon_k|^{3/4}}, |\widetilde{B}_k|, |\widetilde{E}_k|\right) > K_\epsilon\right\}.$$

We call the region defined by the above inequalities on $(\widetilde{A}, \widetilde{B}, \widetilde{C})$ the *good region* for the processes and evaluate separately the probability that we exit this region (i.e. that $L < \tilde{\theta}$) via one of the three processes $\widetilde{A}, \widetilde{B}$ or $\widetilde{E}$. By definition (3.14) of $\tilde{\theta}$ and since we will always take n large enough to have

$$K_\epsilon(1 + \log_2^{3/4}(n)) < \log(n),$$

as long as $k_0 \leqslant k < L$, we can apply the estimates obtained in Propositions 3.4 and 3.5. Specifically, we will decompose the processes $\widetilde{A}, \widetilde{B}$ and $\widetilde{E}$ into their predictable and martingale parts and use Doob's maximal inequality and L^2 estimates to control the martingales.
Let us start with $\widetilde{B}$. We write for $k_0 \leqslant k \leqslant L$,

$$\widetilde{B}_k = \widetilde{B}_{k_0} + \sum_{\ell=k_0}^{k-1} \mathbb{E}\left[\Delta\widetilde{B}_\ell | \mathcal{F}_\ell\right] + M_k^B$$

where $(M^B_{k\wedge L})_{k\geqslant k_0}$ is an $(\mathcal{F}_k)$-martingale which starts from 0 at time k_0. We first evaluate the drift/predictable part. To ease the calculation and readability, we will deliberately drop the integer-part notation $\lfloor\cdot\rfloor$ and introduce scales. Recall that the value of $\eta = \varepsilon_{k_0}$ has been fixed above, but we may decrease it for convenience as long as it is independent of n and ϵ. We start from $k_0 = (t_{\text{ext}} - \eta)n$ and we let

$$k_j = (t_{\text{ext}} - \eta 2^{-j})n,$$

for $0 \leqslant j \leqslant (\frac{2}{5} + \frac{1}{100})\log_2(n)$. In particular we have $j + |\log_2 \eta| \leqslant |\log_2 \varepsilon_k| \leqslant (j+1) + |\log_2 \eta|$ for

[7] To be precise, and stressing the dependence in n, the processes $(\widetilde{B}^n_{[tn]\wedge\theta^n} : t \in [0, t_{\text{ext}}])$ converge in law for the $\|\|_\infty$ distance towards a limiting process $(\mathcal{B}_t : t \in [0, t_{\text{ext}}])$ which is continuous and in particular continuous at t_{ext}. Similarly $(\widetilde{C}^n_{[tn]\wedge\theta^n} : t \in [0, t_{\text{ext}}]) \to (\mathcal{C}_t : t \in [0, t_{\text{ext}}])$ for a random continuous process and furthermore $\mathcal{C}_{t_{\text{ext}}} = \frac{3\sqrt{3}}{2}\mathcal{B}_{t_{\text{ext}}}$.

all $k_j \leqslant k \leqslant k_{j+1}$. With this notation, and using our estimate (3.16), we know that if $k \geqslant k_0$ we have

$$\begin{aligned}
\mathbb{1}_{k\leqslant L}\left|\sum_{\ell=k_0}^{k-1}\mathbb{E}\left[\Delta\widetilde{B}_\ell|\mathcal{F}_\ell\right]\right| &\underset{(3.16)}{\leqslant} \mathsf{Cst}\sum_{\ell=k_0}^{\infty}\mathbb{1}_{\ell<L}\left(\frac{\max\left(|\widetilde{A}_\ell|,|\widetilde{B}_\ell|,|\widetilde{C}_\ell|\right)}{\sqrt{\varepsilon_\ell n}}+\frac{1}{\varepsilon_\ell n}n^{-1/30}\right)\\
&\underset{\text{good region}}{\leqslant} \mathsf{Cst}\cdot K_\epsilon\cdot\sum_{\ell=k_0}^{\infty}\mathbb{1}_{\ell<L}\left(\frac{|\log\varepsilon_\ell|^{3/4}+1}{\sqrt{\varepsilon_\ell n}}+\frac{1}{\varepsilon_\ell n}n^{-1/30}\right)\\
&\underset{\text{scales}}{\leqslant} \mathsf{Cst}\cdot K_\epsilon\cdot\sum_{j=1}^{\log_2(n)}\sum_{\ell=k_{j-1}}^{k_j-1}\left(\frac{(j+|\log_2\eta|)^{3/4}}{\sqrt{\eta 2^{-j}n}}+\frac{1}{\eta 2^{-j}n}n^{-1/30}\right)\\
&\leqslant \mathsf{Cst}\cdot K_\epsilon\cdot\sum_{j=1}^{\log_2(n)}\left(\sqrt{\eta}\frac{(j+|\log_2\eta|)^{3/4}}{2^{j/2}}+n^{-1/30}\right)\\
&\leqslant \mathsf{Cst}\cdot K_\epsilon\cdot|\log\eta|\cdot\sum_{j=1}^{\log_2(n)}\left(\sqrt{\eta}\frac{j+1}{2^{j/2}}+n^{-1/30}\right)\\
&\leqslant K_\epsilon\cdot(\mathsf{Cst}\cdot\sqrt{\eta}|\log\eta|)\,,
\end{aligned}$$

for n large enough where $\mathsf{Cst}>0$ is a constant that may vary from line to line but that does not depend on n, nor on ϵ nor on η as long as it is small. In particular, we may decrease the value of η so that the parenthesis in the last display is smaller than $1/4$ say. We obtain that the sum of the absolute values of the expected conditional drifts of $\widetilde{B}$ between k_0 and L is bounded by $K_\epsilon/4$ (deterministicaly).

We deduce that the event $\{L<\tilde{\theta}$ and $|B_{k_0}|\leqslant K_\epsilon/4$ and $|\widetilde{B}_L|>K_\epsilon\}$ is included in the event $\{L<\tilde{\theta}$ and $|M_L^B|>K_\epsilon/2\}$ so that in particular we can write

$$\begin{aligned}
\mathbb{P}\left(L<\tilde{\theta}\text{ and we exit the region by }\widetilde{B}\right) &\leqslant \mathbb{P}\left(|\widetilde{B}_{k_0}|>K_\epsilon/4\right)+\mathbb{P}\left(L<\tilde{\theta}\text{ and }|M_L^B|>K_\epsilon/2\right)\\
&\leqslant \mathbb{P}\left(|\widetilde{B}_{k_0}|>K_\epsilon/4\right)+\mathbb{P}\left(\sup_{k_0\leqslant k<L}|M_k^B|>K_\epsilon/2\right)\\
&\underset{\text{Doob}}{\leqslant} \mathbb{P}\left(|\widetilde{B}_{k_0}|>K_\epsilon/4\right)+4\frac{\mathbb{E}\left[(M_L^B)^2\right]}{K_\epsilon^2/4}.
\end{aligned}$$

Up to increasing K_ϵ we can bound the first term by ϵ using Lemma 3.2. To bound the second term, we use our variance estimate (3.20) which gives in the good region

$$\mathbb{E}\left[(\Delta M_k^B)^2|\mathcal{F}_k,\,k\leqslant L\right]=\mathrm{Var}\left(\Delta M_k^B|\mathcal{F}_k,\,k\leqslant L\right)=\mathrm{Var}(\Delta\widetilde{B}_k|\mathcal{F}_k,\,k\leqslant L)\underset{(3.20)}{\leqslant}\frac{K}{n}.$$

By the orthogonality of martingale increments in L^2 we deduce that

$$\mathbb{E}[(M_L^B)^2]=\sum_{k=k_0}^{\infty}\mathbb{E}[(\Delta M_k^B)^2\mathbb{1}_{k\leqslant L}|\mathcal{F}_k]\leqslant\frac{K(t_{\mathrm{ext}}n-k_0)}{n}=K\eta.$$

Hence we obtain

$$\mathbb{P}(L<\tilde{\theta}\text{ and we exit the good region by }\widetilde{B})\leqslant\epsilon+16\frac{\mathbb{E}[(M_L^B)^2]}{K_\epsilon^2}\leqslant\epsilon+\frac{16K\eta}{K_\epsilon^2}.$$

If K_ϵ is large enough, the second term is also less than ϵ so that the probability in the left-hand side is small. Conclusion: it is unlikely that we exit first the good region because of the process $\widetilde{B}$.

CASE OF $\widetilde{E}$. The proof is similar, but we shall use more precisely the form of the conditional expected drifts. As before, we write

$$\widetilde{E}_k = \widetilde{E}_{k_0} + \sum_{\ell=k_0}^{k-1} \mathbb{E}\left[\Delta\widetilde{E}_\ell | \mathcal{F}_\ell\right] + M^E_k$$

where $(M^E_{k\wedge L})_{k\geqslant k_0}$ is an $(\mathcal{F}_k)$-martingale which starts from 0 at time k_0. We will bound $\mathbb{P}(L < \tilde{\theta}$ and $\widetilde{E}_L > K_\epsilon)$ and the case $\widetilde{E}_L < -K_\epsilon$ will be treated similarly. Let us introduce L^-_E, the last time before L where $\widetilde{E}$ is smaller that $K_\epsilon/2$. In particular on the event $\{L < \tilde{\theta}$ and $\widetilde{E}_L > K_\epsilon\}$, for $L^-_E < k \leqslant L$ the process $\widetilde{E}$ is larger than $K_\epsilon/2$ and its conditional expected drift therefore satisfies

$$\begin{aligned}
\left|\mathbb{E}\left[\Delta\widetilde{E}_k|\mathcal{F}_k\right] - \frac{1}{\varepsilon_k n}\underbrace{\widetilde{E}_k}_{\geqslant K_\epsilon/2}\right| &\underset{(3.16)\&(3.17)}{\leqslant} \frac{3\sqrt{3}}{2}\left(\frac{K}{\varepsilon_k n}\sqrt{\varepsilon_k}\max\left(|\widetilde{A}_k|,|\widetilde{B}_k|,|\widetilde{C}_k|\right) + \frac{K}{\varepsilon_k n}n^{-1/30}\right)\\
&\quad + \frac{\delta}{\varepsilon_k n}\max\left(|\widetilde{B}_k|,|\widetilde{C}_k|\right) + \frac{K}{\varepsilon_k n}\varepsilon_k^{3/4}|\widetilde{A}_k| + \frac{K}{\varepsilon_k n}n^{-1/30}\\
&\leqslant \frac{(\delta + 3K\sqrt{\varepsilon_k})}{\varepsilon_k n}\max\left(|\widetilde{B}_k|,|\widetilde{C}_k|\right) + \frac{4K}{\varepsilon_k n}\varepsilon_k^{1/2}|\widetilde{A}_k| + \frac{4K}{\varepsilon_k n}n^{-1/30}\\
&\underset{\substack{\text{good region}\\ n \text{ large enough}}}{\leqslant} K_\epsilon\cdot\left(\frac{2(\delta + 3K\sqrt{\varepsilon_k})}{\varepsilon_k n} + \frac{4K\varepsilon_k^{1/2}|\log\varepsilon_k|^{3/4}}{\varepsilon_k n}\right).
\end{aligned}$$

Up to further diminishing η (which forces $\varepsilon_k < \eta$ to be small), we can assume that the right-hand side is smaller than $K_\epsilon/(4\varepsilon_k n)$ for n large enough so that we are sure that the conditional expected drift $\mathbb{E}\left[\Delta\widetilde{E}_k|\mathcal{F}_k\right]$ is less than $-K_\epsilon/(4\varepsilon_k n)$ for $L^-_E < k < L$ and in particular it is negative and pulls back the process towards 0.

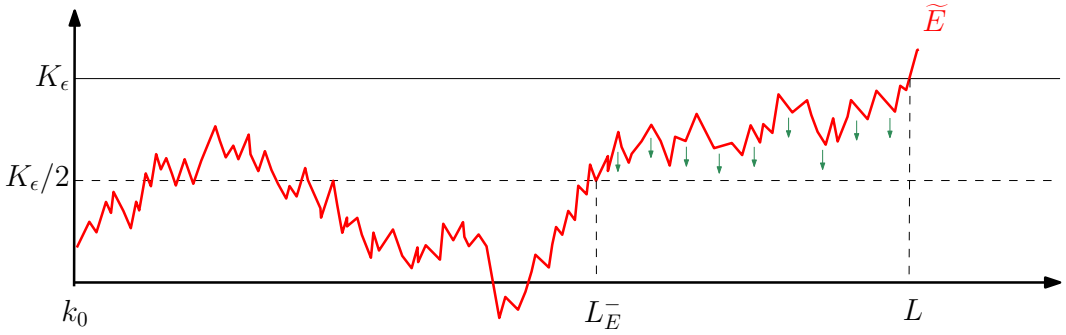


Figure 3.7: Illustration of the proof. If we exit the good region through the process $\widetilde{E}$, then it has a negative drift (green arrows on the figure) over the time interval (L^-_E, L) and this forces its martingale part to vary too much.

We deduce that on the event $\{k_0 < L^-_E < L < \tilde{\theta}$ and $\widetilde{E}_L > K_\epsilon\}$ the variation of the martingale M^E over $[L^-_E, L]$ must be larger than $K_\epsilon/2$ (just because the drift plays against the process in this

region). Hence,

$$\mathbb{P}(L<\tilde{\theta} \text{ and } \widetilde{E}_L > K_\epsilon) \leqslant \mathbb{P}(L_E^- \leqslant k_0) + \mathbb{P}\left(\sup_{k_0\leqslant k\leqslant L} |M^E_{k\wedge L}| > \frac{K_\epsilon}{4}\right)$$

We now use our variance estimates (3.21) and (3.20) in the good region. In particular,

$$\begin{aligned}\mathbb{E}\left[(\Delta M_k^E)^2 \mathbb{1}_{k\leqslant L}\right] &= \mathrm{Var}\left(\Delta M_k^E \mathbb{1}_{k\leqslant L}\right) = \mathrm{Var}\left(\left(\Delta\widetilde{C}_k - \frac{3\sqrt{3}}{2}\Delta\widetilde{B}_k\right)\mathbb{1}_{k\leqslant L}\right)\\ &\leqslant \mathrm{Cst}\cdot\left(\mathrm{Var}(\Delta\widetilde{C}_k\mathbb{1}_{k\leqslant L}) + \mathrm{Var}(\Delta\widetilde{B}_k\mathbb{1}_{k\leqslant L})\right) \underset{(3.20)\&(3.21)}{\leqslant} \frac{\mathrm{Cst}}{\sqrt{\varepsilon_k}n},\end{aligned}$$

where $\mathrm{Cst}>0$ as usual does not depend on n nor on ϵ nor on η. Summing those variances over one scale we obtain

$$\sum_{k=k_i}^{k_{i+1}-1}\frac{\mathrm{Cst}}{\sqrt{\varepsilon_k}n} \leqslant \frac{\mathrm{Cst}(k_{i+1}-k_i)}{\sqrt{\varepsilon_{k_{i+1}}}n} \leqslant \mathrm{Cst}\frac{\eta 2^{-i}n}{\sqrt{\eta 2^{-i}}n} = \mathrm{Cst}\sqrt{\eta}(\sqrt{2})^{-i}.$$

We deduce that

$$\mathbb{P}\left(\sup_{k_0\leqslant k\leqslant L}|M_k^E| > \frac{K_\epsilon}{4}\right) \underset{\text{Doob}}{\leqslant} \frac{16\mathbb{E}[(M_L^E)^2]}{K_\epsilon^2} \leqslant \frac{\mathrm{Cst}}{K_\epsilon^2}\sum_{i=0}^{\infty}\sum_{k=k_i}^{k_{i+1}-1}\mathbb{E}[(\Delta M_k^E)^2\mathbb{1}_{k\leqslant L}] \leqslant \frac{\mathrm{Cst}\sqrt{\eta}}{K_\epsilon^2}.$$

If K_ϵ is large enough, this bound, as well as $\mathbb{P}(L_E^-\leqslant k_0)$ (by Lemma 3.2), are less than ϵ so the probability of the event $\{k_0<L<\tilde{\theta} \text{ and } \widetilde{E}_L > K_\epsilon\}$ is less than 2ϵ. Combined with the symmetric case when $\widetilde{E}_L<-K_\epsilon$, this finishes the case of $\widetilde{E}$.

Let's finally move on to the control of $\widetilde{A}$. Again, we decompose $\widetilde{A}$ as follows

$$\widetilde{A}_k = \widetilde{A}_{k_0} + \sum_{\ell=k_0}^{k-1}\mathbb{E}\left[\Delta\widetilde{A}_\ell|\mathcal{F}_\ell\right] + M_k^A,$$

where $(M^A_{k\wedge L})_{k\geqslant k_0}$ is a martingale for the canonical filtration and starts at 0 at time k_0. Compared to the above cases, we shall look more precisely at the scale of L and introduce

$$J \text{ such that } k_J \leqslant L < k_{J+1}.$$

In particular, recall that if $k_j\leqslant k\leqslant k_{j+1}$ we have $j+|\log_2\eta| \leqslant |\log_2\varepsilon_k| \leqslant (j+1)+|\log_2\eta|$ so that up to losing a multiplicative factor, we may replace $|\log\varepsilon_k|$ by the corresponding scale j in the calculations. As before, let us bound from above the probability that we exit the good region with the process $\widetilde{A}$, that is

$$\mathbb{P}(L<\tilde{\theta} \text{ and } \widetilde{A}_L > K_\epsilon\cdot(J+1)^{3/4})$$

and the case $L<\tilde{\theta}$ and $\widetilde{A}_L<-K_\epsilon\cdot(J+1)^{3/4}$ is symmetric. As for the case of $\widetilde{E}$, we introduce L_A^- the last time before L where $\widetilde{A}$ is smaller than $K_\epsilon(J+1)^{3/4}/2$ and I its corresponding scale (i.e. such that $k_I\leqslant L_A^-<k_{I+1}$), see Figure 3.8. As before, we get from Lemma 3.2 that $L_A^->k_0$ with high

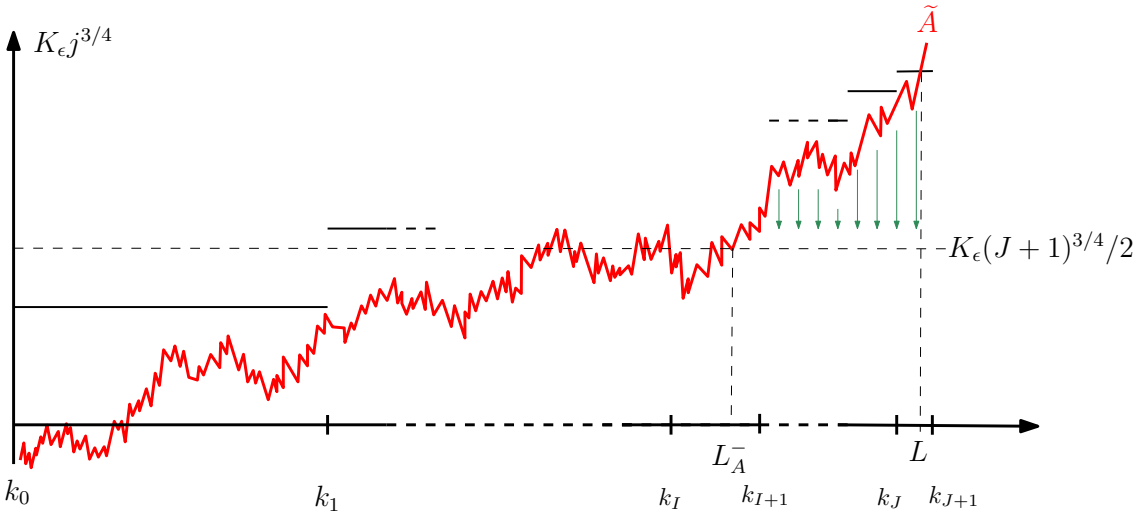


Figure 3.8: Illustration of the proof. If we exit the good region through the process $\widetilde{A}$, then it has a negative drift (green arrows on the figure) over the time interval (L_A^-, L) whose strength is proportional to $J^{3/4}K_\epsilon$ over a scale. As in the above cases, this forces its martingale part to vary too much.

probability when K_ϵ is large. We will now use the fact that the conditional expected drift of $\widetilde{A}$ not only goes against $\widetilde{A}$ but also that its strength is linear in $\widetilde{A}$.

Specifically, when $L_A^- < k < L$, the process $\widetilde{A}$ is larger than $K_\epsilon(J+1)^{3/4}/2$ while the other processes are in absolute value less than K_ϵ so that by (3.15) the predictable drift is negative and of order $-\widetilde{A}_k/(\varepsilon_k n)$:

$$\begin{aligned}\left|\mathbb{E}\left[\Delta\widetilde{A}_k|\mathcal{F}_k\right] - \frac{1}{4\varepsilon_k n}\underbrace{\widetilde{A}_k}_{\geqslant K_\epsilon(J+1)^{3/4}/2}\right| &\underset{(3.15)}{\leqslant} \frac{\delta}{\varepsilon_k n}|\widetilde{A}_k| + \frac{K\varepsilon_k^{1/4}}{\varepsilon_k n}\max\left(|\widetilde{B}_k|,|\widetilde{C}_k|\right) + \frac{K}{\varepsilon_k n}n^{-1/30}\\ &\underset{\text{good region}}{\leqslant} \frac{1/10}{\varepsilon_k n}K_\epsilon(J+1)^{3/4} + \frac{K\varepsilon_k^{1/4}}{\varepsilon_k n}K_\epsilon + \frac{K}{\varepsilon_k n}n^{-1/30}\\ &\leqslant \frac{1}{9\varepsilon_k n}K_\epsilon(J+1)^{3/4}, \end{aligned} \tag{3.38}$$

for n large enough up to diminishing η if necessary. In particular, $\mathbb{E}\left[\Delta\widetilde{A}_k|\mathcal{F}_k\right]$ is less than $-\frac{K_\epsilon}{100}\frac{(J+1)^{3/4}}{\varepsilon_k n}$ and summing the conditional expected drift over all $k \in (L_A^-, L)$ yields total drift smaller than

$$\sum_{k=L_A^-+1}^{L-1}\mathbb{E}[\Delta A_k|\mathcal{F}_k] \leqslant -c\cdot(J-I-1)K_\epsilon(J+1)^{3/4},$$

for some constant $c>0$. Let us first concentrate on the case where $I+1<J$ so that $J-I-1>0$. In particular, the variation of the martingale M^A between L_A^-+1 and L must compensate this drift

and must be larger than $-c(J-I-1)K_\epsilon(J+1)^{3/4}$. Thus, we have

$$\begin{aligned}
&\mathbb{P}\left(k_0 < L_A^- < L < \tilde{\theta} \text{ and } I+1<J \text{ and } \widetilde{A}_L/(J+1)^{3/4} > K_\epsilon\right)\\
&\leqslant \sum_{j=2}^{\log_2(n)}\sum_{i=0}^{j-2}\mathbb{P}\left(\sup_{k_i\leqslant \ell<k_{j+1}\wedge L} M_\ell^A - \inf_{k_i\leqslant \ell<k_{j+1}\wedge L} M_\ell^A > c(j-i-1)K_\epsilon(j+1)^{3/4}\right)\\
&\underset{Doob}{\leqslant} \text{Cst}\sum_{j=2}^{\log_2(n)}\sum_{i=0}^{j-2}\frac{\mathbb{E}[(M^A_{k_{j+1}\wedge L}-M^A_{k_i})^2]}{K_\epsilon^2(j+1)^{3/2}(j-i-1)^2}.
\end{aligned}$$

Thanks to our variance estimates (3.19) we have $\mathbb{E}[(\Delta M_k^A)^2\mathbf{1}_{k<L}]\leqslant \frac{\text{Cst}}{\varepsilon_k n}$ so that after summing over scales we obtain $\mathbb{E}[(M^A_{k_{j+1}\wedge L}-M^A_{k_i})^2]\leqslant \text{Cst}\cdot(j+1-i)$. Plugging this back into the above estimate we deduce

$$\begin{aligned}
&\mathbb{P}\left(k_0 < L_A^- < L < \tilde{\theta} \text{ and } I+1<J \text{ and } \widetilde{A}_L/(J+1)^{3/4} > K_\epsilon\right)\\
&\leqslant \frac{\text{Cst}}{K_\epsilon^2}\sum_{j=2}^{\log_2(n)}\sum_{i=0}^{j-2}\frac{(j+1-i)}{(j+1)^{3/2}(j-i-1)^2}\leqslant \frac{\text{Cst}}{K_\epsilon^2},
\end{aligned}$$

so that this probability can be made arbitrarily small by making K_ϵ large. The case $\widetilde{A}_L/(J+1)^{3/4}<-K_\epsilon$ is treated similarly. As for the case $|I-J|\leqslant 1$, since $k_0<L_A^-$ w.h.p. (by Lemma 3.2), we use that in this case the martingale M^A must have a variation of at least $K_\epsilon(j+1)^{3/4}/2$ over (k_{j-1},k_{j+1}) (we do not use the strength of the drift, but just the fact it plays against us over (L_A^-,L) as for the case of $\widetilde{E}$). By Doob maximal inequality and the above estimate, this probability is bounded from above by $\text{Cst}/((j+1)^{3/2}K_\epsilon^2)$, whose sum over $0\leqslant j\leqslant \log_2(n)$ is $\leqslant \frac{\text{Cst}}{K_\epsilon^2}$. We conclude similarly.

□

3.4.3 This is the end

Using Proposition 3.6 and (3.8), we can conclude as in Lemma 3.1 that the process X_k stays positive at least as long as

$$n\varepsilon_k^2 \gg \sqrt{n}(\varepsilon_k)^{3/4}|\log\varepsilon_k|^{3/4}, \quad \text{i.e. as long as} \quad t_{\text{ext}}n-k\gg n^{3/5}(\log n)^{3/5}.$$

Through a more refined control on $\widetilde{A}$, we shall first prove that we can remove the $\log^{3/5}n$ and prove that $nt_{\text{ext}}-\theta = O_{\mathbb{P}}(n^{3/5})$, see Proposition 3.7. The convergence in law of $n^{-3/5}(nt_{\text{ext}}-\theta)$ will be deduced by doing a SDE approximation for the process $\widetilde{A}_k$ when $\varepsilon_k\approx n^{-2/5}$ in Proposition 3.8.

Since we now take a close look at times $k=t_{\text{ext}}n-O(n^{3/5})$, let us introduce a new piece of notation: for $k\geqslant 0$, we write

$$\boldsymbol{t}_k := n^{-3/5}(t_{\text{ext}}n-k), \quad \text{so that} \quad k=t_{\text{ext}}n-\boldsymbol{t}_k n^{3/5} \quad \text{i.e.} \quad \varepsilon_k=\boldsymbol{t}_k n^{-2/5}.$$

With this notation at hands, we can state a refined control on $\widetilde{A}$.

Proposition 3.7 (Control on $\widetilde{A}$ in the critical region)**.** *For all $\epsilon > 0$ there exists K_ϵ such that with probability at least $1-\epsilon$, for all $k \leqslant t_{\text{ext}} n$ such that $\boldsymbol{t}_k \geqslant K_\epsilon$, we have*

$$|\widetilde{A}_k| < K_\epsilon \boldsymbol{t}_k^{1/8}.$$

In particular, performing the same argument as in the beginning of this subsection, we deduce that X_k stays positive until time $t_{\text{ext}} n - K_\epsilon n^{3/5}$, that is $\theta > t_{\text{ext}} n - K_\epsilon n^{3/5}$ with probability at least $1-\epsilon$.

Proof. The proof is similar to the control of $\widetilde{A}$ in Proposition 3.6. With the notation of the proof of Proposition 3.6, let us introduce

$$T = L \wedge \min\{k \geqslant k_0 : |\widetilde{A}_k| > K_\epsilon \cdot \boldsymbol{t}_k^{\frac{1}{8}}\}$$

and $J \in \{0,1,2,\dots\}$ the corresponding scale, i.e. such that $2^J \geqslant \boldsymbol{t}_T > 2^{(J-1)}$. As for the previous control of $\widetilde{A}$, we will replace $\boldsymbol{t}_T$ by 2^J in the calculation to make the reading easier. Note in particular that $k \mapsto \boldsymbol{t}_k$ is decreasing.

We bound the probability $\mathbb{P}(T = k < L \text{ and } \widetilde{A}_k > K_\epsilon \cdot 2^{\frac{J}{8}})$, the case $\{T = k < L \text{ and } \widetilde{A}_k < -K_\epsilon 2^{\frac{J}{8}}\}$ being similar. For this, let $\alpha > 0$ be a small constant (to be precised later), and let

$$T^- = \sup\left\{k_0 \leqslant k \leqslant T : \widetilde{A}_k \leqslant \alpha(I-J+1)K_\epsilon 2^{J/8} \text{ with } 2^{I-1} < \boldsymbol{t}_k \leqslant 2^I\right\}$$

and $I \geqslant J$ its corresponding scale (notice the slight difference here with the proof of Proposition 3.6 because I enters in the definition of the barrier). As before, Lemma 3.2 will entail that $T^- > k_0$ with high probability as $n \to \infty$ and when $k_0 \leqslant T^- \leqslant k \leqslant T$ and $2^{i-1} < \boldsymbol{t}_k \leqslant 2^i$, we have $\widetilde{A}_k \geqslant \alpha(i-J+1)K_\epsilon 2^{J/8}$. By the same calculation as in (3.38) we have

$$\mathbb{E}\left[\Delta\widetilde{A}_k | \mathcal{F}_k\right] \leqslant -\frac{\alpha}{8}\frac{(i-J+1)K_\epsilon 2^{J/8}}{\boldsymbol{\varepsilon}_k n}.$$

Summing those expected conditional drifts over all $T^- + 1 \leqslant k < T$ yields a total drift smaller than

$$\begin{aligned}
\sum_{k=T^-+1}^{T-1} \mathbb{E}[\Delta\widetilde{A}_k|\mathcal{F}_k] &\underset{\text{scales}}{\leqslant} \sum_{i=J+1}^{I-1}\sum_{k\geqslant 0} \mathbb{1}_{2^i \geqslant \boldsymbol{t}_k > 2^{i-1}} \mathbb{E}[\Delta\widetilde{A}_k|\mathcal{F}_k] \\
&\leqslant -\frac{\alpha}{8}\sum_{i=J+1}^{I-1}(i-J+1)K_\epsilon 2^{J/8}\sum_{k\geqslant 0}\mathbb{1}_{2^i\geqslant \boldsymbol{t}_k > 2^{i-1}}\frac{1}{\boldsymbol{\varepsilon}_k n} \\
&\leqslant -\frac{\alpha}{16}\sum_{i=J+1}^{I-1}(i-J+1)K_\epsilon 2^{J/8} \\
&\leqslant -\frac{\alpha}{16}(I-J-1)^2 K_\epsilon 2^{J/8}.
\end{aligned}$$

Let us first focus on the case $I - J \geqslant 2$: as soon as $T^- > k_0$ the variation of the martingale M^A between T^- and T must compensate this drift plus the difference of the starting and ending values, and so must be larger than

$$K_\epsilon 2^{J/8}\left(\frac{\alpha}{16}(I-J-1)^2 - \alpha(I-J+1) + 2^{-1/8}\right).$$

If α has been chosen small enough (e.g. $\alpha = \frac{1}{100}$), as soon as $I - J \geqslant 2$, this is larger $\frac{1}{32} K_\epsilon 2^{J/8} (I-J)^2$. As in the proof of Proposition 3.6, the sum of the variances of the increments of M^A between scales i and j is bounded above by $\mathsf{Cst}(i-j)$ and so the probability that M^A varies by more than $\frac{1}{32}(i-j)^2 K_\epsilon 2^{j/8}$ over this time interval is bounded above using Doob's inequality by

$$\mathsf{Cst} \frac{i-j}{\left((i-j)^2 K_\epsilon 2^{j/8}\right)^2}.$$

Summing these probabilities over all scales $j_0 \leqslant j \leqslant i$, we deduce that

$$\begin{aligned} &\mathbb{P}\left(k_0 < T^- < T < L \text{ and } I - 1 > J \geqslant j_0 \text{ and } \widetilde{A}_T > K_\epsilon 2^{J/8}\right) \\ &\leqslant \frac{\mathsf{Cst}}{K_\epsilon^2} \sum_{i \geqslant j+2 \geqslant j_0+2} \frac{i-j}{(i-j)^4 2^{j/4}} \leqslant \frac{\mathsf{Cst} \cdot 2^{-j_0/4}}{K_\epsilon^2}, \end{aligned}$$

and this can be made arbitrarily small by taking j_0 large enough. Finally, we treat the case $0 \leqslant I - J \leqslant 1$ similarly, by noting that in this case, if $k_0 < T^-$ (which has high probability by Lemma 3.2), the variation of $\widetilde{A}$ between times T^- and T is at least $(2^{-1/8} - 2\alpha) K_\epsilon 2^{J/8}$. Since the drift is negative, the martingale M^A must have a variation of order $K_\epsilon 2^{J/8}$ (provided $\alpha < \frac{1}{4}$) over the scale J, and the conclusion is the same.

□

In the rest of this subsection we stress back the dependence in n and use $\theta^n \equiv \theta$ for the stopping time of the exploration and study the convergence of

$$\boldsymbol{t}_{\theta^n} \in \mathbb{R} \quad \text{such that} \quad \theta^n = t_{\mathrm{ext}} n - \boldsymbol{t}_{\theta^n} n^{3/5}.$$

Proposition 3.8. *We have the following convergence in distribution as n goes to infinity*

$$\boldsymbol{t}_{\theta^n} \xrightarrow[n\to\infty]{(d)} 3^{-3/5} \cdot 2^{4/5} \cdot \vartheta^{-2},$$

where $\vartheta = \inf\{t \geqslant 0 : W_t = -t^{-2}\}$ with W a standard linear Brownian motion started from 0 at 0.

Proof. Fix $\epsilon > 0$ and let $K_\epsilon > 0$ so that on an event $\mathcal{E}_n$ of probability at least $1 - 3\epsilon$, the conclusions of Lemma 3.2, Proposition 3.7 and Proposition 3.6 hold. Fix $K_\epsilon^{-1} > \xi > 0$ small enough so that $K_\epsilon \xi^{1/8} \leqslant \epsilon$. We shall first focus on the times k satisfying $\xi \leqslant \boldsymbol{t}_k \leqslant \xi^{-1}$ and consider the renormalized process

$$\widetilde{F}_k = \frac{\widetilde{A}_k}{\boldsymbol{t}_k^{1/4}}, \quad 0 \leqslant k \leqslant \theta^n.$$

Let us compute its conditional expected drift and variance: for $k < \tilde{\theta}^n$ with $\xi \leqslant \boldsymbol{t}_k \leqslant \xi^{-1}$, on the event $\mathcal{E}_n$ the assumptions of Proposition 3.4 hold, so that using $\boldsymbol{\varepsilon}_k = n^{-2/5} \boldsymbol{t}_k$ we have

$$\mathbb{E}[\Delta \widetilde{F}_k | \mathcal{F}_k, \mathcal{E}_n] \leqslant \frac{\delta}{\boldsymbol{t}_k n^{3/5}} |\widetilde{A}_k| + \frac{K}{\boldsymbol{t}_k n^{3/5}} n^{-1/30} = \frac{\delta}{\boldsymbol{t}_k^{3/4} n^{3/5}} |\widetilde{F}_k| + \frac{K}{\boldsymbol{t}_k n^{3/5}} n^{-1/30} \tag{3.39}$$

$$\left| \mathrm{Var}\left(\Delta \widetilde{F}_k | \mathcal{F}_k, \mathcal{E}_n\right) - \frac{2\sqrt{3}}{\boldsymbol{t}_k^{3/2} n^{3/5}} \right| = \left| \frac{1}{\boldsymbol{t}_k^{1/2}} \mathrm{Var}\left(\Delta \widetilde{A}_k | \mathcal{F}_k\right) - \frac{2\sqrt{3}}{\boldsymbol{t}_k^{3/2} n^{3/5}} \right| \leqslant \frac{\delta}{\boldsymbol{t}_k^{3/2} n^{3/5}}. \tag{3.40}$$

We now make δ vary with n and take $\delta \equiv \delta_n \xrightarrow[n\to\infty]{} 0$ in the above displays. Indeed, using the notation of Propositions 3.4 and 3.5 we can do so as soon as $\eta(\delta_n) > 1/\xi \cdot n^{-2/5}$. To avoid stopping times issues, we possibly extend $\widetilde{F}$ after time θ^n (in the case $\boldsymbol{t}_{\theta^n} \leqslant \xi$) by a process $\widehat{F}$ whose increments are $\pm(\frac{2\sqrt{3}}{\boldsymbol{t}_k^{3/2} n^{3/5}})^{1/2}$ with probability $1/2$ (in particular independent, centered, with variance $\frac{2\sqrt{3}}{\boldsymbol{t}_k^{3/2} n^{3/5}}$ and whose L^∞-norm tends to 0 uniformly as $n \to \infty$), so that our estimates (3.39) and (3.40) remain true for all $\{k : \xi \leqslant \boldsymbol{t}_k \leqslant \xi^{-1}\}$. Let us recapitulate what we have: with probability at least $1 - 3\epsilon$ for all $\{k : \xi \leqslant \boldsymbol{t}_k \leqslant \xi^{-1}\}$:

$$\begin{cases} |\widehat{F}_{nt_{\mathrm{ext}} - \xi^{-1} n^{3/5}}| < \epsilon, \quad \text{(by Prop. 3.7 and the assumption } K_\epsilon \xi^{1/8} \leqslant \epsilon), \\ \\ \mathbb{E}[\Delta \widehat{F}_k | \mathcal{F}_k] = o(n^{-3/5}) \cdot |\widehat{F}_k| + o(n^{-3/5}), \\ \mathrm{Var}\left(\Delta \widehat{F}_k | \mathcal{F}_k\right) = \dfrac{2\sqrt{3}}{\boldsymbol{t}_k^{3/2} n^{3/5}} + o(n^{-3/5}), \\ \|\Delta \widehat{F}_k\|_\infty = o(1), \end{cases}$$

where the $o(1)$ function is uniform in $\{k : \xi \leqslant \boldsymbol{t}_k \leqslant \xi^{-1}\}$. By standard results in diffusion approximation, see e.g. [127], this implies the following weak convergence for the $\|\|_\infty$-norm:

$$\left(\widehat{F}_{t_{\mathrm{ext}} n - t n^{3/5}} - \widehat{F}_{t_{\mathrm{ext}} n - \xi^{-1} n^{3/5}}\right)_{\xi \leqslant t \leqslant \xi^{-1}} \xrightarrow[n\to\infty]{} (\mathcal{H}_t)_{\xi \leqslant t \leqslant \xi^{-1}},$$

where the process $\mathcal{H}$ satisfies the stochastic differential equation (in reverse time) $\mathrm{d}\mathcal{H}_{-t} = \frac{\sqrt{2\sqrt{3}}}{t^{3/4}} \mathrm{d}B_{-t}$ with initial condition $\mathcal{H}_{\xi^{-1}} = 0$. By Dubbins-Schwarz theorem, the solution of this SDE can be written as

$$2 \cdot 3^{1/4} \left(W_{\frac{1}{\sqrt{t}}} - W_{\frac{1}{\sqrt{\xi^{-1}}}}\right)_{\xi \leqslant t \leqslant \xi^{-1}}$$

where W is a standard linear Brownian motion with $W_0 = 0$. Letting $\epsilon \to 0$ and $\xi \to 0$, we deduce the following convergence weak convergence over all compact subsets of $(0, \infty)$:

$$\left(\widehat{F}_{t_{\mathrm{ext}} n - t n^{3/5}}\right)_{0 < t < \infty} \xrightarrow[n\to\infty]{} \left(W_{\frac{1}{\sqrt{t}}}\right)_{0 < t < \infty}. \tag{3.41}$$

To see that the above convergence implies the convergence of stopping times recall that

$$\begin{aligned} \boldsymbol{t}_{\theta^n} := \sup\{\boldsymbol{t}_k \geqslant 0,\ X_k = 0\} &= \sup\{\boldsymbol{t}_k \geqslant 0,\ \widetilde{F}_k = -n^{4/5} \mathscr{X}(k/n)/\boldsymbol{t}_k\} \\ &= \sup\{\boldsymbol{t}_k \geqslant 0,\ \widehat{F}_k \leqslant -n^{4/5} \mathscr{X}(k/n)/\boldsymbol{t}_k\}. \end{aligned}$$

In particular, the time $\boldsymbol{t}_{\theta^n}$ can be seen as the first time when started from $+\infty$ that the process $\widehat{F}$ crosses the barrier $\mathscr{C}^n$ defined by

$$\mathscr{C}^n(\boldsymbol{t}_k) = -n^{4/5} \mathscr{X}(k/n)/\boldsymbol{t}_k.$$

Recalling (3.8), we have $-n^{4/5} \mathscr{X}(k/n)/\boldsymbol{t}_k \sim -3\boldsymbol{t}_k$, so that the barrier $\mathscr{C}^n$ converges towards the graph $\mathscr{C}$ of the function $t \mapsto -3t$. Since the crossing of $\mathcal{C}$ by $\left(W_{1/\sqrt{t}} : 0 < t < \infty\right)$ when started

from $+\infty$ happens at an almost surely positive time τ and since W immediately takes values strictly above and below $\mathscr{C}$ after hitting it, it follows that

$$\boldsymbol{t}_{\theta^n} \xrightarrow[n\to\infty]{(d)} \tau = \sup\{t \geqslant 0 :\ 2\cdot 3^{1/4}\cdot W_{\frac{1}{\sqrt{t}}} = -3t\}.$$

By scaling we have the equality in distribution

$$\begin{aligned}
\tau &\overset{(d)}{=} \sup\{t \geqslant 0 :\ 2\cdot 3^{1/4}\cdot W_{\frac{1}{\sqrt{t}}} = -3t\}\\
&\underset{u=1/\sqrt{t}}{=} \left(\inf\{u \geqslant 0 : W_u = \frac{3^{3/4}}{2}u^{-2}\}\right)^{-2}\\
&\underset{\alpha>0}{\overset{(d)}{=}} \left(\inf\{u \geqslant 0 : \frac{1}{\sqrt{\alpha}}\cdot W_{\alpha u} = \frac{3^{3/4}}{2}u^{-2}\}\right)^{-2}\\
&\underset{\alpha u=v}{=} \left(\frac{1}{\alpha}\inf\{v \geqslant 0 : W_v = \sqrt{\alpha}\alpha^2\cdot\frac{3^{3/4}}{2}v^{-2}\}\right)^{-2}\\
&\underset{\alpha^{5/2}\cdot\frac{3^{3/4}}{2}=1}{=} \left(\frac{3^{3/4}}{2}\right)^{-4/5}\left(\inf\{v \geqslant 0 : W_v = v^{-2}\}\right)^{-2}.
\end{aligned}$$

The statement follows.

□

3.4.4 Proof of Theorem 3.2: Size and composition of the KS-Core

We have now all the ingredients to prove our main Theorem 3.2. First by Proposition 3.8, the renormalized ending time $\boldsymbol{t}_{\theta^n}$ converges in distribution to $2^{4/5}3^{-3/5}\vartheta^{-2}$ where ϑ is the hitting time of the curve $t \mapsto -t^{-2}$ by a Brownian motion. At this time, by Proposition 3.6 and (3.8) we have

$$Y_{\theta^n} = \underbrace{B_{\theta^n}}_{\underset{\text{Prop.3.6}}{\leqslant}\ \mathrm{Cst}\sqrt{n}\log(n)^{3/4}} + \underbrace{n\mathscr{Y}\left(\frac{\theta^n}{n}\right)}_{\underset{(3.8)}{\sim}\ 4\boldsymbol{t}_{\theta^n}n^{3/5}} \approx 4\boldsymbol{t}_{\theta^n}n^{3/5},$$

$$Z_{\theta^n} = \underbrace{C_{\theta^n}}_{\underset{\text{Prop.3.6}}{\leqslant}\ \mathrm{Cst}\, n^{3/10}\log(n)^{3/4}} + \underbrace{n\mathscr{Z}\left(\frac{\theta^n}{n}\right)}_{\underset{(3.8)}{\sim}\ 4\sqrt{3}\boldsymbol{t}_{\theta^n}n^{2/5}} \approx 4\sqrt{3}\boldsymbol{t}_{\theta^n}^{3/2}n^{2/5}.$$

Moreover using Proposition 3.1, the KS-Core is just obtained by pairing the remaining half-edges uniformly at random. Our theorem follows. Ouff.

3.5 Comments

We conclude this paper with a few perspectives that our work opens.

Near critical heuristics. We believe that our techniques can be used to tackle the near-critical window for the Karp-Sipser core. In particular, this window should be obtained by starting from

$$d^n_{1,c} = n(1 - \frac{\sqrt{3}}{2}) + O(n^{3/5}), \quad 2d^n_{2,c} = O(n^{3/5}), \quad \text{and } 3d^n_{3,c} = n\frac{\sqrt{3}}{2} + O(n^{3/5}),$$

whereas we studied only the critical case (3.3). All these shifts in the starting configuration should result in a shift of order $O(n^{3/5})$ of the absorption time. In a similar vein, one could study the "Phase 2" of the Karp-Sipser algorithm [19] which, in the supercritical case, consists in removing a uniform vertex when there are no leaves left. The analysis of this phase should be intimately connected to the above near-critical dynamics.

Universality. Obviously, we conjecture that the geometry of the critical core and the scaling limits results are independent of the fine details of the model of random graph we started with. In particular, it should hold for the Erdős-Rényi case or for configuration models with small enough degrees. However, proving a general result seems challenging because we heavily rely on the exact form of the fluid-limit of our exploration processes (such results are available for the Erdős–Rényi case, see [19]).

Stopped Markov chain. More generally, we believe that the techniques developed in this paper could be used to understand precisely the exit times of Markov chains from domains. To fix ideas, let $(\mathbb{X}^n : k \geqslant 0)$ be a $\mathbb{Z}^d$-valued Markov chain whose expected conditional drift is well-approximated by $\phi(\mathbb{X}^n/n)$ for some function $\phi : \mathbb{R}^d \to \mathbb{R}^d$. The differential equation method shows that under some mild assumptions $(n^{-1}\mathbb{X}^n_{\lfloor tn \rfloor} : t \geqslant 0)$ converges towards a solution $\mathcal{X}$ to $\mathcal{X}'(t) = \phi(\mathcal{X}(t))$. If Ω is a bounded domain and Ω^n its discrete approximation, it is reasonable to believe that the exist time θ^n of Ω^n by $\mathbb{X}^n$ should converge after normalization towards the exit time t_{ext} of Ω by $\mathcal{X}$. However, the fine fluctuations of θ^n around nt_{ext} should depend on fine properties of ϕ (and its derivatives) near the exit point. We plan on addressing those general questions in future works.

Comparison with the k-core phase transition. Finally, it is interesting to compare our results with the appearance of the k-core in random graphs as studied in [150, 108], where the phase transition is discontinuous.

Recall that the k-core of a graph $\mathfrak{g}$ is the maximal subgraph of $\mathfrak{g}' \subset \mathfrak{g}$ so that the induced degree inside $\mathfrak{g}'$ of each of its vertices is at least k. The emergence of a giant k-core has been studied for the Erdős–Rényi random graph and the configuration model, see [150, 108]. A difference with the Karp–Sipser core is that the phase transition is discontinuous: when the k-core exists asymptotically, its proportion is bounded away from 0. This can be explained heuristically as follows.

Suppose for the discussion that $k = 3$ and that we are interested in the size of the 3-core in a configuration model on vertices of degrees $1, 2, 3$ and 4. As in the case of the Karp–Sipser algorithm, one can reveal the 3-core by iteratively taking a leg attached to a vertex of degree $\leqslant 2$, remove it, and destroy the vertex it is attached to as well as the connection it makes in the graph (hence diminishing the unmatched degree of the vertices in question). As in this paper, if one starts with

some proportions p_1, p_2, p_3, p_4 of legs attached to vertices of degree one, two, three and four, we can write the differential equation governing the fluid limit of this process, see [108]. The main difference with the Karp–Sipser core is that in this case, the number of legs attached to leaves (to be precise to vertices of degree 1 or 2) is not necessarily decreasing. Actually, in the critical case, the fluid limit of the proportion of vertices of degrees $1, 2$ follows a curve which is tangent to the boundary of the domain at some point before diving back into the bulk of the simplexe and dying at the right corner, see Figure 3.9 (and compare with Figure 3.5). This explains the first-order phase transition in this case: a slight perturbation of the initial conditions may push the curve to exit the domain at a very different location.

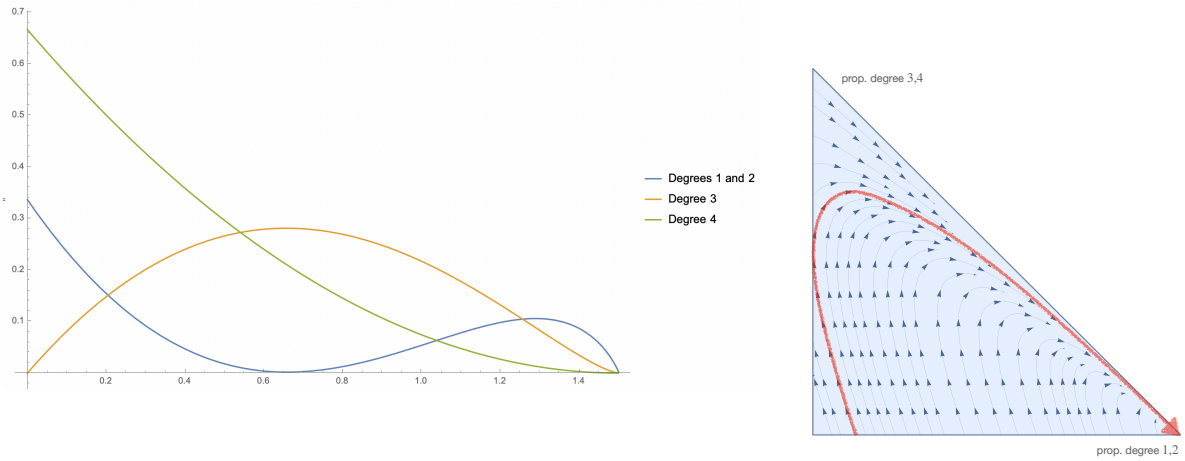


Figure 3.9: Illustration of the fluid limit of the renormalized number of legs attached to vertices of degree $4, 3$ and $\leqslant 2$ in the k-core algorithm at the critical point. In particular, a slight perturbation of the initial conditions may cause a drastic change of the absorption time of the system and this explains why the k-core percolation exhibits a first-order phase transition.

Partie II

Parking dynamique sur des arbres

Chapitre 4 :

Transition de phase via la limite locale

Ce chapitre est très court.

In this chapter, we show the existence of a phase transition for the model of parking on a (random) rooted tree under slight hypothesis. We first recall briefly the parking rules. Given a finite rooted tree $\mathbf{t}$, we assume that the vertices of $\mathbf{t}$ are parking spots and each can accomodate (at most) one car. We imagine now that cars arrive one after the other on the vertices of $\mathbf{t}$. Each car tries to park at its arriving vertex but if it is already occupied, it drives towards the root and takes the first available spot. If no free spot is found during its descent towards the root, then the car exits the tree without parking. Given a rooted tree $\mathbf{t}$, we denote by $(A_x : x \in \mathbf{t})$ the car arrivals on the vertices of $\mathbf{t}$, and we suppose that the $(A_x : x \in \mathbf{t})$'s are independent random variables and that their law only depends on the *outdegree* of the vertex i.e. its number of children. In what follows, we simply write degree instead of outdegree. For all integers $k \geqslant 0$, we consider $(\mu_{(k),\alpha} : 0 \leqslant \alpha \leqslant 1)$ a family of probability measures which is stochastically increasing in α, such that for every $0 \leqslant \alpha \leqslant 1$, the common law of the car arrivals on a vertex of degree k is

$$\mu_{(k),\alpha} \text{ with mean } m_{(k),\alpha} \text{ and variance } \sigma^2_{(k),\alpha}.$$

We let $\varphi_\alpha(\mathbf{t})$ be the outgoing flux of cars i.e. the number of exiting cars.

Concerning the trees, let us consider a sequence of rooted trees $(\mathcal{T}_n, \rho_n)_{n\geqslant 1}$ such that the tree $\mathcal{T}_n$ has n vertices, and an infinite (random) rooted tree $(\mathcal{T}_\infty, \rho_\infty)$ which has almost surely one spine, such that the sequence $(\mathcal{T}_n)$ converges in the sense of Benjamini–Schramm quenched towards $(\mathcal{T}_\infty, \rho_\infty)$. Recall the the Benjamini–Schramm limit of a sequence of trees is a unimodular tree which has by [74, Theorem 13] 1 or 2 spines almost surely. The case of one spine is similar to the one-dimensional case, so that we focus here on the more interesting case of one spine. We have just described the parking procedure on finite rooted trees, and have to be more precise to define it on $\mathcal{T}_\infty$. The point is that the

root ρ_∞ is not "at the bottom" of the tree $\mathcal{T}_\infty$ (think for example of critical Bienaymé–Galton–Watson trees and Aldous' sin-tree, which we described at the end of Section 1.1.2). Thus we orient the edges of $(\mathcal{T}_\infty, \rho_\infty)$ as follows: for the spine of $\mathcal{T}_\infty$ i.e. the (unique) infinite non-backtracking path starting from the root ρ_∞, we orient the edges from the root towards infinity, and the other edges are oriented in the direction of this spine. As a consequence, there is exactly one edge going out from each vertex, and this defines the direction in which the cars can drive. Moreover, for each vertex $x \in \mathcal{T}_\infty$, there is a finite number of vertices from which there exists an oriented path towards x. In other words, the subtree above x is finite, and the car arrivals in this subtree determines the status of x (occupied or free) in the final configuration of parking. Thus the parking procedure is well defined on $(\mathcal{T}_\infty, \rho_\infty)$ (even if we can not speak about outgoing flux in this case) and we can consider the *clusters of parked cars* in $\mathcal{T}_\infty$, which are the trees of the forest obtained by keeping the edges in $\mathcal{T}_\infty$ only if they link two occupied spots in the final configuration. Since the $(\mu_{(k),\alpha}, 0 \leqslant \alpha \leqslant 1)$'s are stochastically increasing in α, we define

$$\alpha_c := \inf\{\alpha > 0 : \ \mathbb{P}\,(\text{there exists an infinite cluster of parked cars in } \mathcal{T}_\infty) > 0\}.$$

We have now all the tools to state our theorem.

Theorem 4.1

We assume that $(\mathcal{T}_n)$ converges in the sense of Benjamini–Schramm quenched towards an infinite (random) rooted tree $(\mathcal{T}_\infty, \rho_\infty)$ which has only one spine almost surely. To ensure a continuity of the family $(\mu_{(k),\alpha} : k \geqslant 1, 0 \leqslant \alpha \leqslant 1)$ with respect to α, we assume that

$$\sum_{k \geqslant 0} \nu_k m_{(k),\alpha} = \alpha,$$

where $\nu_k = \mathbb{P}\,(\deg(\rho_\infty) = k \text{ in } \mathcal{T}_\infty)$, and we suppose there exists a constant K such that $m_{(k),\alpha} < K$ and $\sigma^2_{(k),\alpha} < K$ for all $0 \leqslant \alpha \leqslant 1$ and for all $k \geqslant 0$. We then have

$$\frac{\varphi_\alpha(\mathcal{T}_n)}{n} \xrightarrow[n\to\infty]{(\mathbb{P})} \begin{cases} 0 & \text{if } \alpha < \alpha_c, \\ C_\alpha & \text{if } \alpha > \alpha_c, \end{cases}$$

where $C_\alpha > 0$ is a positive constant if $\alpha > \alpha_c$. Moreover, the constant C_α depends on the $(\mu_{(k),\alpha} : k \geqslant 0)$ and on $(\mathcal{T}_\infty, \rho_\infty)$ but not on $(\mathcal{T}_n)$.

As a corollary of our proof, we will see $\alpha_c \leqslant 1$ because if $\sum_{k\geqslant 0} \nu_k m_{(k),\alpha} = \alpha > 1$, then the renormalized flux of outgoing cars $\varphi_\alpha(\mathcal{T}_n)/n$ also converges towards a constant $C_\alpha > 0$. Indeed, there are more cars than parking spots in this case !

Proof. We fix $\alpha > 0$. We start by proving that the outgoing flux renormalized by n converges in probability. To this end, we have by conservation of cars,

$$\sum_{x \in \mathcal{T}_n} A_x = \varphi_\alpha(\mathcal{T}_n) + \sum_{x \in \mathcal{T}_n} \mathbb{1}_{x \text{ contains a parked car}}.$$

On the one hand, the mean density of arriving cars converges in probability. Indeed, let us compute the first two moments of this quantity.

$$\mathbb{E}\left[\frac{1}{n}\sum_{x\in\mathcal{T}_n} A_x\right] = \mathbb{E}\left[\frac{1}{n}\sum_{x\in\mathcal{T}_n} \mathbb{E}\left[A_x|\deg(x)\right]\right] = \mathbb{E}\left[\frac{1}{n}\sum_{x\in\mathcal{T}_n} m_{(\deg(x)),\alpha}\right] = \mathbb{E}\left[\frac{1}{n}\sum_{x\in\mathcal{T}_n} m_{(\deg(x)),\alpha}\wedge K\right]$$

Moreover, the degree of the root is a continuous function for the local topology, and thus, for all α, the function $(\mathbf{g},\rho)\mapsto m_{(\deg(\rho)),\alpha}\wedge K$ is continuous and bounded for the local topology. Since $(\mathcal{T}_n)$ converges in the sense of Benjamini–Schramm towards $\mathcal{T}_\infty$, we obtain

$$\mathbb{E}\left[\frac{1}{n}\sum_{x\in\mathcal{T}_n} A_x\right] \underset{n\to\infty}{\to} \mathbb{E}\left[m_{(\deg(\rho_\infty)),\alpha}\right] = \sum_{k\geqslant 0}\nu_k m_{(k),\alpha} = \alpha.$$

The computation is very similar for the second moment:

$$\begin{aligned}\mathbb{E}\left[\left(\frac{1}{n}\sum_{x\in\mathcal{T}_n} A_x\right)^2\right] &= \mathbb{E}\left[\frac{1}{n^2}\sum_{x\in\mathcal{T}_n} A_x^2 + \frac{1}{n^2}\sum_{x\in\mathcal{T}_n}\sum_{y\neq x\in\mathcal{T}_n} A_xA_y\right]\\ &= \mathbb{E}\left[\frac{1}{n^2}\sum_{x\in\mathcal{T}_n}\sigma^2_{(\deg(x)),\alpha} + m^2_{(\deg(x)),\alpha} + \frac{1}{n^2}\sum_{x\in\mathcal{T}_n}\sum_{y\neq x\in\mathcal{T}_n} m_{(\deg(x)),\alpha}m_{(\deg(y)),\alpha}\right]\\ &= \mathbb{E}\left[\frac{1}{n^2}\sum_{x\in\mathcal{T}_n}\sigma^2_{(\deg(x)),\alpha} + \frac{1}{n^2}\sum_{x\in\mathcal{T}_n} m_{(\deg(x)),\alpha}\sum_{y\in\mathcal{T}_n} m_{(\deg(y)),\alpha}\right].\end{aligned}$$

Using now the quenched Benjamini–Schramm convergence, we have

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{x\in\mathcal{T}_n} A_x\right)^2\right] \underset{n\to\infty}{\to} 0 + \mathbb{E}\left[m_{(\deg(\rho)),\alpha}\right]^2 = \alpha^2.$$

Lastly, using Bienaymé–Tchebychev's inequality, we have for all $\varepsilon>0$,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{x\in\mathcal{T}_n} A_x - \alpha\right| > \varepsilon\right) \leqslant \frac{\mathbb{E}\left[\left(\frac{1}{n}\sum_{x\in\mathcal{T}_n} A_x\right)^2\right] - \mathbb{E}\left[\frac{1}{n}\sum_{x\in\mathcal{T}_n} A_x\right]^2}{\left(\varepsilon + \left|\mathbb{E}\left[\left(\frac{1}{n}\sum_{x\in\mathcal{T}_n} A_x\right)\right] - \alpha\right|\right)^2} \underset{n\to\infty}{\to} 0.$$

On the other hand, we can show similarly that for all $R>0$,

$$\frac{1}{n}\sum_{x\in\mathcal{T}_n}\mathbb{1}_{x \text{ contains a car coming from distance}\leqslant R} \underset{n\to\infty}{\overset{(\mathbb{P})}{\to}} \mathbb{P}\left(\rho_\infty \text{ contains a car coming from distance} \leqslant R \text{ in } \mathcal{T}_\infty\right).$$

Moreover, since we assume that $\mathcal{T}_\infty$ has almost surely one spine, and this spine is "below" the root ρ_∞, for all $\varepsilon>0$, we can find $R_\varepsilon\geqslant 0$ such that with probability at least $1-\varepsilon$ the part "above" the root ρ_∞ in $\mathcal{T}_\infty$ has total height $\leqslant R_\varepsilon$. Thus,

$$\left|\mathbb{P}\left(\rho_\infty \text{ contains a car coming from distance} \leqslant R_\varepsilon\right) - \mathbb{P}\left(\rho_\infty \text{ contains a car}\right)\right| \leqslant \varepsilon.$$

Since the function $\mathbf{t} \mapsto \mathbb{1}_{\{\text{the part above the root of } \mathbf{t} \text{ has height } \leqslant R_\varepsilon\}}$ is continuous (and bounded) for the local convergence, we deduce from the Benjamini–Schramm convergence of $(\mathcal{T}_n)$ that we can find n_0 such that for all $n \geqslant n_0$,

$$\mathbb{P}\left(\frac{1}{n}\sum_{x\in\mathcal{T}_n} \mathbb{1}_{\text{the part above } x \text{ has height} > R_\varepsilon \text{ in } \mathcal{T}_n} \leqslant \varepsilon\right) \geqslant 1 - 2\varepsilon.$$

Thus,

$$\mathbb{P}\left(\frac{1}{n}\sum_{x\in\mathcal{T}_n} \mathbb{1}_{x \text{ contains a car coming from height} > R_\varepsilon} \leqslant \varepsilon\right) \geqslant 1 - 2\varepsilon.$$

We obtain

$$\frac{1}{n}\sum_{x\in\mathcal{T}_n} \mathbb{1}_{x \text{ contains a car}} \xrightarrow[n\to\infty]{(\mathbb{P})} \mathbb{P}\left(\rho_\infty \text{ contains a car}\right),$$

and this, together with the first part, shows that

$$\frac{\varphi_\alpha(\mathcal{T}_n)}{n} \xrightarrow[n\to\infty]{(\mathbb{P})} \alpha - \mathbb{P}\left(\rho_\infty \text{ contains a car}\right).$$

Note that if $\alpha > 1$, then the right-hand side is necessarily a positive constant. Now, it suffices to show that the almost sure existence of an infinite cluster of cars in $\mathcal{T}_\infty$ is equivalent to the fact that $\alpha > \mathbb{P}\left(\rho_\infty \text{ contains a car}\right)$. We introduce the transport function

$$f(\mathbf{t}, x, y) = \mathbb{1}_{\{\text{a car arriving on } x \text{ parks on } y \text{ in } \mathbf{t}\}}.$$

The main difficulty is to give a proper definition of this function. To do this, we fix a relative ordering between the vertices which is coherent layers by layers, and we imagine that each car need a time 1 one to go through each edge.

For $\alpha \leqslant \alpha_c$, we just park the cars time by time, which means that a car coming from x parks at y if x is the first vertex in the ordering among the arriving vertices of the cars arriving at first time at y, see Figure 4.1. If $\alpha > \alpha_c$, we introduce a little subtlety and apply this rule twice. More precisely, we introduce $\alpha_c < \alpha' := (\alpha + \alpha_c)/2 < \alpha$ and decompose for every $x \in \mathbf{t}$ we decompose the car arrival $A_x \sim \mu_{(\deg(x)),\alpha}$ at the vertex x so that

$$A_x = A_x^{(1)} + A_x^{(2)},$$

where $A_x^{(1)}$ has law $\mu_{(\deg(x)),\alpha'}$ and $A_x^{(2)} \geqslant 0$ almost surely. Then, to determine the value of $f(\mathbf{t}, x, y)$, we first park the cars $(A_x^{(1)}, x \in \mathbf{t})$ (with the above rule) and then the cars $(A_x^{(2)}, x \in \mathbf{t})$. Notice that given x and y two vertices of $\mathbf{t}$, the value of $f(\mathbf{t}, x, y)$ only depends on the subtree above y (and its value is 0 if x is not in this subtree), which is always almost surely finite.

We now apply the mass transport principle to this function:

$$\mathbb{E}\left[\sum_{x\in\mathcal{T}_\infty} f(\mathcal{T}_\infty, x, \rho_\infty)\right] = \mathbb{E}\left[\sum_{x\in\mathcal{T}_\infty} f(\mathcal{T}_\infty, \rho_\infty, x)\right].$$

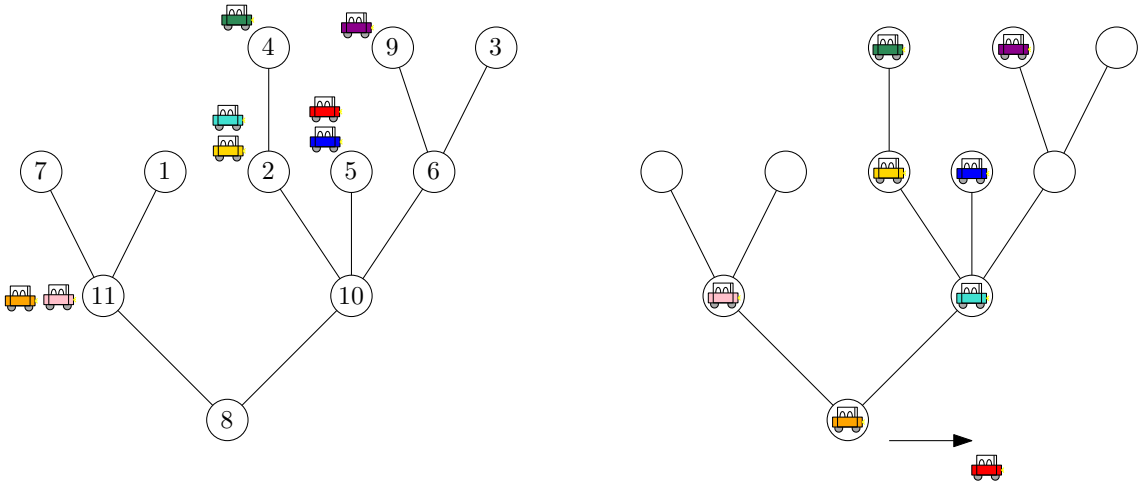


Figure 4.1: On the left, a rooted tree with a car arrival configuration together with a relative ordering of the vertices which defines a priority rule for parking. At time 0, we park one car at each vertex (if there are cars arriving on it). At time 1, we allow cars driving through one edge towards the root. For exemple, the orange car parks on the root vertex. If there are cars coming from two different vertices arriving at the same vertex, then we use the priority rule. For exemple the turquoise car parks on the vertex with priority 10 and not the red car. Note that the red car can not park since the orange car arrive at the root before it.

On the left-hand side, we simply have $\mathbb{E}\left[\sum_{x\in\mathcal{T}_\infty} f(\mathcal{T}_\infty, x, \rho_\infty)\right] = \mathbb{P}\left(\rho_\infty \text{ contains a car}\right)$. On the right-hand side, we have

$$\mathbb{E}\left[\sum_{x\in\mathcal{T}_\infty} f(\mathcal{T}_\infty, \rho_\infty, x)\right] = \mathbb{E}\left[\text{number of cars arriving at } \rho_\infty \text{ which park on } \mathcal{T}_\infty\right].$$

If there is no infinite cluster almost surely, then all cars park, thus this quantity is simply $\mathbb{E}\left[A_{\rho_\infty}\right] = \alpha$. This proves our theorem when $\alpha < \alpha_c$. Otherwise i.e. when $\alpha > \alpha_c$, we use a sprinkling argument : there is a chance that there is an infinite cluster and that ρ_∞ belongs to this cluster, and more than that, there is a positive probability that $A^{(1)}_{\rho_\infty} < A_{\rho_\infty}$ and ρ_∞ is in an infinite cluster even with the car arrivals $(A^{(1)}_x, x \in \mathbf{t})$ (with α'. In this case, there is no vertex x such that the cars corresponding to $A^{(2)}_{\rho_\infty}$ park on x. Thus $\mathbb{E}\left[\sum_{x\in\mathcal{T}_\infty} f(\mathcal{T}_\infty, \rho_\infty, x)\right] < \alpha$, which concludes the proof.

□

This theorem shows a phase transition for the flux of outgoing cars between a sublinear and a linear regime, which coincides with the existence of an infinite cluster of parked cars with positive probability. We wonder if it exists an infinite cluster of cars at criticality. We believe that there exists a tailor-made non-unimodular graph for which there exists an infinite cluster at α_c.

Open Question. Is there a unimodular tree $\mathcal{T}_\infty$ such that the probability of the existence of an infinite cluster of cars in $\mathcal{T}_\infty$ is positive at $\alpha = \alpha_c$?

Chapitre 5 :

Sharpness of the phase transition for parking on random trees

Les résultats de ce chapitre sont issus de l'article [64] publié dans Random Structures and Algorithm.

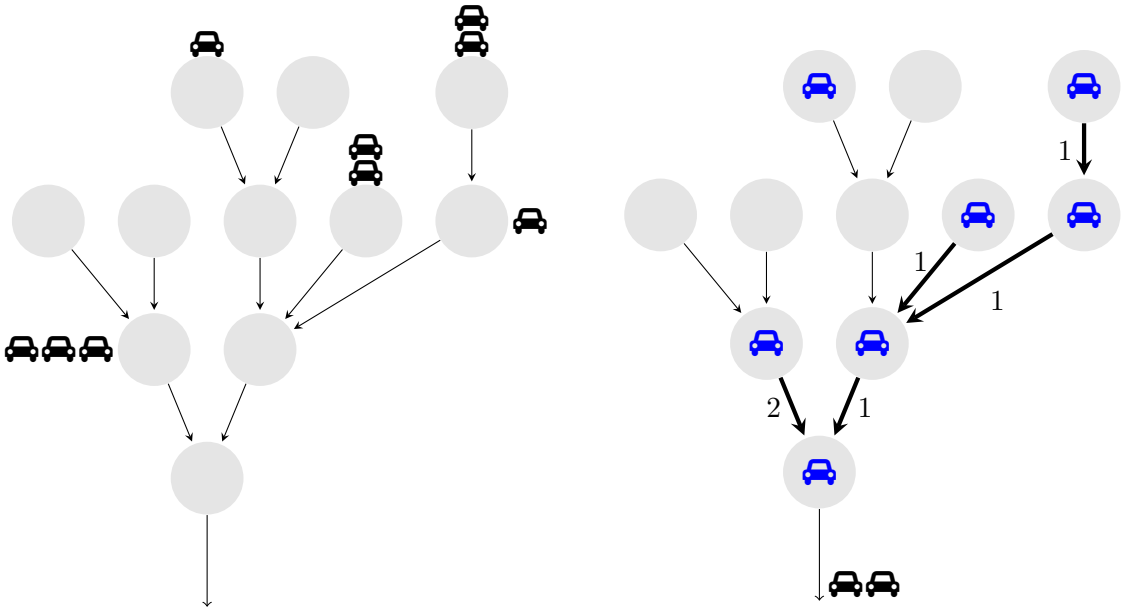


Figure 5.1: Illustration of the parking process of 9 cars on a tree. On the left: a rooted tree together with a configuration of cars trying to park. On the right, the resulting parking configuration with flux on the edges and with two cars which did not manage to park on the tree.

Recently, a phase transition phenomenon has been established for parking on random trees in [60, 100, 111, 137, 148]. We extend the results of [75] on general Bienaymé–Galton–Watson trees and allow different car arrival distributions depending on the vertex outdegrees. We then prove that this phase transition is *sharp* by establishing exponential bounds for the flux of exiting cars. This has consequences on the offcritical geometry of clusters of parked spots which displays similarities with the classical Erdős-Renyi random graph model.

Contents

5.1 Introduction

Parking functions on the line are combinatorial objects first introduced by Konheim and Weiss in [122] in the context of collision in hashing functions. Since then, many generalizations of the parking procedure have been studied, most notably on plane trees. On critical Bienaymé–Galton–Watson trees with i.i.d. car arrivals on the vertices, Curien and Hénard proved in [75] that the parking procedure undergoes a phase transition: when the "density" of cars is small, then the probability that all cars can park is large, whereas when the density is too large, then with high probability, there is at least one car that will not manage to park. This transition was first observed by Lackner and Panholzer [137], then by Goldschmidt and Przykucki [100] on Cayley trees with cars arriving uniformly on the vertices. Other particular cases have been studied in [60, 148, 148]. The phase transition was also proved in related models: see [22, 100] for the case of supercritical Bienaymé–Galton–Watson trees and [61] for a similar framework on regular trees.

In this work, we generalize the results of [75] by allowing the distributions of car arrivals to depend on the vertex's outdegree. But most importantly we show that this phase transition is *sharp*. Establishing sharpness of phase transition in statistical mecanics models is a crucial step in the understanding of the transition and in particular offcritical regimes, see [5, 80]. In our case, this sharpness will appear as exponential bounds for the flux of cars in the subcritical and supercritical cases (see Theorem 5.2) and will have direct consequences on the geometry of clusters in these regimes (Corollary 1).

General Phase Transition. Let us first recall the parking procedure on a rooted tree $\mathbf{t}$ i.e. a tree with a distinguished vertex called *the root* and denoted by $\varnothing$. We assume that the edges of $\mathbf{t}$ are oriented towards the root and consider the vertices of $\mathbf{t}$ as parking spots. Imagine now that cars arrive one after the other on the vertices of $\mathbf{t}$. Each car tries to park at its arriving parking spot. If the parking spot is empty, the car stops there. If not, the driver follows the edges towards the root and takes the first available space, if there is one. If not, the car leaves without parking (see Figure 5.1). An important property of this model is its Abelian property: changing the order of the car arrivals does not affect the final configuration and the number of cars that exit the tree.

We consider here a slightly more general model than in [75] by allowing the law of car arrivals to depend on the outdegree of the vertex. Specifically, given a rooted tree $\mathbf{t}$, we suppose that the arrivals of the cars on each vertex of $\mathbf{t}$ are independent random variables and that their law only depends on the outdegree of the vertex i.e. its number of children. In what follows, we simply write degree instead of outdegree (note the special role of the root vertex). We denote by $(L_x : x \in \mathbf{t})$ the car arrivals on the vertices of $\mathbf{t}$. The common law of the arrival of cars on a vertex of degree k is

$$\mu_{(k)} \text{ with mean } m_{(k)} \geqslant 0 \text{ and finite variance } \sigma^2_{(k)}. \tag{5.1}$$

In this chapter we shall only deal with (rooted) *plane* trees (i.e. such that the children of a given vertex are ranked from left to right), which are versions of critical Bienaymé–Galton–Watson tree with offspring distribution

$$\nu = \sum_{k=0}^{\infty} \nu_k \delta_k \text{ with mean 1 and finite variance } \Sigma^2, \tag{5.2}$$

the classical Bienaymé–Galton–Watson tree $\mathcal{T}$, the Bienaymé–Galton–Watson tree conditioned to have n vertices $\mathcal{T}_n$[1] and the Bienaymé–Galton–Watson tree conditioned to survive forever $\mathcal{T}_\infty$. We assume throughout the chapter that the number of cars arriving on a "typical" vertex has exponential tails, i.e.

$$F(z) = \sum_{k \geqslant 0} \nu_k \sum_{i \geqslant 0} \mu_{(k)}(i) z^i \tag{H_{exp}}$$

has a radius of convergence strictly larger than 1. We also suppose that $\nu \neq \delta_1$ and that there exists $k \geqslant 1$ such that $\nu_k > 0$ and $\mu_{(k)} \neq \delta_1$. Let $\varphi(\mathbf{t})$ be the flux of the parking process on $\mathbf{t}$ i.e. the number of exiting cars. Given a vertex x of $\mathbf{t}$, we sometimes denote by $\varphi_x(\mathbf{t})$ the flux at vertex x of $\mathbf{t}$, i.e. the outgoing flux of the parking process on $\mathsf{Top}(\mathbf{t}, x)$ the subtree of the descendants of x in $\mathbf{t}$. To characterize the location of the phase transition, we introduce the size-biased distribution $\overline{\nu}(k) = k\nu_k$ for $k \geqslant 1$ and the quantities

$$\mathbb{E}_{\overline{\nu}}[m] := \sum_{k=0}^{\infty} k\nu_k\, m_{(k)}, \quad \mathbb{E}_{\nu}[m] := \sum_{k=0}^{\infty} \nu_k\, m_{(k)} \quad \text{and} \quad \mathbb{E}_{\nu}[\sigma^2 + m^2 - m] := \sum_{k=0}^{\infty} \nu_k \left(\sigma^2_{(k)} + m^2_{(k)} - m_{(k)}\right).$$

[1] In all this chapter, we shall implicitly restrict to the values of n for which $\mathbb{P}(|\mathcal{T}| = n) > 0$.

Theorem 5.1 (*Phase transition for parking*)

We assume $\mathbb{E}_{\overline{\nu}}[m] \leqslant 1$. The parking process on Bienaymé–Galton–Watson tree undergoes a phase transition which depends on the sign of the quantity

$$\Theta := (1 - \mathbb{E}_{\overline{\nu}}[m])^2 - \Sigma^2 \mathbb{E}_{\nu}[\sigma^2 + m^2 - m]. \tag{5.3}$$

More precisely, we have three regimes classified as follows:

	subcritical $\Theta > 0$	*critical* $\Theta = 0$	*supercritical* $\Theta < 0$
$\varphi(\mathcal{T}_n)$ *as* $n \to \infty$	*converges in law*	$\xrightarrow[n\to\infty]{(\mathbb{P})} \infty$ *but is* $o(n)$	$\sim cn$ *with* $c > 0$
$\Sigma^2 \mathbb{E}[\varphi(\mathcal{T})] + \mathbb{E}_{\overline{\nu}}[m] - 1$	$-\sqrt{\Theta}$	0	∞
$\mathbb{P}$ ($\varnothing$ *is parked in* $\mathcal{T}$)	$\mathbb{E}_{\nu}[m]$	$\mathbb{E}_{\nu}[m]$	$\mathbb{E}_{\nu}[m] - c$

As an example of application of Theorem 5.1, if the cars can arrive only on the leaves with law $\mu_{(0)}$, then the phase transition occurs for $\Theta_{\text{leaf}} = 1 - \Sigma^2 \nu_0(\sigma^2_{(0)} + m^2_{(0)} - m_{(0)})$, where $m_{(0)}$ and $\sigma^2_{(0)}$ are respectively the expectation and the variance of the number of arrivals at a leaf. However, if the cars arrive with the same global density but spreaded on every vertex i.e. if the distribution of the car arrivals is $\mu = \nu_0 \mu_{(0)} + (1 - \nu_0)\delta_0$ and does not depend on the degree of the vertex, then $\Theta_{\text{unif}} = (1 - \nu_0 m_{(0)})^2 - \Sigma^2 \nu_0(\sigma^2_{(0)} + m^2_{(0)} - m_{(0)}) \leqslant \Theta_{\text{leaf}}$. This means that with the same density of cars, the parking can be subcritical if the cars arrive only on the leaves but supercritical if the cars arrive uniformly on every vertex.

A natural assumption for the parking process to be subcritical is that $\mathbb{E}_{\nu}[m] \leqslant 1$, so that there are typically fewer cars than parking spots. The assumption $\mathbb{E}_{\overline{\nu}}[m] \leqslant 1$ may sound unnatural but comes for the fact that the number of children of the vertices in a "typical" branch has size-biased law $\overline{\nu}$ [100]. Indeed, the parking process is supercritical when $\mathbb{E}_{\overline{\nu}}[m] > 1$, as we will see in Section 5.2.1 (see the remark before Proposition 5.2). As a consequence of this phase characterization, we can deduce that if $\mathbb{E}_{\overline{\nu}}[m] \leqslant 1$ and $\Theta > 0$, the parking process is subcritical and therefore $\mathbb{E}_{\nu}[m] \leqslant 1$, and conversly, if $\mathbb{E}_{\overline{\nu}}[m] \leqslant 1$ and $\mathbb{E}_{\nu}[m] > 1$, then the parking process is supercritical and hence $\Theta < 0$. However these implications are not derived by easy algebraic manipulations.

As we said above, Theorem 5.1 generalizes the result of [75] and its proof follows the same lines. It is presented in Section 5.2.

Sharpness and exponential bounds. Our main contribution in this chapter consists in showing that the phase transition established in Theorem 5.1 is sharp. More precisely, we shall reinforce the first line in the table of Theorem 5.1 by proving exponential bounds:

Theorem 5.2 (*Exponential bounds for the flux*)

Let $\varepsilon > 0$. In the supercritical regime i.e. if $\Theta < 0$, there exists $\delta > 0$, and $n_0 \geqslant 0$ such that for all $n \geqslant n_0$,

$$\mathbb{P}\left(|\varphi(\mathcal{T}_n) - cn| \geqslant \varepsilon n\right) \leqslant e^{-\delta n},$$

where $c > 0$ is as in Theorem 5.1. In the subcritical regime, there exists $\delta > 0$, and $n_0 \geqslant 0$ such that for all $n \geqslant n_0$,

$$\mathbb{P}\left(|\varphi(\mathcal{T})| \geqslant \varepsilon n\right) \leqslant e^{-\delta n}.$$

Notice that the second item of Theorem 5.2 applies to the unconditioned tree $\mathcal{T}$. It holds also for $\mathcal{T}_n$ after changing the constants since $\mathbb{P}(|\mathcal{T}| = n)$ has polynomial probability. Our proof of Theorem 5.2 is very different in the supercritical and subcritical cases. In the supercritical case, the exponential bounds will be established by showing first exponential bounds for the fringe subtree distribution of $\mathcal{T}_n$ in Section 5.2.2. This may be a result of independent interest which complements the law of large numbers and the Central Limit Theorem of Aldous [7] and Janson [106, 105]. In the subcritical case, we adopt a very different analytic point of view. Following [100] the flux at the root of $\mathcal{T}$ satisfies a recursive distributional equation which turns into an analytic equation on its generating function $z \mapsto W(z)$. However, the equation has a singularity at $z = 1$ and $W(1) = 1$. By employing Newton–Puiseux expansion we are able to resolve this singularity and prove that in the subcritical case $z \mapsto W(z)$ has radius of convergence strictly larger than 1. This is the object of Section 5.3.

Offcritical geometry. We will give an application of this exponential decay for the flux to the size of the connected components after the parking procedure, that is the clusters of occupied parking spots in $\mathcal{T}_n$ in the subcritical and supercritical phases. We notice that this geometry shares many similarities with the size of the connected components of the Erdős-Renyi random graph: only logarithmic clusters in the subcritical case and a giant component in the supercritical phase. This is actually not a mere coincidence and in forthcoming works we shall exhibit a strong link between parking on random trees and random graph processes [66, 67].

Corollary 1 (Offcritical geometry)**.** *Let $|C_{\max}(n)|$ be the size of the largest parked connected component in $\mathcal{T}_n$, and $|C_2(n)|$ be the size second largest connected component. Then,*

$$\textit{(supercritical } \Theta < 0\textit{)} \qquad \frac{|C_{\max}(n)|}{n} \xrightarrow[n\to\infty]{(\mathbb{P})} C \quad \textit{and} \quad \mathbb{P}(|C_2(n)| \geqslant A\ln(n)) \underset{n\to\infty}{\to} 0,$$

$$\textit{(subcritical } \Theta > 0\textit{)} \qquad \mathbb{P}(|C_{\max}(n)| \geqslant A\ln(n)) \underset{n\to\infty}{\to} 0,$$

where $C \in (0,1)$ and $A > 0$ are constants that depend on the laws ν and $\mu_{(k)}$ for $k \geqslant 0$.

Acknowledgments. I would like to thank warmly Nicolas Curien for precious suggestions and corrections. I am also very grateful to Olivier Hénard for many interesting discussions. I also thank the two anonymous referees for their valuable suggestions that helped clarifying the paper.

5.2 Phase transition and fringe subtrees

In this section, we generalize the phase transition result of [75] to our case, that is when the distribution of the car arrivals depends on the degree of the vertex. The strategy of proof is very similar and we will only highlight the necessary adaptations. The crux is the adaptation of [75, Proposition 1] into Proposition 5.1.

5.2.1 The mean flux and the probability that the root is parked in $\mathcal{T}$

We first obtain the expected outgoing flux of the unconditioned Bienaymé–Galton–Watson tree using a differential equation. To this purpose, we let the cars arrive according to random times A_x uniform in $[0,1]$ independently on each vertex x of the tree $\mathcal{T}$. More precisely, conditionally on $\mathcal{T}$, we define a family $(A_x)_{x\in\mathcal{T}}$ of i.i.d. random variables with law $\mathsf{Unif}[0,1]$ independently of car arrivals $(L_x)_{x\in\mathcal{T}}$. We denote by $\varphi(\mathcal{T},t)=\varphi(t)$ the outgoing flux on the root of $\mathcal{T}$ after the parking procedure with car arrivals $L_x^{(t)}=\mathbb{1}_{A_x\leqslant t}L_x$ on each vertex $x\in\mathcal{T}$ conditionally on $\mathcal{T}$. Note that conditionally on $\mathcal{T}$, the car arrivals $(L_x^{(t)})_{x\in\mathcal{T}}$ are independent with law $\mu_{(k)}^{(t)}=(1-t)\delta_0+t\mu_{(k)}$ if x has $k\geqslant 0$ children.

Proposition 5.1 (Phase transition for the mean flux)**.** *For $t\in[0,1]$, we denote by $\Phi(t)=\mathbb{E}[\varphi(\mathcal{T},t)]$ the mean flux at the root of $\mathcal{T}$ with car arrivals with law $\mu_{(k)}^{(t)}$. Let $t_{\min}$ be the smallest solution to $(1-\mathbb{E}_{\overline{\nu}}[m]t)^2=\Sigma^2\mathbb{E}_\nu[\sigma^2+m^2-m]t$ in $[0,1]$ (set $t_{\min}=+\infty$ if there is no such solution). Then, for $t\in[0,1]$*

$$\Phi(t)=\begin{cases}\dfrac{(1-\mathbb{E}_{\overline{\nu}}[m]t)-\sqrt{(1-\mathbb{E}_{\overline{\nu}}[m]t)^2-\Sigma^2\mathbb{E}_\nu[\sigma^2+m^2-m]t}}{\Sigma^2} & \text{if}\quad t\leqslant t_{\min}\\ +\infty & \text{if}\quad t>t_{\min}.\end{cases}\tag{5.4}$$

Proof. We use the same notation as in [75, Proposition 1]: if x is a vertex of $\mathcal{T}$, we denote by $I^x(s)$ the number of cars that arrived at time s on the vertex x which contribute to $\varphi(t)$, i.e. those that did not manage to park at their arrival time $s\leqslant t$. For $t\in[0,1]$, we have

$$\Phi(t)=\mathbb{E}\left[\sum_{x\in\mathcal{T}}I^x(A_x)\mathbb{1}_{0\leqslant A_x\leqslant t}\right]=\mathbb{E}\left[\sum_{x\in\mathcal{T}}\sum_{k=0}^{\infty}I^x(A_x)\mathbb{1}_{0\leqslant A_x\leqslant t}\mathbb{1}_{\{x\text{ has degree }k\}}\right]$$

Recall that the degree of a vertex x in a tree $\mathbf{t}$ is a function of $\mathsf{Top}(x,\mathbf{t})$ the subtree of the descendants of x. We use the many-to-one formula (see e.g. [75, Formula 3]) and integrate on $s=A_x$ to obtain

$$\Phi(t)=\int_0^t \mathrm{d}s\sum_{k=0}^{\infty}\sum_{h=0}^{\infty}\mathbb{E}\left[I(s,h)\mathbb{1}_{\{S_h\text{ has degree }k\}}\right],$$

where $I(s,h)$ is obtained as follows: First recall the construction of Kesten tree $\mathcal{T}_\infty$. Consider a semi-infinite line $S_0,S_1,\dots$, rooted at S_0, called the *spine*, and graft independently on each S_i a random number $Y-1$ of independent Bienaymé–Galton–Watson trees with offspring distribution μ where $Y\sim\overline{\nu}$, and consider a random uniform ordering of the children of S_i. Here, we define a tree $\mathcal{T}(h)$,

for $h \geqslant 0$, by considering only a finite line $S_0, S_1, \dots, S_h$ and grafting independently on each S_i a random number $Y-1$ of independent Bienaymé–Galton–Watson trees where $Y \sim \overline{\nu}$ for $0 \leqslant i < h$, and consider a random uniform ordering of the children of S_i and furthermore X independent copies of $\mathcal{T}$ on S_h where $X \sim \nu$ (see Figure 5.2). This tree is decorated by letting cars arrive with law $\mu^{(s)}_{(l)}$ at each vertex of degree l independently, except on the vertex S_h where we put an independent number of cars distributed as $\mu_{(k)}$ (instead of $\mu^{(s)}_{(k)}$) when S_h has degree k. Then $I(s,h)$ is the number of those cars arriving on S_h that do not manage to park after all other cars of $\mathcal{T}(h)$ have parked.

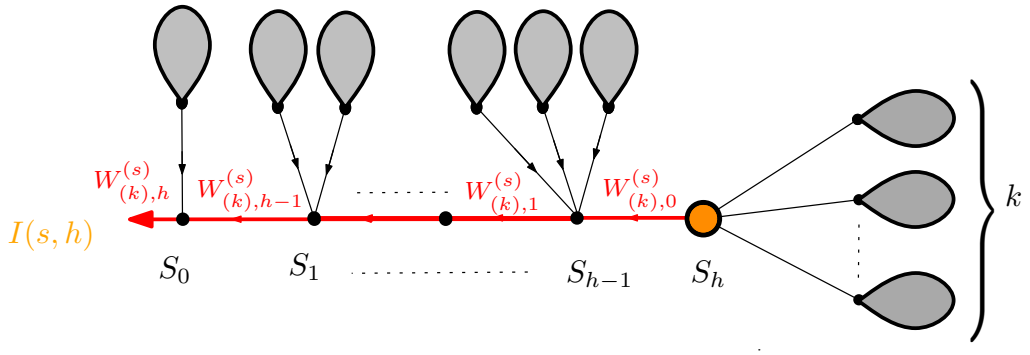


Figure 5.2: The tree $\mathcal{T}(h)$ conditioned by $\{S_h \text{ has degree } k\}$.

To compute $\mathbb{E}\left[I(s,h)\mathbb{1}_{S_h \text{ has degree } k}\right]$, we use the fact that at time s and on the event $\{S_h$ has degree $k\}$, the outgoing flux from the vertex S_{h-i} (before parking the cars arriving on S_h) is given by a random walk $W^{(s)}_{(k),i}$ minus its current infimum, where the random walk $W^{(s)}_{(k),i}$ has length h and i.i.d. increments of law $Z^{(s)}-1$ with

$$Z^{(s)} = \sum_{i=1}^{Y-1} F^{(s)}_i + P^{(s)}_{(Y)}$$

where $Y \sim \overline{\nu}$, the $F^{(s)}_i$ are copies of $F^{(s)} \sim \varphi(\mathcal{T}, s)$ for $i \geqslant 0$, the $P^{(s)}_{(k)}$ have law $\mu^{(s)}_{(k)}$ for $k \geqslant 0$ and all the variables are independent. The starting point of the random walk $W^{(s)}_{(k),0}$ is distributed as the sum of k independent copies of $\varphi(\mathcal{T}, s)$ minus 1. We define $T^{(s)}_{(k),-1}$ to be the first hitting time of -1 by the walk $W^{(s)}_{(k)}$ and we write $\mathbb{P}_x$ for the law of $W^{(s)}_{(k)}$ started at x and $\mathbb{E}_x$ for the corresponding expectation, for $x \in \mathbb{Z}$. When specifying its starting point, the random walk $W^{(s)}_{(k)}$ does not depend on k so that we simply write $W^{(s)}$ or $T^{(s)}_{-1}$. Summing over h, we obtain in the same way as in [75, Proposition 1]

$$\sum_{h=0}^{\infty} \mathbb{E}\left[I(s,h)\mathbb{1}_{S_h \text{ has degree } k}\right] = \nu_k\left(\frac{1}{2}(\sigma^2_{(k)} + m^2_{(k)} - m_{(k)}) + k m_{(k)}\Phi(s)\right)\mathbb{E}_0\left[T^{(s)}_{-1}\right].$$

On the one hand, if $\mathbb{E}[Z^{(s)}-1] \geqslant 0$, the random walk $W^{(s)}_{(k)}$ has a nonnegative drift, hence $\mathbb{E}_0\left[T^{(s)}_{-1}\right] = \infty$. On the other hand, if $\mathbb{E}[Z^{(s)}-1] < 0$, the random walk $W^{(s)}_{(k)}$ has a strictly negative drift and by Wald equality

$$\mathbb{E}_0\left[T^{(s)}_{-1}\right] = \frac{1}{\mathbb{E}[1-Z^{(s)}]} = \frac{1}{1-\mathbb{E}_{\overline{\nu}}[m]s - \Sigma^2\Phi(s)}.$$

Summing over $k \geqslant 0$, we obtain $\Phi(0) = 0$, and for all $t \leqslant t_c$,

$$\Phi(t) = \int_0^t \mathrm{d}s \frac{\frac{1}{2}\mathbb{E}_\nu[\sigma^2 + m^2 - m] + \mathbb{E}_{\overline{\nu}}[m]\Phi(s)}{1 - s\mathbb{E}_{\overline{\nu}}[m] - \Sigma^2\Phi(s)},$$

where $t_c = \inf\{t \geqslant 0 : 1 - \mathbb{E}_{\overline{\nu}}[m]t - \Sigma^2\Phi(t) < 0\}$. We can easily check that the function defined on the right-hand side of (5.4) satisfies this equation. It remains to check that both functions "blow up" at the same time $t_{\min} = t_c$. This is done in the proof of Proposition 1 in [75] using monotone and dominated convergence.

□

When $\mathbb{E}_{\overline{\nu}}[m] \leqslant 1$, then $t \mapsto (1 - \mathbb{E}_{\overline{\nu}}[m]t)^2 - \Sigma^2\mathbb{E}_\nu[\sigma^2 + m^2 - m]t$ is decreasing over $[0, 1]$. Hence, we obtain the following phase characterization for parking: when $t_{\min} < 1$ then $\Theta < 0$ (supercritical regime), when $t_{\min} = 1$ then $\Theta = 0$ (critical regime) and when $t_{\min} > 1$ then $\Theta > 0$ (subcritical regime).

Remark. When $\mathbb{E}_{\overline{\nu}}[m] > 1$, the function $t \mapsto (1 - \mathbb{E}_{\overline{\nu}}[m]t)^2 - \Sigma^2\mathbb{E}_\nu[\sigma^2 + m^2 - m]t$ is positive at $t = 0$ and negative at $t = 1/\mathbb{E}_{\overline{\nu}}[m] < 1$, so that $t_{\min} < 1/\mathbb{E}_{\overline{\nu}}[m] < 1$, and the parking process is supercritical.

Using the same technique, we now control the probability that the root of an unconditioned Bienaymé–Galton–Watson tree is parked as in [75, Proposition 2].

Proposition 5.2 (Probability that the root is parked)**.** *We have*

$$\mathbb{P}\left(\varnothing \textit{ is parked in } \mathcal{T}\right) = \begin{cases} \mathbb{E}_\nu[m] & \textit{if} \quad \Theta \geqslant 0, \\ < \mathbb{E}_\nu[m] & \textit{if} \quad \Theta < 0. \end{cases} \tag{5.5}$$

Proof. We proceed as in the previous proposition and let the cars arrive according to random times A_x independently for each vertex x on the tree $\mathcal{T}$. Let $p_t = \mathbb{P}(\varnothing \text{ contains a car in } (\mathcal{T}, t))$. Then,

$$p_t = \int_0^t \mathrm{d}s \sum_{h \geqslant 0} \mathbb{P}(P(s,h))$$

where $P(s,h)$ is the event that in the labeled tree $\mathcal{T}(h)$, one car arriving on the vertex S_h at time s goes down the spine and manages to park on the empty root $\varnothing$. Moreover, conditionally on $\{S_h \text{ has } k \text{ children}\}$, the event $P(s,h)$ is $\cup_{i=1}^{L_{(k)}} \left\{T_{-i}^{(s)} = h\right\}$ under $\mathbb{P}_{W_{(k),0}^{(s)}}$ where $L_{(k)} \sim \mu_{(k)}$. Thus,

$$\begin{aligned} \sum_{h=0}^{\infty} \mathbb{P}(P(s,h) \cap \{S_h \text{ has } k \text{ children}\}) &= \nu_k \mathbb{E}\left[\sum_{i=1}^{L_{(k)}} \mathbb{P}_{W_0^{k,(s)}}(T_{-i}^{(s)} < \infty)\right] \\ &= \nu_k \mathbb{E}\left[\sum_{i=1}^{L_{(k)}} \mathbb{P}_0(T_{-1}^{(s)} < \infty)^{W_0^{k,(s)}+i}\right] \\ &= \begin{cases} \nu_k m_{(k)} & \text{if} \quad s \leqslant t_{\min} \\ < \nu_k m_{(k)} & \text{if} \quad s > t_{\min}. \end{cases} \end{aligned}$$

Summing over k and integrating over s, we get the desired result.

□

5.2.2 Fringe subtrees and weak law of large numbers for the flux

We now want to show a convergence result for the flux of the conditioned Bienaymé–Galton–Watson tree $\mathcal{T}_n$. To this end, a useful tool will be the Łukasiewicz walk (in the depth-first order) of the Bienaymé–Galton–Watson tree (see [129]), decorated with the car arrivals on each vertex. Therefore in the rest of the chapter, we consider (S, L) a random process where S is a random walk with starting point $S_0 = 0$ and i.i.d. increments of law $\mathbb{P}(S_1 = k) = \nu_{k+1}$ for $k \geqslant -1$, and conditionally on (S), the $(L_i)_{i\geqslant 0}$ are independent of law $\mu_{(S_{i+1}-S_i+1)}$. Note that the $(L_i)_{i\geqslant 0}$ are i.i.d. and $\mathbb{E}[L_0] = \mathbb{E}_\nu[m]$. We also define T the first hitting time of -1 by the walk S. Then the law of $(S_i, L_i)_{0\leqslant i\leqslant n}$ conditionally on $\{T = n\}$ is the law of the Łukasiewicz walk of a Bienaymé–Galton–Watson tree $\mathcal{T}_n$ conditioned to have n vertices "decorated" with the car arrivals on each vertex. Therefore, we couple (S, L) with $\mathcal{T}_n$ and $\mathcal{T}$ so that $(S, L)_{0\leqslant i\leqslant T}$ (resp. $(S, L)_{0\leqslant i\leqslant n}$) is the Łukasiewicz walk of $\mathcal{T}$ (resp. $\mathcal{T}_n$) decorated with the car arrivals conditionally on T (resp. $\{T = n\}$).

Proposition 5.3 (Law of large numbers for the flux)**.** *The flux at the root of $\mathcal{T}_n$ satisfies*

$$\frac{\varphi(\mathcal{T}_n)}{n} \xrightarrow[n\to\infty]{(\mathbb{P})} \mathbb{E}_\nu[m] - \mathbb{P}\left(\varnothing \text{ is parked in } \mathcal{T}\right).$$

Proof. By conservation of cars, the total number of cars arriving on $\mathcal{T}_n$ is

$$\left(\sum_{x\in\mathcal{T}_n} L_x\right) = \varphi(\mathcal{T}_n) + \sum_{x\in\mathcal{T}_n} \mathbb{1}_{x \text{ is parked}}.$$

We first prove that the proportion of arriving cars per vertex in $\mathcal{T}_n$ converges in probability towards $\mathbb{E}_\nu[m]$ as n goes to ∞. Recall that we defined a random walk S with i.i.d increments of law ν together with L the car arrival "decoration" and T the first hitting time of -1, so that conditionally on $\{T = n\}$ and on $\mathcal{T}_n$, the car arrival on v_k, the kth vertex of $\mathcal{T}_n$ in the depth-first order, is $L_{v_k} = L_{k-1}$ and has law $\mu_{(S_k-S_{k-1}+1)}$ for $1 \leqslant k \leqslant n$. Then, for all $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\sum_{x\in\mathcal{T}_n} L_x - \mathbb{E}_\nu[m]n\right| \geqslant \varepsilon n\right) \leqslant \frac{\mathbb{P}\left(\left|\sum_{k=0}^{n-1}(L_k - \mathbb{E}_\nu[m])\right| \geqslant \varepsilon n\right)}{\mathbb{P}(T = n)}. \tag{5.6}$$

Since the L_k are i.i.d. and their common law has by assumption ($H_{\exp}$) an exponential tail, we can bound the above numerator by $e^{-\delta n}$ for some $\delta > 0$. Moreover we recall the classical asymptotic

$$\mathbb{P}(T = n) \sim Cn^{-3/2} \tag{5.7}$$

for some $C > 0$ as n goes to $+\infty$ (at least along values for which $\mathbb{P}(T = n) > 0$). Hence the probability on the left-hand side converges to 0 as n goes to ∞ and does so exponentially fast.

Then we observe that the degree of a vertex x in a given tree $\mathbf{t}$ is a function of $\mathrm{Top}(x, \mathbf{t})$, the subtree of the descendants of x. Therefore we can use the theorem of Janson [106, Theorem 1.3, formula (1.11)] which states that the fringe subtree distribution of a conditioned Bienaymé–Galton–Watson tree $\sum_{x\in\mathcal{T}_n} \delta_{\mathrm{Top}(x,\mathcal{T}_n)}/n$ converges in probability to the ν-Galton–Watson measure. Adding

car arrivals decoration on each vertex, this implies that

$$\frac{1}{n}\sum_{x\in\mathcal{T}_n}\mathbb{1}_{x \text{ is parked}} \xrightarrow[n\to\infty]{(\mathbb{P})} \mathbb{P}\left(\varnothing \text{ is parked in } \mathcal{T}\right).$$

The desired result follows.

□

We have seen above that the total number of cars arriving on $\mathcal{T}_n$ concentrates around its (unconditioned) expectation $\mathbb{E}_\nu[m]n$. We will see in the sequel that we can also obtain exponential bounds for the outgoing flux of cars but let us first sketch the proof of Theorem 5.1.

Sketch of Proof of Theorem 5.1. The second line of Theorem 5.1 can be easily derived from Proposition 5.1. Moreover Proposition 5.2 already gives us the third line of the table of Theorem 5.1 and together with Proposition 5.3, we obtain that the flux is linear when $\Theta < 0$ and sublinear when $\Theta \geqslant 0$. There only remains to check that the flux converges in law in the subcritical case and diverges in the critical case. The proof is an adapation of [75, Section 4] and in particular [75, Lemma 3], which shows that no car coming from far away in $\mathcal{T}_n$ contributes to $\varphi(\mathcal{T}_n)$.

□

To obtain exponential bounds for the flux (at least in the case of supercritical parking process), we also need a exponential bounds for the fringe subtree distribution. We establish such a result in a more general context, which concerns not only the subtree of the descendants of the vertices, but a more general local neighborhood. Let $\mathbf{t}$ be a plane tree, $x \in \mathbf{t}$ be a vertex of $\mathbf{t}$ and $k \geqslant 0$. We define $H_k(\mathbf{t}, x) = \mathsf{Top}(\mathbf{t}, x_k)$ where x_k is the kth ancestor of x if there is one. Otherwise, we just say $H_k(\mathbf{t}, x) = \Diamond$. When $\mathbf{t} = \mathcal{T}_n$ and $x = \mathfrak{u}_n$ is a uniform vertex of $\mathcal{T}_n$, Aldous [7] (see also Stufler [163]) has proved that $\mathcal{T}_n$, seen from the vertex $\mathfrak{u}_n$ converges in distribution (for the local topology) towards an infinite plane tree with almost surely one spine $\mathcal{T}^*$ called the random sin-tree: Consider $(u_k)_{k\geqslant 0}$ a semi-infinite path, "pointed" at u_0, such that u_{k+1} is the ancestor of u_k for all $k \geqslant 0$. Then graft independently on u_0 a random number X of independent Bienaymé–Galton–Watson trees where $X \sim \nu$ and on each u_k for $k \geqslant 1$ a random number $Y - 1$ of independent Bienaymé–Galton–Watson trees where $Y \sim \overline{\nu}$, and consider a random uniform ordering of the children of u_k. Note that for all $k \geqslant 0$, the subtree $\mathsf{Top}(\mathcal{T}^*, u_k) = H_k(\mathcal{T}^*, u_0)$ has law $\mathcal{T}(k)$ (see Figure 5.2). Aldous' sin-tree $\mathcal{T}^*$ is closely related to the Kesten tree $\mathcal{T}_\infty$: whereas $\mathcal{T}_\infty$ describes the local limit of $\mathcal{T}_n$ near the root vertex, $\mathcal{T}^*$ describes its local limit near a "typical" vertex.

Proposition 5.4 (Exponential bounds for the fringe subtrees)**.** *Let $\mathbf{t}$ be a (fixed) plane finite tree and $k \geqslant 0$ an integer such that the height of $\mathbf{t}$ is at least k. For every $\varepsilon > 0$, there exists $\delta > 0$ and $n_0 \geqslant 0$ such that for all $n \geqslant n_0$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{x\in\mathcal{T}_n}\mathbb{1}_{H_k(\mathcal{T}_n,x)=\mathbf{t}} - \mathbb{P}(H_k(\mathcal{T}^*, u_0) = \mathbf{t})\right| \geqslant \varepsilon\right) \leqslant e^{-\delta n},$$

Proof. We shall prove the result using the Łukasiewicz walk of $\mathcal{T}_n$ which is almost a random walk and for which exponential bounds for density of patterns is easy to see. Let $L^{\mathbf{t}} = (L_i^{\mathbf{t}})_{0\leqslant i\leqslant|\mathbf{t}|}$ be the Łukasiewicz walk of $\mathbf{t}$ and $(v_i)_{0\leqslant i\leqslant|\mathbf{t}|-1}$ be the vertices of $\mathbf{t}$ in the depth-first order, so that $L_{i+1}^{\mathbf{t}} - L_i^{\mathbf{t}} + 1$ is the number of children of v_i in $\mathbf{t}$ for $0 \leqslant i \leqslant |\mathbf{t}| - 1$. We denote by M the maximum of the increments of $L^{\mathbf{t}}$.

We extend the walk (S) as a bi-infinite walk by setting $S_j = 1$ for $j < 0$. For $j \in \mathbb{Z}$, we write $(S_i^{(j)} = S_{i+j} - S_j)_{i\in\mathbb{Z}}$ for the walk shifted at time j. We claim that there exists a function $f_{\mathbf{t}}$ defined over bi-infinite paths and taking values in $\{0,1\}$ such that conditionally on $\{T = n\}$,

$$H_k(\mathcal{T}_n, v_j) = \mathbf{t} \qquad \Longleftrightarrow \qquad f_{\mathbf{t}}(S^{(j)}) = 1, \tag{5.8}$$

where $(v_k)_{0\leqslant k\leqslant n-1}$ are the vertices of $\mathcal{T}_n$ listed in the depth-first order. Specifically, $f_{\mathbf{t}}$ is defined as follows: considering $W \in \mathbb{Z}^{\mathbb{Z}}$, we first define $\tau_0 = 0$, then $\tau_1 = \sup\{j < \tau_0, W_j \leqslant W_{\tau_0}\}$ if it exists ($-\infty$ otherwise), and so on up to $\tau_k = \sup\{j < \tau_{k-1}, W_j \leqslant W_{\tau_{k-1}}\}$, so that if $\tau_k > -\infty$, the τ_i's correspond to the locations of the ancestors of the vertex "0". We then set $f_{\mathbf{t}}(W) = 1$ if and only if $\tau_k \geqslant -|\mathbf{t}|$ and $(W_j^{(\tau_k)})_{0\leqslant j\leqslant|\mathbf{t}|} = L^{\mathbf{t}}$. In particular, the value of $f_{\mathbf{t}}(W)$ only depends on $(W_k)_{-|\mathbf{t}|\leqslant k\leqslant|\mathbf{t}|}$. More than that, if $\tau_k \geqslant -|\mathbf{t}|$ and $W_{k+1} - W_k > M$ for some $\tau_k \leqslant k \leqslant \tau_k + |\mathbf{t}|$, then $f_{\mathbf{t}}(W) = 0$ (and changing the values of W_k for $k \leqslant \tau_k$ or $k \geqslant \tau_k + |\mathbf{t}|$ does not change the value of $f_{\mathbf{t}}$). Therefore we consider the random walk $(\widetilde{S}_k)_{k\in\mathbb{Z}}$ such that $\widetilde{S}_k = S_k$ for $k \leqslant 0$ and $\widetilde{S}_{k+1} - \widetilde{S}_k = (S_{k+1} - S_k) \wedge (M+1)$, and the corresponding shifted walk $\widetilde{S}^{(j)}$.

Using (5.8), we obtain

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{n}\sum_{x\in\mathcal{T}_n} \mathbb{1}_{H_k(\mathcal{T}_n,x)=\mathbf{t}} - \mathbb{P}(H_k(\mathcal{T}^*,u_0)=\mathbf{t})\right| \geqslant \varepsilon\right)\\
&= \mathbb{P}\left(\left|\frac{1}{n}\sum_{j=0}^{n-1} f_{\mathbf{t}}(\widetilde{S}^{(j)}) - \mathbb{P}(H_k(\mathcal{T}^*,u_0)=\mathbf{t})\right| \geqslant \varepsilon \,\middle|\, T = n\right)\\
&\leqslant \frac{\mathbb{P}\left(\left|\frac{1}{n}\sum_{j=0}^{n-1} f_{\mathbf{t}}(\widetilde{S}^{(j)}) - \mathbb{P}(H_k(\mathcal{T}^*,u_0)=\mathbf{t})\right| \geqslant \varepsilon\right)}{\mathbb{P}(T=n)}.
\end{aligned}$$

Apart from the first $0 \leqslant j \leqslant |\mathbf{t}|$ values, the function $j \mapsto f_{\mathbf{t}}(\widetilde{S}^{(j)}) = f_{\mathbf{t}}((\widetilde{S}_k^{(j)})_{-|\mathbf{t}|\leqslant k\leqslant|\mathbf{t}|})$ is a function of the underlying Markov chain $((\widetilde{S}_k^{(j)})_{-|\mathbf{t}|\leqslant k\leqslant|\mathbf{t}|})_{j\geqslant 0}$. Furthermore, by Aldous [7, Proposition 10] we can see that $\mathbb{P}(H_k(\mathcal{T}^*,u_0)=\mathbf{t})$ is the probability that $f_{\mathbf{t}}(\widetilde{S}) = 1$ under the stationary distribution of the Markov chain $((\widetilde{S}^{(j)})_{-|\mathbf{t}|\leqslant k\leqslant|\mathbf{t}|})_{j\geqslant|\mathbf{t}|}$ which is simply the law of the two-sided random walk with increments $\nu_{(\cdot+1)\wedge(M+1)}$. Since $\nu_{(\cdot+1)\wedge(M+1)}$ has finite support, this is a Markov chain with finite state space, and Sanov's theorem [77, Section 6.2] implies an exponential decay for the numerator, which is bounded above by $e^{-\delta n}$ for some $\delta > 0$ for n large enough. Using (5.7), we get the desired result.

□

5.3 Exponential bounds for the parking process

5.3.1 Supercritical parking process

We are now able to prove Theorem 5.2 in the supercritical case, i.e when $\Theta < 0$.

Proof of Theorem 5.2, supercritical case. As in the proof of Proposition 5.3, by conservation of cars, we have

$$\varphi(\mathcal{T}_n) = \sum_{x\in\mathcal{T}_n} L_x - \sum_{x\in\mathcal{T}_n} \mathbb{1}_{x \text{ is parked}}.$$

Hence,

$$\mathbb{P}\left(|\varphi(\mathcal{T}_n) - cn| \geqslant \varepsilon n\right) \leqslant \mathbb{P}\left(\left|\sum_{x\in\mathcal{T}_n} L_x - \mathbb{E}_\nu[m]n\right| \geqslant \frac{\varepsilon}{2}n\right) + \mathbb{P}\left(\left|\sum_{x\in\mathcal{T}_n} \mathbb{1}_{x \text{ is parked}} - (\mathbb{E}_\nu[m] - c)\, n\right| \geqslant \frac{\varepsilon}{2}n\right).$$

Using the bound (5.6) for the total number of cars, the first term of the right-hand has exponential decay. We now want to have exponential bounds for the number of occupied parking spots. By Theorem 5.1, the probability $\mathbb{P}(\varnothing \text{ is parked in } \mathcal{T})$ is $\mathbb{E}_\nu[m] - c$. We therefore can choose $M \geqslant 0$ such that

$$\mathbb{P}(|\mathcal{T}| \leqslant M \text{ and } \forall x \in \mathcal{T}, L_x \leqslant M) \geqslant 1 - \varepsilon/8 \qquad \text{and} \tag{5.9}$$

$$\mathbb{E}_\nu[m] - c \geqslant \mathbb{P}(\varnothing \text{ is parked in } \mathcal{T}, |\mathcal{T}| \leqslant M \text{ and } \forall x \in \mathcal{T}, L_x \leqslant M) \geqslant \mathbb{E}_\nu[m] - c - \varepsilon/8. \tag{5.10}$$

Then

$$\mathbb{P}\left(\left|\sum_{x\in\mathcal{T}_n} \mathbb{1}_{x \text{ is parked}} - (\mathbb{E}_\nu[m] - c)\, n\right| \geqslant \frac{\varepsilon}{2}n\right) \leqslant \mathbb{P}\left(\left|\sum_{x\in\mathcal{T}_n} \mathbb{1}_{|\operatorname{Top}(\mathcal{T}_n,x)|>M \text{ or } \exists y\in\operatorname{Top}(\mathcal{T}_n,x) \text{ s.t. } L_y>M}\right| \geqslant \frac{\varepsilon}{4}n\right)$$
$$+\mathbb{P}\left(\left|\sum_{x\in\mathcal{T}_n} \mathbb{1}_{x \text{ is parked}}\mathbb{1}_{|\operatorname{Top}(\mathcal{T}_n,x)|\leqslant M}\mathbb{1}_{\forall y\in\operatorname{Top}(\mathcal{T}_n,x),L_y\leqslant M} - (\mathbb{E}_\nu[m] - c)\, n\right| \geqslant \frac{\varepsilon}{4}n\right).$$

For the first term, using the bound (5.9) and applying an easy extension of the exponential bounds result of Proposition 5.4 with car arrivals decoration on the finitely many configurations of tree $\mathbf{t}$ such that $|\mathbf{t}| \leqslant M$ and such that all vertices carry less than M cars, we get $\delta_2 > 0$ and $n_2 \geqslant 0$ such that for all $n \geqslant n_2$,

$$\mathbb{P}\left(\sum_{x\in\mathcal{T}_n} \mathbb{1}_{|\operatorname{Top}(\mathcal{T}_n,x)|\leqslant M}\mathbb{1}_{\forall y\in\operatorname{Top}(\mathcal{T}_n,x),L_y\leqslant M} < \left(1 - \frac{\varepsilon}{4}\right) n\right)$$
$$\leqslant \mathbb{P}\left(\left|\sum_{x\in\mathcal{T}_n} \mathbb{1}_{|\operatorname{Top}(\mathcal{T}_n,x)|\leqslant M}\mathbb{1}_{\forall y\in\operatorname{Top}(\mathcal{T}_n,x),L_y\leqslant M} - n\mathbb{P}(|\mathcal{T}| \leqslant M \text{ and } \forall x \in \mathcal{T}, L_x \leqslant M)\right| > \frac{\varepsilon}{8}n\right) \leqslant e^{-\delta_2 n}.$$

For the second term, using (5.10) and (an extension of) Proposition 5.4 again, we obtain

$$\mathbb{P}\left(\left|\sum_{x\in\mathcal{T}_n}\mathbb{1}_{x \text{ is parked}}\mathbb{1}_{|\operatorname{Top}(\mathcal{T}_n,x)|\leqslant M}\mathbb{1}_{\forall y\in\operatorname{Top}(\mathcal{T}_n,x),L_y\leqslant M}-(\mathbb{E}_\nu[m]-c)\,n\right|\geqslant\frac{\varepsilon}{4}n\right)$$
$$\leqslant\mathbb{P}\left(\left|\sum_{x\in\mathcal{T}_n}\mathbb{1}_{x \text{ is parked}}\mathbb{1}_{|\operatorname{Top}(\mathcal{T}_n,x)|\leqslant M}\mathbb{1}_{\forall y\in\operatorname{Top}(\mathcal{T}_n,x),L_y\leqslant M}\right.\right.$$
$$\left.\left.-\,\mathbb{P}(\varnothing \text{ is parked in } \mathcal{T},|\mathcal{T}|\leqslant M \text{ and } \forall x\in\mathcal{T},L_x\leqslant M)n\right|\geqslant\frac{\varepsilon}{8}n\right)\leqslant e^{-\delta_3 n}.$$

for some $\delta_3>0$ and for n large enough. We get the desired result by combining theses inequalities.

□

5.3.2 Subcritical parking process

We now prove Theorem 5.2 in the subcritical case, i.e when $\Theta>0$. Our strategy of proof is very different from the supercritical case and requires analytical and geometric arguments. Let X be the number of cars that visit the root of $\mathcal{T}$ and W its generating function, i.e.

$$W(z)=\sum_{k=0}^{+\infty}z^k\mathbb{P}(k \text{ cars visit the root})$$

Since $\varphi(\mathcal{T})=(X-1)_+=\sup(X-1,0)$, it suffices to show that X has an exponential tail. More precisely, we will show that W is a convergent series and has radius of convergence strictly larger than 1 so that the probability $\mathbb{P}(k \text{ cars visit the root})$ has exponential decay. Using the branching property at the root of $\mathcal{T}$ we see that X is a solution to the following recursive distributional equation:

$$X\overset{(d)}{=}\sum_{i=1}^{N}(X_i-1)_+ + P_N, \tag{DE}$$

where $N\sim\nu$, the P_k have law $\mu_{(k)}$ for $k\geqslant 0$, the X_i for $i\geqslant 0$ are i.i.d. copies of the variable X and all variables on the right-hand side are independent. Therefore, W satisfies the following equation at least in terms of formal power series:

$$W(z)=\sum_{k\geqslant 0}\nu_k A_k(z)\left(\frac{1}{z}(W(z)-p_0)+p_0\right)^k, \tag{EQ}$$

where A_k is the generating function of $P_k\sim\mu_{(k)}$ and $p_0=\mathbb{P}\left(\varnothing \text{ is parked not in } \mathcal{T}\right)$. This equation on W was used in [100] and [60] in the case of Poisson Bienaymé–Galton–Watson trees and geometric or Poisson arrivals of cars. In these cases, they gave an explicit solution for W in the subcritical case. Notice in passing that (EQ) or equivalently (DE) characterizes the law of X and in particular, the quantity p_0 which appears in (EQ) is determined by (EQ). In the subcritical case however, we

already proved $p_0 = 1 - \mathbb{E}_\nu[m]$ in Proposition 5.2 by probabilistic means. Plugging this into (EQ) we deduce that $z \mapsto W(z)$ is solution of $F(z-1, W(z)-1) = 0$ where

$$F(x,y) = \sum_{k\geqslant 0} \nu_k A_k(1+x)\left(\frac{1}{1+x}(y+1-p_0)+p_0\right)^k - y - 1 = \sum_{i,j\geqslant 0} a_{i,j}x^i y^j,$$

for some family $(a_{i,j})_{i,j\geqslant 0}$, which is analytic around $(0,0)$ in both variables by the assumption (H_{exp}). Since W is a generating function, it is locally well defined around $z=0$ and is also the unique solution of $F(z-1, w(z)-1)\times z = 0$ and $w(0) = p_0$. Its radius of convergence is at least 1 and we have $W(1) = 1$. The problem is that $z = 1$ and $W(1) = 1$ may be a singularity for the equation: indeed, we have

$$\partial_y F(x,y) = -1 + \sum_{k\geqslant 1} k\nu_k \frac{A_k(1+x)}{1+x}\left(\frac{y+1-p_0}{1+x}+p_0\right)^{k-1},$$

$$\partial_x F(x,y) = \sum_{k\geqslant 1} \nu_k A_k'(1+x)\left(\frac{y+1-p_0}{1+x}+p_0\right)^k - k\nu_k A_k(1+x)\frac{y+1-p_0}{(1+x)^2}\left(\frac{y+1-p_0}{1+x}+p_0\right)^{k-1}.$$

Since A_k is the generating function of $\mu_{(k)}$, we have $A_k(1) = 1$ and $A_k'(1) = m_{(k)}$. Using the fact that $p_0 = 1 - \mathbb{E}_\nu[m]$ (Proposition 5.2), we obtain $\partial_x F(0,0) = 0$ and $\partial_y F(0,0) = 0$, so one cannot use the implicit function theorem at this point to extend $W(z)$. In the subcritical case we will resolve this singularity and prove that $F(z,w)$ has two analytic branches at $(0,0)$.

Lemma 5.1. *The origin is a double point of the section defined by $F(x,y) = 0$: that means that in a neighborhood of $(0,0)$, the equation $F(x, y(x)) = 0$ has two analytic branches.*

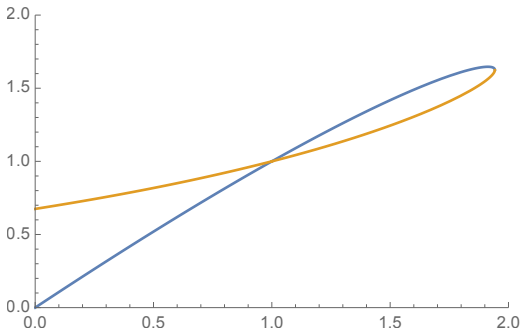


Figure 5.3: The curve defined by $F(x+1, y+1) = 0$ when the offspring distribution is geometric and the car arrivals are i.i.d with Poisson distribution of parameter $\alpha = 0.325 \leqslant \alpha_c = \sqrt{2}-1$ (see [60]). Notice the two analytic branches around $(1,1)$. The orange one is the generating function W.

Proof. We apply Newton's method to determine the Newton–Puiseux expansion at the point $(0,0)$ [92, p498-500]. The Newton–Puiseux expansion of a solution y of $F(x, y(x)) = 0$ around $(0,0)$ shows

that any solution can be expressed as a locally convergent series (see [121, Remark 1.2 (2) p6]) of the form

$$y(x) = \sum_{k \geqslant k_0} c_k x^{k/d}$$

where $k_0 \in \mathbb{Z}$ and d is a positive integer. Inductively we will show that $d = 1$, $k_0 \geqslant 0$ and that we have only two choices for the sequence $(c_k)_{k \geqslant k_0}$ in our case and so that F has two solutions $y_\pm$ which are analytic around 0. We look for a solution to $F(x, y(x)) = 0$ of the form $y(x) = cx^a + o(x^a)$ with $c \neq 0$. We expand F up to order 2 and we obtain

$$\begin{aligned} 0 &= a_{2,0}x^2 + a_{0,2}c^2x^{2a} + a_{1,1}cx^{1+a} + o(x^{2a\wedge 2}) \\ &= \left(\mathbb{E}_\nu[m^2 + \sigma^2 - m] + \mathbb{E}_{\overline{\nu}}[m]^2(\Sigma^2 - 2) - 2\mathbb{E}_{\overline{\nu}}[m]\right)\frac{x^2}{2} + \frac{\Sigma^2 c^2}{2}x^{2a} \\ &\quad + c\left(\mathbb{E}_{\overline{\nu}}[m](1-\Sigma^2) - 1\right)x^{1+a} + o(x^{2a\wedge 2}) \end{aligned}$$

Since the equation has to be identically satisfied, the main asymptotic should be 0. This can only happen if two or more of the exponents in $\{2, 1+a, 2a\}$ coincide and the coefficients of the corresponding monomial in F are zero. We obtain here $a = 1$ and c satisfies a quadratic equation that has two different real solutions which are

$$c_\pm = \frac{-\left(\mathbb{E}_{\overline{\nu}}[m](1-\Sigma^2) - 1\right) \pm \sqrt{\Theta}}{\Sigma^2} \tag{5.11}$$

We choose one of the two solutions for c_1 and suppose that we have a solution of the form $y(x) = c_1 x + \cdots + c_{k-1}x^{k-1} + o(x^{k-1})$ (where some of the c_i's may vanish) and look for a solution of the form $y(x) = c_1 x + \cdots + c_{k-1}x^{k-1} + c_k x^a + o(x^a)$ where $k - 1 < a \leqslant k$. Expanding F up to order $k+1$ and we obtain

$$\begin{aligned} 0 &= \sum_{1 \leqslant i,j \leqslant k+1} a_{i,j} x^i \left(c_1 x + \cdots + c_{k-1}x^{k-1} + c_k x^a + o(x^a)\right)^j + o(x^{k+1}) \\ &= c_k\left(c_1\frac{\Sigma^2}{2} + \left(\mathbb{E}_{\overline{\nu}}[m](1-\Sigma^2) - 1\right)\right)x^{1+a} + x^{k+1}\sum_{1 \leqslant i,j \leqslant k+1} a_{i,j} \sum_{\substack{1 \leqslant l_1,\ldots,l_j \leqslant k \\ \sum l_n = k+1-i}} c_{l_1}\ldots c_{l_j} + o(x^{(1+a)\wedge(k+1)}) \\ &= \pm\sqrt{\Theta}c_k x^{1+a} + x^{k+1}\sum_{1 \leqslant i,j \leqslant k+1} a_{i,j} \sum_{\substack{1 \leqslant l_1,\ldots,l_j \leqslant k \\ \sum l_n = k+1-i}} c_{l_1}\ldots c_{l_j} + o(x^{(1+a)}), \end{aligned}$$

depending on the choice of c_1. If $a < k$ and since the main asymptotic should be zero, then $c_k = 0$. Therefore we can take $a = k$, and the above equation implies

$$c_k = \pm\frac{1}{\sqrt{\Theta}} \sum_{2 \leqslant i,j \leqslant k+1} a_{i,j} \sum_{\substack{1 \leqslant l_1,\ldots,l_j \leqslant k \\ \sum l_n = k+1-i}} c_{l_1}\ldots c_{l_j},$$

which can possibly be zero. This recursively shows the Puiseux expansion of y has only integer powers, and concludes the proof.

□

Proof of Theorem 5.2, subcritical case. By Lemma 5.1, the equation $F(z-1, w(z)-1) = 0$ has two analytic branches around $(z, w) = (1, 1)$ (Figure 5.3). One of these branches coincides with the generating function of the number of cars which visit the root W defined in (5.3.2) in a neighborhood of 1^-. Therefore W can be extended analytically in a complex neighborhood of $z = 1$. Moreover since W is a generating function, its power expansion around $z = 0$ has non-negative coefficients. Hence, by Pringhsheim's theorem [92, Theorem IV.6 p240], the radius of convergence of W around 0 is greater than 1 and the desired result follows.

□

5.4 Application to the size of the connected components

In this section we want to study the size of the connected components in the final configuration i.e. the clusters of vertices that contain a car after the parking procedure and prove Corollary 1. We say that $x \in \mathbf{t}$ is free (resp. parked) if it contains (resp. does not contain) a car after parking.

5.4.1 Supercritical Case: Giant component

In this section we suppose that we are in the supercritical case i.e. $\Theta < 0$ and prove Corollary 1 in this phase. We first prove that the second largest connected component is of size $O(\ln(n))$. This can be easily deduced from the following proposition.

Proposition 5.5. *When $\Theta < 0$, there exists $A_0 \geqslant 0$, such that for all $A > A_0$,*

$$\mathbb{P}(\exists x \in \mathcal{T}_n \text{ s.t. } |\operatorname{Top}(\mathcal{T}_n, x)| \geqslant A\ln(n) \text{ and } x \text{ is a free spot}) \xrightarrow[n\to\infty]{} 0$$

Proof. We let v_k be the kth vertex in the depth-first exploration of $\mathcal{T}_n$ for $0 \leqslant k \leqslant n-1$. We use the fact that conditionally on $|\operatorname{Top}(\mathcal{T}_n, v_k)| = N$, the tree $\operatorname{Top}(\mathcal{T}_n, v_k)$ has law $\mathcal{T}_N$: for every fixed tree $\mathbf{t}$,

$$\mathbb{P}\Big(\operatorname{Top}(\mathcal{T}_n, v_k) = \mathbf{t} \Big| |\operatorname{Top}(\mathcal{T}_n, v_k)| = N\Big) = \mathbb{P}(\mathcal{T}_N = \mathbf{t}). \tag{5.12}$$

We first bound the probability in the proposition by the expectation of the number of such vertices x. Therefore, the probability in the proposition satisfies

$$\begin{aligned}
&\mathbb{P}(\exists x \in \mathcal{T}_n \text{ s.t. } |\operatorname{Top}(\mathcal{T}_n, x)| \geqslant A\ln(n) \text{ and } x \text{ is a free spot})\\
&\leqslant \sum_{k=1}^{n} \mathbb{E}\left[\mathbb{1}_{|\operatorname{Top}(\mathcal{T}_n, v_k)| \geqslant A\ln(n) \text{ and } v_k \text{ is a free spot}}\right]\\
&\leqslant \sum_{k=1}^{n} \sum_{s \geqslant A\ln(n)} \mathbb{E}\left[\mathbb{1}_{v_k \text{ is a free spot}} \Big| |\operatorname{Top}(\mathcal{T}_n, v_k)| = s\right] \mathbb{P}(|\operatorname{Top}(\mathcal{T}_n, v_k)| = s)\\
&\underset{(5.12)}{\leqslant} n \sup_{s \geqslant A\ln(n)} \mathbb{P}(\varnothing \text{ is a free spot in } \mathcal{T}_s)\\
&\underset{\text{Thm } 5.2}{\leqslant} n \times e^{-\delta A \ln(n)},
\end{aligned}$$

where δ is independent of A (and n). Therefore, if $A \geqslant 2/\delta$, this quantity converges to 0 as n goes to ∞.

□

Recall that $|C_{\max}(n)|$ is the size of the largest parked connected component of $\mathcal{T}_n$. Recall Aldous' sin-tree $\mathcal{T}^*$ from Section 5.2.2 with spine $\{u_0, u_1, \dots\}$.

Proposition 5.6. *We have the following convergence*

$$\frac{|C_{\max}(n)|}{n} \xrightarrow[n\to\infty]{(\mathbb{P})} \mathbb{P}\left(\forall k \geqslant 0, u_k \text{ is parked in } \mathcal{T}^*\right),$$

where the probability on the righthand side is computed by imagining that we perform the parking on $\mathcal{T}^*$ (rooted at infinity) with the same rules for car arrivals as for $\mathcal{T}$.

Proof. We chose $A > A_0$ and we work on the complement event of that of Proposition 5.5 i.e. on $\mathcal{E}_n = \{\forall x \in \mathcal{T}_n \text{ s.t. } |\operatorname{Top}(\mathcal{T}_n, x)| \geqslant A\ln(n), x \text{ is parked}\}$. Then, when n is large enough, the root is parked and its parked component contains all vertices x such that $|\operatorname{Top}(\mathcal{T}_n, x)| \geqslant A\ln(n)$ and is therefore with high probability the only parked component of size larger than $A\ln(n)$. Hence, we can decompose the vertices of $\mathcal{T}_n$ according to whether they have an ancestor which is a free parking spot and how far this ancestor is: on $\mathcal{E}_n$, when n is large enough and with high probability,

$$\begin{aligned} |C_{\max}(n)| \;=\; & n - |\{x \in \mathcal{T}_n \text{ s.t. } x \text{ has an ancestor at distance } < A\ln(n), \text{ which is a free spot}\}| \\ & - |\{x \in \mathcal{T}_n \text{ s.t. } x \text{ has an ancestor at distance } \geqslant A\ln(n), \text{ which is a free spot}\}|. \end{aligned}$$

Since $A \geqslant A_0$ as defined in Proposition 5.5, then on $\mathcal{E}_n$ i.e. with high probability,

$$|\{x \in \mathcal{T}_n, x \text{ has an ancestor at distance } \geqslant A\ln(n), \text{ which is a free spot }\}| = 0.$$

Recall that conditionally on the tree $\mathcal{T}_n$ (or $\mathcal{T}^*$) the car arrivals are independent on each vertex (and their law only depends on the tree through their arrival vertex degree). Therefore, we can extend [163, Theorem 5.2] to trees with car decorations by extending the couplings given by the total variation distance in such a way that the car arrivals are the same on $H_{k_n}(\mathcal{T}_n, x)$ and $H_{k_n}(\mathcal{T}^*, u_0)$ whenever both trees are equal. Since in our case $\ln(n) = o(\sqrt{n})$, we deduce that

$$\begin{aligned} \Big| \frac{1}{n} |\{x \in \mathcal{T}_n \text{ s.t. } x \text{ has an ancestor at distance } < A\ln(n), \text{ which is a free spot }\}| \\ - \mathbb{P}\left(\exists 0 \leqslant k \leqslant A\ln(n), u_k \text{ is not parked in } \mathcal{T}^*\right) \Big| \xrightarrow[n\to\infty]{(\mathbb{P})} 0. \end{aligned}$$

As the probability $\mathbb{P}(\mathcal{E}_n)$ converges to 0 and $\mathbb{P}\left(\exists 0 \leqslant k \leqslant A\ln(n), u_k \text{ is not parked in } \mathcal{T}^*\right)$ converges to $\mathbb{P}\left(\exists k \geqslant 0, u_k \text{ is not parked in } \mathcal{T}^*\right)$ as n goes to ∞, we get the desired result.

□

5.4.2 Subcritical parking

In this section we suppose that we are in the subcritical case i.e. $\Theta > 0$ and prove Corollary 1 in this case.

Proof. [Proof of Corollary 1] The proof is based on the sprinkling method which consists in adding cars while staying in the subcritical phase. Since $\Theta > 0$, there exists $\varepsilon > 0$ such that

$$\Theta' = \Theta + \varepsilon^2 - 2\varepsilon(1 - \mathbb{E}_{\overline{\nu}}[m] - \Sigma^2 \mathbb{E}_{\nu}[m]) > 0,$$

This means that the parking process on $\mathcal{T}_n$ or $\mathcal{T}$ with offspring distribution ν and car arrivals with distribution $\tilde{\mu}^{\varepsilon}_{(k)}$ such that $\tilde{\mu}^{\varepsilon}_{(k)}(j) = (1-\varepsilon)\mu_{(k)}(j) + \varepsilon\mu_{(k)}(j-1)$ (that is we add a car with probability ε on each vertex independently) is still subcritical. We denote by $\tilde{\varphi}$ the corresponding flux. Recall that $\varphi_x(\mathbf{t})$ is the flux at vertex x i.e. on $\mathrm{Top}(\mathbf{t}, x)$. Imagine that in $\mathcal{T}_n$ with arrivals μ, we have a large parked component C of size larger than $A\ln(n)$. If we further let cars arrive on each vertex with probability ε, then by the law of large numbers, the root x of C gets a flux $\tilde{\varphi}_x(\mathcal{T}_n) \geqslant A\varepsilon\ln(n)/2$ with probability at least $1/2$ when n is large enough. Therefore, for n large enough,

$$\begin{aligned}\mathbb{P}\left(|C_{\max}(n)| \geqslant A\ln(n)\right) &\leqslant 2\mathbb{P}\left(\exists x \in \mathcal{T}_n, \tilde{\varphi}_x(\mathcal{T}_n) \geqslant \frac{\varepsilon A}{2}\ln(n)\right) \\ &\leqslant 2\sum_{M=1}^{n}\sum_{k=1}^{n}\mathbb{P}\Big(\tilde{\varphi}_{v_k}(\mathcal{T}_n) \geqslant \frac{\varepsilon A}{2}\ln(n)\Big|\,|\mathrm{Top}(\mathcal{T}_n, v_k)| = M\Big)\mathbb{P}\left(|\mathrm{Top}(\mathcal{T}_n, v_k)| = M\right).\end{aligned}$$

Now we use the conditional law (5.12), the asymptotic (5.7) and Theorem 5.2 so that

$$\begin{aligned}\mathbb{P}\left(|C_{\max}(n)| \geqslant A\ln(n)\right) &\underset{(5.12)}{\leqslant} 2\sum_{M=1}^{n}\sum_{k=1}^{n}\mathbb{P}\left(|\mathrm{Top}(\mathcal{T}_n, v_k)| = M\right)\frac{\mathbb{P}\left(\tilde{\varphi}(\mathcal{T}) \geqslant \frac{\varepsilon A}{2}\ln(n)\right)}{\mathbb{P}\left(|\mathcal{T}| = M\right)} \\ &\underset{(5.7)}{\leqslant} 2C\sum_{M\geqslant 1}\sum_{k=1}^{n}\mathbb{P}\left(|\mathrm{Top}(\mathcal{T}_n, v_k)| = M\right) \times n^{3/2} \times \mathbb{P}\left(\tilde{\varphi}(\mathcal{T}) \geqslant \frac{\varepsilon A}{2}\ln(n)\right) \\ &\underset{\text{Thm } 5.2}{\leqslant} 2Cn^{5/2} \times n^{-\delta\varepsilon A/2},\end{aligned}$$

for some constant $C > 0$ and some $\delta > 0$. Therefore, choosing $A > 2/(\delta\varepsilon) + 5/2$, we obtain the desired result.

□

Chapitre 6 :

Parking on Cayley trees & Frozen Erdős–Rényi

Les résultats de ce chapitre sont issus de l'article [67], écrit en collaboration avec Nicolas Curien et accepté pour publication dans The Annals of Probability.

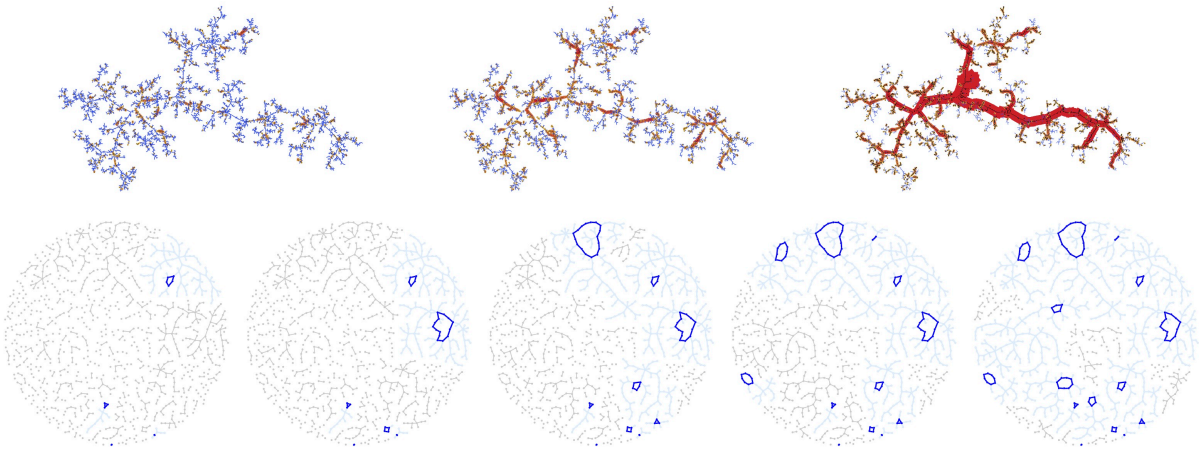


Figure 6.1: First line: Parking on a random Cayley tree with 10000 vertices when resp. 4000, 5000 and 6000 cars have arrived (color and thickness indicate the flux of cars along the edges). Second line: The frozen Erdős–Rényi process at stages 400, 500, 600, 700 and 800 on a graph with 1000 vertices.

Consider a uniform rooted Cayley tree T_n with n vertices and let m cars arrive sequentially, independently, and uniformly on its vertices. Each car tries to park on its arrival node, and if the spot is already occupied, it drives towards the root of the tree and parks as soon as possible. Lackner & Panholzer [128] established a phase transition for this process when $m \approx \frac{n}{2}$. In this work, we couple this model with a variant of the classical Erdős–Rényi random graph process. This enables us to describe the phase transition for the size of the components of parked cars using a modification of the multiplicative coalescent which we name the *frozen multiplicative coalescent*. The geometry of critical parked clusters is also studied. Those trees are very different from Bienaymé–Galton–Watson trees and should converge towards the growth-fragmentation trees canonically associated to the 3/2-stable process that already appeared in the study of random planar maps.

Contents

Introduction

In this paper we establish a connection between the parking process on a random Cayley tree and a certain modification of the classical Erdős–Rényi random graph obtained by freezing or more precisely "slowing down" components with surplus. This unexpected relationship enables us to understand the phase transition for parking established in [128] and in return gives a new point of view on the Erdős–Rényi random graph and the multiplicative coalescent process. Our coupling works by redirecting and discarding certain edges in the random graph process in order to construct step-by-step the underlying tree to accommodate the parking process (using a Markovian or "peeling" construction). The geometry of the parked components at criticality is built by a "multiplicative" merging similar to the construction of the minimal spanning tree [4] but gives rise to random trees which we believe to converge towards the growth-fragmentation trees [30] that already appeared in the study of random planar maps [31, 32]. This conjecture is further supported by deep analogies between the enumeration of planar maps and that of fully parked trees with outgoing flux.

Parking on random trees. Let us first recall the model of parking on a Cayley tree first studied in [128]. Consider a finite tree $\mathfrak{t}$ with a root vertex. We interpret the vertices of $\mathfrak{t}$ as being parking spots (each vertex can accommodate only one car) and we let cars arrive sequentially, independently and uniformly over the vertices of $\mathfrak{t}$. Each car tries to park on its arrival node, unless the spot is taken in which case it drives towards the root of the tree in search of the first available parking spot. If during its descent to the root vertex no free spot is found, then the car exits the tree without parking, see Figure 6.2.

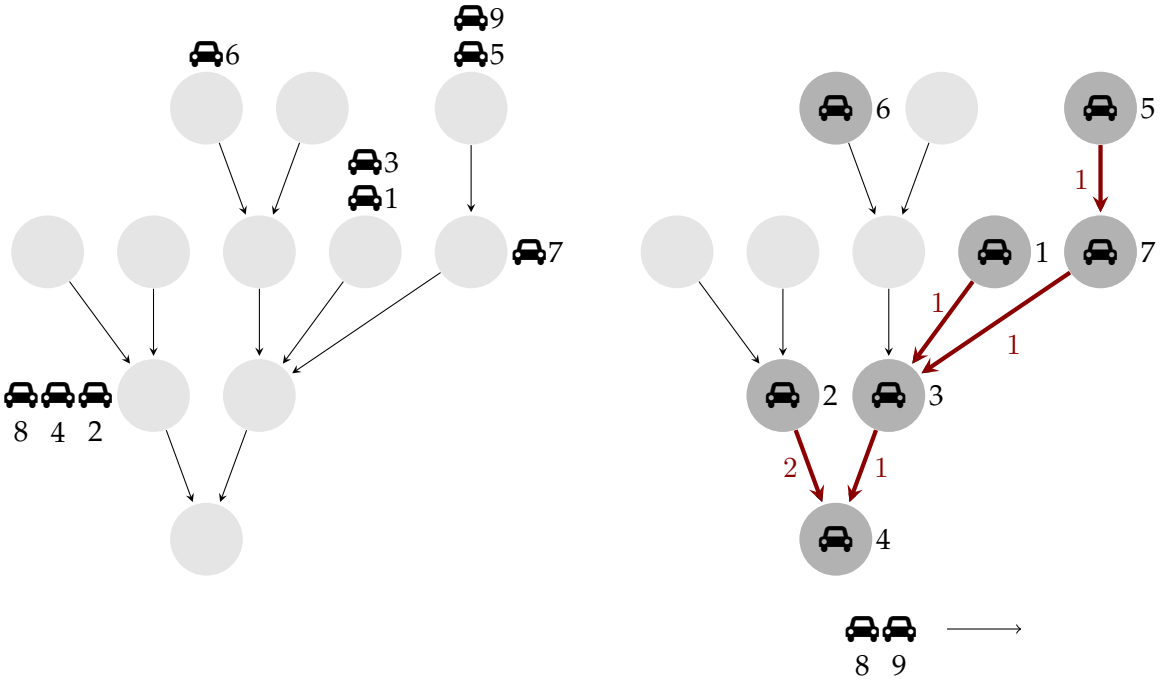


Figure 6.2: On the left, a rooted tree with 11 vertices (the root vertex is the bottom vertex) where all edges are oriented towards the root vertex, together with 9 cars arriving on its vertices. On the right, the result of the (sequential) parking of the 9 cars. The flux of cars along each edge is indicated. Notice that two cars did not manage to park and exited the tree.

Of course when the underlying tree is a discrete line, this corresponds to the famous one-dimensional parking process of Konheim & Weiss [122] which is now part of the folklore in probability [57]. The study of parking on more general trees was only recently initiated by Lackner & Panholzer [128] where the underlying tree was a uniform Cayley tree of fixed size rooted at a uniform vertex (see also [53, 118, 119] for related works in combinatorics). Recall that a Cayley tree of size n is a (unordered) tree over the labeled vertices $\{1, 2, \dots, n\}$. This model was later studied from a probabilistic angle in [100]. Since then, a body of work with an increasing level of generality has emerged [21, 60, 75, 148] ultimately considering critical conditioned Bienaymé–Galton–Watson tree (with finite variance) for the underlying tree and independent car arrivals whose laws may depend on the degree of the vertices [64]. See [22, 61] for the case of supercritical trees. In this broad context, it was shown that a sharp *phase transition* appears for the parking process: there is a critical "density" of cars (depending on the combinatorial details of the model) such that below this density, almost all cars manage to park, whereas above this density, a positive proportion of cars do not find a parking spot. See [64, 75] for precise statements. The goal of this work is to provide scaling limits for the *critical* and *near-critical* dynamics of the parking process in the special case of uniform Cayley trees with i.i.d uniform car arrivals where the critical density is $\frac{1}{2}$, see [128]. Perhaps surprisingly, this will

be done by relating the model to the ubiquitous Erdős–Rényi random graph.

Frozen Erdős–Rényi. Fix $n \geqslant 1$. Over the vertex set $\{1, 2, \dots, n\}$, consider for $i \geqslant 1$ independent identically distributed oriented edges $\vec{E}_i = (X_i, Y_i)$ where both endpoints are independent and uniform over $\{1, 2, \dots, n\}$. We denote by E_i the unoriented version of the oriented edge $\vec{E}_i$. Notice in particular that we may have $X_i = Y_i$ and $\vec{E}_i = \vec{E}_j$ for $i \neq j$. For $m \geqslant 0$, the *Erdős-Rényi* random graph[1] is the random multigraph $G(n, m)$ whose vertex set is $\{1, 2, \dots, n\}$ and whose unoriented edge set is the multiset $\{\{E_i : 1 \leqslant i \leqslant m\}\}$.

We now define the *frozen* Erdős–Rényi process $(F(n, m) : m \geqslant 0)$, which is obtained from the above graph process $(G(n, m) : m \geqslant 0)$ by "freezing" or more precisely slowing down the components which are not trees. The vertices of $F(n, m)$ will be of two types: standard "white", or frozen "blue" vertices. The blue vertices constitute the *freezer* of $F(n, m)$. Initially $F(n, 0)$ is made of the n labeled white vertices $\{1, 2, \dots, n\}$. As in the $(G(n, m) : m \geqslant 0)$ process, we let the (same) edges $\vec{E}_i = (X_i, Y_i)$ arrive sequentially for $i \geqslant 1$ but discard some of them and color the vertices in $F(n, \cdot)$ according to the following rule, see Figure 6.3: for $m \geqslant 1$

- if both endpoints of E_m are white vertices then the edge E_m is added to $F(n, m-1)$ to form $F(n, m)$. If this addition creates a cycle in the graph then the vertices of its component are declared frozen and colored in blue.
- if both endpoints of E_m are blue (frozen vertices), then E_m is discarded.
- if E_m connects a white and a blue vertex, then E_m is kept if $\vec{E}_m$ goes from the white to the blue vertex. If so, the new connected component is declared frozen and colored in blue.

A more general version of the frozen process depending on a parameter $p \in [0, 1]$ can be defined (see Section 6.9.1) by keeping edges between white and blue components with probability p. Different models of "frozen" percolation have already been considered on the Erdős–Rényi random graph [71, 152, 153] or on other graphs [11, 78, 120], but to the best of our knowledge, the above random graph processes are new. One interesting feature of the frozen process is that for any $m \geqslant 0$, conditionally on the frozen part of $F(n, m)$, the "forest part" made of the white components is a uniform forest given its number of vertices and edges, see Proposition 6.8. We shall refer to this property as the *free forest property*. The geometry of large critical uniform random forests has been studied in particular by Luczak [135] using counting results of Rényi and Britikov [47, 156] and more recently by Martin & Yeo [138] using an exploration process converging to an inhomogeneous diffusion with reflecting boundary. We shall revisit and shed new light on those results using random walks coding and (conditioned) 3/2-stable processes, see Section 6.7.4.

In the case of the Erdős–Rényi random graph, Aldous proved in a famous paper [10] that the process of the component sizes in $G(n, m)$ exhibits a phase transition in the critical window $m =$

[1] Commonly in the literature, the Erdős–Rényi random graph is a simple graph where self-loops and multiple edges are forbidden, but this small variant is more natural probabilistically as it was noticed already in [91], [107, Section 1] [36, Section 2.3.1] or [132].

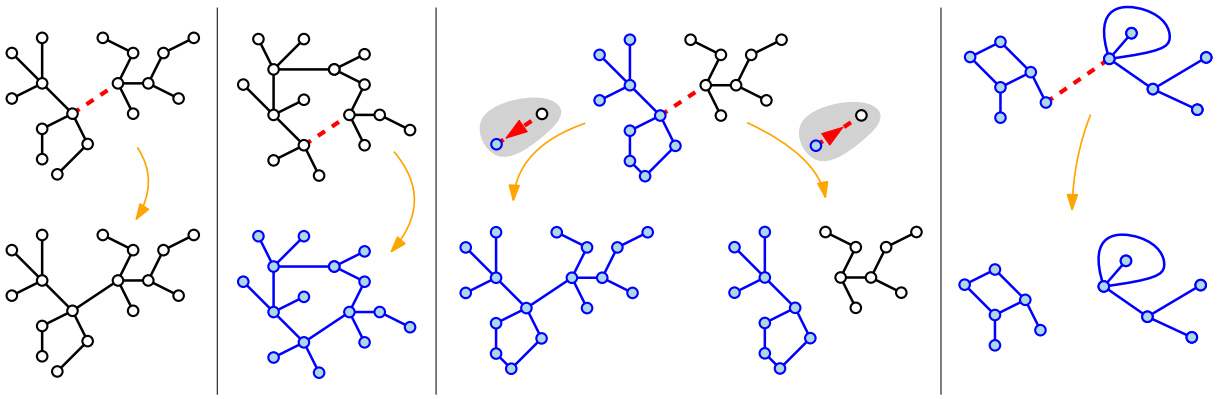


Figure 6.3: Illustration of the transitions in the frozen Erdős-Rényi process. The new edge to be examined is in dotted red. If this edge appears between two white tree-components, it is kept (first and second figures on the left). When a cycle is created, the component is colored in blue and becomes frozen (second figure). An edge appearing between a frozen blue and a white component is kept if it goes from white to blue and the entire new component is declared frozen. All edges between frozen components are discarded.

$\frac{n}{2} + \frac{\lambda}{2} n^{2/3}$ for $\lambda \in \mathbb{R}$. The same critical window will appear in this work and so to lighten notation, when we have a discrete process $(X(n,m) : m \geqslant 0)$ where n denotes the fixed "size" of the system and $m = 1, 2, 3, \ldots$ is an evolving parameter, we shall denote its continuous time analog by a mathrm letter

$$\mathrm{X}_n(\lambda) = X\left(n, \left\lfloor \frac{n}{2} + \frac{\lambda}{2} n^{2/3} \right\rfloor \vee 0\right), \qquad \text{for } \lambda \in \mathbb{R}. \tag{6.1}$$

The parameter λ will often be called the "time" parameter and will enable us to compare processes of different sizes in the same time window. This will e.g. apply to $G(n,m)$ and $F(n,m)$ to yield $\mathrm{G}_n(\lambda)$ and $\mathrm{F}_n(\lambda)$. With this notation, Aldous proved that after renormalizing the component sizes of $\mathrm{G}_n(\lambda)$ by $n^{-2/3}$, the resulting process converges to the multiplicative coalescent which is a random càdlàg process $(\mathscr{M}(\lambda) : \lambda \in \mathbb{R})$ with values in ℓ^2 intuitively starting from "dust" as time $-\infty$ and such that every pair of particles of mass x and y merges to a new particle of mass $x+y$ at a rate xy, see Figure 6.4. Using Aldous' work [10] and its extensions [36, 50], we are able to prove (Theorem 6.2) a similar result for the component sizes in $\mathrm{F}_n(\lambda)$ and refer to the scaling limit $(\mathscr{FM}(\lambda) : \lambda \in \mathbb{R})$ as the *frozen multiplicative coalescent*. This however requires careful cutoffs and controls since the dynamics of the frozen Erdős–Rényi is not "monotonous". Similar ideas have been used by Rossignol in [157] to define a split/merge stationary dynamics on the scaling limit of critical random graphs.

To be a bit more precise, the particles of the frozen multiplicative coalescent $\mathscr{FM}(\lambda)$ at time λ are of two types: the frozen (blue) particles whose decreasing masses are in ℓ^1 and non-frozen (white) particles whose decreasing masses form a sequence in ℓ^2. Then $\mathscr{FM}$ is a càdlàg process with values in $\ell^1 \times \ell^2$ which evolves heuristically according to the same dynamics as that of $\mathrm{F}_n(\cdot)$: every pair of white particles of mass x and y merge to a new white particle of mass $x+y$ at a rate xy,

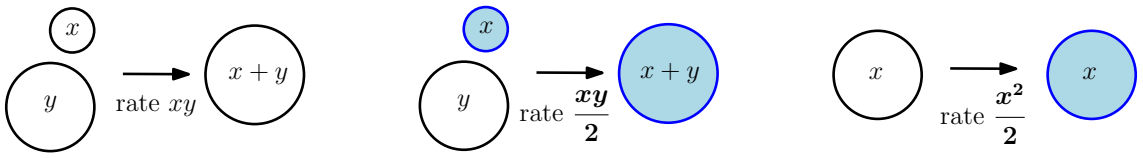


Figure 6.4: Dynamics of the frozen multiplicative coalescent $\mathscr{FM}$. The interaction between standard "white" particles is the same as in the multiplicative coalescent (left), but the interaction between white and frozen "blue" particles is slowed down (middle). Besides, a white particle can become blue at rate proportional to its mass squared (right).

whereas a blue particle of mass x merges with a white particle of mass y to form a blue particle of mass $x+y$ at a rate $\frac{xy}{2}$. Also, a white particle of mass x becomes frozen "if it creates an internal cycle" which appears with a rate $\frac{x^2}{2}$, see Figure 6.4. We stress that this "infinitesimal transition" is only a heuristic description of $\mathscr{FM}$ and its actual definition (at least in the present work) is given by Theorem 6.2 as the limit of the discrete frozen Erdős–Rényi processes. The process $\mathscr{FM}$ also has a Markovian property and in particular the process of the total mass of the frozen particles is a Feller pure-jump process with an explicit jump kernel close to that of a $\frac{1}{2}$-stable subordinator, see Proposition 6.15. Since $\mathscr{FM}$ is naturally coupled with the multiplicative coalescent, it gives a new perspective on the multiplicative coalescent (see Part 7.6). We also introduce generalized frozen multiplicative coalescents depending on a parameter $p \in [0,1]$, the case $p = 1/2$ being the one above, whereas for $p = 1$ the dynamics heuristically corresponds to removing the edges which would create surplus larger than 2 in the Erdős–Rényi model, see Section 7.6. We wonder whether the dynamics of the frozen multiplicative coalescent $\mathscr{FM}$ can be described by "merging the excursion lengths of random functions" as it is the case for the multiplicative coalescent [17, 50, 133, 169] or its version with linear deletion [139]. We also leave open the question of "entrance laws" or the behavior at $-\infty$ of those processes, [133].

Coupling parking on Cayley trees and the frozen Erdős–Rényi. As announced above, the main input of this paper is to construct an explicit coupling between the dynamical parking process on a uniform rooted Cayley tree T_n and the frozen Erdős–Rényi process $F(n,\cdot)$ so that the components match up. On the tree side, this coupling consists in considering the underlying tree T_n as unknown and exploring its oriented edges one after the other to park the cars. To do this we develop a general Markovian or "peeling" exploration of Cayley trees (Section 6.2.1) similar to that of [65, 72] and which may have further applications. To be a bit more precise, for $m \geqslant 0$ consider

$$T_{\mathrm{near}}(n,m) \subset T_n$$

the subforest of T_n spanned by the m edges emanating directly from a vertex containing one of the first m cars (recall that the edges are oriented towards the root), see Figure 6.7. Then we prove in Proposition 6.6 that we can couple the parking process on T_n with the frozen Erdős–Rényi $F(n,\cdot)$ so that after merging the frozen components of $F(n,m)$ we get the same components as $T_{\mathrm{near}}(n,m)$

–but the geometry inside the components is totally different–. See Section 6.2.2 for details. The construction is easier to understand when the underlying tree T_n is replaced by a uniform random mapping M_n and we start with this case in Section 6.1.3. In particular, in our construction, the discovery of the first cycle in $G(n,\cdot)$, or equivalently in $F(n,\cdot)$, corresponds to a car parking at the root of T_n, and this enables us to prove the remarkable identity:

Proposition 6.1 (Complete parking and acyclicity of $G(n,m)$)**.** *For $n \geqslant 1$ and $m \geqslant 0$ we have*

$$\mathbb{P}(m \text{ i.i.d. uniform cars manage to park on } T_n)\left(1-\frac{m}{n}\right) = \mathbb{P}(G(n,m) \text{ is acyclic}).$$

Combining this proposition with classical tree enumeration going back to Rényi [156] and Britikov [47], we recover the counting results of Lackner & Panholzer [128, Theorems 3.2 & 4.5 & 4.6] which were derived using (sometimes delicate) analytic combinatorics and singularity analysis, see Section 6.4. Another consequence concerns the scaling limit of the component sizes in the parking process: let us denote by $C_i(n,m)$ for $i \geqslant 1$ the non-increasing sizes (number of vertices) of the components of $T_{\text{near}}(n,m)$ of T_n when m cars have arrived. We put the component of the root vertex aside and denote its size by $C_*(n,m)$. We also write $D(n,m)$ for the number of cars among the first m that did not manage to park (the letter D stands for "discarded"). With our convention (6.1) we prove:

Theorem 6.1 (*Dynamical scaling limit for the component sizes and the outgoing flux*)

We have the following convergence in distribution for the Skorokhod topology on $\mathbb{D}\text{adlag}(\mathbb{R}, \ell^2 \times \mathbb{R}_+ \times \mathbb{R}_+)$

$$\begin{pmatrix} n^{-2/3}\cdot \mathrm{C}_{n,i}(\lambda)\,, & i\geqslant 1\\ n^{-2/3}\cdot \mathrm{C}_{n,*}(\lambda)\\ n^{-1/3}\cdot \mathrm{D}_n(\lambda)\end{pmatrix}_{\lambda\in\mathbb{R}} \xrightarrow[n\to\infty]{(d)} \begin{pmatrix} \mathscr{C}_i(\lambda), & i\geqslant 1\\ \mathscr{C}_*(\lambda)\\ \mathscr{D}(\lambda)\end{pmatrix}_{\lambda\in\mathbb{R}}.$$

The processes $\mathscr{C}_i, \mathscr{C}_*$ and $\mathscr{D}$ are built from the frozen multiplicative coalescent as follows:

- *$(\mathscr{C}_i(\lambda) : i \geqslant 1)$ is the non-increasing sequence of masses of the white particles in $\mathscr{FM}(\lambda)$,*
- *$\mathscr{C}_*(\lambda)$ is the sum of the masses of the blue particles in $\mathscr{FM}(\lambda)$,*
- *$\mathscr{D}(\lambda) = \frac{1}{2}\int_{-\infty}^{\lambda} \mathrm{d}s\, \mathscr{C}_*(s)$.*

Notice in particular that, in the critical window $m = \frac{n}{2} + O(n^{2/3})$, the flux of cars that did not manage to park in T_n is of order $n^{1/3}$ whereas the size of the largest cluster of parked cars is of order $n^{2/3}$. See Figure 6.5 for a simulation of a critical parking and its decomposition into parked components. Our theorem also holds for different versions of components e.g. if we only keep the edges between parked vertices, see Section 6.7.2.

Remark (Dynamical parking and coalescence). It is striking to notice that Konheim & Weiss' parking on the line is related to the *additive* coalescent [57, 50], whereas the essence of our findings is that the parking process on random Cayley trees obeys a modified *multiplicative* coalescence rule.

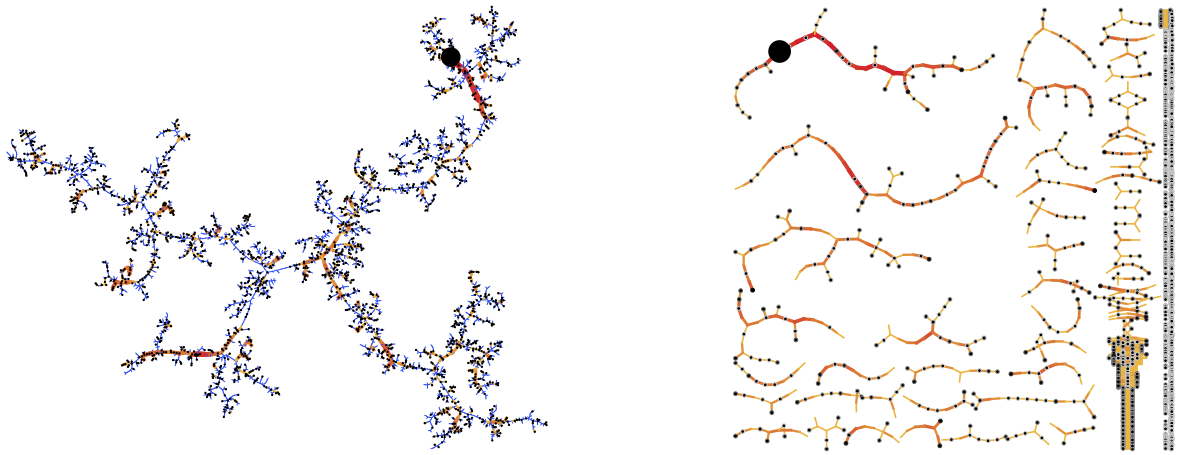


Figure 6.5: (Left) A simulation of a critical parking on T_{5000} with 2500 cars. The colors and widths of the edges indicate the flux of cars going through them. The root of the tree is represented by a black disk. (Right) The decomposition of the same tree into its parked components.

Geometry of fully parked trees and Bertoin's growth-fragmentation processes. Theorem 6.1 describes the phase transition of the parking in terms of the *sizes* of the parked components and outgoing flux of cars in T_n. But one can wonder about the *geometry* of the parked components and the flux of cars on its edges. It is not hard to see (see Proposition 6.12) that except for the component of the root vertex, conditionally on their sizes N, those components are (after relabeling of the vertices and cars) uniform *fully parked trees*, i.e. random uniform rooted Cayley tree T_N with N vertices carrying N labeled cars conditioned on the (unlikely) event that all cars successfully park on T_N. In what follows, we shall consider a slight variant of this model and denote by P_N a uniform *nearly* parked tree of size $N \geqslant 1$ which is a uniform rooted Cayley trees of size N carrying $N-1$ labeled cars conditioned on the event that the root ρ stays void after parking, see Figure 6.6 and Figure 6.7.

The conditioning imposed on the parking configuration makes the geometry of P_N very different from that of a uniform Cayley tree T_N: heuristically they are more elongated or path-like. When restricted to nearly parked trees, our coupling gives a construction of a nearly parked tree P_N from a uniform Cayley tree T_N of size N whose edges are labeled from 1 up to $N-1$ and oriented randomly (see Section 6.5.2.1 for details). Using this we are able to prove:

Proposition 6.2 (Typical height of P_N)**.** *The mean height of a nearly parked tree of size N is*

$$\frac{1}{N}\mathbb{E}\left[\sum_{x\in \mathrm{Vertices}(P_N)} \mathrm{d}_{\mathrm{gr}}^{P_N}(\rho,x)\right] = \sum_{h=1}^{N-1}\binom{N}{h+1}\left(\frac{h+1}{N}\right)^2 N^{1-h}\left(\frac{1}{2}\right)_h \underset{N\to\infty}{\sim} \frac{\Gamma(3/4)}{2^{1/4}\sqrt{\pi}}\cdot N^{3/4},$$

where $(x)_a = x(x+1)\cdots(x+a-1)$ is the Pochhammer symbol.

A nearly parked tree P_N naturally comes with a labeling $(\phi_N(e) : e \in \mathrm{Edges}(P_N))$ on its edges

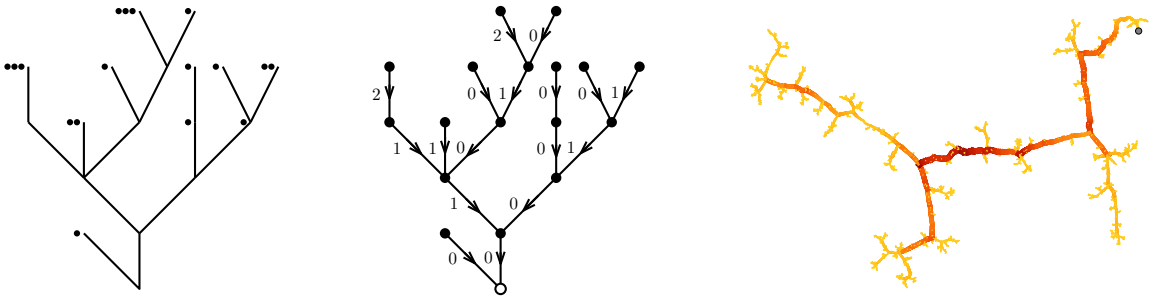


Figure 6.6: A nearly parked tree with 18 vertices and 17 cars that manage to park while leaving the root empty (the labels of the vertices and cars are not displayed for the sake of clarity). Right: A simulation of a large uniform nearly parked tree of size 15000, where the thickness and color of the edges indicate the flux of cars going through them.

counting the number of cars going through that edge in the parking process, see Figure 6.6. An Abelian property actually shows that this labeling does not depend on the order in which we have parked the cars. In particular, the sum $\sum \phi_N(e)$, corresponding to the total distance travelled by the cars, is also invariant under relabeling of the cars. We compute the expectation of this quantity:

Proposition 6.3 (Total traveled distance $N^{5/4}$)**.** *The mean total distance travelled by the cars in a uniform nearly parked tree P_N is*

$$\mathbb{E}\left[\sum_{e\in \mathrm{Edges}(P_N)} \phi_N(e)\right] = \frac{1}{2}\sum_{h=1}^{N-2} \binom{N-1}{h+1}(h+2)N^{-h}\left(\frac{1}{2}\right)_h \underset{N\to\infty}{\sim} \frac{\Gamma(1/4)}{2^{5/4}\sqrt{\pi}}\cdot N^{5/4}.$$

The heuristic picture suggested by the above two results is that a uniform nearly parked tree P_N is of height $N^{3/4}$ and that the flux of cars along "long branches" of P_N is of order $N^{1/2}$ so that $N^{3/4}\cdot N^{1/2} = N^{5/4}$ is the total distance driven by the cars. This is coherent with the fact that in the critical window, the outgoing flux at the bottom of the root component of size $n^{2/3}$ is of order $n^{1/3} = (n^{2/3})^{1/2}$ by Theorem 6.1.

In fact, we believe that rescaled uniform nearly parked trees converge after normalization towards the growth-fragmentation trees that already appeared in the study of scaling limits of random planar maps and the Brownian sphere, see [31, 32, 130] or [72, Chapter 14.3.2]. Those are "labeled continuum random trees" describing the genealogy of the masses of individuals in a family of living cells. These cells evolve independently one from the other, and the dynamics of the mass of a typical cell is governed by (a variant of) a 3/2-stable spectrally negative Lévy process. Each negative jump-time for the mass is interpreted as a birth event, in the sense that it is the time at which a daughter cell is born, whose initial mass is precisely given by the absolute height of the jump (so that conservation of mass holds at birth events). With some work, one can define a version $(\mathcal{T}, (\phi(x) : x \in \mathcal{T}))$ of those random labeled trees conditioned to start from a single cell of mass 0 and to have a total "volume" of 1, see [34, 33] for details. We propose the following conjecture:

Conjecture 1. *We have the following convergence in distribution for some* $c_1, c_2 > 0$

$$\left(\left(P_N, \frac{\mathrm{d}_{\mathrm{gr}}^{P_N}}{N^{3/4}}\right); \left(\frac{\phi_N(e)}{\sqrt{N}} : e \in \mathrm{Edges}(P_N)\right)\right) \xrightarrow[N\to\infty]{(d)} (c_1 \cdot \mathcal{T}, (c_2 \cdot \varphi(x) : x \in \mathcal{T}))\,,$$

see [34] for the topology one may want to use (which in particular implies the Gromov–Hausdorff convergence on the first coordinate).

Apart from the above propositions, one strong support for this conjecture is the fact that fully parked trees satisfy a Tutte-like equation by splitting the flux at the root which is reminiscent of that appearing in the realm of planar maps, see Section 6.8. In particular, if correct, combining the above conjecture with our coupling construction would uncover a "dynamical" construction of growth-fragmentation trees which is similar in spirit to that of the minimal spanning tree [4] but with a redirection of the edges.

The paper is organized in two main parts. The first one is purely in the discrete setting and presents the coupling construction as well as its enumerative and geometric consequences. The second one focuses on scaling limits and involves the multiplicative coalescent of Aldous as well as stable Lévy processes. For the reader's convenience, we provide an index of the main notations at the end of the paper.

Acknowledgements. We acknowledge support from ERC 740943 *GeoBrown*. We thank Linxiao Chen, Armand Riera and especially Olivier Hénard for several motivating discussions during the elaboration of this work. We thank Svante Janson and Cyril Banderier for comments about the first version of this work. Last, but not least, we warmly thank the anonymous referee for a thorough reading of our paper and precious comments which were greatly appreciated.

I Discrete constructions

This part is devoted to the discrete constructions and couplings. We consider non-necessarily connected finite *multigraphs* $\mathfrak{g}$, i.e. self-loops and multiple edges are allowed. The number of vertices of $\mathfrak{g}$ will be denoted by $\|\mathfrak{g}\|_\bullet$, its number of edges by $\|\mathfrak{g}\|_{\bullet\text{-}\bullet}$ and the vertex set is often taken to be $\{1, 2, \dots, \|\mathfrak{g}\|_\bullet\}$. The vertices of our graphs will often be colored in two colors, white (standard) or blue (frozen), and we denote by $\|\mathfrak{g}\|_\circ$ and $\|\mathfrak{g}\|_{\color{blue}\bullet}$ the number of vertices of each color and by $[\mathfrak{g}]_\circ$ and $[\mathfrak{g}]_{\color{blue}\bullet}$ the graphs induced on vertices of each color. The surplus of a *connected* multigraph $\mathfrak{g}$ is defined as $\|\mathfrak{g}\|_{\bullet\text{-}\bullet} - \|\mathfrak{g}\|_\bullet + 1$ and corresponds to the number of "cycles" created when building $\mathfrak{g}$. The subgraph made of the components without surplus is called the *forest part* of $\mathfrak{g}$ and denoted by $[\mathfrak{g}]_{\mathrm{tree}}$.

In the rest of the paper T_n is a uniform rooted Cayley tree with n labeled vertices $\{1, 2, \dots, n\}$. We shall always see the edges of T_n as oriented towards the root vertex. For $i \geqslant 1$, we let X_i, Y_i be i.i.d. uniform points of $\{1, 2, \dots, n\}$ so that $\vec{E}_i = (X_i, Y_i)$ can be seen as i.i.d. uniform oriented edges (self-loops are allowed). In the sequel T_n will always be independent of $(X_i : i \geqslant 1)$ but not of $(Y_i : i \geqslant 1)$...

6.1 Warmup

In this section we introduce the main ingredients for the coupling of the parking process on Cayley trees with the Erdős–Rényi random graph. We shall first describe the different notions of components in the parking process. We then present the coupling in the case of the parking on random mappings (Proposition 6.4) for which the proof is easier to understand. Note that Lackner & Panholzer [128] already noticed striking similarities between parking on mappings and parking on Cayley trees.

6.1.1 Components and versions of parked trees

Fix a random uniform rooted Cayley tree T_n with n labeled vertices $\{1, 2, \dots, n\}$ and independently of it, let $X_i \in \{1, 2, \dots, n\}$ be uniform i.i.d. car arrivals for $i \geqslant 1$. For $m \geqslant 0$, we proceed to the parking of the first m cars as explained in the introduction and consider the clusters of parked cars. There are several possible notions to define those clusters and let us go from the more restrictive to the more permissive, see Figure 6.7:

- If we only keep the edges (and neighboring vertices) having a positive flux of cars (that is through which at least one car had to go), then we obtain the strong components, i.e. a subforest $T_{\mathsf{strong}}(n, m) \subset T_n$. The components of $T_{\mathsf{strong}}(n, m)$ different from the component containing the root vertex are[2] either isolated empty vertices or **strongly parked trees** which are rooted Cayley trees of size N carrying N labeled cars, so that all cars manage to park (outgoing flux 0), and such that all edges have a positive flux of cars.

- If we only keep the edges so that both extremities are occupied spots, then we obtain the full components $T_{\mathsf{full}}(n, m)$. The components of $T_{\mathsf{full}}(n, m)$ different from the component containing the root vertex are either isolated empty vertices or **fully parked trees** which are rooted Cayley trees of size N carrying N labeled cars and so that all cars manage to park (outgoing flux 0).

- Finally, if we only keep the edges emanating from the occupied vertices, then we obtain the near components $T_{\mathsf{near}}(n, m)$. The components of $T_{\mathsf{near}}(n, m)$ different from that of the root vertex are **nearly parked trees** i.e. rooted Cayley trees of size N carrying $N - 1$ labeled cars and so that the root vertex stays empty after parking the cars.

Of course we have $T_{\mathsf{strong}}(n, m) \subset T_{\mathsf{full}}(n, m) \subset T_{\mathsf{near}}(n, m)$ in terms of edge sets. The component of the root vertex in those forests may not be a strong/fully/nearly parked tree since a positive flux of car may exit through the root (or in the case of near components, the root vertex may contain a car). When the component of the root *is neither* a strongly/fully/nearly parked tree nor an empty vertex, we shall *color it in blue*. On a high level, the starting observation of this paper is that for n fixed the processes

$$m \mapsto T_\star(n, m) \quad \text{for } \star \in \{\mathsf{strong}, \mathsf{full}, \mathsf{near}\} \quad \text{are Markov processes,}$$

[2] after an increasing relabeling of its vertices and cars

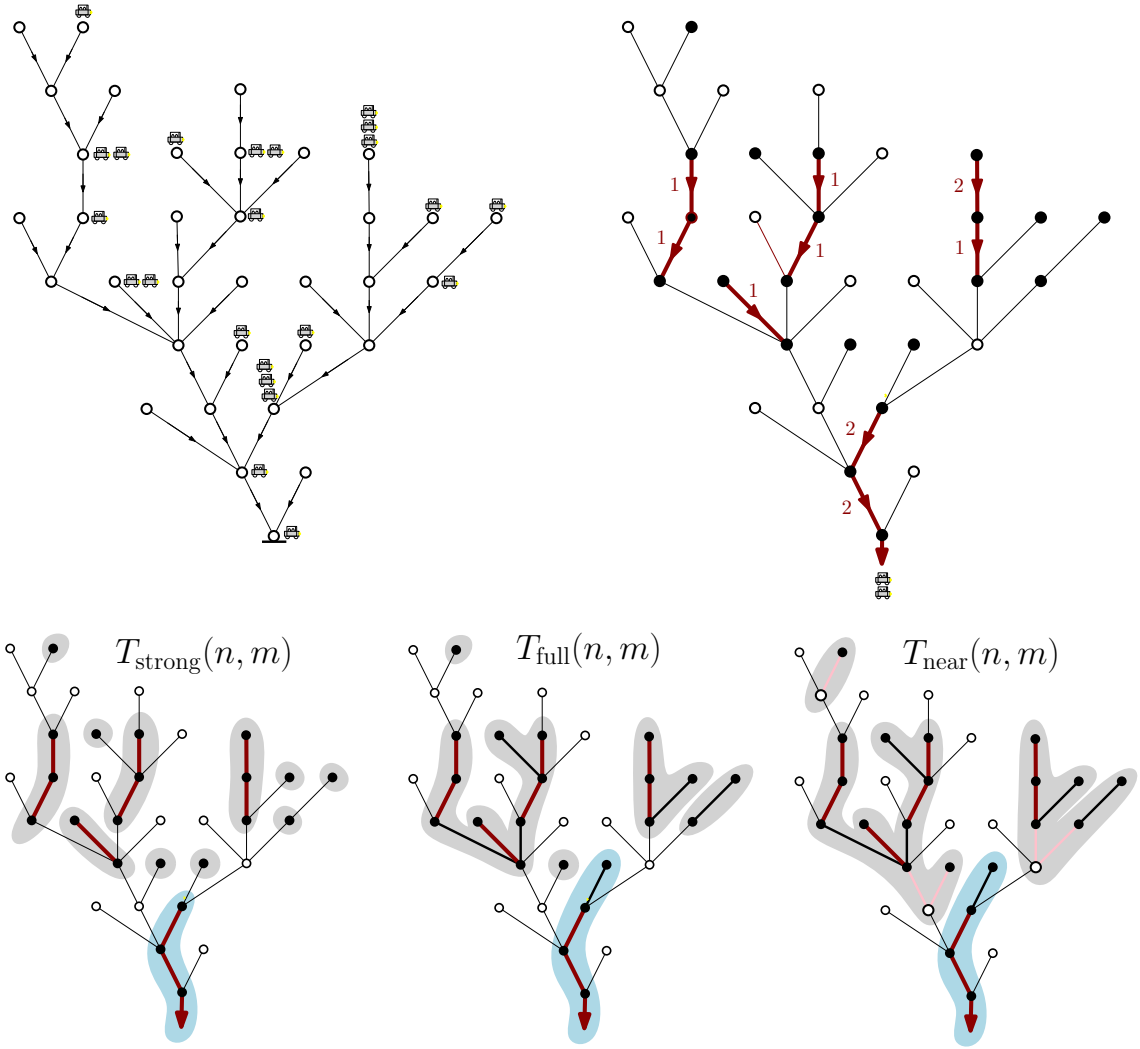


Figure 6.7: Illustration of the different notions of clusters of parked cars: from left to right the strong, full and near components. The labels of the vertices and of the cars are not displayed for better readability. After $m = 23$ car arrivals, the black vertices contain a car and the red edges have seen at least one car going through them. The grey components are, from left to right, strongly/fully/nearly parked trees. The components of the root vertex is not of the same type and is thus colored in blue.

see Proposition 6.6 and Section 6.2.3. Although the notions of strong or full components seem more natural than the notion of near components, we shall see in the next sections that the evolution of $m \mapsto T_{\text{near}}(n,m)$ is very close to the evolution of the Erdős–Rényi random graph, which constitutes the basis of our work. One key feature is that $T_{\text{near}}(n,m+1)$ is obtained from $T_{\text{near}}(n,m)$ by adding at most one edge (which is not the case for the two other notions of components).

Remark (Versions of parked trees). Fully parked trees have been considered by Lackner & Panholzer in [128] and strongly parked trees by King & Yan in [119]. Both works provide enumeration formulas

which we shall recover in Section 6.4. Different versions of fully parked trees have recently been investigated by Chen [59] and Panholzer [148], see Section 6.8.2 for more details.

6.1.2 Frozen and Erdős-Rényi random graphs

Recall from the introduction the definition of the random graph process $(G(n,m) : m \geqslant 0)$ obtained by adding sequentially i.i.d. uniform unoriented edges $E_i = \{X_i, Y_i\}$ where the oriented edges $\vec{E}_i = (X_i, Y_i)$ have i.i.d. uniform endpoints over $\{1, 2, \ldots, n\}$.

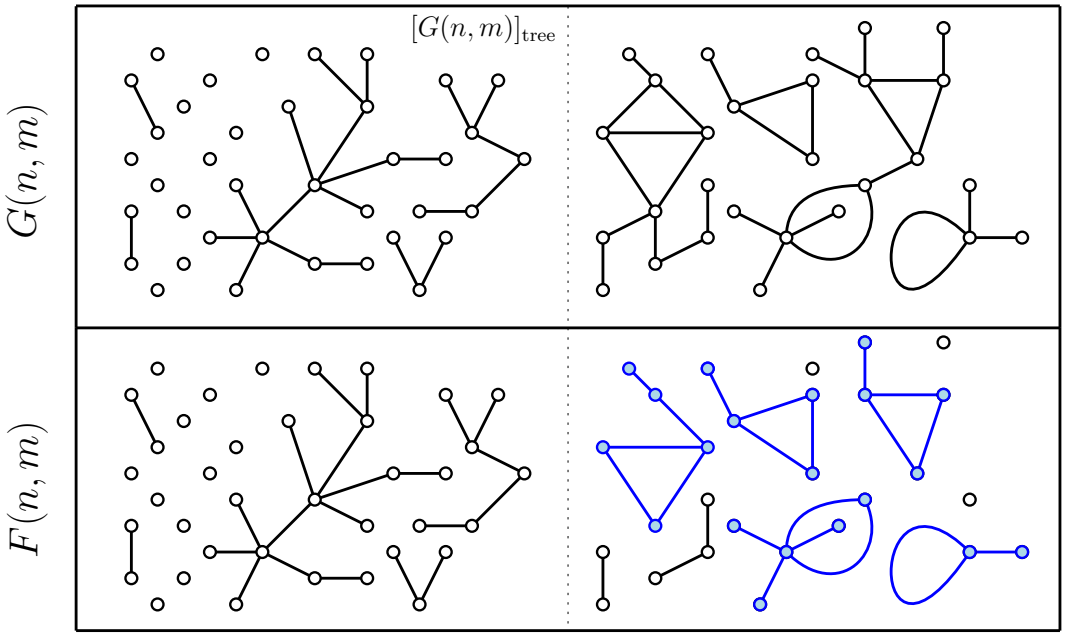


Figure 6.8: Illustration of the inclusion of $F(n,m)$ inside $G(n,m)$. The two processes coincide on connected components of $G(n,m)$ that do not contain surplus (left column) and $F(n,m)$ is obtained by further splitting the remaining components in $G(n,m)$.

The frozen process $(F(n,m) : m \geqslant 0)$ is constructed by discarding certain of those edges and coloring the vertices in blue or white (see Figure 6.3). In particular $\|F(n,m)\|_{\bullet\text{-}\bullet} \leqslant m$ and the inequality may be strict. The vertices in the frozen components of $F(n,m)$ will be colored in blue while the others stay white. In that construction, the process $F(n,\cdot)$ lives inside $G(n,\cdot)$ and in particular for every $n, m \geqslant 0$ we have

$$F(n,m) \subset G(n,m) \tag{6.2}$$

in terms of edge set. Moreover, it is easy to see by induction on $m \geqslant 0$ that $F(n,m)$ and $G(n,m)$ coincide on the forest part $[G(n,m)]_{\text{tree}}$, see Figure 6.8.

6.1.3 Parking on random mapping and the frozen Erdős–Rényi

A *mapping* is a graph over the n labeled vertices $\{1,2,\dots,n\}$ with oriented edges and so that each vertex has exactly one edge pointing away from it, see Figure 6.9. Equivalently, the oriented edges of the graph can be seen as $i \to \sigma(i)$ where σ is a map $\{1,2,\dots,n\} \to \{1,2,\dots,n\}$, hence the name "mapping". In particular, if M_n is a uniform random mapping on $\{1,2,\dots,n\}$ then the *targets* $\sigma(i)$ i.e. the vertices to which point the edges emanating from $1,2,\dots,n$ are just i.i.d. uniform on $\{1,2,\dots,n\}$.

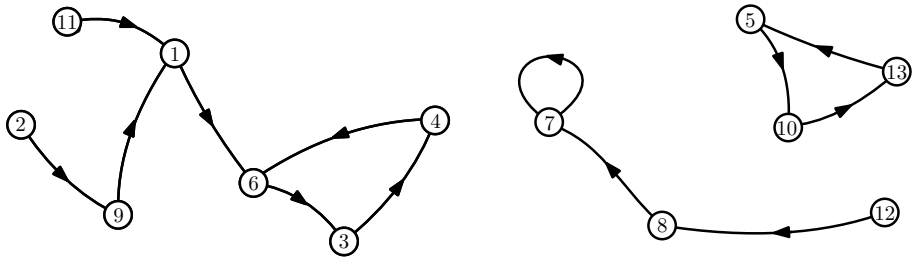


Figure 6.9: An example of a mapping over $\{1,2,\dots,13\}$.

The parking process can be extended from a rooted tree to a mapping (see [128]): Given the random mapping M_n, we consider independent uniform car arrivals $X_i \in \{1,2,\dots,n\}$. Each car tries to park on its arrival vertex and stops there if the parking spot is empty. Otherwise the car follows the oriented edges of M_n and takes the first available space, if there is one. If the car is caught in an endless loop, then it exits without parking.

As for the parking on T_n, when m cars have arrived we can define submappings (subgraphs of a mapping)

$$M_{\text{strong}}(n,m) \subset M_{\text{full}}(n,m) \subset M_{\text{near}}(n,m),$$

by keeping respectively the oriented edges with positive flux of cars, the oriented edges linking two occupied spots, or the oriented edges emanating from occupied spots in the parking process. In the remainder of this section we shall focus on $M_{\text{near}}(n,m)$. When an (oriented) cycle is discovered in $M_{\text{near}}(n,m)$ we shall color the entire non-oriented component in blue.

In the above construction, the car arrivals X_i's are independent of the uniform random mapping M_n. The main observation of this section is that one can in fact couple the oriented edges $\vec{E}_i = (X_i, Y_i)$ from which we constructed the process $F(n,\cdot)$ with M_n so that M_n has the correct law and furthermore that $F(n,m)$ has the same components (in terms of subsets of vertices) as $M_{\text{near}}(n,m)$. In particular:

- M_n will be constructed from $(X_i, Y_i)_{i\geqslant 1}$.
- $M_{\text{near}}(n,\cdot)$ is constructed by additionally taking $(X_i : i \geqslant 1)$ to be the car arrival process.
- In fact, the mapping M_n ends up being independent of $(X_i : i \geqslant 1)$ in the construction.

Of course, the components of $F(n,m)$ and $M_{\text{near}}(n,m)$ are not the same subgraphs since we will need to redirect[3] the edges of $F(n,m)$ to obtain $M_{\text{near}}(n,m)$. More precisely:

Proposition 6.4 (Coupling of parking on mapping with the frozen Erdős–Rényi)**.** *We can couple the uniform random mapping M_n with the X_i's and the Y_i's in such a way that*

(Parking on mapping) *The graph M_n is a uniform random mapping on $\{1,2,\dots,n\}$ independent of the car arrivals $(X_i : i \geqslant 1)$,*

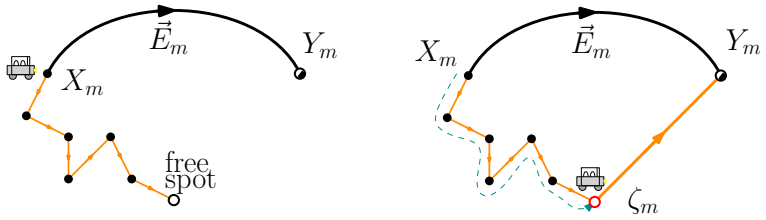
(Coupling with $F(n,\cdot)$) *For each $m \geqslant 0$, the subgraph $M_{\text{near}}(n,m)$ has the same (unoriented) connected components as $F(n,m)$. More precisely:*

- *The blue components of $F(n,m)$ correspond to components with surplus in $M_{\text{near}}(n,m)$,*
- *The indices of the discarded edges in $F(n,\cdot)$ correspond to the indices of the cars that do not manage to park on M_n.*

Proof. We will construct the mapping M_n by prescribing the targets $\sigma(i)$ of its vertices using the oriented edges $\vec{E}_m$'s according to the following rule:

From oriented edges to parking on mapping. The starting points $(X_i : i \geqslant 1)$ of the edges $\vec{E}_i$ are the i.i.d. arrivals of the cars over $\{1,2,\dots,n\}$. We use them to construct iteratively an increasing sequence of oriented graphs $(M(n,m) : m \geqslant 0)$ where $M(n,0)$ is the graph over $\{1,2,\dots,n\}$ with no edge. For $m \geqslant 1$, we use the edges of $M(n,m-1)$ to (try to) park the mth car arrived on X_m. If we manage to park it, we denote by $\zeta_m \in \{1,2,\dots,n\}$ its parking spot, otherwise we set $\zeta_m = \dagger$. When $\zeta_m \neq \dagger$, we add the edge $\zeta_m \to Y_m$ to $M(n,m-1)$ to form $M(n,m)$, equivalently we put

$$\sigma(\zeta_m) = Y_m \quad \text{when } \zeta_m \neq \dagger. \tag{6.3}$$



It is perhaps not clear for the reader how the above rule serves as recipe to construct a mapping, so let us make a couple of remarks and refer to Figure 6.10 for a step-by-step illustration. First notice that Y_m is (obviously) independent of $M(n,m-1)$, but then $M(n,m)$ depends (obviously!) on Y_m. It is easy to see by induction that every vertex in $M(n,m)$ has at most one edge pointing away from it and that the vertices having an emanating edge are those which already accommodate a car. If the mth car is trapped in an endless loop the graph does not evolve and we have $M(n,m) = M(n,m-1)$.

[3] note that when we redirect an oriented edge, we may change its starting vertex (as opposed to just changing its orientation)

When $\{X_i : 1 \leqslant i \leqslant m\}$ has spanned $\{1, 2, \dots, n\}$ (which holds for large enough m when the $\vec{E}_i$'s are i.i.d. uniform oriented edges), the graph $M(n, m)$ is constant and we *define*

$$M_n := \bigcup_{m \geqslant 1} M(n, m),$$

which is a random mapping of size n. With this construction, it is clear that we have

$$M_{\text{near}}(n, m) = M(n, m), \qquad \text{for every } m \geqslant 0.$$

The second point of the proposition is then easy to check by induction (see Figure 6.10) in particular the edges emanating from a blue component in $F(n, \cdot)$ correspond, via the coupling, to cars arriving on a component already containing an oriented loop: such cars will be trapped in an endless loop and contribute to the outgoing flux in the parking, whereas the corresponding edges are discarded in the frozen process. The non-trivial probabilistic point consists in showing the first point of the proposition, i.e. that this coupling reproduces the parking on a uniform random mapping or in other words, that M_n, made of the edges

$$\zeta_i \to Y_i, \qquad \text{when } \zeta_i \neq \dagger$$

forms a uniform mapping, independent of the X_i's (but not of the Y_i's !!!). Fix $m \geqslant 1$ and notice that ζ_m is determined by $M(n, m-1)$ and X_m, and in particular is independent of Y_m. We deduce that conditionally on $\zeta_m \neq \dagger$, its target $\sigma(\zeta_m) = Y_m$ is independent of the edges already constructed in $M(n, m-1)$, also independent of $(X_i : i \geqslant 1)$, and is uniform over $\{1, 2, \dots, n\}$. Since $\{\zeta_i : i \geqslant 1\}$ spans $\{1, 2, \dots, n\}$ almost surely, conditionally on the X_i's, the ζ_i (different from $\dagger$) can be seen as a way to sample the vertices of $\{1, 2, \dots, n\}$ (and they all will be sampled) and at each step they are assigned an independent random uniform target. A moment's thought shows that the targets of all vertices are i.i.d. uniform over $\{1, 2, \dots, n\}$ and independent of $(X_i : i \geqslant 1)$.

□

6.2 Coupling of parking on Cayley trees with the frozen Erdős–Rényi

We shall now perform a similar coupling between the frozen Erdős–Rényi process and the parking process on a uniform Cayley tree. Although the main idea (considering T_n as unknown and revealing $T_{\text{near}}(n, m)$ step-by-step in a Markovian way) is the same, the Markovian exploration of Cayley trees is a little more complicated than in the case of random mappings and we shall need some extra randomness to perform the coupling.

6.2.1 Markovian exploration of rooted Cayley trees

We present the Markovian explorations of uniform Cayley trees which are adapted from [65]. We call these "peeling explorations" by analogy with the peeling process of random planar maps [72]. We

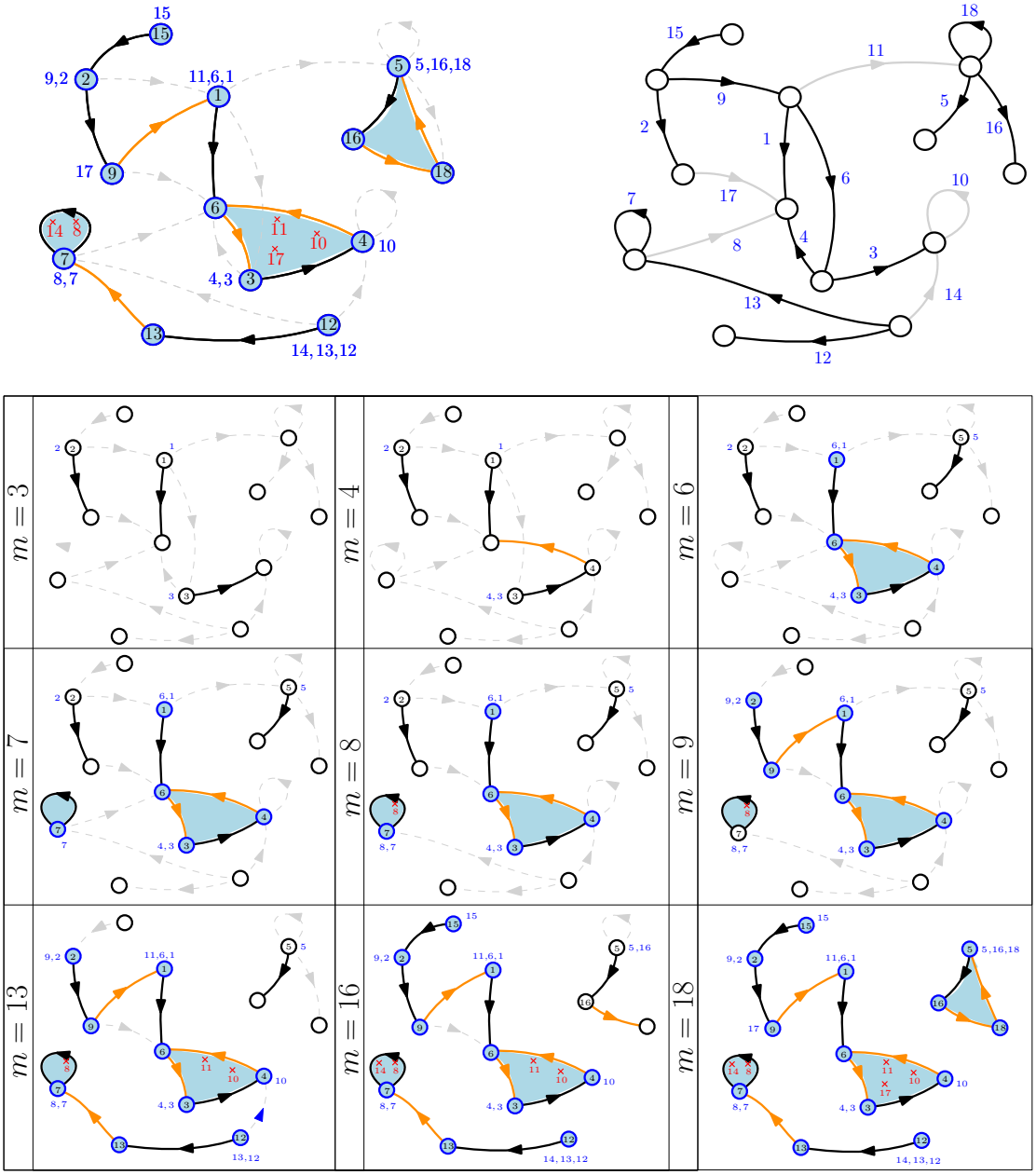


Figure 6.10: Illustration of the redirections of the edges $(\vec{E}_i : 0 \leqslant i \leqslant m)$ to obtain $M(n, m)$. On the top right, a possible value for $G(13, 18)$ and its corresponding $F(13, 18)$ if one only keeps the black edges. On the left, the corresponding $M(13, 18)$. The redirected edges which are different from those in $F(n, \cdot)$ are in orange. The labels of the vertices are not displayed for better visibility, but the labels of the cars are present (blue for their arrival vertices, black for the parking spot, and red if the car does not manage to park). In the tabular, we represented a step-by-step construction by displaying $M(13, m)$ for $m = 3, 4, 6, 7, 8, 9, 13, 16$ and 18.

will later tailor those explorations to the parking process using a specific peeling algorithm and this will yield the coupling with the frozen Erdős–Rényi, see Proposition 6.6.

Recall that a rooted Cayley tree $\mathbf{t}$ is an unordered tree over the n labeled vertices $\{1, 2, \dots, n\}$ where one of its vertices has been distinguished and called the root. This root enables us to orient all edges of $\mathbf{t}$ towards it. As in the case of mappings, this allows us to speak of the target $\sigma(i)$ of each vertex $i \in \{1, 2, \dots, n\}$ which is the vertex to which points the edge emanating from i. A difference with the previous section is that in the case of trees no loop can be created and the root vertex $\mathbf{r}$ of $\mathbf{t}$ has no target which we write as $\sigma(\mathbf{r}) = \varnothing$. The information on $\mathbf{t}$ is thus encoded by the n "instructions"

$$\{i \to \sigma(i)\} \quad \text{where } \sigma(i) \in \{1, 2, \dots, n\} \cup \{\varnothing\} \text{ for } i \in \{1, 2, \dots, n\}.$$

An *exploration* of $\mathbf{t}$ can be seen as revealing those n instructions one by one by discovering the target of one vertex at a time. A set S of instructions is said to be *compatible* if it corresponds to a subset of instructions of some tree. Any such set can be interpreted as a forest of rooted trees by connecting the vertices to their revealed targets, see Figure 6.11. If the target of the root is "revealed" (one should probably better say that the root vertex is revealed) then we record this information by coloring the corresponding tree in blue, the other trees being referred to as white.

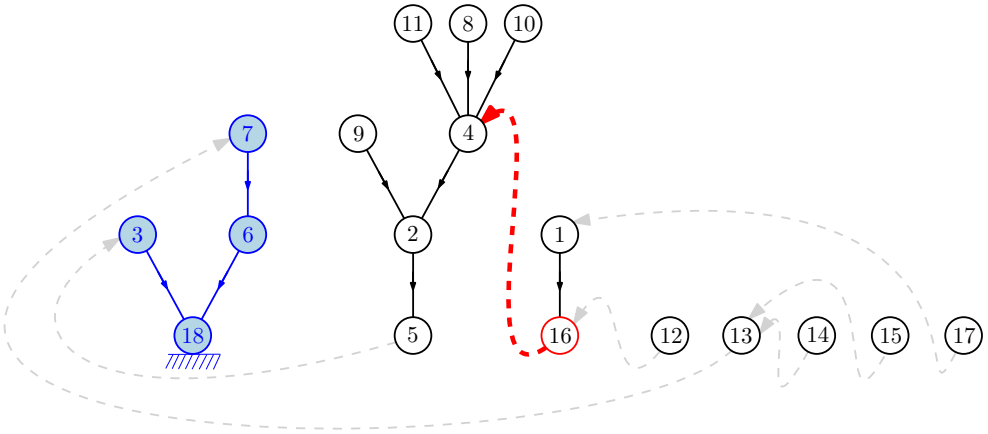


Figure 6.11: Illustration of the rooted forest obtained from the explored subset $S = \{3 \to 18, 18 \to \varnothing, 7 \to 6, 6 \to 18, 11 \to 4, 8 \to 4, 10 \to 4, 4 \to 2, 9 \to 2, 2 \to 5, 1 \to 16\}$. Notice that the root vertex has been revealed thanks to the presence of $18 \to \varnothing$ and we have colored the corresponding tree in blue. In this example the next vertex to be peeled is $\mathfrak{a}(S) = 16$ and its target is the vertex 4. The remaining edges of the underlying tree are displayed in dotted gray.

Of course, a given rooted Cayley tree $\mathbf{t}$ with n vertices can be explored in $n!$ different ways and we shall choose one using a function $\mathfrak{a}$, called the *peeling algorithm*, which associates any subset S of compatible instructions (which does not yet form a tree) with a vertex $\mathfrak{a}(S)$ whose target is not revealed yet (in particular $\mathfrak{a}$ depends only on S and not on the underlying tree, and $\mathfrak{a}(S)$ must be a

root of a standard white tree of the forest associated with S). The peeling of $\mathbf{t}$ with algorithm $\mathfrak{a}$ is then the sequence

$$\mathbf{S}_0^{\mathfrak{a}} \subset \mathbf{S}_1^{\mathfrak{a}} \subset \cdots \subset \mathbf{S}_n^{\mathfrak{a}} = \bigcup_{i=1}^{n} \{i \to \sigma(i)\},$$

where $\mathbf{S}_0^{\mathfrak{a}}$ is the empty set and $\mathbf{S}_{i+1}^{\mathfrak{a}} = \mathbf{S}_i^{\mathfrak{a}} \cup \{\mathfrak{a}(\mathbf{S}_i^{\mathfrak{a}}) \to \sigma(\mathfrak{a}(\mathbf{S}_i^{\mathfrak{a}}))\}$ for all $i \leqslant n-1$. We shall call those explorations "peeling explorations" or "Markovian explorations", indeed, when the peeling algorithm $\mathfrak{a}$ is deterministic and when the underlying tree $\mathbf{t}$ is a uniform rooted Cayley tree, this exploration is a Markov chain with explicit probability transitions:

Proposition 6.5 (Markov transitions for peeling exploration of uniform Cayley trees)**.** *Fix a peeling algorithm $\mathfrak{a}$. If T_n is a uniform rooted Cayley tree with n vertices, then the exploration $(\mathbf{S}_i^{\mathfrak{a}})_{0\leqslant i\leqslant n}$ of T_n with algorithm $\mathfrak{a}$ is a Markov chain whose probability transitions are described as follows. Conditionally on $\mathbf{S}_i^{\mathfrak{a}}$ and on $\mathfrak{a}(\mathbf{S}_i^{\mathfrak{a}})$, in the forest representation of $\mathbf{S}_i^{\mathfrak{a}}$ we denote by $k \geqslant 1$ the number of vertices of the tree of root $\mathfrak{a}(\mathbf{S}_i^{\mathfrak{a}})$ and by $\ell \geqslant 0$ the number of vertices of the blue tree (if any) then:*

- *If $\ell = 0$, with probability $\frac{k}{n}$ we have $\sigma(\mathfrak{a}(\mathbf{S}_i^{\mathfrak{a}})) = \varnothing$ (i.e. the vertex we peel is the root of the underlying Cayley tree), otherwise $\sigma(\mathfrak{a}(\mathbf{S}_i^{\mathfrak{a}}))$ is a uniform vertex not belonging to the tree of root $\mathfrak{a}(\mathbf{S}_i^{\mathfrak{a}})$.*
- *If $\ell \geqslant 1$, with probability $\frac{\ell+k}{n}$ the target $\sigma(\mathfrak{a}(\mathbf{S}_i^{\mathfrak{a}}))$ is a uniform vertex of the blue tree of $\mathbf{S}_i^{\mathfrak{a}}$, otherwise it is a uniform vertex of the remaining trees except the tree of root $\mathfrak{a}(\mathbf{S}_i^{\mathfrak{a}})$.*

The proof is similar to that of [65, Proposition 1] and relies on counting formulas established in [65, Lemma 5] based on Pitman's approach [149, Lemma 1]. Specifically we have:

Lemma 6.1. *If $\mathbf{f}$ is a forest of white rooted trees on $\{1, 2, \ldots, n\}$ with m edges, then the number of rooted Cayley trees containing $\mathbf{f}$ is n^{n-m-1}. If $\mathbf{f}^*$ is a forest of rooted trees on $\{1, 2, \ldots, n\}$ with m edges containing a blue tree with $\ell \geqslant 1$ vertices, then the number of rooted Cayley trees containing $\mathbf{f}^*$ with the root being the root of the blue tree is ℓn^{n-m-2}.*

To be precise, [65] considers Cayley trees as rooted at the vertex n whereas we allow the root vertex to be any vertex of $\{1, 2, \ldots, n\}$ hence the factor n difference between the numbers appearing in the above lemma and those of [65, Lemma 5].

Proof of Proposition 6.5. Given the last lemma, the proof is easy to complete. It suffices to notice that since the underlying tree T_n is uniform over all rooted Cayley trees with n vertices, for all $i \geqslant 0$, conditionally on $\mathbf{S}_i^{\mathfrak{a}}$, the tree T_n is a uniform tree among those which contain the forest associated to $\mathbf{S}_i^{\mathfrak{a}}$ (with or without a blue tree depending whether the root vertex has been revealed or not). Hence, for every (compatible) target $v \in \{1, 2, \ldots, n\} \cup \{\varnothing\}$,

$$\mathbb{P}\Big(\sigma(\mathfrak{a}(\mathbf{S}_i^{\mathfrak{a}})) = v \Big| \mathbf{S}_i^{\mathfrak{a}}, \mathfrak{a}(\mathbf{S}_i^{\mathfrak{a}})\Big) = \frac{\#\{\mathbf{t} \text{ containing the forest associated with } \mathbf{S}_i^{\mathfrak{a}} \cup \{\mathfrak{a}(\mathbf{S}_i^{\mathfrak{a}}) \to v\}\}}{\#\{\mathbf{t} \text{ containing the forest associated with } \mathbf{S}_i^{\mathfrak{a}}\}}.$$

Using Lemma 6.1, we recognize the transition probabilities given in Proposition 6.5 and obtain the desired result.

□

The interest of Proposition 6.5 is that different peeling algorithms can be used to explore a uniform Cayley tree. See [65, Sections 3 and 4] for applications to the greedy independent set, the Aldous–Broder or Pitman algorithms.

6.2.2 The near exploration

Recall the notion of near components defined in Section 6.1.1. We shall see that $T_{\mathrm{near}}(n,\cdot)$ can be interpreted as a peeling process of T_n using an algorithm (called $\mathfrak{a}_{\mathrm{near}}$ below) tailored to the parking process. Furthermore, as in the last section, we shall make a coupling of T_n with the oriented edges $\vec{E}_i$'s so that T_n stays independent of the car arrivals X_i's but in such a way that $T_{\mathrm{near}}(n,\cdot)$ is closely related to the frozen process $F(n,\cdot)$.

Proposition 6.6 (The main coupling)**.** *We can couple T_n with the X_i's and the Y_i's so that:*

- ***(Parking on Cayley tree)*** *The tree T_n is a uniform rooted Cayley tree independent of the car arrivals $(X_i : i \geqslant 1)$.*

- ***(Coupling with $F(n,\cdot)$)*** *For each $m \geqslant 0$, the subforest $T_{\mathrm{near}}(n,m)$ has the same (unoriented) connected components (in terms of subsets of vertices) as $F(n,m)$ where all the frozen components have been joined. More precisely:*

  - *The white components of $F(n,m)$ are the connected components of $T_{\mathrm{near}}(n,m)$ with a possible exception for the component containing the root if $T_{\mathrm{near}}(n,m)$ has a blue component,*

  - *The vertices of the blue components of $F(n,m)$ correspond to the vertices of the (unique) blue component of $T_{\mathrm{near}}(n,m)$,*

  - *The indices of the discarded edges in $F(n,\cdot)$ correspond to the indices of the cars that do not manage to park on T_n.*

Proof. As in the proof of Proposition 6.4, we shall construct T_n using the $\vec{E}_i$'s. The main difference being that the appearance of the first cycle in $G(n,m)$ corresponds to the detection of the root vertex in the Cayley tree and that we need an additional randomization to redirect some of the edges $\vec{E}_i$ (whereas in the case of mapping, the redirection was a measurable function of the X_i and Y_i).

From oriented edges to parking on trees. The starting points $(X_i : i \geqslant 1)$ of the edges $\vec{E}_i$ are the i.i.d. arrivals of the cars over $\{1, 2, \dots, n\}$. We use them to construct iteratively an increasing sequence of compatible instructions $(\mathbf{S}_m^{\text{park}} : m \geqslant 0)$ or equivalently of growing forests $(T(n, m) : m \geqslant 0)$ with zero or one blue tree. Initially $\mathbf{S}_0^{\text{park}}$ is the empty set and for $m \geqslant 1$, we use the edges of $\mathbf{S}_{m-1}^{\text{park}}$ to (try to) park the mth car arrived on X_m. If we manage to park it, we denote by $\zeta_m \in \{1, 2, \dots, n\}$ its parking spot, otherwise set $\zeta_m = \dagger$. If $\zeta_m = \dagger$ then $\mathbf{S}_m^{\text{park}} = \mathbf{S}_{m-1}^{\text{park}}$. Otherwise

- if the addition of the edge $\zeta_m \to Y_m$ does not create a cycle in $T(n, m-1)$, then add it to $\mathbf{S}_{m-1}^{\text{park}}$ to form $\mathbf{S}_m^{\text{park}}$,
- if the addition of the edge $\zeta_m \to Y_m$ creates a cycle in $T(n, m-1)$ then
  - If $T(n, m-1)$ has no blue tree (the root vertex is not revealed), then add $\zeta_m \to \varnothing$ to form $\mathbf{S}_m^{\text{park}}$,
  - Otherwise add $\zeta_m \to U_m$ where U_m is a uniform point over the blue tree of $T(n, m-1)$ sampled independently of the past to form $\mathbf{S}_m^{\text{park}}$, see Figure 6.12.

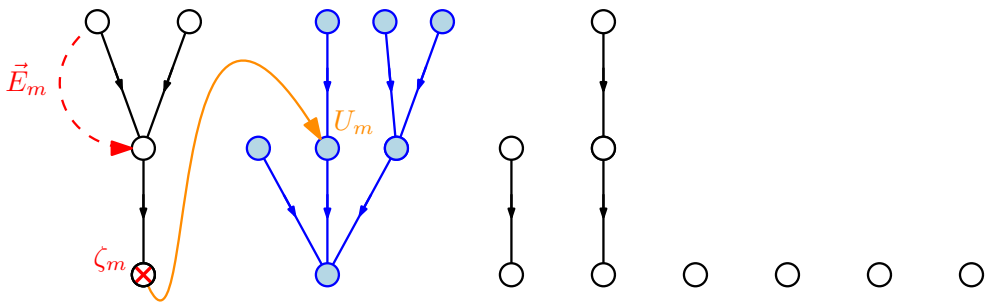


Figure 6.12: When a new cycle is created by the addition of the edge $\vec{E}_m$, the target of ζ_m is chosen uniformly in the blue tree.

As in the proof of Proposition 6.4, the increasing forests $T(n, m)$ eventually stabilize to form a (blue) tree and we put

$$T_n := \bigcup_{m \geqslant 0} T(n, m).$$

With this definition, it is clear that we have

$$T_{\text{near}}(n, m) = T(n, m) \quad \text{for all } m \geqslant 0,$$

and the deterministic properties of the coupling between the parking on T_n and $F(n, \cdot)$ are easy to prove by induction. It thus remains to prove that T_n is indeed a uniform rooted Cayley tree independent of $(X_i : i \geqslant 1)$. To see this, we shall interpret the Markov chain $(\mathbf{S}_m^{\text{park}} : m \geqslant 0)$ as a peeling exploration of a uniform Cayley tree. Specifically, given $(X_i : i \geqslant 1)$ we construct a peeling algorithm $\mathfrak{a}_{\text{near}}$ as follows. At $m = 0$, we start from the empty set $\mathbf{S}_0^{\mathfrak{a}_{\text{near}}}$ and for $m \geqslant 1$, if $\mathbf{S}_{m-1}^{\mathfrak{a}_{\text{near}}}$ is

the current status of the exploration, we let a car arrive on vertex X_m. The car follows the oriented edges already present in $\mathbf{S}_{m-1}^{\mathfrak{a}_{\text{near}}}$ to find its parking spot ζ_m. As in the case of random mapping, if the car does not park (i.e. exits through the root of the tree) then we put $\zeta_m = \dagger$ and do not trigger a peeling step, i.e. move to step $m+1$. In the case $\zeta_m \neq \dagger$ we put

$$\mathfrak{a}_{\text{near}}(\mathbf{S}_{m-1}^{\mathfrak{a}_{\text{near}}}) = \zeta_m, \tag{6.4}$$

that is we reveal the target $\sigma(\zeta_m) \in \{1, 2, \dots, n\} \cup \{\varnothing\}$ and include $\zeta_m \to \sigma(\zeta_m)$ to form $\mathbf{S}_m^{\mathfrak{a}_{\text{near}}}$. The process $(\mathbf{S}_m^{\text{park}} : m \geqslant 0)$ has the same law as the peeling exploration $(\mathbf{S}_m^{\mathfrak{a}_{\text{near}}} : m \geqslant 0)$ with the random algorithm $\mathfrak{a}_{\text{near}}$: indeed the probability transitions of $\mathbf{S}^{\text{park}}$ described above are the same as those of Proposition 6.5. Conditionally on the X_i's, the function $\mathfrak{a}_{\text{near}}$ can be seen as a deterministic peeling algorithm, so by Proposition 6.5, the tree T_n constructed this way is indeed uniform. In particular, the tree T_n is independent of the $(X_i : i \geqslant 1)$ which are themselves i.i.d. uniform on $\{1, 2, \dots, n\}$. Our claim follows

□

Remark (Other couplings between mappings and Cayley trees.). By combining the constructions in the previous two sections, we get a coupling between a uniform mapping M_n and a uniform rooted Cayley tree T_n which is different from the one [13] based on Joyal's bijection.

Convention

In the rest of the paper we shall always suppose that the tree T_n the car arrivals $(X_i : i \geqslant 0)$ and the frozen Erdős–Rényi process $F(n, \cdot)$ are built from the sequence $(\vec{E}_i = (X_i, Y_i) : i \geqslant 1)$ as in the proof of Proposition 6.6.

6.2.3 The strong exploration

We saw above in the proof of Proposition 6.6 that the process $m \mapsto T_{\text{near}}(n, m)$ can be seen as a peeling exploration of the underlying tree T_n with the algorithm $\mathfrak{a}_{\text{near}}$ that reveals the targets of the parked vertices. In a similar vein, one can interpret $m \mapsto T_{\text{strong}}(n, m)$ as a peeling exploration where we reveal the target of a vertex when a car leaves that vertex (for the first time). More precisely, we let the cars arrive one by one on the vertices X_i and peel the vertices when the cars need to *move* and find their potential parking spot (as opposed to the former near algorithm where we peeled the vertex on which the ith car parked). In particular, the arrival of a car may result in no peeling step (e.g. if the car parks on its arrival vertex) or to several peeling steps, see Figure 6.13. We do not formalize further and hope it is clear for the reader. After m cars have arrived, this exploration has revealed the strong components

$$T_{\text{strong}}(n, m)$$

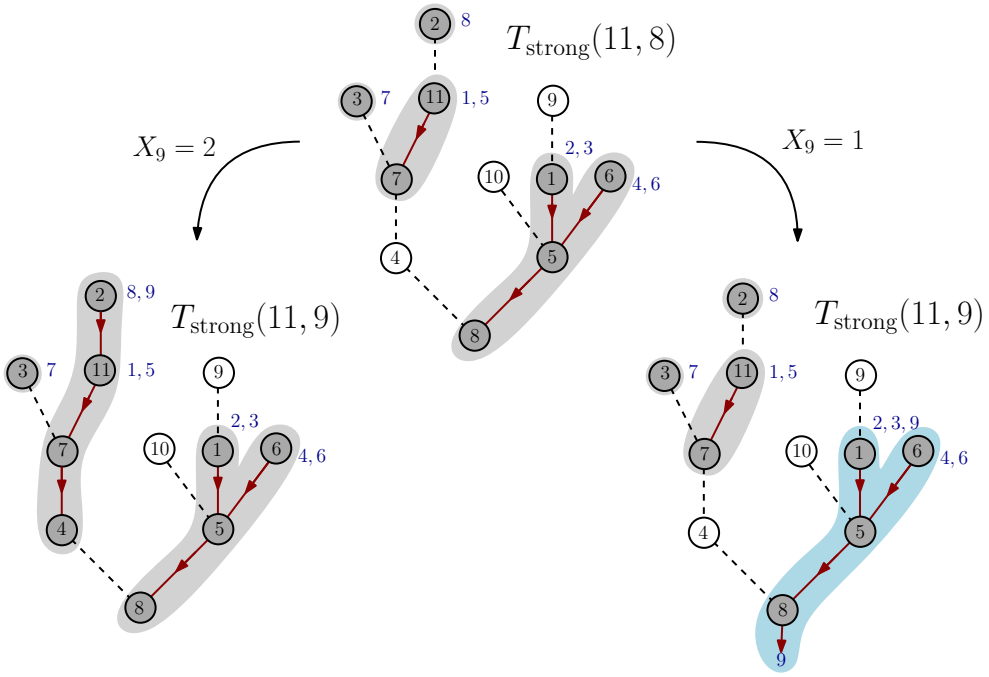


Figure 6.13: Illustration of the strong parking peeling algorithm. On the top, the current status $\{11 \to 7, 1 \to 5, 6 \to 5, 5 \to 8\}$ after 8 cars have arrived triggering in total 4 peeling steps. The available spots are in white whereas the gray vertices already contain a car, the red edges have positive flux. If the next car arrives on vertex 2, it triggers two peeling steps resulting in $2 \to 11$ and $7 \to 4$ before parking on vertex 4. If the next car arrives on vertex 1, it follows the edges, triggers the step $8 \to \varnothing$ and cannot park. The root components then becomes blue because we discovered the root of the underlying tree.

which we defined in Section 6.1.1. Recall also that if the outgoing flux of cars is positive then the tree carrying the root vertex in $T_{\text{strong}}(n, m)$ is seen as a blue tree (and indeed we discovered the root vertex during the peeling exploration).

This peeling exploration enables to see $T_{\text{strong}}(n, m)$ (together with its coloring) as a Markov chain. We shall not describe its probability transitions, but we shall use it to relate the probability that the root of T_n contains a car to the probability that the outgoing flux in T_n is equal to 0. Recall from the introduction that $D(n, m)$ is the number of cars that did not manage to park among the first m cars.

Lemma 6.2. *For $n \geqslant 1$ and $0 \leqslant m \leqslant n$ we have*

$$\mathbb{P}(\textit{the root of } T_n \textit{ is not occupied by one of the first } m \textit{ cars } | D(n, m) = 0) = 1 - \frac{m}{n}.$$

Proof. Let us explore the underlying tree T_n using the strong parking peeling algorithm until we manage to park $m \leqslant n$ cars (notice that the number of peeling steps is between 0 and $m - 1$). On the event $\{D(n, m) = 0\}$ all peeling steps performed so far have not revealed the root vertex of T_n

(we did not need to peel the root vertex since no car was emanating from it) so the corresponding forest $T_{\text{strong}}(n,m)$ is made of white rooted trees (no blue tree) containing $n-m$ isolated vertices which do not yet accommodate a car. By the proof of Proposition 6.5 and Lemma 6.1, conditionally on $T_{\text{strong}}(n,m)$, the probability that the root vertex of T_n (which is yet undiscovered) is a given root of a tree $\mathbf{t}$ of $T_{\text{strong}}(n,m)$ is proportional to the number of vertices of $\mathbf{t}$. Hence, the probability that the root vertex of T_n does not contain one of the first m cars is $\frac{n-m}{n}$ as desired.

□

Remark. We saw above that $m \mapsto T_{\text{strong}}(n,m)$ and $m \mapsto T_{\text{near}}(n,m)$ can be seen as peeling explorations of T_n. It does not seem to be the case for $m \mapsto T_{\text{full}}(n,m)$ although it is a Markov process and might alternatively be used to prove the above proposition.

6.3 Free forest property

In this section we gather several results about (random uniform) labeled (unrooted unordered) forests over $\{1,2,\dots,n\}$. We first recall their enumeration from classical results of Rényi and Britikov.

6.3.1 Uniform (unrooted) forest

Let $\mathfrak{F}(n,m)$ be the set of all unrooted unordered forests over the n labeled vertices $\{1,2,\dots,n\}$ with $n-m$ components (hence m edges in total). To enumerate such forests it is better to considered the trees as *indexed* by $\{1,2,\dots,n-m\}$ and consider the set of all unrooted, unordered forests of $\{1,2,\dots,n\}$ with $n-m$ components indexed by $1,2,\dots,n-m$. The number of such forests with components of sizes $(k_1,\dots,k_{n-m})$ is equal to

$$\binom{n}{k_1,\dots,k_{n-m}} \prod_{i=1}^{n-m} k_i^{k_i-2}, \tag{6.5}$$

the binomial coefficient $\binom{n}{k_1,\dots,k_{n-m}}$ counts for the number of choices to partition the n vertices in a list of $n-m$ subsets of $k_1,\dots,k_{n-m}$ vertices and on each subset there are $k_i^{k_i-2}$ ways to choose a spanning tree (Cayley's formula). To manipulate those numbers, let us introduce

$$\mathbf{T}(z) = \sum_{n\geqslant 1} \frac{n^{n-2}}{n!} z^n \tag{6.6}$$

the exponential generating function of (unrooted) Cayley trees. Summing (6.5) over all choices of $k_1,\dots,k_{n-m}$ and dividing by $(n-m)!$ to remove the indexation of the components we deduce that

$$\#\mathfrak{F}(n,m) = \frac{n!}{(n-m)!}[x^n]\mathbf{T}^{n-m}(x), \tag{6.7}$$

where we recall the standard notation $[x^n]\sum_{i\geqslant 0} a_i x^i = a_n$. Based on (6.7) Rényi [156] showed

$$\#\mathfrak{F}(n,m) = \frac{1}{(n-m)!}\sum_{i=0}^{n-m}\binom{n-m}{i}\left(-\frac{1}{2}\right)^i (n-m+i)n^{m-i-1}\frac{n!}{(m-i)!}. \tag{6.8}$$

Note that the power series $\mathbf{T}(z)$ is convergent when $|z| \leqslant \mathrm{e}^{-1}$, and for $z = \mathrm{e}^{-1}$ we have $\mathbf{T}(\mathrm{e}^{-1}) = \frac{1}{2}$ (it follows from (6.20) below) so that $2\mathbf{T}(z/\mathrm{e})$ is the generating function of a probability measure

$$\mu(k) \quad := \quad 2 \cdot \frac{k^{k-2}}{\mathrm{e}^k \cdot k!} \qquad \text{for } k \in \{1, 2, \dots\}, \tag{6.9}$$

of expectation $2z\partial_z \mathbf{T}(z)|_{z=\mathrm{e}^{-1}} = 2$ which has furthermore a heavy tail $\mu(k) \sim \sqrt{\frac{2}{\pi}} \cdot k^{-5/2}$ as $k \to \infty$. The following proposition is the probabilistic translation of the above combinatorial results:

Proposition 6.7. *Let $\mathcal{C}_1, \dots, \mathcal{C}_{n-m}$ be the components indexed from 1 to $n-m$ in a uniform manner of a uniform unrooted unordered forest over $\{1, 2, \dots, n\}$ with m edges. The vector of the sizes*

$$(\|\mathcal{C}_i\|_\bullet : 1 \leqslant i \leqslant n-m)$$

has the same law as the increments of a random walk $(S_i : 0 \leqslant i \leqslant n-m)$ started from $S_0 = 0$ with i.i.d. increments of law μ and conditioned on $\{S_{n-m} = n\}$. Furthermore, conditionally on their sizes $(\|\mathcal{C}_i\|_\bullet : 1 \leqslant i \leqslant n-m)$ the (increasing relabeling of the) trees $\mathcal{C}_i$ are independent (unrooted) uniform Cayley trees.

Proof. The fact that conditionally on the vertices in each component, their increasing relabeled versions are independent Cayley trees is clear already in our way to obtain (6.5). The same property holds true if we condition on the sizes of the components only. For the first point, notice that the probability that the increments of the walk are $k_1, \dots, k_{n-m}$ with $k_1 + \cdots + k_{n-m} = n$ is equal to $2^{n-m}\mathrm{e}^{-n} \prod_{i=1}^{n-m} \frac{k_i^{k_i-2}}{k_i!}$ which is proportional to (6.5) and where the proportionality factor only depends on n and m. This proves the proposition. Note for the record that we have

$$\mathbb{P}(S_{n-m} = n) = \frac{2^{n-m}(n-m)!}{\mathrm{e}^n n!} \cdot \#\mathfrak{F}(n, m). \tag{6.10}$$

□

The above proposition still holds if we consider a random walk with step distribution generating function given by $z \mapsto \mathbf{T}(z_0)^{-1} \cdot \mathbf{T}(z \cdot z_0)$ for any $0 < z_0 \leqslant \mathrm{e}^{-1}$. However, our choice of $z_0 = \mathrm{e}^{-1}$ is the "correct" probabilistic choice in the critical window $m = \frac{n}{2} + O(n^{2/3})$ and yields a measure μ with a heavy tail in the domain of attraction of the 3/2-stable law. More precisely, we shall consider the stable Lévy process $(\mathscr{S}_t)_{t \geqslant 0}$ with index 3/2 and only positive jumps, which starts from 0 and normalized so that its Lévy measure is

$$\frac{1}{\sqrt{2\pi}} |x|^{-5/2} \mathbf{1}_{x>0}, \quad \text{or equivalently} \quad \mathbb{E}[\exp(-\ell \mathscr{S}_t)] = \exp(\tfrac{2^{3/2}}{3} t \ell^{3/2})$$

for any $\ell, t \geqslant 0$, see [28, Section VIII]. We chose this normalization so that $n^{-2/3}(S_{n/2} - n) \xrightarrow[n\to\infty]{(d)} \mathscr{S}_1$. By standard results [177], for any $t > 0$ the variable $\mathscr{S}_t$ –which is distributed as a 3/2-stable totally asymmetric spectrally positive random variable– has a density with respect to the Lebesgue measure on $\mathbb{R}$ which we denote by $p_t(x)$ for $x \in \mathbb{R}$ and $t > 0$. By the scaling property of $(\mathscr{S})$ we have

$$p_t(x) = t^{-2/3} p_1(x \cdot t^{-2/3}),$$

with

$$p_1(x) = -\frac{1}{2}e^{x^3/12}\left(x\mathrm{Ai}\left(\frac{x^2}{4}\right) + 2\mathrm{Ai}'\left(\frac{x^2}{4}\right)\right),$$

where Ai is the Airy function. In particular,

$$p_1(0) = \frac{3^{1/6}\Gamma\left(\frac{2}{3}\right)}{2\pi}.$$

The function $p_1(x)$ (see Figure 6.14) is sometimes called the (map)-Airy distribution as in [23] (in the notation of [23, Definition 1] we have $p_1(-x) = c\mathcal{A}(cx)$ with $c = \frac{1}{2}$ and in the notation of [138] we have $p_1(x) = cg(cx)$ with $c = 2^{2/3}$). In particular, it is a smooth positive function tending to 0 at $\pm\infty$ and from [23, Eq. (3)] we have the following asymptotics

$$p_1(\lambda) \sim \begin{cases} \frac{1}{\sqrt{2\pi}}\sqrt{|\lambda|}\exp\left(-\frac{|\lambda|^3}{6}\right) & \text{if} \quad \lambda \to -\infty \\ \frac{1}{\sqrt{2\pi}}|\lambda|^{-5/2} & \text{if} \quad \lambda \to +\infty. \end{cases} \tag{6.11}$$

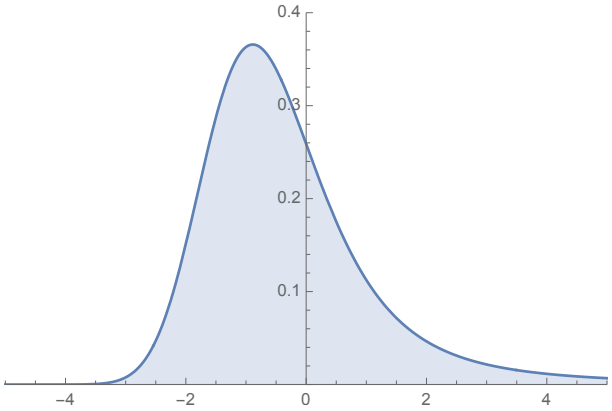


Figure 6.14: Plot of the density $p_1(\cdot)$ over $[-5,5]$. The function is rapidly decreasing to 0 as $x \to -\infty$ and polynomially decreasing to 0 as $x \to \infty$. It is smooth and unimodal: increasing from $-\infty$ to ≈ -0.886 and then decreasing up to ∞.

Using this notation and equipped with (6.8), Britikov [47] computed the asymptotic of $\#\mathfrak{F}(n,m)$ as n and m go to ∞. Those results are recalled here:

Lemma 6.3 (Britikov [47])**.** *If $n, m \to \infty$, then*

$$\#\mathfrak{F}(n,m) \sim \begin{cases} \frac{n^{2m}}{2^m m!}\sqrt{1-\frac{2m}{n}} & \text{if} \quad \frac{2m-n}{n-m}n^{1/3} \to -\infty \\ \frac{n^{n-1/6}}{2^{n-m}(n-m)!}p_1(\lambda)\sqrt{2\pi} & \text{if} \quad m = \frac{n}{2} + \frac{\lambda}{2}n^{2/3} \\ \frac{n^{n-2}}{2^{n-m-1}(n-m-1)!}\left(\frac{2m}{n}-1\right)^{-5/2} & \text{if} \quad \frac{2m-n}{n-m}n^{1/3} \to +\infty. \end{cases}$$

Those asymptotics are better understood on the variable $\mathbb{P}(S_{n-m} = n)$ which is related to the number of forests by (6.10). Indeed, writing $m = \frac{n}{2} + \frac{\lambda}{2}n^{2/3}$ and using the asymptotics on $p_1(\lambda)$

from (6.11), we see that as long[4] as $|\lambda| \ll n^{1/3}$ then

$$n^{2/3} \cdot \mathbb{P}(S_{n-m} = n) \sim p_1(\lambda), \tag{6.12}$$

which can be seen as a strong form of the local central limit for random variables in the domain of attraction of stable laws due to Gnedenko.

6.3.2 Free forest property

In this section we establish a Markovian property of the frozen Erdős–Rényi process $F(n,m)$. Recall that $D(n,m) = m - \|F(n,m)\|_{\bullet\text{-}\bullet}$ stands for the number of discarded edges up to time m (i.e. the edges that have not been added in $F(n,m)$ because their starting point was in a frozen component) and recall that $\|F(n,m)\|_\bullet$ is the number of vertices in the frozen (blue) components of $F(n,m)$.

Proposition 6.8 (Free forest property)**.** *For any $n \geqslant 1, m \geqslant 0$, conditionally on $\|F(n,m)\|_{\bullet\text{-}\bullet}$ and $\|F(n,m)\|_\bullet$ the (increasing relabeling of the) forest part $[F(n,m)]_{\text{tree}}$ is uniformly distributed over*

$$\mathfrak{F}\big(n - \|F(n,m)\|_\bullet, \|F(n,m)\|_{\bullet\text{-}\bullet} - \|F(n,m)\|_\bullet\big).$$

The proof of the proposition follows from two invariance properties of the law of a uniform random forest which are described as follows. We shall write $W(n,m)$ for a uniform forest of $\mathfrak{F}(n,m)$.

- *Size-biased removal.* Pick $X \in \{1,2,\dots,n\}$ uniformly and independently of $W(n,m)$ and denote by K the number of vertices of the tree containing the vertex X in $W(n,m)$. Then conditionally on K, the forest obtained by removing the tree containing X and relabeling the vertices in increasing order has the same law as $W(n-K, m-K+1)$.
- *Addition of one edge.* Pick $(X,Y) \in \{1,2,\dots,n\}^2$ uniformly and independently of $W(n,m)$ and let us add the edge $E = \{X,Y\}$ to the forest $W(n,m)$. If the addition of this edge creates a cycle, let us denote by K the number of vertices of this component. Otherwise put $K = 0$. Then conditionally on K, the forest obtained by adding E to $W(n,m)$, and removing the corresponding component if this addition creates a cycle has the same law as $W(n-K, m-K+1)$ (as usual up to an order-preserving relabeling of the vertices).

The proof of these two facts is easily seen by counting arguments, see [138, p 957 after Lemma 6.1] for a proof of the second one. In the first case, we call this operation a size-biased removal because the tree of size K removed from $W(n,m)$ is not uniform over all components but biased by its number of vertices.

Proof. We prove the proposition by induction on $m \geqslant 0$. For $m = 0$ there is nothing to prove. We decompose the effect of the (tentative) addition of the edge $\vec{E}_{m+1} = (X_{m+1}, Y_{m+1})$ to $F(n,m)$ in a two steps procedure. First, conditionally on $\|F(n,m)\|_\bullet$ and $\|F(n,m)\|_{\bullet\text{-}\bullet}$ we decide whether:

1. $X_{m+1}, Y_{m+1} \in [F(n,m)]_\bullet$ with probability $n^{-2}\|F(n,m)\|_\bullet^2$,

[4] that is $\lambda \equiv \lambda_n$ may depend on n but $n^{-1/3} \cdot \lambda_n \to 0$ as $n \to \infty$

2. $X_{m+1} \in [F(n,m)]_\bullet$ and $Y_{m+1} \in [F(n,m)]_\circ$ with probability $n^{-2}\|F(n,m)\|_\bullet \cdot \|F(n,m)\|_\circ$,
3. $X_{m+1} \in [F(n,m)]_\circ$ and $Y_{m+1} \in [F(n,m)]_\bullet$ with probability $n^{-2}\|F(n,m)\|_\bullet \cdot \|F(n,m)\|_\circ$,
4. or $X_{m+1}, Y_{m+1} \in [F(n,m)]_\circ$ with probability $n^{-2}\|F(n,m)\|_\circ^2$

In this first two cases the edge E_{m+1} is not added and $F(n,m+1) = F(n,m)$. Conditionally on case 3, the point X_{m+1} is uniformly distributed over $[F(n,m)]_\circ$ and the addition of the edge E_{m+1} will link the component of X_{m+1} to a frozen component, freezing it. Since by induction, $[F(n,m)]_\circ$ was a uniform forest, we conclude by invariance under size-biased removal that $[F(n,m+1)]_\circ$ is again a uniform forest of $\mathfrak{F}\big(n - \|F(n,m+1)\|_\bullet, \|F(n,m+1)\|_{\bullet\text{-}\bullet} - \|F(n,m+1)\|_\bullet\big)$. Case 4 is similar and we argue as above using the invariance property under addition of one edge.

□

Corollary 2 (Transitions of the size of the freezer and discarded edges)**.** *For every fixed $n \geqslant 1$, the process*

$$\big(\|F(n,m)\|_\bullet, \|F(n,m)\|_{\bullet\text{-}\bullet} : m \geqslant 0\big)$$

is a (inhomogeneous) Markov chain with transitions

$$\mathbb{P}\left(\begin{array}{rcl|l} \Delta\|F(n,m)\|_\bullet & = & 0 & \|F(n,m)\|_\bullet \\ \Delta\|F(n,m)\|_{\bullet\text{-}\bullet} & = & 1 & \|F(n,m)\|_{\bullet\text{-}\bullet} \end{array}\right) = \frac{\|F(n,m)\|_\circ^2}{n^2},$$

$$\mathbb{P}\left(\begin{array}{rcl|l} \Delta\|F(n,m)\|_\bullet & = & 0 & \|F(n,m)\|_\bullet \\ \Delta\|F(n,m)\|_{\bullet\text{-}\bullet} & = & 0 & \|F(n,m)\|_{\bullet\text{-}\bullet} \end{array}\right) = \frac{\|F(n,m)\|_\bullet}{n},$$

and writing $n' = \|F(n,m)\|_\circ = n - \|F(n,m)\|_\bullet$ and $m' = \|F(n,m)\|_{\bullet\text{-}\bullet} - \|F(n,m)\|_\bullet$, for $k \geqslant 1$,

$$\begin{aligned}\mathbb{P}\left(\begin{array}{rcl|l} \Delta\|F(n,m)\|_\bullet & = & k & \|F(n,m)\|_\bullet \\ \Delta\|F(n,m)\|_{\bullet\text{-}\bullet} & = & 1 & \|F(n,m)\|_{\bullet\text{-}\bullet} \end{array}\right) &= \binom{n'}{k}\frac{k^{k-2}\#\mathfrak{F}(n'-k, m'-k+1)}{\#\mathfrak{F}(n',m')}\left(\frac{k^2+k\|F(n,m)\|_\bullet}{n^2}\right) \\ &= \mu(k)(n'-m')\frac{\mathbb{P}(S_{n'-m'-1} = n'-k)}{\mathbb{P}(S_{n'-m'} = n')}\left(\frac{k^2+k\|F(n,m)\|_\bullet}{n^2}\right).\end{aligned}$$

In particular if $k = yn^{2/3}$, $\|F(n,m)\|_\bullet = xn^{2/3}$, $m = \frac{n}{2} + \frac{\lambda}{2}n^{2/3}$ for $x, y \geqslant 0$, $\lambda \in \mathbb{R}$ and $m - \|F(n,m)\|_{\bullet\text{-}\bullet} = o(n^{2/3})$, using the asymptotic on the tail of μ given after (6.9) together with (6.12) we deduce that if $y > 0$, the last probability transitions are asymptotic to

$$\frac{1}{y^{3/2}\sqrt{2\pi}} \cdot g_{x,\lambda}(y) \cdot n^{-4/3} \quad \text{where } g_{x,\lambda}(y) := (y+x)\frac{p_1(\lambda - x - y)}{p_1(\lambda - x)}, \tag{6.13}$$

and this asymptotic is uniform as long as $x, y, |\lambda| \ll n^{1/3}$ uniformly and $k \to \infty$. We will meet again the function $g_{x,\lambda}(y)$ in Section 6.7.3 when dealing with the scaling limit of the process $\|F(n,m)\|_\bullet$ in the critical window.

6.4 Enumerative consequences

In this section, we derive enumerative consequences of the coupling between the parking process on Cayley trees and the frozen Erdős–Rényi. In particular we recover much of the results of [128]. The reader may also find a discussion about enumeration of (strongly or fully) parked trees with outgoing flux at the end of the paper (Section 6.8).

We denote by $\mathrm{PF}(n,m)$ (resp. $\mathrm{PF}_{\mathrm{root}}(n,m)$) the number of configurations made of m labeled cars arriving on the vertices of a Cayley tree over $\{1,2,\dots,n\}$ so that all cars can park i.e. no outgoing flux (resp. so that the root vertex does not contain a car after the parking process). These numbers thus count the parking functions on Cayley trees [128, 148]. In particular, the number of fully parked trees of size n is $\mathrm{PF}(n,n)$ whereas the number of nearly parked trees of size n is $\mathrm{PF}_{\mathrm{root}}(n,n-1)$. Also for $m \leqslant n-1$ we have that $\mathrm{PF}_{\mathrm{root}}(n,m) \leqslant \mathrm{PF}(n,m)$ and actually Lemma 6.2 shows that

$$\mathrm{PF}_{\mathrm{root}}(n,m) = \left(1-\frac{m}{n}\right)\cdot \mathrm{PF}(n,m). \tag{6.14}$$

6.4.1 Exact counting and asymptotics for parking functions

We start by proving Proposition 6.1 stated in the Introduction: By Proposition 6.6, the probability that the root of a uniform Cayley tree of size n is not parked after m i.i.d uniform car arrivals is the probability that $F(n,m)$ (hence $G(n,m)$) contains no cycle (i.e. no frozen blue component). In that case, the graph $G(n,m)$ must be an unrooted forest. Therefore we have

$$\mathbb{P}(\text{the root of } T_n \text{ is not occupied by one of the first } m \text{ cars}) = \mathbb{P}(G(n,m) \text{ has no cycle}), \tag{6.15}$$

and so Proposition 6.1 follows by combining the last display with (6.14). We can actually go further and give a formula for the number $\mathrm{PF}_{\mathrm{root}}(n,m)$:

Proposition 6.9. *For $0 \leqslant m \leqslant n-1$ we have*

$$\mathrm{PF}_{\mathrm{root}}(n,m) = \frac{n^{n-m-1}m!n!}{(n-m)!}\sum_{i=0}^{n-m}\binom{n-m}{i}\frac{(-1)^i 2^{m-i}n^{m-i}(n+i-m)}{(m-i)!}. \tag{6.16}$$

In particular, when $m=n-1$, the number of nearly parked trees of size n (see Section 6.1.1) is equal to

$$\mathrm{PF}_{\mathrm{root}}(n,n-1) = 2^{n-1}(n-1)!n^{n-2}. \tag{6.17}$$

Proof. Equation (6.15) can be rewritten as

$$\begin{aligned}\mathrm{PF}_{\mathrm{root}}(n,m) &= n^{n-1}\cdot n^m \mathbb{P}(G(n,m) \text{ has no cycle}),\\ &= n^{n+m-1}\sum_{f\in\mathfrak{F}(n,m)}\frac{1}{n^{2m}}m!2^m\\ &= n^{n-m-1}m!2^m\cdot \#\mathfrak{F}(n,m)\end{aligned}$$

and the result follows after plugging in (6.8).

□

By combining the above proposition with (6.14) we find an exact expression of $\mathrm{PF}(n,m)$ for $m \leqslant n-1$, see [128, Theorem 4.5] for a different[5] expression. Plugging the asymptotics of Section 6.3.1 we also recover (and extend) the asymptotics of [128, Theorem 4.6] namely

$$\left(1-\frac{m}{n}\right)\cdot \mathrm{PF}(n,m) \sim \begin{cases} n^{n+m-1}\sqrt{1-\dfrac{2m}{n}} & \text{if } \ \frac{2m-n}{n-m}n^{1/3} \to -\infty \\ \dfrac{2^{2m-n}m!}{(n-m)!}n^{2n-m-1}p_1(\lambda)\sqrt{2\pi}n^{-1/6} & \text{if } \ m=\frac{n}{2}+\frac{\lambda}{2}n^{2/3} \\ \dfrac{2^{2m-n+1}n^{2n-m-3}m!}{(n-m-1)!}\left(\dfrac{2m}{n}-1\right)^{-5/2} & \text{if } \ \frac{2m-n}{n-m}n^{1/3} \to +\infty. \end{cases}$$

6.4.2 Enumeration of parked trees

In the case $m=n$, Equation (6.14) is meaningless and does not enable us to compute $\mathrm{PF}(n,n)$. To do so, we shall use a decomposition at the root of a nearly parked tree and recover [128, Theorem 3.2]. We introduce the (exponential) generating functions for nearly parked trees, fully parked trees and strongly parked trees

$$\mathbf{N}(x)=\sum_{n\geqslant 1}\frac{\mathrm{PF}_{\mathrm{root}}(n,n-1)}{n!(n-1)!}x^n, \quad \mathbf{F}(x)=\sum_{n\geqslant 1}\frac{\mathrm{PF}(n,n)}{(n!)^2}x^n, \quad \mathbf{S}(x)=\sum_{n\geqslant 1}\frac{\mathrm{SP}(n,n)}{(n!)^2}x^n, \tag{6.18}$$

where $\mathrm{SP}(n,n)$ is the number of strongly parked tree of size n. By Proposition 6.9 we have $\mathbf{N}(x)=\frac{1}{2}\mathbf{T}(2x)$.

Lemma 6.4. *The number of fully parked trees with n vertices and n cars is*

$$\mathrm{PF}(n,n)=((n-1)!)^2\sum_{j=0}^{n-1}\frac{(n-j)\cdot(2n)^j}{j!}.$$

Proof. Performing a decomposition at the root of a nearly parked tree (the trees attached to the root of a nearly parked tree are fully parked trees up to an order-preserving relabeling of the vertices and of the cars) we obtain for $n\geqslant 1$:

$$\begin{aligned} \mathrm{PF}_{\mathrm{root}}(n,n-1) &= \sum_{k\geqslant 0}\frac{1}{k!}\sum_{\substack{\sum_{i=1}^k n_i=n-1\\ n_i\geqslant 1}}\binom{n-1}{n_1,\dots,n_k}\binom{n}{1,n_1,\dots,n_k}\prod_{i=1}^k \mathrm{PF}(n_i,n_i) \\ &= \sum_{k\geqslant 0}\frac{1}{k!}\sum_{\substack{\sum_{i=1}^k n_i=n-1\\ n_i\geqslant 1}} n!(n-1)!\prod_{i=1}^k\frac{\mathrm{PF}(n_i,n_i)}{(n_i!)^2} \end{aligned}$$

Here k denotes the number of subtrees attached to the root and the factor $1/k!$ corresponds to the $k!$ reorderings of the subtrees that represent the same tree. This is equivalent to the following equation on generating functions

$$\mathbf{N}(x)=x\cdot\exp(\mathbf{F}(x)). \tag{6.19}$$

[5] Obviously the two expressions coincide numerically, but we have not been able to transform one into the other

Recalling that $\frac{1}{2}\mathbf{T}(2x) = \mathbf{N}(x)$ and using the classical relations (see for instance [143])

$$\mathbf{T}(x) = x\mathbf{T}'(x) - \frac{1}{2}(x\mathbf{T}'(x))^2 \quad \text{and} \quad \mathbf{T}'(x) = \exp(x\mathbf{T}'(x)), \tag{6.20}$$

we deduce that

$$\mathbf{F}(x) = 2x\mathbf{T}'(2x) + \ln\left(1 - x\mathbf{T}'(2x)\right).$$

This relation is the same as that obtained by Lackner & Panholzer in [128, Equation 5]. Using Lagrange inversion formula on $x\mathbf{T}'(x)$ via (6.20) (right) we obtain

$$[x^n] \ln\left(1 - x\mathbf{T}'(2x)\right) = -\frac{1}{n}\sum_{j=0}^{n-1} \frac{(2n)^j}{j!},$$

and straightforward calculations yield the result.

□

In turn the enumeration of fully parked trees can be used to count strongly parked trees by a simple substitution operation. This was already done by King & Yan in [119] but we recall it to prepare the reader to similar decompositions in Section 6.5.1.

Proposition 6.10 (King & Yan [119]). *For $n \geqslant 1$ we have $\mathrm{SP}(n,n) = (2n-2)!$.*

Proof. A fully parked tree can be decomposed into the strong component of the root vertex (which can be reduced to a single vertex) on which fully parked trees are attached. This decomposition translates into the following equation for $n \geqslant 1$,

$$\mathrm{PF}(n,n) = \sum_{n_0=1}^{n} \mathrm{SP}(n_0,n_0) \sum_{\substack{k_1,k_2,\dots,k_{n_0}\geqslant 0 \\ \sum k_i = K}} \prod_{j=1}^{n_0} \frac{1}{k_j!} \sum_{\substack{n_1,\dots,n_K\geqslant 1 \\ \sum n_i = n-n_0}} \binom{n}{n_0, n_1, \dots, n_K} \prod_{j=1}^{K} \mathrm{PF}(n_j,n_j).$$

Summing over $n \geqslant 1$, we obtain

$$\mathbf{F}(x) = \mathbf{S}(x \cdot \exp(\mathbf{F}(x))), \tag{6.21}$$

see [119, Section 3]. Solving the above equation (see [119]), we obtain $\mathbf{S}(x) = 1 - \ln(2) - \sqrt{1-4x} + \ln\left(1+\sqrt{1-4x}\right)$, whose derivative is simply the usual generating function of the Catalan numbers $\binom{2n}{n}/(n+1)$ i.e. $\mathbf{S}'(x) = (1-\sqrt{1-4x})/(2x)$, hence $\mathrm{SP}(n,n) = (2n-2)!$.

□

In Section 6.8 we show how the above results can be extended to enumerate exactly and asymptotically fully/strongly parked trees with a positive outgoing flux at the root. In particular, those problems are very similar to the enumeration of random planar maps with a boundary.

6.5 Geometry of parked trees

In this section we study the *geometry* of the components, specifically the near components, in the parking process. We prove that uniform nearly parked trees of size N have height of order $N^{3/4}$ and total flux of order $N^{5/4}$. We expect that those large scale properties are shared by the fully or strongly parked trees (and that versions of Conjecture 1 hold for them). We start by describing the decomposition of a uniform nearly parked tree of size N into strongly/fully parked components.

6.5.1 Law of large numbers for components

Recall from Section 6.1.1 (see Figure 6.7) the definition of nearly/fully/strongly parked trees as the components (different from the root component and possibly from isolated vertices) of the subforests

$$T_{\text{strong}}(n,m) \subset \mathrm{T}_{\text{full}}(n,m) \subset T_{\text{near}}(n,m).$$

We saw in the proof of Lemma 6.4 that a nearly parked tree can be decomposed at the root into a forest of fully parked trees. Going further, we can decompose each fully parked tree into a forest of strongly parked trees after removing the edges without flux, see Figure 6.15. In this decomposition, each nearly parked tree $\mathfrak{n}$ is associated with a bitype rooted tree $\mathsf{Bitype}(\mathfrak{n})$ such that the vertices at even generation are disks $\circ/\bullet$ and those at odd generations are squares $\square$:

- Each parked vertex of $\mathfrak{n}$ corresponds to a disk vertex $\bullet$ in $\mathsf{Bitype}(\mathfrak{n})$, and the empty root vertex of $\mathfrak{n}$ corresponds to the root vertex $\circ$ of $\mathsf{Bitype}(\mathfrak{n})$.
- The children of each disk vertex in $\mathsf{Bitype}(\mathfrak{n})$ are square vertices which correspond to the strongly parked components of $\mathfrak{n}$ linked to this vertex by edges with zero flux.
- The children of each square vertex in $\mathsf{Bitype}(\mathfrak{n})$ correspond to the vertices of the strongly parked component of $\mathfrak{n}$ above the corresponding square.

For convenience, the tree $\mathsf{Bitype}(\mathfrak{n})$ is given a plane orientation by fixing independently for each vertex an order on its children. Recall that by Proposition 6.9 we have $\mathbf{N}(x) = \frac{1}{2}\mathbf{T}(2x)$ and combining it with (6.19) and (6.21), we get

$$\mathbf{N}\left(\frac{1}{2\mathrm{e}}\right) = 1/4 < \infty \quad \text{and} \quad \mathbf{F}\left(\frac{1}{2\mathrm{e}}\right) = \mathbf{S}\left(\frac{1}{4}\right) = 1 - \ln(2) < \infty. \tag{6.22}$$

Therefore, we can define a random nearly parked tree P (whose size is not fixed) under the critical Boltzmann distribution, i.e. with law

$$\mathbb{P}\left(P = \mathfrak{n}\right) = \frac{4}{(\|\mathfrak{n}\|_\bullet - 1)!\|\mathfrak{n}\|_\bullet!}\left(\frac{1}{2\mathrm{e}}\right)^{\|\mathfrak{n}\|_\bullet}. \tag{6.23}$$

Lemma 6.5. *The tree* $\mathsf{Bitype}(P)$ *has the law of a bitype alternating Bienaymé–Galton–Watson (BGW) tree where (disk) vertices at even generations have Poisson offspring distribution* $\nu_\bullet$ *with mean* $\mathbf{F}((2\mathrm{e})^{-1})$ *and where (square) vertices at odd height have offspring distribution* $\nu_\square$ *given by*

$$\nu_\square(n) = \frac{[x^n]\mathbf{S}(x/4)}{\mathbf{S}(1/4)} = \frac{4^{-n}}{1-\ln(2)}\frac{(2n-2)!}{(n!)^2}, \qquad \text{for } n \geqslant 1.$$

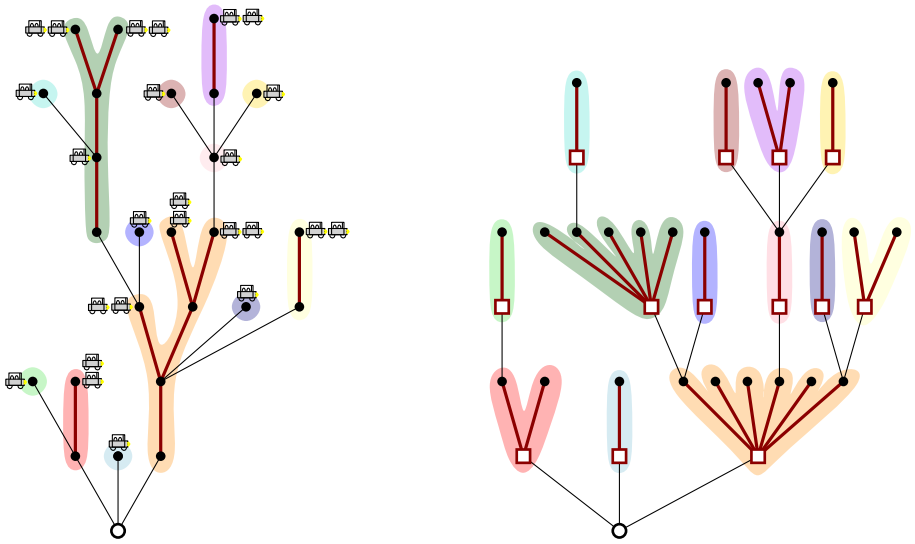


Figure 6.15: (Left) A nearly parked tree $\mathfrak{n}$ with 26 vertices. (Right) Its decomposition $\mathrm{Bitype}(\mathfrak{n})$ into the tree of strong components. Notice that the square vertices (at odd generations in $\mathrm{Bitype}(\mathfrak{n})$) correspond to the strong components and their degrees are the sizes of the components. Whereas $\mathfrak{n}$ is unordered, the tree $\mathrm{Bitype}(\mathfrak{n})$ is ordered for convenience.

Proof. Let $\mathfrak{t}$ be a fixed bitype alternating plane rooted tree starting at a disk vertex and let us denote by $(\square_i)_{1\leqslant i\leqslant n_\square}$ its square vertices, by $(k_{\square_i})_{1\leqslant i\leqslant n_\square}$ their respective number of children all of which should be positive, by $(\bullet_i)_{1\leqslant i\leqslant n_\bullet}$ its disk vertices and by $(k_{\bullet_i})_{1\leqslant i\leqslant n_\bullet}$ their respective number of children. Notice that

$$\sum_{i=1}^{n_\bullet} k_{\bullet_i} = n_\square \quad \text{and} \quad \sum_{i=1}^{n_\square} k_{\square_i} = n_\bullet - 1. \tag{6.24}$$

The probability that the BGW tree described in the lemma equals $\mathfrak{t}$ is

$$\prod_{i=1}^{n_\bullet} \mathrm{e}^{-\mathbf{F}((2\mathrm{e})^{-1})} \frac{\mathbf{F}((2\mathrm{e})^{-1})^{k_{\bullet_i}}}{k_{\bullet_i}!} \prod_{i=1}^{n_\square} \frac{(1/4)^{k_{\square_i}} \mathrm{SP}(k_{\square_i}, k_{\square_i})}{\mathbf{S}(1/4)(k_{\square_i}!)^2}.$$

By counting the number of ways to partition the vertices $\{1, 2, \ldots, n_\bullet\}$ and assign a strongly parked tree to each square vertex of $\mathfrak{t}$, recalling (6.23), we deduce that the probability that $\mathrm{Bitype}(P) = \mathfrak{t}$ is equal to

$$\frac{4 \cdot (2\mathrm{e})^{-n_\bullet}}{(n_\bullet - 1)! n_\bullet!} \times \binom{n_\bullet}{1, k_{\square_1}, \ldots, k_{\square_{n_\square}}} \binom{n_\bullet - 1}{k_{\square_1}, \ldots, k_{\square_{n_\square}}} \prod_{i=1}^{n_\square} \mathrm{SP}(k_{\square_i}, k_{\square_i})(k_{\square_i})! \times \prod_{i=1}^{n_\square} \frac{1}{k_{\square_i}!} \prod_{i=1}^{n_\bullet} \frac{1}{k_{\bullet_i}!}.$$

Using (6.24), (6.21) and (6.22) it can be easily checked that the above two probabilities are the same and we get the desired result.

□

This decomposition is used in the following lemma which states that inside a large uniform nearly parked tree of size N, there is an essentially unique fully parked tree of size $N - O_{\mathbb{P}}(1)$ containing an essentially unique strongly parked tree of size $N/2 + o_{\mathbb{P}}(N)$. The proof is based on a condensation phenomenon for conditioned subcritical Bienaymé–Galton–Watson trees and is similar to the approach of Addario-Berry to block size in random planar maps [3]. This will be used in the proof of Theorem 6.1 when dealing with full components.

Proposition 6.11. *Let P_N be a uniform nearly parked tree of size N and consider $MF(P_N)$ the fully parked tree of maximal size above its root and $MS(P_N)$ the strongly parked component of maximal size included in P_N. Then we have*

$$\frac{\|MF(P_N)\|_\bullet}{N} \xrightarrow[N\to\infty]{(\mathbb{P})} 1, \qquad \frac{\|MS(P_N)\|_\bullet}{N} \xrightarrow[N\to\infty]{(\mathbb{P})} \frac{1}{2},$$

furthermore the second largest fully parked tree is of size $O_{\mathbb{P}}(1)$ and the second largest strongly parked tree is of size $O_{\mathbb{P}}(N^{2/3})$.

During the proof we shall need a well-known "big-jump" lemma for which we provide some details for the reader's convenience. See [16, Lemma 2.5] or [76, Lemma 3.3] for similar results and [18, 20] for generalizations.

Lemma 6.6 (Single big-jump in a random sum)**.** *Let $Z_1, \dots, Z_i, \dots$ be i.i.d. random variables of law ν having a heavy tail $\nu(n) \sim cn^{-\alpha}$ for some $c > 0$ and $\alpha > 1$. We let K be a random variable independent of the Z_i's and having some exponential moment $\mathbb{E}[e^{\delta K}] < \infty$ for some $\delta > 0$. We consider the random sum*

$$\mathfrak{S} = \sum_{i=1}^{K} Z_i.$$

Then conditionally on $\{\mathfrak{S} = N\}$, if we remove the largest term Z_i for $i \in \{1, 2, \dots, K\}$ from $(Z_1, \dots, Z_K)$, then the remaining random vector converges in law towards $(Z_1, \dots, Z_{\overline{K}-1})$ where $\overline{K}$ is the size-biased variable K independent of the Z_i's. In particular $N - \max_{1\leqslant i\leqslant K} Z_i$ converges in law as $N \to \infty$.

Proof. Since ν is a regular polynomial tail and K has exponential moments, it follows from [20, Theorem 3 (i)] that

$$\lim_{N\to\infty} \frac{\mathbb{P}(\mathfrak{S} = N)}{\mathbb{P}(Z_1 = N)} = \mathbb{E}[K], \tag{6.25}$$

(in our cases of applications, this can directly be checked by a calculation using generating functions). Now, fix $k \geqslant 1$, fix values $n_1, \dots, n_{k-1}$ and denote by $\tilde{Z}_1, \dots, \tilde{Z}_{K-1}$ the re-indexed variables $\{Z_i : 1 \leqslant i \leqslant K \text{ with } i \neq \mathsf{argmax}_{1\leqslant i\leqslant K} Z_i\}$. Then for N large we have

$$\begin{aligned}
&\mathbb{P}(K = k \text{ and } \tilde{Z}_j = n_j \text{ for } 1 \leqslant j \leqslant k-1 \mid \mathfrak{S} = N) \\
= \quad &\frac{1}{\mathbb{P}(\mathfrak{S} = N)} k\mathbb{P}(K = k)\nu(n_1)\cdots\nu(n_{k-1})\nu\left(N - \sum_{i=1}^{k-1} n_i\right) \\
\xrightarrow[N\to\infty]{(6.25)} \quad &\frac{1}{\mathbb{E}[K]} k\mathbb{P}(K = k)\nu(n_1)\cdots\nu(n_{k-1}).
\end{aligned}$$

Since the above probabilities sum to 1, this implies the desired convergence in law.

□

Proof of Proposition 6.11. Let us start with the case of the fully parked tree of maximal size. By the decomposition of nearly parked trees at the root vertex (proof of Lemma 6.4), the size of the critical Boltzmann nearly parked tree P can be written as $1+\sum_{i=1}^{K} Z_i$ where K is a Poisson random variable of mean $1-\ln(2)$ independent of $Z_1, Z_2, \dots, Z_i, \dots$ which are the sizes of i.i.d. critical Boltzmann fully parked trees, i.e. with $\mathbb{P}(Z_i = n) = [x^n]\mathbf{F}(x/(2\mathrm{e}))/(1-\ln(2)) \sim \sqrt{\frac{2}{\pi}}\frac{1}{1-\ln 2} n^{-5/2}$ as $n \to \infty$. We can thus directly apply Lemma 6.6 and deduce that when we condition $1+\sum_{i=1}^{K} Z_i$ to be equal to N, then as $N \to \infty$ with high probability one of the Z_i is of order $N - O_{\mathbb{P}}(1)$. This translates into the desired result on $\|MF(P_N)\|_\bullet$.

Let us now move to the case of strongly parked tree. By Lemma 6.5 the variable $\|MS(P_N)\|_\bullet$ is equal in law to the maximal degree of a square vertex in the alternating bitype BGW tree with offspring distribution $(\nu_\bullet, \nu_\square)$ conditioned to have N disk vertices in total. We shall first consider the monotype Bienaymé–Galton–Watson tree obtained by "skipping" the odd generations i.e. with offspring distribution ξ given by $\sum_{i=1}^{K} S_i$ where K has Poisson distribution with mean $1-\ln(2)$ independent of the S_i's which are i.i.d. with distribution $\nu_\square$. This BGW tree is subcritical since

$$\mathbb{E}[K]\cdot\mathbb{E}[S_1] = x\mathbf{S}'(x)\big|_{x=1/4} = \frac{1}{2},$$

and furthermore it has a regular varying heavy tail $\mathbb{P}(\xi = n) \sim \frac{1-\ln(2)}{4\sqrt{\pi}} n^{-5/2}$ as $n \to \infty$. Here also a "big-jump" or "condensation" phenomenon appears [105, 123] and it is known, that when its number of vertices N goes to infinity, the maximal degree of such a tree is of order $N/2$, whereas the second largest is of order $N^{2/3}$ with high probability. Let us now focus on this largest degree vertex whose degree we denote by D. Remember from the construction that this degree has been obtained as $D = \sum_{i=1}^{K} S_i$. After conditioning on D, we can apply Lemma 6.6 and deduce that when D is large, the largest degree of the square vertices contributing to D is $D - O_{\mathbb{P}}(1)$ as desired.

□

After all these combinatorial decompositions, the following should come as no surprise:

Proposition 6.12. *Conditionally on their component sizes and after relabeling, the white strong (resp. full, resp. near) components in T_n after cars $X_1, \dots, X_m$ have parked, are independent uniform strongly (resp. fully, resp. nearly) parked trees.*

Proof (sketch). To fix ideas, let us consider the case of the full components. Fix $n \geqslant 1$ and $m \geqslant 0$ and let us condition on everything except the internal structure of the fully parked trees (obtained after relabeling of the vertices and cars as usual) of $T_{\mathrm{full}}(n, m)$. That is, we reveal the partition of $\{1, 2, \dots, n\}$ into the full components, the induced partition of the cars $\{1, 2, \dots, m\}$, the edges of T_n between empty vertices, as well as the possible blue tree of $T_{\mathrm{full}}(n, m)$ containing the root. It should be clear then that any fully parked of the proper size can appear in each component, so that the

probability of seeing a given configuration is proportional to $\prod_{i=1}^{k} \mathrm{PF}(n_i, n_i)$. The result follows. The case of near or strong components is similar.

□

6.5.2 Height and Flux

In this section we use our coupling construction of Section 6.2.2 specified to the case of nearly parked trees to deduce some geometric information on the latter. We use the letter N to denote the size of the nearly parked tree not to confuse with the size n of the underlying Cayley tree.

6.5.2.1 "Coupling construction" of nearly parked trees

Fix $N \geqslant 1$ and denote by P_N a uniform random nearly parked tree with N vertices (chosen among the $2^{N-1}(N-1)!N^{N-2}$ possibilities, according to Proposition 6.9). This is a random rooted Cayley tree over $\{1, 2, \ldots, N\}$ which carries $N-1$ cars arrivals $X_i \in \{1, 2, \ldots, N\}$ so that after the parking process, all cars are parked and the root of the tree is free. We can obtain such a random tree by applying the coupling construction of Section 6.2.2 with the oriented edges $(\vec{E}_i : 1 \leqslant i \leqslant N-1)$ on the event when the unoriented edges $(E_i : 1 \leqslant i \leqslant N-1)$ do not create any cycle. In such case, the graph $G(N, N-1)$ is simply a uniform (unrooted) Cayley tree T_N^u (u stands for unrooted) and the edges $(\vec{E}_i : 1 \leqslant i \leqslant N-1)$ can be obtained by labeling the edges of T_N^u by $\{1, 2, \ldots, N-1\}$ uniformly at random and given random independent orientations. We shall denote by $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_N$ the labeled oriented edges of T_N^u (they correspond to $\vec{E}_1, \ldots, \vec{E}_N$ on the appropriate event) and by $\vec{r}_1, \ldots, \vec{r}_N$ their redirections which form the nearly parked tree P_N (where the oriented edges are directed towards its root). In this special case, the coupling presented in the proof of Proposition 6.6 (or Proposition 6.4) is very simple: With the same notation, the ith car arriving on vertex X_i will *always* find a parking spot ζ_i and we redirect the edge $\vec{e}_i = (X_i, Y_i)$ into $\vec{r}_i = (\zeta_i, Y_i)$. Since we never encounter loops, we never have $\zeta_i = \dagger$ and never create any "blue" component. See Figure 6.16.

In the above construction, for $j > i$ let us describe the event on which the car number $j > i$ needs to go through the redirection $\vec{r}_i$ of the edge $\vec{e}_i$ to find its parking spot. To do this we introduce $i = \ell_1, \ldots, \ell_s = j$ the labels of the edges on the path between the edge $\vec{e}_i$ and $\vec{e}_j$ (both included) in T_N^u, see Figure 6.17. We consider the record times $1 = \tau_1 < \tau_2 < \cdots < \tau_k$ associated with the strict ascending records $i = b_1 < b_2 < \cdots < b_{k-1} < b_k$ of the process $\ell_1, \ell_2, \ldots, \ell_{s-1}, \ell_s$. That is we put $\tau_1 = 1$, set $b_1 = \ell_{\tau_1} = i$ and recursively define

$$b_{i+1} = \ell_{\tau_{i+1}} \text{ where } \tau_{i+1} = \inf\{\tau_i < t \leqslant s : \ell_t > b_i\}.$$

Lemma 6.7. *With the above notation, the jth car goes through the redirection $\vec{r}_i$ of the edge $\vec{e}_i$ in P_N if and only if $b_k = j$ (that is j is the maximal record) and all edges with labels $b_1, b_2, \ldots, b_{k-1}$ on the path from $\vec{e}_i$ to $\vec{e}_j$ in T_N^u point away from $\vec{e}_j$, and furthermore $\vec{e}_j$ points away from $\vec{e}_i$.*

Proof. Let us show the proposition by induction on N, see Figure 6.18. Fix $i < j$ and consider the path going from $\vec{e}_i$ to $\vec{e}_j$ (included) in T_N^u. We write v_i and v_j for the vertices at the extremities of this

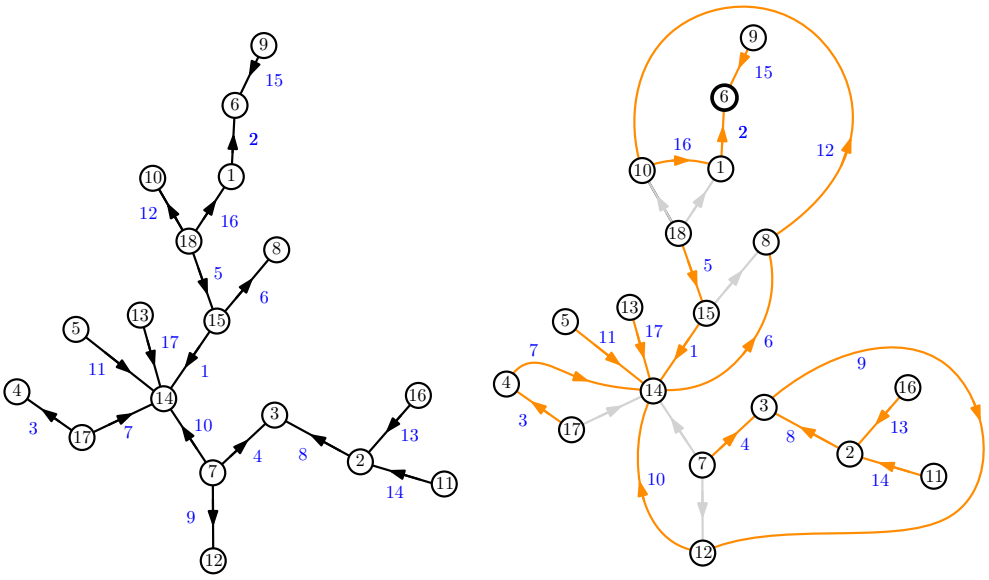


Figure 6.16: Illustration with $N = 18$ of the construction of a nearly parked tree from a uniform unrooted Cayley tree whose edges are uniformly labeled and oriented. The black edges represent the $\vec{e}_i$'s and the orange edges are their redirections $\vec{r}_i$'s. The root of P_{18} is here the vertex 6.

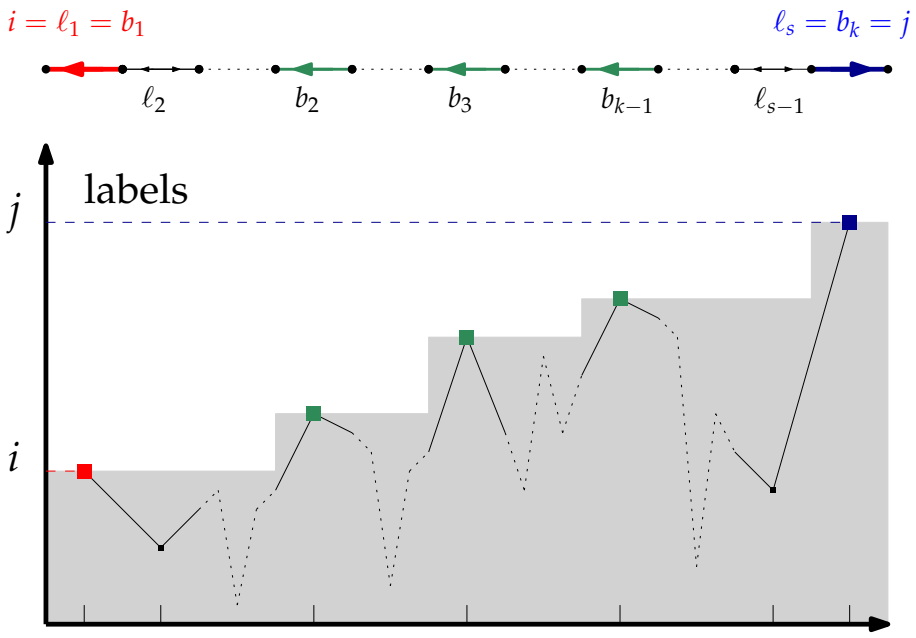


Figure 6.17: Labeling of the edges on the branch from $\vec{e}_i$ to $\vec{e}_j$ in T_N^u. The jth car goes through the redirection of $\vec{e}_i$ in P_N if and only if j is a record on the branch from $\vec{e}_i$ to $\vec{e}_j$ and the edges of records are oriented accordingly.

branch, the point v_i being closer to $\vec{e}_i$ and v_j to $\vec{e}_j$. Imagine that we build the tree P_N by re-orienting the edges $\vec{e}_1, \dots, \vec{e}_n$ one after the other and first contemplate the situation when we examine the edge $\vec{e}_{b_k}$. This edge connects two nearly parked trees (which may be reduced to single free spots) made of some of the edges $\vec{r}_1, \dots, \vec{r}_{b_k-1}$ that we already re-oriented. We denote those nearly parked trees $\mathfrak{P}_i$ and $\mathfrak{P}_j$ where $\mathfrak{P}_i$ contains v_i and $\mathfrak{P}_j$ contains v_j. Clearly if the edge $\vec{e}_{b_k}$ separates $\vec{e}_i$ from $\vec{e}_j$ then the jth car is already parked in $\mathfrak{P}_j$ and did not go through $\vec{r}_i$. So the interesting case is when $b_k = j$. In this case we must further have that $\vec{e}_{b_k}$ is oriented from $\mathfrak{P}_i$ to $\mathfrak{P}_j$ for otherwise the jth car arrives on $\mathfrak{P}_j$ and parks at its root without going through $\vec{r}_i$. Let us now go backward in time and examine the situation when we constructed $\vec{r}_1, \dots, \vec{r}_{b_{k-1}-1}$ and were about to re-orient $\vec{e}_{b_{k-1}}$. Similarly as above, at that time, the edge $\vec{e}_{b_{k-1}}$ connects two nearly parked trees $\tilde{\mathfrak{P}}_i$ containing the vertex v_i and another one $\tilde{\mathfrak{P}}'$ which may not contain v_j. A reasoning similar to the one above shows that it is necessary for the jth car to go through $\vec{r}_i$ that $\vec{e}_{b_{k-1}}$ points towards $\tilde{\mathfrak{P}}_i$. In this case, when the jth car arrives, it lands on some vertex of $\tilde{\mathfrak{P}}'$, follows the oriented edges to its root and then go through $\vec{r}_{b_{k-1}}$ to reach the target of $\vec{e}_{b_{k-1}}$ in $\tilde{\mathfrak{P}}_i$. Asking whether that car goes through $\vec{r}_i$ is equivalent to asking whether a car arrival corresponding to the edge $\overleftarrow{e}_{b_{k-1}}$ (with reversed orientation) would go through $\vec{r}_i$ inside $\tilde{\mathfrak{P}}_i$. Since the size of the system strictly decreased, we can then apply the induction hypothesis and deduce the condition presented in the lemma.

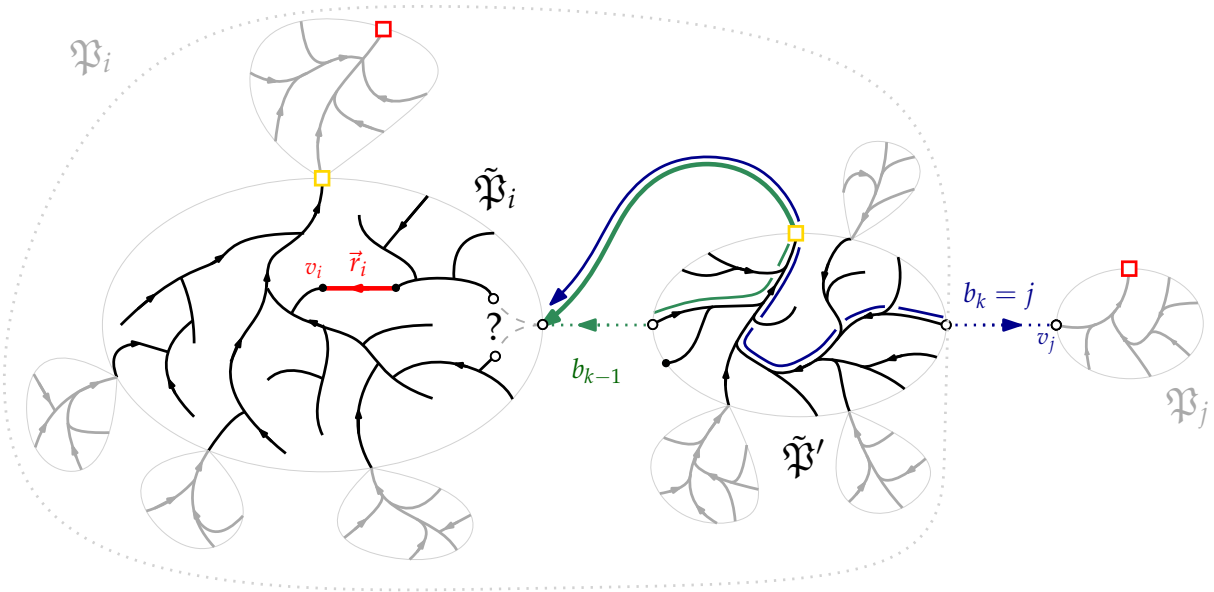


Figure 6.18: Illustration of the proof of Lemma 6.7. The edges $\vec{r}_1, \dots, \vec{r}_{b_{k-1}-1}$ are in black. The edge $\vec{r}_{b_{k-1}}$ is in green as well as the journey of the b_{k-1}th car. The edges $\vec{r}_{b_{k-1}+1}, \dots, \vec{r}_{b_k-1}$ are in gray. The root of the trees $\tilde{\mathfrak{P}}_i$ and $\tilde{\mathfrak{P}}'$ are in yellow, and those of $\mathfrak{P}_i$ and $\mathfrak{P}_j$ are in red. The car of index $b_k = j$ goes through $\vec{r}_i$ if and only if $\vec{e}_{b_k}$ and $\vec{e}_{b_{k-1}}$ respectively points away from and towards v_i and if a car landing on the target of $\vec{e}_{b_{k-1}}$ would go through $\vec{r}_i$ inside $\tilde{\mathfrak{P}}_i$. The beginning of the journey of the jth car is in blue.

□

6.5.3 Typical height

We can now prove Proposition 6.2.

Proof of Proposition 6.2. We suppose that P_N is constructed from T_N^u as in the preceding section. Independently of T_N^u, we let $V \in \{1, 2, \dots, N\}$ be a uniform point and $I \in \{1, 2, \dots, N-1\}$ be an independent (label of an) oriented edge. Then we have

$$\frac{1}{N}\mathbb{E}\left[\sum_{x \in P_N} \mathrm{d}_{\mathrm{gr}}^{P_N}(\rho, x)\right] = (N-1)\mathbb{P}\left(\vec{r}_I \text{ contributes to the height of } V \text{ in } P_N\right).$$

Notice that the number H_N of edges on the path between V and $\vec{e}_I$ (included) in T_N^u has the same law as the length of the branch of T_N^u between two uniform distinct points. By [143, Theorem 7.8 p76] we have for $1 \leqslant h \leqslant N-1$,

$$\mathbb{P}(H_N = h) = \frac{h+1}{N-1}\frac{N(N-1)\cdots(N-h)}{N^{h+1}},$$

To see whether the Ith car will contribute to the height of V in P_N we can graft an imaginary oriented edge $\vec{e}_{N+1}$ on V oriented away from $\vec{e}_I$ and apply Lemma 6.7 with $j = N+1$ to ask whether that fictive $N+1$th car would go through $\vec{e}_I$. We deduce that the necessary and sufficient condition is that all oriented edges on the path from $\vec{e}_I$ to V in T_N^u corresponding to strict ascending record for their labels are oriented away from V. Since conditionally on H_N, the order preserving relabeling of the edges on the branch is uniform, we deduce that the number of such records is equal in law to the number of cycles with disjoint support of a uniform permutation σ_{H_N} over H_N elements, see [92, Example II.16 p140]. By [92, Example III.2 p155] we have

$$\mathbb{E}\left[\left(\frac{1}{2}\right)^{\#\mathrm{Cycles}(\sigma_h)}\right] = \prod_{j=0}^{h-1}\frac{1/2+j}{j+1} = \frac{(h+1)}{(h+1)!}\left(\frac{1}{2}\right)_h. \tag{6.26}$$

Combining these lines, we obtain

$$\begin{aligned}\frac{1}{N}\mathbb{E}\left[\sum_{x \in P_N} \mathrm{d}_{\mathrm{gr}}^{P_N}(\rho, x)\right] &= (N-1)\sum_{h=1}^{N-1}\mathbb{P}(H_N = h)\mathbb{E}\left[\left(\frac{1}{2}\right)^{\#\mathrm{Cycles}(\sigma_h)}\right] \\ &= \sum_{h=1}^{N-1}\frac{N(N-1)\cdots(N-h)}{N^{h+1}}\cdot\frac{(h+1)}{h!}\left(\frac{1}{2}\right)_h\end{aligned}$$

and we get the desired result. The asymptotic of this sum is done by standard estimates: the main contribution appears for $h \approx x\sqrt{N}$ with $x \geqslant 0$ for which the terms are of order

$$\binom{2h}{h}\frac{h}{4^h}\prod_{i=1}^{h}\left(1 - \frac{i}{N}\right) \sim N^{1/4}\cdot\sqrt{\frac{x}{\pi}}\mathrm{e}^{-\frac{x^2}{2}},$$

and a series-integral comparison yields the asymptotic $N^{3/4}\int_0^\infty \mathrm{d}x\sqrt{\frac{x}{\pi}}\mathrm{e}^{-\frac{x^2}{2}} = N^{3/4}\frac{\Gamma(3/4)}{2^{1/4}\sqrt{\pi}}$.

□

6.5.3.1 Total traveled distance

Proof of Proposition 6.3. The proof is similar to that of Proposition 6.2. If I, J are two distinct uniform edge labels of $\{1, 2, \dots, N-1\}$ then we have

$$\mathbb{E}\left[\text{Total Distance Traveled in } P_N\right] = (N-1)(N-2)\mathbb{P}(J\text{th car goes through } I\text{th edge}),$$

where $(N-1)(N-2)$ is the number of distinct pairs of edges. Since choosing 2 different edges in a tree is the same as choosing two vertices at distance at least 2, by [143, Theorem 7.8] again, the length $\tilde{H}_N$ of the branch from $\vec{e}_I$ to $\vec{e}_J$ in T_N^u is distributed as

$$\mathbb{P}(\tilde{H}_N = h) = \frac{N}{N-2} \cdot \frac{h+1}{N-1} \frac{N(N-1)\cdots(N-h)}{N^{h+1}}.$$

Since conditionally on $\tilde{H}_N$ the increasing reordering of the labels on the branch is uniform, by Lemma 6.7 and using (6.26) again, conditionally on $\tilde{H}_N = h$, the probability that the Jth car passes through $\vec{r}_I$ is equal to $\frac{1}{h}$ (the probability that J is a record) time

$$\frac{1}{2}\frac{h}{h!}\left(\frac{1}{2}\right)_{h-1},$$

where the additional $1/2$ comes from requiring the good orientation for $\vec{e}_J$. Combining those lines gives the desired result. The asymptotic of the sum is done as in the preceding proof and is left to the reader.

□

II Scaling limits

This part is devoted to scaling limits in the critical regime $m = \frac{n}{2} + O(n^{2/3})$. We first use known results on the (standard augmented) multiplicative coalescent to show the convergence of the component sizes in the frozen Erdős–Rényi process (Theorem 6.2). Thanks to our coupling construction (Section 6.2.2) these translate into results on the parking process on Cayley trees (Theorem 6.1). We then take another point of view on the limiting processes, and in icular on the total mass of the frozen components, using the Markovian properties of $F(n, \cdot)$. On the way we describe the scaling limit of component sizes of a critical random forest using conditioned stable Lévy processes thus giving an alternative (and shorter) approach to the results of Martin & Yeo [138]. Recall convention (6.1).

6.6 The frozen multiplicative coalescent

In this section we establish a scaling limit for the component sizes in the frozen Erdős–Rényi $\mathrm{F}_n(\cdot)$ in the critical window. This is deduced from known results on the multiplicative coalescent but requires some care because the inclusion of the frozen process in the Erdős–Rényi process is "non-monotonous". We use cutoffs and controls which are similar to those of [157].

For $q \geqslant 1$, we let

$$\ell^q_\downarrow := \left\{ (x_1, x_2, \cdots) : x_1 \geqslant x_2 \geqslant \cdots \geqslant 0 \text{ and } \sum_{i \geqslant 1} x_i^q < \infty \right\},$$

be the space of non-increasing ℓ^q sequences. It has a natural norm inherited from the ℓ^q space and is a closed subspace of ℓ^q. In the following we denote by $\mathcal{E} = \ell^1_\downarrow \times \ell^2_\downarrow$ which is a Polish space when endowed with the distance $\mathrm{d}_\mathcal{E}$ defined by

$$\mathrm{d}_\mathcal{E}\left((\mathbf{x}, \mathbf{y}), (\mathbf{x}', \mathbf{y}')\right) = \sum_{i \geqslant 1} |x_i - x'_i| + \left(\sum_{i \geqslant 1} |y_i - y'_i|^2 \right)^{1/2}.$$

An element $(\mathbf{x}, \mathbf{y})$ of $\mathcal{E}$ will be interpreted as the masses of the particles of a system, the particles whose masses are $x_1, x_2, \ldots$ will be called the *frozen* or *blue* particles and their total mass is finite, whereas the particles whose masses are $y_1, y_2, \ldots$ will be called the standard or white particles and their total mass may be infinite. With this interpretation in mind, and in accordance with the notation for graphs, we put for $\mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \mathcal{E}$

$$[\mathbf{z}]_\bullet = \mathbf{x} \quad \text{and} \quad [\mathbf{z}]_\circ = \mathbf{y}, \quad \text{and} \quad \|\mathbf{z}\|_\bullet = \sum_{i \geqslant 1} x_i.$$

Recall from the Introduction the definition of the frozen Erdős–Rényi random graph $(F(n, m) : m \geqslant 0)$ and its continuous time counterpart $(\mathrm{F}_n(\lambda) : \lambda \in \mathbb{R})$. We shall denote by

$$\mathbb{F}_n(\lambda) \in \mathcal{E}$$

the decreasing sequence of the sizes of the frozen blue components (completed with zeros) of $\mathrm{F}_n(\lambda)$ renormalized by $n^{-2/3}$, followed by the decreasing sequence of sizes of the white components also renormalized by $n^{-2/3}$ (also completed with zeros). If $I \subset \mathbb{R}$ is an interval and Pol some Polish space, we denote by $\mathbb{C}\mathrm{adlag}(I, \mathrm{Pol})$ the set functions $f : I \to \mathrm{Pol}$ which are right-continuous with left limits at every point, endowed with the Skorokhod J_1 topology on every compact interval of I. The main theorem of this section is:

Theorem 6.2 (*Scaling limit for component sizes of the frozen Erdős–Rényi*)

We have the following convergence in distribution for the Skorokhod topology on $\mathbb{C}\mathrm{adlag}(\mathbb{R}, \mathcal{E})$

$$\left(\mathbb{F}_n(\lambda)\right)_{\lambda \in \mathbb{R}} \xrightarrow[n \to \infty]{(d)} \left(\mathscr{FM}(\lambda)\right)_{\lambda \in \mathbb{R}}. \tag{6.27}$$

The process $\mathscr{FM}$ is called the frozen multiplicative coalescent.

Remark. The above result *defines* the process $\mathscr{FM}$, although it will follow from the proof that $\mathscr{FM}$ can be built from the (augmented) multiplicative coalescent of Aldous [10] by taking an appropriate cutoff procedure.

The proof of Theorem 6.2 occupies the rest of this section. To fix ideas, we shall restrict to a fixed compact time interval and prove the convergence (6.27) for $\lambda \in [-1, 0]$ (since convergence in law of

stochastic process indexed by $\mathbb{R}$ is equivalent to convergence over all compact intervals $I \subset \mathbb{R}$). We first prove a convergence in distribution using the (weaker) supremum norm

$$\mathrm{d}_{\sup}\left((\mathbf{x},\mathbf{y}),(\mathbf{x}',\mathbf{y}')\right) = \sup_{i\geqslant 1}|x_i - x_i'| + \sup_{i\geqslant 1}|y_i - y_i'|, \tag{6.28}$$

and then lift it for the $\mathrm{d}_{\mathcal{E}}$ distance by proving the required tightness (see Proposition 6.9). Recall from the Introduction that the dynamics between standard white particles in the frozen multiplicative coalescent is the same as in the multiplicative coalescent, but the interaction between standard and frozen particles is different. Our first difficulty in this program is that the frozen part is always present, i.e. $\|\mathscr{FM}(\lambda)\|_\bullet > 0$ for all $\lambda \in \mathbb{R}$. In the next section, we shall prove however that we can neglect the effect of the frozen part that is "old enough" in the ℓ^1-sense in the time-window $\lambda \in [-1,0]$. Then, we approximate the remaining frozen process by a process on finitely many "particles" for which the convergence in distribution is obvious, see Figure 6.19. These cutoff procedures are of course reminiscent of the original construction of Aldous [10] and of the more recent work of Rossignol on dynamical percolation [157]. We first present deterministically the two cutoff procedures in the following section and then prove the necessary estimates using the relations with the multiplicative coalescent process (Proposition 6.13).

6.6.1 Getting rid of old cycles

We first start with a control that enables us to get rid of the frozen part that is "old enough". For $\lambda_0 < -1$, let us denote by

$$\mathrm{O}_n(\lambda_0) \tag{6.29}$$

the union of the components of $\mathrm{G}_n(0) = G(n, \lfloor \frac{n}{2} \rfloor)$ which have a surplus that appeared before time λ_0 i.e. before $\left\lfloor \frac{n}{2} + \frac{\lambda_0}{2} n^{2/3} \right\rfloor$ edges have been added, see Figure 6.19. We will see in Proposition 6.13 that $n^{-2/3} \cdot \|\mathrm{O}_n(\lambda_0)\|_\bullet$ is small provided that λ_0 is negative enough. We shall compare our usual frozen process $\mathrm{F}_n(\lambda)$ for $\lambda \in [\lambda_0, 0]$ to the process

$$\mathrm{F}_n^{[\lambda_0]}(\lambda) : \lambda \in [\lambda_0, 0]$$

which is started from time λ_0 without any frozen part and obtained as follows. Let us consider the graph $\mathrm{G}_n(\lambda_0)$ and remove from it the components with surplus to get its forest part $[\mathrm{G}_n(\lambda_0)]_{\text{tree}}$. We then let the remaining edges arrive as in the $G(n,m)$ process and only examine those that connect points of $[\mathrm{G}_n(\lambda_0)]_{\text{tree}}$ and apply the rule of the frozen process (Figure 6.3) to get $\mathrm{F}_n^{[\lambda_0]}(\lambda)$ for $\lambda \in [\lambda_0, 0]$, see Figure 6.19 (third line).

Of course, the process F_n and $\mathrm{F}_n^{[\lambda_0]}$ are not identical, but it is clear that their possible differences are only located on $\mathrm{O}_n(\lambda_0)$. Since the supremum distance $\mathrm{d}_{\sup}$ defined in (6.28) decreases under the non-increasing re-arrangement of both parts, it follows from the above remark that for all $\lambda \in [\lambda_0, 0]$

$$\mathrm{d}_{\sup}\left(\mathbb{F}_n(\lambda); \mathbb{F}_n^{[\lambda_0]}(\lambda)\right) \;\leqslant\; 2 \cdot n^{-2/3} \cdot \|\mathrm{O}_n(\lambda_0)\|_\bullet, \tag{6.30}$$

where $\mathbb{F}_n^{[\lambda_0]}(\lambda)$ is the pair of renormalized non-increasing sizes of frozen components followed by the renormalized sizes of the standard components of $\mathrm{F}_n^{[\lambda_0]}(\lambda)$.

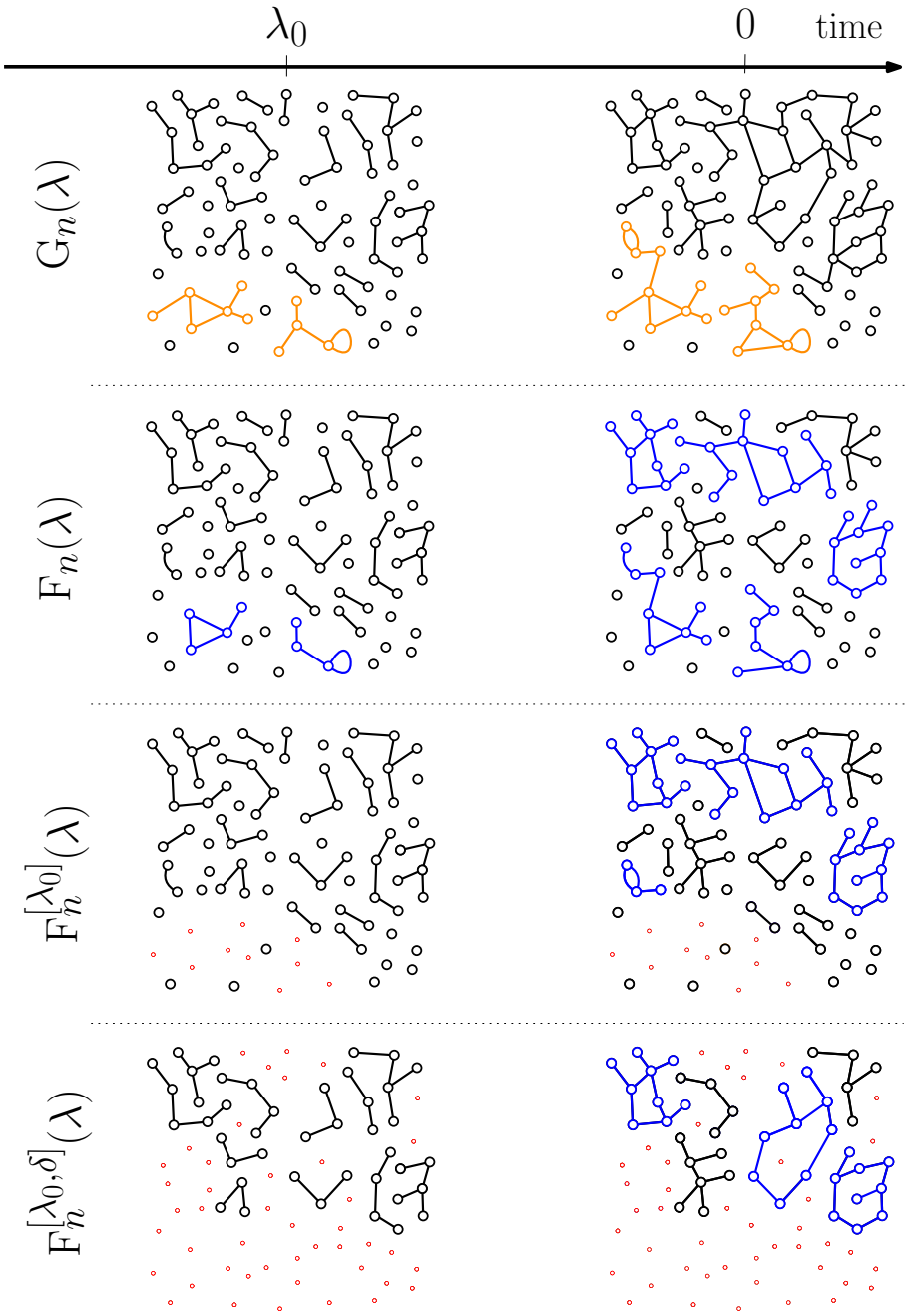


Figure 6.19: The two cutoff procedures: Comparisons of the processes F_n with $\mathrm{F}_n^{[\lambda_0]}$ and $\mathrm{F}_n^{[\lambda_0,\delta]}$. The orange part on the top right corner is $\mathrm{O}_n(\lambda_0)$. Third line: The process $\mathrm{F}_n^{[\lambda_0]}$ is obtained by starting from $G(n, \left\lfloor \frac{n}{2} + \frac{\lambda_0}{2} n^{2/3} \right\rfloor)$, removing the components having a surplus, and then applying the rules of the construction of the frozen Erdős–Rényi with the remaining edges. Fourth line: The process $\mathrm{F}_n^{[\lambda_0,\delta]}$ is further obtained by restricting to components of size at least $\delta n^{2/3}$ at time λ_0.

6.6.2 Approximation by the η-skeleton

Following the above notation rule, we write $\mathrm{G}_n^{[\lambda_0]}(\cdot)$ the Erdős–Rényi process started from $[\mathrm{G}_n(\lambda_0)]_{\mathrm{tree}}$ at time λ_0 and keeping only the remaining edges that belong to $[\mathrm{G}_n(\lambda_0)]_{\mathrm{tree}}$. Our goal now is to approximate the process $(\mathrm{F}_n^{[\lambda_0]}(\lambda) : \lambda \in [\lambda_0, 0])$ by a process with a number of particles that stays bounded as $n \to \infty$. More precisely, in the following, we shall call the components of $[\mathrm{G}_n(\lambda_0)]_{\mathrm{tree}} = \mathrm{F}_n^{[\lambda_0]}(\lambda_0)$ the "**specks**" and say that their **masses** are given by their number of vertices renormalized by $n^{-2/3}$. Those specks will be seen as "macroscopic vertices" on which acts our dynamics. Indeed, notice that in the processes $\mathrm{G}_n^{[\lambda_0]}$ or $\mathrm{F}_n^{[\lambda_0]}$, for $\lambda \in [\lambda_0, 0]$, all the vertices belonging to the same speck share the same color. This enables us to define the "speck version" of this processes denoted respectively by $(\mathsf{SG}_n^{[\lambda_0]}(\lambda) : \lambda_0 \leqslant \lambda \leqslant 0)$ and $(\mathsf{SF}_n^{[\lambda_0]}(\lambda) : \lambda_0 \leqslant \lambda \leqslant 0)$ obtained by contracting all vertices and edges belonging to the same speck: the vertex set of those processes is thus made of the specks of $[\mathrm{G}_n(\lambda_0)]_{\mathrm{tree}}$. Alternatively, those processes can be constructed sequentially: initially all specks are disjoint and we interpret the incoming edges between vertices as incoming edges between the corresponding specks and following the usual rules to get SG and SF.

For every $\delta > 0$, we denote by $(\mathrm{F}_n^{[\lambda_0,\delta]}(\lambda) : \lambda \in [\lambda_0, 0])$ the frozen process started from $\mathrm{F}_n^{[\lambda_0]}(\lambda_0)$ and obtained by only examining those edges between or inside specks of mass at least δ, see Figure 6.19 fourth line. Similarly as above we denote by $(\mathsf{SF}_n^{[\lambda_0,\delta]}(\lambda) : \lambda \in [\lambda_0, 0])$ the associate process on specks after identifying all vertices belonging to the same speck. The fact that we discarded some edges may affect the colors and the connections of the vertices (resp. specks) due to the non-monotonicity of the frozen dynamics. However, we shall see that if δ is small enough the dynamics are coherent (Lemma 6.8) on a large part of the graph. We now establish deterministic inclusions between all these processes.

The η-skeleton. Consider the graph $\mathsf{SG}_n^{[\lambda_0]}(0)$ on the specks. This graph has a certain number of non-trivial cycles (including self-loops and multiple edges) involving certain specks. We fix $\eta > 0$ and define the

$$\eta\text{-skeleton}$$

as the set of all specks in $\mathsf{SG}_n^{[\lambda_0]}(0)$ that belong to a non-backtracking path whose extremities are either a speck of a cycle of $\mathsf{SG}_n^{[\lambda_0]}(0)$ or a speck of mass at least $\eta > 0$, see Figure 6.20. In particular, all specks on (non-backtracking) paths in $\mathsf{SG}_n^{[\lambda_0]}(0)$ between specks of the η-skeleton actually belong to the η-skeleton[6]. We then denote by

$$\gamma = \gamma_n(\lambda_0, \eta) \tag{6.31}$$

the minimal mass of a speck on the η-skeleton.

The key is to show that as soon as $\delta \leqslant \gamma$ the induced frozen Erdős–Rényi process $\mathsf{SF}_n^{[\lambda_0,\delta]}$ is constant on the η-skeleton. More precisely:

[6] in other words, two vertices of the η-skeleton which are connected in $\mathrm{G}_n^{[\lambda_0]}(0)$ are connected within the η-skeleton and all non-trivial cycles of $\mathrm{G}_n^{[\lambda_0]}(0)$ are inside the η-skeleton

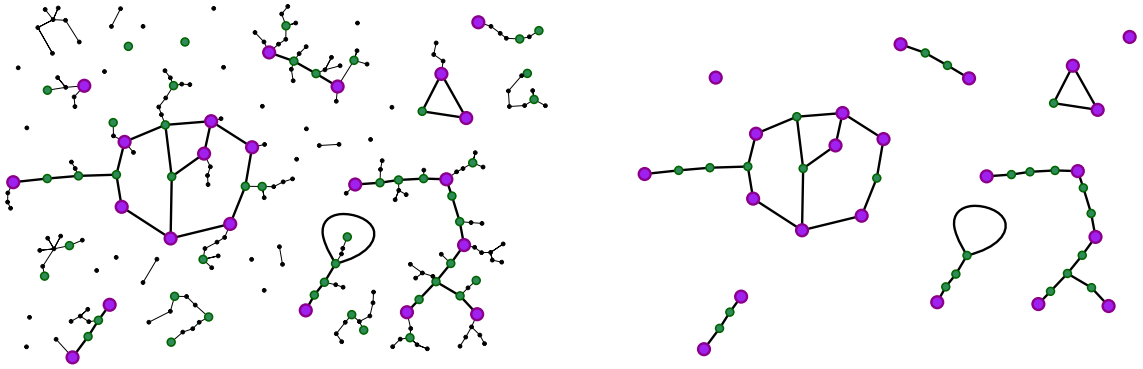


Figure 6.20: On the left the graph $\mathrm{SG}_n^{[\lambda_0]}(0)$ and its η-skeleton on the right. The points (black, green, purple) represent the specks, i.e. the components of $[\mathrm{G}_n(\lambda_0)]_{\mathrm{tree}}$. Specks of mass larger than η (i.e. with more than $\geqslant \eta n^{2/3}$ vertices) are displayed in purple, those of mass in-between γ and η are in green, and those of mass smaller than γ are in black.

Lemma 6.8. *With the above notation for any $\lambda \in [\lambda_0, 0]$ and any $\delta \in [0, \gamma)$, the edges between vertices belonging to specks of the η-skeleton and the color of these vertices are the same in $\mathrm{F}_n^{[\lambda_0,\delta]}(\lambda)$ and in $\mathrm{F}_n^{[\lambda_0]}(\lambda)$.*

Proof. We prove the lemma by induction, adding the edges one by one. We focus in the proof on the speck graphs, and the statement concerning the underlying graphs on the vertices $\{1, 2, ..., n\}$ will follow. Fix $\delta < \gamma$ and consider the status of the edges and the color of the specks of the η-skeleton in $\mathrm{SF}_n^{[\lambda_0,\delta]}(\lambda)$ and $\mathrm{SF}_n^{[\lambda_0]}(\lambda)$. Clearly, at time $\lambda = \lambda_0$ they match up: all specks are isolated and white. By induction, suppose that at some time $\lambda \in [\lambda_0, 0]$ we examine the status of an edge E_i between two specks and that at time λ^- the induced graph on the η-skeleton is the same in $\mathrm{SF}_n^{[\lambda_0,\delta]}(\lambda^-)$ and in $\mathrm{SF}_n^{[\lambda_0]}(\lambda^-)$:

- Suppose first that E_i is an edge of the η-skeleton. In particular, both endpoints are located on a speck of mass at least γ and this edge gets examined both in $\mathrm{SF}_n^{[\lambda_0,\delta]}(\lambda^-)$ and in $\mathrm{SF}_n^{[\lambda_0]}(\lambda^-)$ (since $\delta \leqslant \gamma$). Since the colors and the connections of the vertices in the η-skeleton are the same in both processes at time λ^-, applying the rules of the construction of the frozen process yields the same transformations for the specks, colors and edges of the η-skeleton in both cases.

- If E_i is not an edge of the η-skeleton, one may think that we do not care whether we add it or not in the frozen processes: Indeed, we saw above that the non-backtracking paths between specks of the η-skeleton stay within the skeleton so that adding that edge does not change the connections between specks of the η-skeleton. However its addition might change the color of some specks of the η-skeleton. But this can only happen if this edge creates a cycle or relates a white component carrying a speck of the η-skeleton to a frozen blue component. Since the frozen blue components necessarily contain a cycle one can check that E_i must belong to the η-skeleton and so we are back to the previous item.

□

Let us use the above lemma and more generally the relations between $\mathrm{G}_n(\cdot), \mathrm{G}_n^{[\lambda_0]}(\cdot)$ and $\mathrm{F}_n(\cdot), \mathrm{F}_n^{[\lambda_0]}(\cdot), \mathrm{F}_n^{[\lambda_0,\delta]}(\cdot)$ to derive bounds on the $\mathrm{d}_{\mathcal{E}}$ distance. For a given vertex $x \in \{1, 2, \dots, n\}$, the cluster of x in $\mathrm{F}_n^{[\lambda_0,\delta]}(\lambda)$ and in $\mathrm{F}_n^{[\lambda_0]}(\lambda)$ may differ[7], but when $\delta < \gamma$, by Lemma 6.8 they contain the same specks of the η-skeleton, so their difference is in particular supported by vertices of specks of mass $\leqslant \eta$ and belonging to the component of x in $\mathrm{G}_n^{[\lambda_0]}(0)$. If we denote by $\Delta_n(\lambda_0, \eta)$ the maximal difference (number of vertices) between a component in $\mathrm{G}_n^{[\lambda_0]}(0)$ and its subcomponent made of (vertices of) specks of mass $\geqslant \eta$ then we have for $\delta \leqslant \gamma$ and $\lambda \in [\lambda_0, 0]$ with an obvious notation

$$\mathrm{d}_{\mathrm{sup}}\left(\mathbb{F}_n^{[\lambda_0,\delta]}(\lambda); \mathbb{F}_n^{[\lambda_0]}(\lambda)\right) \leqslant n^{-2/3} \cdot \Delta_n(\lambda_0, \eta). \tag{6.32}$$

6.6.3 Estimates via the augmented multiplicative coalescent

Recall from the previous section the definition of the variables $O_n(\lambda_0), \gamma_n(\lambda_0, \eta)$ and $\Delta_n(\lambda_0, \eta)$. We provide the necessary (asymptotic) controls on those variables to apply the above cutoffs. These are derived using known estimates on the (augmented) multiplicative coalescent [10, 36, 157].

Proposition 6.13. *For any $\varepsilon > 0$ one can find*

$$\lambda_0 < -1, \quad \eta \in (0, 1), \quad \delta \in (0, \eta),$$

and $n_0 \geqslant 1$ so that for all $n \geqslant n_0$ with probability at least $1 - \varepsilon$ we have

$$n^{-2/3} \cdot \|\mathrm{O}_n(\lambda_0)\|_\bullet \leqslant \varepsilon, \tag{6.33}$$

$$n^{-2/3} \cdot \Delta_n(\lambda_0, \eta) \leqslant \varepsilon, \tag{6.34}$$

$$\gamma_n(\lambda_0, \eta) \geqslant \delta. \tag{6.35}$$

Proof. The proof will follow from the convergence of the component sizes and surplus of $G(n, m)$ towards the augmented multiplicative coalescent and some of its basic properties. To help the reader, let us sketch in which order the variables will be chosen

$$\varepsilon \longrightarrow \begin{array}{c} \lambda_0 < 0 \\ \text{so that w.h.p.} \\ n^{-2/3} \cdot \|\mathrm{O}_n(\lambda_0)\|_\bullet \leqslant \varepsilon \end{array} \longrightarrow \begin{array}{c} \eta > 0 \\ \text{so that w.h.p.} \\ n^{-2/3} \cdot \Delta_n(\lambda_0, \eta) \leqslant \varepsilon \end{array} \longrightarrow \begin{array}{c} \delta > 0 \\ \text{so that w.h.p.} \\ \delta \leqslant \gamma_n(\lambda_0, \eta), \end{array}$$

where w.h.p. indicates with high probability.

We first recall the construction of the standard augmented multiplicative coalescent following Broutin & Marckert [50]. Let $(B_t : t \geqslant 0)$ be a linear Brownian motion and Ξ be an independent Poisson point process on $\mathbb{R}_+ \times \mathbb{R}_+$ with unit intensity. For $\lambda \in \mathbb{R}$, we consider the process $B^{(\lambda)}$ obtained by reflecting $t \mapsto B_t + \lambda t - \frac{t^2}{2}$ above its running infimum (i.e. by subtracting the running infimum process). The processes $B^{(\lambda)}$ are coupled with respect to λ since they involve the same Brownian motion B. Each excursion of $B^{(\lambda)}$ is then seen as a particle of mass given by its length

[7] we do not have deterministic inclusion of one inside the other

and the surplus of this particle is the number of atoms of Ξ that fall under this excursion (i.e. such that $B_\tau^{(\lambda)} \geqslant y$ if the atom lies at (τ, y)). After ranking the particles in decreasing mass and recording their surplus, we get an element $((\mathscr{M}_i(\lambda))_{i\geqslant 1}, (\mathscr{S}_i(\lambda))_{i\geqslant 1})$ of

$$\mathbb{U}^{\downarrow} = \left\{ (x,s) \in \ell^2_{\downarrow} \times \mathbb{Z}^{\infty}_{\geqslant 0} : \sum_{i\geqslant 1} x_i s_i < \infty \text{ and } s_i = 0 \text{ whenever } x_i = 0 \right\},$$

in particular the total mass of the particles with positive surplus is almost surely finite [36]. The process $\lambda \mapsto (\mathscr{M}(\lambda), \mathscr{S}(\lambda))$ is the augmented multiplicative coalescent introduced in [36] and appears as the scaling limit[8] of the renormalized component sizes and surplus in $G_n(\lambda)$, see [36, 50].

This convergence holds for the Skorokhod topology on $\mathbf{Cadlag}(\mathbb{R}, \mathbb{U}^{\downarrow})$ where $\mathbb{U}^{\downarrow}$ is endowed with the metric

$$\mathrm{d}_{\mathbb{U}^{\downarrow}}\left((x,s),(x',s')\right) = \left(\sum_{i\geqslant 1} |x_i - x'_i|^2\right)^{1/2} + \sum_{i\geqslant 1} |x_i s_i - x'_i s'_i|.$$

Since $(x,s) \in \mathbb{U}^{\downarrow} \mapsto \sum x_i \mathbf{1}_{s_i>0}$ is continuous for this topology, the previous convergence implies the convergence of the total renormalized size $n^{-2/3} \cdot \|\mathrm{O}_n(0)\|_\bullet$ of all components at time $\lambda = 0$ carrying a surplus at time 0, towards its continuous counterpart

$$n^{-2/3} \cdot \|\mathrm{O}_n(0)\|_\bullet \xrightarrow[n\to\infty]{(d)} \sum_{i\geqslant 1} \mathscr{M}_i(0) \mathbf{1}_{\mathscr{S}_i(0)>0}. \tag{6.36}$$

Furthermore, we have seen above that every atom (τ, y) of Ξ corresponds, in the augmented coalescent to a surplus of one in a particle and we can define the time of appearance of this surplus as the smallest $\lambda \in \mathbb{R}$ so that $B_\tau^{(\lambda)} \leqslant y$, notice that $\lambda > -\infty$ almost surely for each atom. It follows from these observations, and the proof of Theorem 4 in [50, Section 7.2] that we have the convergence in law for each $\lambda_0 < 0$ fixed

$$n^{-2/3} \cdot \|\mathrm{O}_n(\lambda_0)\|_\bullet \xrightarrow[n\to\infty]{(d)} |\mathscr{C}(\lambda_0)|, \tag{6.37}$$

where $|\mathscr{C}(\lambda_0)|$ is the total mass of all particles of $\mathscr{M}(0)$ which have a surplus appeared before time λ_0. Also, $|\mathscr{C}(\lambda_0)| \to 0$ as $\lambda_0 \to -\infty$ almost surely by dominated convergence. Together with the last display, this proves the first point of the proposition and gives the existence of λ_0 and n_0. Estimates of that flavor were first obtained by enumerative techniques, see [94, Theorem 2.20].

For the second and third item, notice that once λ_0 has been fixed, the convergence to the augmented multiplicative coalescent [37] implies that the masses of the specks (i.e. the renormalized sizes of the components of $\mathrm{G}_n^{[\lambda_0]}(\lambda_0) = [\mathrm{G}_n(\lambda_0)]_{\mathsf{tree}}$) converge in distribution in the ℓ^2 sense to

$$\left(\mathscr{M}_i(\lambda_0)\mathbf{1}_{\mathscr{S}_i(\lambda_0)=0} : i \geqslant 1\right)^{\downarrow}. \tag{6.38}$$

After an inoffensive Poissonization of the time (i.e. by letting the edges arrive according to a Poisson process instead of discrete time steps) the second point is a consequence of the Feller property of the

[8] Actually, we deal with a slightly different version of the $G(n,p)$ model since we have a fixed number m of edges and we allow self-loops and multiple edges, but this model is considered in [36, 132] and the result applies.

multiplicative coalescent [10, Proposition 5] together with the last display: in words the renormalized component sizes of $\mathrm{G}_n^{[\lambda_0]}(0)$ are well approximated by restricting to specks of mass $\geqslant \eta$ in the ℓ^2 sense (uniformly in n) hence in the ℓ^∞ sense so that $\sup_{n\geqslant 1} n^{-2/3}\Delta_n(\lambda_0,\eta) \to 0$ in probability as $\eta \to 0$. The third convergence is similar and follows from the Feller property of the augmented coalescent [37, Theorem 3.1]. See also [157, Corollary 5.6].

□

6.6.4 Proof of Theorem 6.2

We now gather the deterministic controls established in Sections 6.6.1 and 6.6.2 together with the probabilistic estimates of Proposition 6.13 to prove Theorem 6.2. As announced, we start with a weaker convergence for the supremum norm.

Convergence for the supremum norm. We consider $\mathcal{E}_0 = \ell^\infty_{\downarrow,0} \times \ell^\infty_{\downarrow,0}$ where $\ell^\infty_{\downarrow,0}$ is the space of non-increasing sequences tending to 0 endowed with $\mathrm{d}_{\sup}$ (see (6.28)) which is a Polish space. Clearly $\mathcal{E} \subset \mathcal{E}_0$ and the convergence for the $\mathrm{d}_{\mathcal{E}}$ distance is stronger than for $\mathrm{d}_{\sup}$.

Fix $\varepsilon > 0$ and find $\lambda_0, \eta, \delta > 0$ and $n_0 \geqslant 1$ as in Proposition 6.13. On the event described in this proposition of probability at least $1-\varepsilon$ for n large enough we have for every $\lambda \in [\lambda_0, 0]$

$$\begin{aligned}
&\mathrm{d}_{\sup}\left(\mathbb{F}_n(\lambda); \mathbb{F}_n^{[\lambda_0,\delta]}(\lambda)\right)\\
\underset{\text{trig. ineg}}{\leqslant}\ &\mathrm{d}_{\sup}\left(\mathbb{F}_n(\lambda); \mathbb{F}_n^{[\lambda_0]}(\lambda)\right) + \mathrm{d}_{\sup}\left(\mathbb{F}_n^{[\lambda_0]}(\lambda); \mathbb{F}_n^{[\lambda_0,\delta]}(\lambda)\right)\\
\underset{(6.30),(6.32)}{\leqslant}\ &n^{-2/3}\left(\|\mathrm{O}_n(\lambda_0)\|_\bullet + \Delta_n(\lambda_0,\eta)\right)\\
\underset{\text{Prop. } 6.13}{\leqslant}\ &2\varepsilon.
\end{aligned} \tag{6.39}$$

On the other hand, by (6.38), the starting configuration of $\mathbb{F}_n^{[\lambda_0,\delta]}$ converges in law towards some vector having a finite number of non-zero components. Since there are only finitely many particles to take care of, applying the dynamics of the frozen coalescent it should be clear that for fixed $\lambda_0 < 0$ and $\delta > 0$

$$\left(\mathbb{F}_n^{[\lambda_0,\delta]}(\lambda) : \lambda \in [\lambda_0, 0]\right)$$

converges in distribution for the Skorokhod topology on $\mathbb{C}\mathbf{adlag}([\lambda_0,0], \mathcal{E}_0)$. If d_{LP} denotes the Lévy–Prokhorov distance associated to the convergence in law for the Skorokhod topology on $\mathbb{C}\mathbf{adlag}([\lambda_0,0], \mathcal{E}_0)$ then restricting (6.39) to $\lambda \in [-1, 0]$ we deduce

$$\mathrm{d}_{LP}\left(\left(\mathbb{F}_n(\lambda)\right)_{\lambda\in[-1,0]}; \left(\mathbb{F}_n^{[\lambda_0,\delta]}(\lambda)\right)_{\lambda\in[-1,0]}\right) \leqslant 2\varepsilon,$$

for all $n \geqslant n_0$ (this actually holds for the supremum norm which is stronger than the Skorokhod distance). Since $(\mathbb{F}_n^{[\lambda_0,\delta]}(\lambda) : \lambda \in [-1,0])$ is converging in law as $n \to \infty$, we can combine this with the last display to deduce that $(\mathbb{F}_n(\lambda))_{\lambda\in[-1,0]}$ is Cauchy for d_{LP} and so converges as desired. Its limit is obtained by first letting $n \to \infty$, then $\delta \to 0$ and finally $\lambda_0 \to -\infty$ in the process $n^{-2/3} \cdot \mathrm{F}_n^{[\lambda_0,\delta]}$.

□

Convergence in $\ell^1_\downarrow \times \ell^2_\downarrow$. To upgrade the previous convergence for the $\mathrm{d}_{\mathrm{sup}}$ distance to a convergence for the distance $\mathrm{d}_{\mathcal{E}}$, we need to prove tightness i.e. to control uniformly over $\lambda \in [-1, 0]$ the cumulative effect of the (rescaled) small component sizes in our frozen coalescent processes. More precisely, for any $\xi \geqslant 0$ and $\mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \mathcal{E}$ we denote by

$$R_{\bullet,\xi}(\mathbf{z}) = \sum_{i\geqslant 1} x_i \mathbf{1}_{x_i \leqslant \xi} \quad \text{and} \quad R_{\circ,\xi}(\mathbf{z}) = \sum_{i\geqslant 1} y_i^2 \mathbf{1}_{y_i \leqslant \xi},$$

respectively the sum of the masses of the blue particles and sum of the squares of the masses of the white particles of mass smaller than ξ. We also put $R_\xi(\mathbf{z}) = R_{\xi,\bullet}(\mathbf{z}) + R_{\xi,\circ}(\mathbf{z})$. We then have :

Lemma 6.9 (Towards tightness of $\mathbb{F}_n(\cdot)$)**.** *For any $\varepsilon > 0$ there exists $\xi > 0$ and $n_0 \geqslant 1$ such that for all $n \geqslant n_0$ we have with probability at least $1 - \varepsilon$*

$$\sup_{\lambda\in[-1,0]} \left(R_\xi(\mathbb{F}_n(\lambda))\right) \leqslant \varepsilon.$$

Proof. Let us begin with the ℓ^2-part. By the inclusion of the frozen exploration process in the standard Erdős–Rényi process, all the components of $\mathrm{F}_n(\lambda)$ for $\lambda \in (-\infty, 0]$ -frozen or not- are contained in $\mathrm{G}_n(0)$. Next, if $0 \leqslant f_1, \ldots, f_k \leqslant \xi$ and $f_1 + \cdots + f_k \leqslant y$ then we have

$$f_1^2 + \cdots + f_k^2 \leqslant (f_1 + \cdots + f_k) \cdot (\xi \wedge y) \leqslant (\xi \cdot y) \wedge y^2. \tag{6.40}$$

We apply this inequality when $f_1, \ldots, f_k$ are the renormalized sizes of the components in $\mathrm{F}_n(\lambda)$ which are included in the same component of $\mathrm{G}_n(0)$ with renormalized size y and this for each component of $\mathrm{G}_n(0)$ which is made of small components of $\mathrm{F}_n(\lambda)$: if we denote by $\boldsymbol{\Upsilon}^{(n)} = (\Upsilon_i^{(n)} : i \geqslant 1)$ the decreasing sizes of the components of $\mathrm{G}_n(0)$ renormalized by $n^{-2/3}$ then we have

$$\sup_{\lambda<0} R_{\circ,\xi}(\mathbb{F}_n(\lambda)) \leqslant \sum_{i\geqslant 1} \left(\xi \cdot \Upsilon_i^{(n)}\right) \wedge \left(\Upsilon_i^{(n)}\right)^2.$$

By the result of Aldous [10], the sequence $(\Upsilon_i^{(n)} : i \geqslant 1)$ converges in distribution for the $\ell^2_\downarrow$ distance towards the multiplicative coalescent $\mathscr{M}(0)$ at time 0. Since $\psi_\xi : \ell^2_\downarrow \to \mathbb{R}_+$ defined by $\psi_\xi((y_i : i \geqslant 1)) = \sum_{i\geqslant 1} y_i \cdot (y_i \wedge \xi)$ is continuous for the $\ell^2_\downarrow$-distance we deduce from the previous convergence that $\psi_\xi(\boldsymbol{\Upsilon}^{(n)})$ converges in law towards $\psi_\xi(\mathscr{M}(0))$. Furthermore, by dominated convergence we have $\psi_\xi(\mathscr{M}(0)) \to 0$ a.s. as $\xi \to 0$. We deduce that

$$\forall \varepsilon > 0, \quad \sup_{n\geqslant 1} \mathbb{P}\left(\sum_{i\geqslant 1} \left(\xi \cdot \Upsilon_i^{(n)}\right) \wedge \left(\Upsilon_i^{(n)}\right)^2 \geqslant \varepsilon\right) \xrightarrow[\xi\to 0]{} 0,$$

and this takes care of the $R_{\circ,\cdot}$ part of the lemma.

The ℓ^1-part is a bit trickier. Recall from Section 6.6.1 that for any $\lambda \in (-\infty, 0]$, the frozen components of $\mathrm{F}_n(\lambda)$ are included in $\mathrm{O}_n(0)$, the union of the components of $\mathrm{G}_n(0)$ that have a surplus. Notice that if $k \geqslant 1$ frozen components of $\mathrm{F}_n(\lambda)$ belong to the same component of $\mathrm{G}_n(0)$, then this component must have surplus at least k (recall that each frozen component contains exactly

one cycle). Hence if $\boldsymbol{X}^{(n)} = (X_i^{(n)} : i \geqslant 1)$ are the decreasing sizes of the components of $\mathrm{O}_n(0)$ renormalized by $n^{-2/3}$ and if K_n is the maximum surplus of a component in $\mathrm{G}_n(0)$ then we have for all $\lambda \in (-\infty, 0]$

$$R_{\bullet,\xi}(\mathbb{F}_n(\lambda)) \leqslant K_n \cdot \sum_{i\geqslant 1} \left(X_i^{(n)} \wedge \xi\right).$$

By [136, Theorem 1], the sequence $(K_n : n \geqslant 1)$ is tight. From the discussion just before (6.36) we get that $\boldsymbol{X}^{(n)}$ converges in law for the $\ell^1_{\downarrow}$-topology towards the masses $\left(\mathscr{M}_i(0)\mathbf{1}_{\mathscr{S}_i(0)>0}\right)^{\downarrow}$ of the particles in the augmented multiplicative coalescent at time 0 that carry a surplus. By the same argument as above we deduce that for every $\varepsilon > 0$, there exists $\xi > 0$ such that $\mathbb{P}\left(K_n \cdot \sum_{i\geqslant 1}\left(\xi \cdot X_i^{(n)}\right) \geqslant \varepsilon\right) \leqslant \varepsilon$ for all $n \geqslant 1$ and this finishes the proof of the lemma.

□

We can now finish the proof of Theorem 6.2: Recall that $\mathbb{F}_n(\lambda)$ is the renormalized sizes of the frozen and standard components in $\mathrm{F}_n(\lambda)$. In a nutshell, the tightness of $(\mathbb{F}_n(\lambda) : \lambda \in [-1, 0])$ for the Skorokhod topology with values in $\mathcal{E}_0$ together with the last lemma establishes the tightness of $(\mathbb{F}_n(\lambda) : \lambda \in [-1, 0])$ for the Skorokhod topology with values in $\mathcal{E}$. Since the convergence in $\mathcal{E}_0$ determines the law, we are done. Let us provide some details. Recall that we already proved that

$$(\mathbb{F}_n(\lambda) : \lambda \in [-1, 0]) \xrightarrow[n\to\infty]{(d)} (\mathscr{FM}(\lambda) : \lambda \in [-1, 0]), \tag{6.41}$$

for the Skorokhod topology on the space $\mathbb{C}\mathrm{adlag}(\mathbb{R}, \mathcal{E}_0)$ of càdlàg functions with values in $\mathcal{E}_0$ endowed with $\mathrm{d}_{\mathrm{sup}}$. By Skorokhod representation theorem, we can then assume that for each $n \geqslant 1$, the processes $\mathbb{F}_n$ and $\mathscr{FM}$ are coupled in such a way that the Skorokhod distance between $\mathbb{F}_n(\cdot)$ and $\mathscr{FM}$ converges almost surely to 0 for $\mathrm{d}_{\mathrm{sup}}$ as $n \to \infty$. This means that we can find increasing time shifts $\psi_n : [-1, 0] \to [-1, 0]$ with $\|\psi_n - \mathrm{Id}\| \to 0$ a.s. and such that

$$\sup_{\lambda\in[-1,0]} \mathrm{d}_{\mathrm{sup}}(\mathbb{F}_n(\lambda), \mathscr{FM}(\psi_n(\lambda)) \xrightarrow[n\to\infty]{a.s.} 0. \tag{6.42}$$

Recalling the notation introduced before Lemma 6.9 and using this lemma, up-to rebuilding a new coupling, one can furthermore suppose that we have for every $\xi > 0$

$$\limsup_{\xi\downarrow 0} \sup_{n\geqslant 1} \sup_{\lambda\in[-1,0]} R_\xi(\mathbb{F}_n(\lambda)) \xrightarrow[n\to\infty]{a.s.} 0.$$

In particular, by Fatou's lemma, this implies a similar estimate for $\mathscr{FM}$ namely,

$$\limsup_{\xi\downarrow 0} \sup_{\lambda\in[-1,0]} R_\xi(\mathscr{FM}(\lambda)) \xrightarrow[n\to\infty]{a.s.} 0.$$

One can now use our coupling to evaluate the $\mathcal{E}$-distance between $\mathbb{F}_n$ and $\mathscr{FM}$, namely

$$\mathrm{d}_{\mathcal{E}}(\mathbb{F}_n(\lambda), \mathscr{FM}(\psi_n(\lambda))) \leqslant 2 \sup_{\lambda\in[-1,0]} (R_\xi(\mathscr{FM}(\lambda)) + R_\xi(\mathbb{F}_n(\lambda))) + \mathrm{d}_{\mathcal{E}}(\mathbb{F}_n^{\{\xi\}}(\lambda), \mathscr{FM}^{\{\xi\}}(\psi_n(\lambda))).$$

Using the second and third to last displays, the first term on the right-hand side can be made small uniformly in n and $\lambda \in [-1, 0]$ by choosing ξ small enough. The second term also tends to 0

thanks to (6.42): since $\mathscr{FM}$ is càdlàg with values in $\mathcal{E}_0$, the maximal size of a particle in $\mathscr{FM}$ (and in $\mathbb{F}_n(\cdot)$) and the maximal number of particles of mass $> \xi$ is bounded over $[-1,0]$. We have indeed proved that in this coupling we have $\mathbb{F}_n \to \mathscr{FM}$ almost surely for the $\mathrm{d}_{\mathcal{E}}$ metric. This implies the desired result.

6.7 Markovian properties of the freezer and the flux

Since the mapping $\mathbf{z} \in \mathcal{E} \to \|\mathbf{z}\|_\bullet$ is continuous for the topology induced by $\mathrm{d}_{\mathcal{E}}$, our Theorem 6.2 implies that

$$(n^{-2/3} \cdot \|\mathrm{F}_n(\lambda)\|_\bullet : \lambda \in \mathbb{R}) \xrightarrow[n\to\infty]{(d)} (\|\mathscr{FM}(\lambda)\|_\bullet : \lambda \in \mathbb{R}), \tag{6.43}$$

for the Skorokhod topology where we recall that $\|\mathscr{FM}(\lambda)\|_\bullet$ is the total mass of the frozen particles in the frozen multiplicative coalescent. In this section, we use the Markov properties of the process $\|F(n,\cdot)\|_\bullet$ –or more precisely of $(\|F(n,\cdot)\|_\bullet, \|F(n,\cdot)\|_{\bullet\bullet})$– given in Proposition 6.8 to prove (Proposition 6.14) the joint convergence of the number of discarded edges $D(n,\cdot)$ in the scale $n^{1/3}$ which, thanks to our coupling construction, will give us the flux of outgoing cars in the parking process on Cayley trees.

Using this Markovian point of view, we also give in Proposition 6.15 a new and perhaps more concrete construction of the process $\lambda \mapsto \|\mathscr{FM}(\lambda)\|_\bullet$ as a pure-jump Feller process with a time-inhomogeneous infinite jump measure

$$\mathbf{n}_\lambda(x, \mathrm{d}y) = \frac{1}{2} \cdot \frac{\mathrm{d}y}{\sqrt{2\pi y^3}} g_{x,\lambda}(y), \tag{6.44}$$

where $g_{x,\lambda}(y)$ was defined in (6.13). We complete this alternative Markovian description of the frozen multiplicative coalescent by computing the law of $[\mathscr{FM}(\lambda)]_\circ$ given $\|\mathscr{FM}(\lambda)\|_\bullet$ (see Proposition 6.16): By passing Proposition 6.8 to the scaling limit, conditionally on $\|\mathscr{FM}(\lambda)\|_\bullet$, the ℓ^2-part of $\mathscr{FM}(\lambda)$ has the law of the scaling limit of component sizes in a critical random forest. This law has been described in [138] using excursion lengths of a time inhomogeneous diffusion with a reflection term. We give here an alternative (and quicker) description of this law using conditioned 3/2-stable Lévy processes. The last two results are not used for the parking process but we include them to motivate further the study of the frozen Erdős–Rényi process, see Part 7.6 for perspectives.

6.7.1 Scaling limits for the freezer and flux

The main result of this section is the joint convergence of the renormalized number of discarded edges $D(n,m)$ together with (6.43), see Figure 6.21.

Proposition 6.14 (Joint convergence of discarded edges)**.** *Jointly with the convergence of Theorem 6.2 we have the following convergence in distribution for the uniform topology on $\mathcal{C}(\mathbb{R},\mathbb{R})$*

$$\left(n^{-1/3} \cdot \mathrm{D}_n(\lambda)\right)_{\lambda\in\mathbb{R}} \xrightarrow[n\to\infty]{(d)} \left(\frac{1}{2}\int_{-\infty}^{\lambda} \mathrm{d}s\, \|\mathscr{FM}(s)\|_\bullet\right)_{\lambda\in\mathbb{R}}.$$

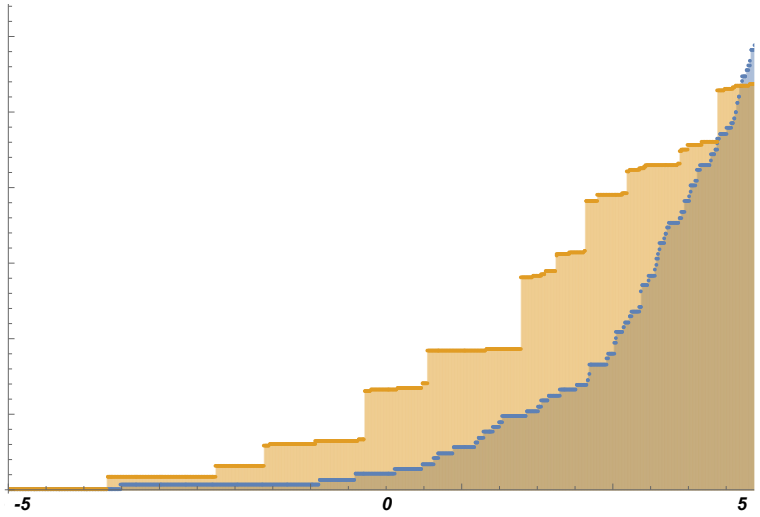


Figure 6.21: A simulation of the renormalized processes $n^{-2/3} \cdot \|\mathrm{F}_n(\lambda)\|_\bullet$ (in orange) and $n^{-1/3} \cdot \mathrm{D}_n(\lambda)$ (in blue) for $\lambda \in [-5, 5]$ and $n = 3000$. Notice that the first process is pure jump in the scaling limit and that the second one is the integral of the first.

The main ingredient to prove the joint convergence of the number of discarded edges together with Theorem 6.2 is the following consequence of Corollary 2

$$\mathbb{P}\left(\Delta D(n,m) = 1 \middle| \|F(n,m)\|_\bullet\right) = \frac{\|F(n,m)\|_\bullet}{n}, \tag{6.45}$$

so that the above result formally follows from integrating and passing to the limit. To make this precise, we shall start with a lemma controlling the flux at the bottom of the critical window:

Lemma 6.10. *There exists a constant $C > 0$ such that for any $n \geqslant 1$ and any $\lambda < -1$*

$$\mathbb{E}[\mathrm{D}_n(\lambda)] \leqslant \frac{C}{|\lambda|} n^{1/3}.$$

Proof. In this proof, for $n \geqslant 1$ and $m \geqslant 0$ we denote by $\mathrm{C}\ell(n,m)$ the cluster of the vertex 1 in $G(n,m)$ and by $\mathrm{Spl}(n,m) = \|\mathrm{C}\ell(n,m)\|_{\bullet\bullet} - \|\mathrm{C}\ell(n,m)\|_\bullet + 1$ its surplus. Recall from (6.2) that the number of blue vertices in $F(n,m)$ is less than the number of vertices which belong to a component which has a cycle in $G(n,m)$. By (6.45) and taking expectations we deduce that

$$\mathbb{E}\left[\Delta D(n,m)\right] = \frac{1}{n}\mathbb{E}\left[\|F(n,m)\|_\bullet\right] \leqslant \mathbb{P}\left(\mathrm{Spl}(n,m) \geqslant 1\right) \leqslant \mathbb{E}\left[\mathrm{Spl}(n,m)\right]. \tag{6.46}$$

Now we provide an upper bound for $\mathbb{E}\left[\mathrm{Spl}(n,m)\right]$ following the proof of [88, Theorem 1.2]. Note that $\mathrm{Spl}(n,m)$ is bounded from above by the number of vertex-disjoint cycles (including self-loops) in $\mathrm{C}\ell(n,m)$. Given a graph $\mathfrak{g}$ and k distinct vertices $(v_1, \ldots, v_k)$, we say that the graph $\mathfrak{g}$ *contains the cycle* $(v_1, \ldots v_k)$ of length $k \geqslant 1$ if $\mathfrak{g}$ contains the (unoriented) edges $(v_i, v_{i+1 \bmod k})$ for $1 \leqslant i \leqslant k$. For $k = 1$, the expected number of self-loops (or cycle of length 1) in $\mathrm{C}\ell(n,m)$ is bounded above by

$m\mathbb{E}\left[\|\mathrm{C}\ell(n,m)\|_\bullet\right]/n^2$. Since every cycle of length $k \geqslant 3$ (resp. 2) corresponds to exactly $2k$ (resp. 2) k-uplet $(v_1,\dots,v_k)$, we have for $2m < n$,

$$\begin{aligned}
&\sum_{k=2}^{n} \sum_{L_k \text{ cycle of length } k} \mathbb{P}\left(G(n,m) \text{ contains } L_k \text{ and } 1 \text{ is connected to } L_k\right)\\
&\leqslant \sum_{k=2}^{n} \sum_{L_k=(v_1,\dots,v_k)} \frac{1}{k} \sum_{I\subset\{1,\dots,m\},|I|=k} \mathbb{P}((\vec{E}_i)_{i\in I} \text{ form } L_k \text{ and } (\vec{E}_i)_{i\leqslant m, i\notin I} \text{ connect } 1 \text{ with } \{v_1,\dots v_k\})\\
&\leqslant \sum_{k=2}^{n} \sum_{(v_1,\dots,v_k)} \frac{1}{k} \sum_{I\subset\{1,\dots,m\},|I|=k} k!\left(\frac{2}{n}\right)^k \cdot \frac{2k}{n}\mathbb{E}\left[\|\mathrm{C}\ell(n,m)\|_\bullet\right]\\
&= \sum_{k=2}^{n} \frac{n!}{k(n-k)!}\binom{m}{k} k!\left(\frac{2}{n^2}\right)^k \cdot \frac{2k}{n}\mathbb{E}\left[\|\mathrm{C}\ell(n,m)\|_\bullet\right]\\
&\leqslant \frac{2}{n}\sum_{k=2}^{n}\left(\frac{2m}{n}\right)^k \mathbb{E}\left[\|\mathrm{C}\ell(n,m)\|_\bullet\right] \leqslant \frac{C}{n\left(1-\frac{2m}{n}\right)}\mathbb{E}\left[\|\mathrm{C}\ell(n,m)\|_\bullet\right],
\end{aligned}$$

for some constant $C > 0$ that may vary in the following lines. An easy adaptation of [109, Theorem 1.1] using [109, Remark 1.6] to our model $G(n,m)$ shows that

$$\mathbb{E}\left[\|\mathrm{C}\ell(n,m)\|_\bullet\right] \leqslant \frac{C}{\left(1-\frac{2m}{n}\right)}, \tag{6.47}$$

so that combining these inequalities we deduce that for $2m < n$

$$\mathbb{E}[\mathrm{Spl}(n,m)] \leqslant \frac{C}{n(1-\frac{2m}{n})^2}. \tag{6.48}$$

Coming back to D and writing $m_n(\lambda) = 0 \wedge \lfloor \frac{n}{2} + \frac{\lambda}{2}n^{2/3}\rfloor$, we obtain using (6.46) that for $\lambda < -1$,

$$\mathbb{E}\left[\mathrm{D}_n(\lambda)\right] = \mathbb{E}\left[D(n,\mathrm{m}_n(\lambda))\right] = \sum_{m=0}^{\mathrm{m}_n(\lambda)-1} \mathbb{E}\left[\Delta D(n,m)\right] \leqslant C\sum_{m=0}^{\mathrm{m}_n(\lambda)-1} \frac{1}{n(1-\frac{2m}{n})^2} \leqslant C\frac{n^{1/3}}{|\lambda|},$$

which concludes the proof.

□

Proof of Proposition 6.14. Recall that the convergence of Theorem 6.2 implies the convergence of of $(n^{-2/3}\|\mathrm{F}_n(\lambda)\|_\bullet : \lambda \in \mathbb{R})$. We now prove the joint convergence of $(n^{-1/3}\mathrm{D}_n(\lambda) : \lambda \in \mathbb{R})$ using the probability transitions given in (6.45). Indeed, writing $m_n(\lambda) = 0 \wedge \lfloor \frac{n}{2} + \frac{\lambda}{2}n^{2/3}\rfloor$ as above, we have

$$\mathrm{D}_n(\lambda) - \mathrm{D}_n(\lambda_0) = \sum_{m=\mathrm{m}_n(\lambda_0)}^{\mathrm{m}_n(\lambda)-1} \Delta D(n,m),$$

and for all $m \geqslant 0$, conditionally on $\|F(n,m)\|_\bullet$, the variable $\Delta D(n,m)$ is a Bernoulli variable with parameter $\|F(n,m)\|_\bullet/n$. By the convergence of the process $(n^{-2/3}\|\mathrm{F}_n(\lambda)\|_\bullet : \lambda \in \mathbb{R})$ to $(\|\mathscr{FM}(\lambda)\|_\bullet : \lambda \in \mathbb{R})$ in the Skorokhod sense, we have for every $-\infty < \lambda_0 < \lambda < +\infty$,

$$\sum_{m=\mathrm{m}_n(\lambda_0)}^{\mathrm{m}_n(\lambda)-1} n^{-4/3}\|F(n,m)\|_\bullet = \frac{1}{2}n^{-2/3}\int_{n^{-2/3}(\mathrm{m}_n(\lambda_0)-n/2)}^{n^{-2/3}(\mathrm{m}_n(\lambda)-n/2)} \|\mathrm{F}_n(s)\|_\bullet \mathrm{d}s \xrightarrow[n\to\infty]{(d)} \frac{1}{2}\int_{\lambda_0}^{\lambda} \mathrm{d}s\, \|\mathscr{FM}(s)\|_\bullet.$$

In addition, since $(\Delta D(n,m) - \|F(n,m)\|_\bullet/n : m \geqslant 0)$ are the increments of a martingale,

$$\begin{aligned}
&\mathbb{E}\left[\left(n^{-1/3}\sum_{m=\mathrm{m}_n(\lambda_0)}^{\mathrm{m}_n(\lambda)-1}\Delta D(n,m) - \sum_{m=\mathrm{m}_n(\lambda_0)}^{\mathrm{m}_n(\lambda)-1} n^{-4/3}\|F(n,m)\|_\bullet\right)^2\right]\\
&= n^{-2/3}\sum_{m=\mathrm{m}_n(\lambda_0)}^{\mathrm{m}_n(\lambda)-1}\mathbb{E}\left[\left(\Delta D(n,m) - \frac{\|F(n,m)\|_\bullet}{n}\right)^2\right]\\
&= n^{-2/3}\sum_{m=\mathrm{m}_n(\lambda_0)}^{\mathrm{m}_n(\lambda)-1}\mathbb{E}\left[\frac{\|F(n,m)\|_\bullet}{n}\left(1-\frac{\|F(n,m)\|_\bullet}{n}\right)^2 + \left(1-\frac{\|F(n,m)\|_\bullet}{n}\right)\cdot\left(\frac{\|F(n,m)\|_\bullet}{n}\right)^2\right]\\
&\leqslant n^{-2/3}\sum_{m=\mathrm{m}_n(\lambda_0)}^{\mathrm{m}_n(\lambda)-1}\mathbb{E}\left[\frac{\|F(n,m)\|_\bullet}{n}\right]
\end{aligned}$$

which converges to 0 as $n\to\infty$ by the estimates of (the proof of) the previous lemma. It follows that

$$\mathrm{D}_n(\lambda) - \mathrm{D}_n(\lambda_0) \xrightarrow[n\to\infty]{(d)} \frac{1}{2}\int_{\lambda_0}^{\lambda} \mathrm{d}s\, \|\mathscr{FM}(s)\|_\bullet.$$

Now we use Lemma 6.10, which shows that $\mathbb{E}\left[n^{-1/3}\mathrm{D}_n(\lambda_0)\right]$ can be made arbitrarily small if we choose λ_0 small enough, and this uniformly in n. Hence, by Fatou's lemma, we obtain for all $\lambda_0 \in (-\infty, \lambda]$,

$$\frac{1}{2}\int_{\lambda_0}^{\lambda} \mathrm{d}s\, \|\mathscr{FM}(s)\|_\bullet < \infty \qquad \text{a.s.},$$

and letting $\lambda_0 \to -\infty$, we get

$$\mathrm{D}_n(\lambda) \xrightarrow[n\to\infty]{(d)} \frac{1}{2}\int_{-\infty}^{\lambda} \mathrm{d}s\, \|\mathscr{FM}(s)\|_\bullet = \mathscr{D}(\lambda). \tag{6.49}$$

The above reasoning can be extended to prove that, jointly with the convergence of the first coordinate in Proposition 6.14, for each $-\infty < \lambda_1 < \lambda_2 < \cdots < \lambda_k < \infty$ we have $n^{-1/3}\cdot \mathrm{D}_n(\lambda_i) \to \mathscr{D}(\lambda_i)$. Since the processes $\mathrm{D}_n(\cdot)$ are increasing and since $\lambda \mapsto \mathscr{D}(\lambda)$ is continuous and increasing as well, this is sufficient to imply the joint convergence of $n^{-1/3}\cdot \mathrm{D}_n(\cdot)$ to $\mathscr{D}(\cdot)$ for the uniform norm over every compact of $\mathbb{R}$.

□

6.7.2 Proof of Theorem 6.1

Near components. Recall the notation $\mathbb{F}_n(\lambda) \in \mathcal{E}$ for the renormalized components sizes (first frozen, followed by the standard ones) in $\mathrm{F}_n(\lambda)$. Accordingly, we write $\mathbb{T}_{\mathrm{near},n}(\lambda)$ for the vector

$$\left(n^{-2/3}\mathrm{C}_{*,n}(\lambda); \left(n^{-2/3}\cdot \mathrm{C}_{i,n}(\lambda) : i \geqslant 1\right)\right),$$

where $\mathrm{C}_{*,n}(\lambda)$ and $\mathrm{C}_{i,n}(\lambda)$ are the sizes of the (blue) root component followed by the other components in decreasing order of size in $\mathrm{T}_{\mathrm{near},n}(\lambda)$. Recall that by Proposition 6.6, the white components of

$F(n,m)$ are exactly the components of $T_{\text{near}}(n,m)$ which do not contain the root and the blue vertices of $F(n,m)$ are exactly the vertices of the parked component of the root. Furthermore, the flux of outgoing cars $D(n,m)$ corresponds to the number of discarded edges in $F(n,m)$. Since the mapping $\mathbf{z} \in \mathcal{E} \mapsto (\|\mathbf{z}\|_\bullet, [\mathbf{z}]_\circ) \in \mathbb{R}_+ \times \ell^2_\downarrow$ is continuous, we can combine Theorem 6.2 and Proposition 6.14 to deduce that

$$\left(n^{-1/3} \cdot \mathrm{D}_n(\lambda); \mathbb{T}_{\text{near},n}(\lambda)\right)_{\lambda\in\mathbb{R}} \xrightarrow[n\to\infty]{(d)} (\mathscr{D}(\lambda), \|\mathscr{FM}(\lambda)\|_\bullet, [\mathscr{FM}(\lambda)]_\circ)_{\lambda\in\mathbb{R}},$$

for the Skorokhod topology on $\mathbb{C}\mathrm{adlag}(\mathbb{R}\times\mathbb{R}\times\ell^2_\downarrow)$.

□

Full components. Let us sketch how to obtain the equivalent of Theorem 6.1 for the full components of $\mathrm{T}_{\text{full}}(n,\cdot)$ rather than the near components. To extend the above convergence to the case of full components, notice first that $\mathrm{D}_n(\lambda)$ stays the same for near and full components, and that as soon as $\mathrm{D}_n(\lambda) > 0$ (which is the case with high probability in the whole critical window) we have with an obvious notation

$$\|\mathbb{T}_{\text{near},n}(\lambda)\|_\bullet = \|\mathbb{T}_{\text{full},n}(\lambda)\|_\bullet,$$

since the blue components of the root with flux are the same in T_{near} and in T_{full}. We just have to show that for any fixed compact time interval I we have

$$\sup_{\lambda\in I} \mathrm{d}_{\ell^2_\downarrow}([\mathbb{T}_{\text{near},n}(\lambda)]_\circ, [\mathbb{T}_{\text{full},n}(\lambda)]_\circ) \xrightarrow[n\to\infty]{(\mathbb{P})} 0,$$

with an obvious notation. For any fixed λ, the fully parked trees of $\mathrm{T}_{\text{full},n}(\lambda)$ are obtained by splitting the nearly parked trees of $\mathrm{T}_{\text{near},n}(\lambda)$ at their root vertices. Since by Proposition 6.12, conditionally on their sizes, those are uniform nearly parked trees (this can also be seen by combining our coupling construction with Proposition 6.8 and Proposition 6.7), we deduce from Proposition 6.11 that each large nearly parked tree of $\mathrm{T}_{\text{near},n}(\lambda)$ contains a unique fully parked tree of roughly the same size. Since $[\mathbb{T}_{\text{near},n}(\lambda)]_\circ$ converges in $\ell^\infty_{\downarrow,0}$ we easily deduce that for each $\lambda \in \mathbb{R}$ we have

$$\mathrm{d}_{\ell^\infty_{\downarrow,0}}([\mathbb{T}_{\text{near},n}(\lambda)]_\circ, [\mathbb{T}_{\text{full},n}(\lambda)]_\circ) \xrightarrow[n\to\infty]{(\mathbb{P})} 0.$$

Actually, the last display also holds for any stopping time Λ which belongs to some fixed time interval. Combining this we the monotony property of the processes T_{near} and T_{full}, standard but tedious arguments (which we shamefully leave to the reader) show that we in fact have

$$\sup_{\lambda\in I} \mathrm{d}_{\ell^\infty_{\downarrow,0}}([\mathbb{T}_{\text{near},n}(\lambda)]_\circ, [\mathbb{T}_{\text{full},n}(\lambda)]_\circ) \xrightarrow[n\to\infty]{(\mathbb{P})} 0.$$

To boostrap the above convergence by replacing the $\ell^\infty_{\downarrow,0}$ metric with the $\ell^2_\downarrow$ metric, we use Lemma 6.9 on $[\mathbb{T}_{\text{near},n}(\lambda)]_\circ$ and remark that the proof straightforwardly extend to $[\mathbb{T}_{\text{full},n}(\lambda)]_\circ$.

□

Strong components. Obviously a version of Theorem 6.1 holds if we consider the strong components in the parking process to the cost of multiplying the scaling limit of the components sizes by $1/2$, since by Proposition 6.11 each large nearly parked tree contains a giant strongly parked tree of roughly half its size. To be precise, one would need to establish the same behavior for the component of the root (which may have some outgoing flux). We refrain from doing so to keep the paper's length acceptable.

6.7.3 The freezer as a Lévy-type process

In this section we give an alternative description of the process $\|\mathscr{FM}(\cdot)\|_\bullet$ by "passing Corollary 2 to the scaling limit". Since we shall not use this in the rest of the paper, the proofs are only sketched and this section can be skipped at first reading. Recall the function $g_{x,\lambda}(y)$ from (6.13).

Proposition 6.15 (A pure-jump description of $\|\mathscr{FM}(\lambda)\|_\bullet$)**.** *The process $\lambda \mapsto \|\mathscr{FM}(\lambda)\|_\bullet$ is a Markov Feller processs with inhomogeneous jump measure*

$$\mathbf{n}_\lambda(x, \mathrm{d}y) = \frac{1}{2}\frac{\mathrm{d}y}{\sqrt{2\pi y^3}} g_{x,\lambda}(y)$$

started "from 0 at time $-\infty$".

Heuristically, this means that the process $\lambda \mapsto \|\mathscr{FM}(\lambda)\|_\bullet$ has no drift, no Brownian part and jumps according to a modification of the (infinite) measure $y^{-3/2}\mathrm{d}y\mathbf{1}_{y>0}$ depending on time λ and location x. This is an example of a so-called Lévy-type process (quite simple in our case since we only have positive jumps) we refer to the monograph [42] for survey. We shall rather see it as the solution of a pure-jump stochastic differential equation driven by some Poisson measure.

Sketch of proof. Let us first see why we can define a Feller Markov process $\mathscr{P}$ with the above jump kernel over a time interval $[\lambda_0, \lambda_1] \subset \mathbb{R}$ starting from the initial value $x_0 \geqslant 0$ at time λ_0. To do this, we consider a Poisson point process Π over $\mathbb{R}_+ \times \mathbb{R}_+ \times [\lambda_0, \lambda_1]$ with (infinite) intensity

$$\frac{1}{2}\frac{\mathrm{d}y}{\sqrt{2\pi y^3}} \cdot \mathrm{d}z\mathbf{1}_{z\geqslant 0} \cdot \mathrm{d}\lambda.$$

We then consider the solution $\mathscr{P}$ to a pure-jump stochastic differential equation driven by Π, obtained by starting from x_0 at time λ_0 and from every atom (y, z, λ) of Π, the process $\mathscr{P}$ has a jump of height y at time λ^- i.e. $\mathscr{P}_\lambda = \mathscr{P}_{\lambda-} + y$ if

$$z \leqslant g_{\mathscr{P}_{\lambda^-},\lambda}(y),$$

so that the jump kernel is indeed given by $\mathbf{n}_\lambda(x, \mathrm{d}y)$. We now verify the usual Lipschitz conditions so that strong solution and pathwise uniqueness holds. For this, we shall first gather a few remarks on the function p_1:

(P1) The function $x \mapsto p_1(x)$ is bimodal: increasing from $-\infty$ to $x_{\max} \approx -0.886$ and then decreasing from then on.

(P2) For all $\lambda \in [\lambda_0, \lambda_1]$ and $x, y \geqslant 0$, the ratio $p_1(\lambda - x - y)/p_1(\lambda - x)$ is bounded by a constant $C = 1 \wedge (p_1(x_{\max})/p_1(\lambda_1)) > 0$ depending only on λ_1.

(P3) The function $g_{x,\lambda}(y)$ is a smooth function of any of its variable $x \geqslant 0, y \geqslant 0$ and $\lambda \in \mathbb{R}$.

Those properties are easily proven using a Math software such as Mathematica or Maple. In particular, using $(P2)$ we see that

$$\forall \lambda \in [\lambda_0, \lambda_1], \forall x \geqslant 0, \quad \int_0^1 y \cdot \mathbf{n}_\lambda(x, \mathrm{d}y) \leqslant c_1(1+x),$$

for some $c_1 > 0$ depending on λ_1 only so that the "linear growth condition is satisfied" and the process does not explode in finite time. By $(P3)$, it follows that for any $A > 0$ we have

$$\forall \lambda \in [\lambda_0, \lambda_1], \forall x, x' \in [0, A], \quad \int_0^1 y \cdot \frac{\mathrm{d}y}{\sqrt{2\pi y^3}} |g_{x,\lambda}(y) - g_{x',\lambda}(y)| \leqslant c_2 |x - x'|,$$

for some $c_2 > 0$ depending on λ_1 and A. We are thus in the classical Lipschitz and linear growth condition so that we have strong solution and pathwise uniqueness for $\mathscr{P}$, see [103, Theorem 9.1, page 245] or [104, Chap III.2.c, page 155]. It is also easy to check that the resulting process is a Feller Markov process $\mathscr{P}$. Furthermore, the process $\mathscr{P}$ is the scaling limit of the chain $\|F(n,m)\|_\bullet$ in the sense that if we start the Markov chain $(\|F(n,m)\|_\bullet, \|F(n,m)\|_{\bullet\!-\!\bullet})$ from $m = \frac{n}{2} + \frac{\lambda_0}{2} n^{2/3}$ with $\|F(n,m)\|_\bullet = x_0 n^{2/3}$ and $m - \|F(n,m)\|_{\bullet\!-\!\bullet} = o(n^{2/3})$ then

$$(n^{-2/3} \mathrm{F}_n(\lambda) : \lambda \in [\lambda_0, \lambda_1]) \xrightarrow[n\to\infty]{} (\mathscr{P}(\lambda) : \lambda \in [\lambda_0, \lambda_1]) \text{ with } \mathscr{P}(\lambda_0) = x_0 \tag{6.50}$$

in the sense of Skorokhod. Indeed, the asymptotics (6.13) shows that the jump kernels of $(n^{-2/3}\|\mathrm{F}_n(\lambda)\|_\bullet, n^{-2/3}\mathrm{D}_n($ converge towards $(\mathbf{n}_\lambda(x, \mathrm{d}y), \mathbf{0})$ and for any $\varepsilon > 0$, all $m = \frac{n}{2} + \frac{\lambda}{2} n^{2/3}$ for $\lambda \in [\lambda_0, \lambda_1]$ and n large enough

$$\mathbb{E}\left[\begin{array}{r} \min(\Delta \|F(n,m+1)\|_\bullet, \varepsilon n^{2/3}) \\ \Delta(m - \|F(n,m)\|_{\bullet\!-\!\bullet}) \end{array} \middle| \begin{array}{l} \|F(n,m)\|_\bullet = x n^{2/3} \\ m - \|F(n,m)\|_{\bullet\!-\!\bullet} \leqslant n^{2/3} \end{array} \right] \leqslant \begin{pmatrix} C\sqrt{\varepsilon} \\ 2(1+x) n^{-1/3} \end{pmatrix},$$

for some constant $C > 0$ independent of n. The convergence (6.50) is then a consequence of general convergence results on Feller processes, see [114, Chapter 19] or [104, Chapter IX, 4]. We leave the verifications to the reader.

Finally, let us see why we can define the Feller process $\mathscr{P}$ by starting from 0 at time $-\infty$. To prove this, we need to show convergence of $\mathscr{P}$ at a fixed time, say $\lambda = 0$, when $\mathscr{P}$ is started from 0 at a very negative time $\lambda_0 \ll 0$. This can be deduced by rather tedious calculations using $\mathbf{n}_\lambda$ and asymptotics of p_1 but let us sketch another route using our cutoff construction of Section 6.6.1. Specifically, recall the construction of the process $\mathrm{F}_n^{[\lambda_0]}(\lambda)$ obtained by throwing all components with cycles in $\mathrm{G}_n(\lambda_0)$ and starting the construction of the frozen process from there. We shall use a variant of this construction by considering a random stopping time Λ_0 (with an implicit dependence in n) associated to $M_0 = \frac{n}{2} + \frac{\Lambda_0}{2} n^{2/3}$ defined as follows

$$M_0 = \inf\left\{ m \geqslant 0 : 2\left(m - \frac{n}{2}\right) - \|G(n,m)\|_\bullet \geqslant \lambda_0 n^{2/3} \right\}.$$

In words, M_0 is the first instant m where the felt time-parameter in the forest part $[G(n,m)]_{\text{tree}}$ is above λ_0. We first claim that for λ_0 negative enough, $M_0 \leqslant n/2$ and is actually close to $\frac{n}{2} + \frac{\lambda_0}{2} n^{2/3}$ with high probability: indeed it follows from (6.48) that $n^{-2/3} \cdot \|\mathrm{G}_n(\lambda_0)\|_\bullet$ is of order λ_0^{-2} and so the function $m \mapsto 2\left(m - \frac{n}{2}\right) - \|G(n,m)\|_\bullet \geqslant \lambda_0 n^{2/3}$ crosses $\lambda_0 n^{2/3}$ around time $\lambda_0 \pm \lambda_0^{-2}$. On this event, by Proposition 6.8, the process $\mathrm{F}_n^{[\Lambda_0]}(\cdot + \Lambda_0)$ has the same transitions as F started from 0 at time $m' = \frac{n'}{2} + \frac{\lambda_0}{2} n'^{2/3} + o(n^{2/3})$ over $n' = n - \|G(n, M_0)\|_\bullet$ vertices with a slight time change coming from the fact that certain edges are discarded (this does not persist in the limit). So by (6.50) it converges after scaling towards the process $\mathscr{P}(\cdot + \lambda_0)$ started from 0 at time λ_0. The convergence of $n^{-2/3} \cdot (\mathrm{F}_n^{[\Lambda_0]}(\lambda) : \lambda \geqslant 0)$ proved in Section 6.6.1 together with the fact that $n^{-2/3}\|\mathrm{G}_n(\Lambda_0)\|_\bullet \to 0$ as $\lambda_0 \to \infty$ and the above convergence imply that the law of the process $\mathscr{P}$ started from 0 at time λ_0 does converge as $\lambda_0 \to \infty$ and this enables us to start $\mathscr{P}$ from 0 at time $-\infty$. Combining those observations we deduce that the process $\mathscr{P}$ started from 0 at time $-\infty$ has the same law as $\|\mathscr{FM}(\cdot)\|_\bullet$. We leave the many details to the fearless reader.

□

6.7.4 Scaling limit of random forest

In this section we revisit the result of Martin and Yeo [138] to complete the Markovian description of the scaling limit of the frozen multiplicative coalescent. As for the preceding section, the results are not used in the rest of the paper and so this part may be skipped at first reading.

As in the proof of Corollary 6.8, for $n \geqslant 1$ and $m \geqslant 0$ we denote by $W(n,m) \in \mathfrak{F}(n,m)$ a uniform random forest over the n labeled vertices $\{1, 2, \ldots, n\}$ with m edges in total. We chose the letter W for the German "Wald" because there are already too many f's in the paper. In particular, the forest $W(n,m)$ has $n - m$ components. Although there is *no* obvious coupling of $W(n,m)$ for varying $m \geqslant 0$ (see [138, Section 1.4.2]), we shall use our usual notation (6.1) and write $\mathrm{W}_n(\lambda)$ for a random forest with $m = \lfloor \frac{n}{2} + \frac{\lambda}{2} n^{2/3} \rfloor$ edges and by $\mathbb{W}_n(\lambda) \in \ell^\infty_{\downarrow,0}$ the renormalized sequence of its component sizes in non-increasing order.

Recall from Section 6.3.1 that $(\mathscr{S}_t)_{t \geqslant 0}$ denotes the stable Lévy process with index $3/2$ and only positive jumps, which starts from 0 and normalized so that $\mathbb{E}[\exp(-\ell \mathscr{S}_t)] = \exp(\frac{2^{3/2}}{3} t \ell^{3/2})$ for any $\ell, t \geqslant 0$, see Figure 6.22 for a simulation. The density of $\mathscr{S}_t$ is $p_t(\cdot)$ for $t > 0$. For any $\lambda \in \mathbb{R}$ we can use this function to define the process $(\mathscr{S}_t^\lambda : 0 \leqslant t \leqslant 1)$ called the $(0,0) \to (1,\lambda)$ bridge, obtained by conditioning $(\mathscr{S}_t : 0 \leqslant t \leqslant 1)$ to be equal to λ at time 1. Of course this is a degenerate conditioning, but it can be obtained by performing an inhomogeneous h-transform with respect to the function

$$\frac{p_{1-t}(\lambda - \mathscr{S}_t)}{p_1(\lambda)},$$

see [131, Theorem 4].

Proposition 6.16 (Another route towards critical random forests). *Fix $\lambda \in \mathbb{R}$. For all $\varepsilon > 0$, we have the following convergence in distribution for the $\ell^{3/2+\varepsilon}_{\downarrow}$- topology*

$$\mathbb{W}_n(\lambda) \xrightarrow[n\to\infty]{(d)} \left(\Delta \mathscr{S}^{\lambda}_t : 0 \leqslant t \leqslant 1\right)^{\downarrow} \tag{6.51}$$

where $(x_i : i \geqslant 1)^{\downarrow}$ is the non-increasing rearrangement of the sequence $(x_i : i \geqslant 1)$.

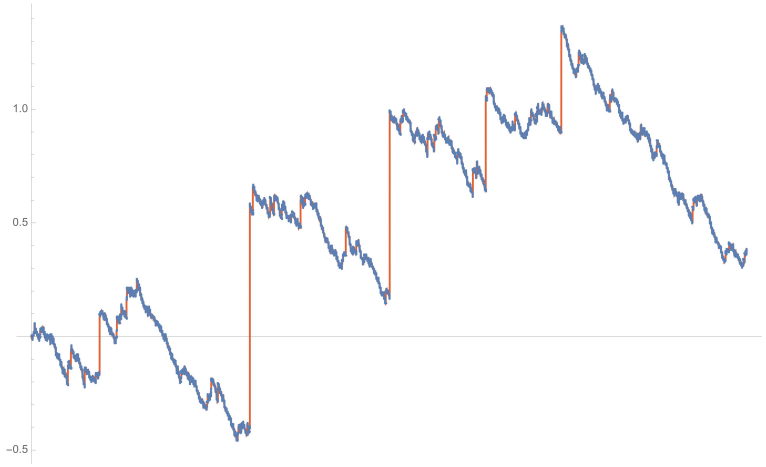


Figure 6.22: A simulation of a $\frac{3}{2}$-stable spectrally positive Lévy process over the time interval $[0,1]$. The jumps are displayed in orange.

Proof. From Proposition 6.7, the sizes of the components in the random forest $W(n,m)$ has the same law as the increments $+2$ of a random walk $(\tilde{S}_i : 0 \leqslant i \leqslant n-m)$ started from 0, conditioned to hit $n - 2(n-m) = \lambda n^{2/3} + o(n^{2/3})$ at time $n-m$ and with independent increments of law $\mu(k+2)$ for $k \in \{-1, 0, 1, 2, \dots\}$ introduced in (6.9). Recall that the variable $\tilde{S}_1$ is centered and in the domain of attraction of the $3/2$-stable spectrally positive random variable. The following convergence for the Skorokhod topology on $\mathbb{C}\mathrm{adlag}([0,1],\mathbb{R})$ follows from the conditional invariance principle of Liggett [131]

$$\left(n^{-2/3}\tilde{S}_{[(n-m)t]} : 0 \leqslant t \leqslant 1\right) \xrightarrow[n\to\infty]{(d)} \left(\mathscr{S}^{\lambda}_t : 0 \leqslant t \leqslant 1\right).$$

In particular by [104, Corollary 2.8], the random point measure $\sum_{0\leqslant t\leqslant 1} \delta_{n^{-2/3}\cdot\Delta\tilde{S}_{[(n-m)t]}}$ converges weakly towards $\sum_{0\leqslant t\leqslant 1}\delta_{\Delta\mathscr{S}^{\lambda}_t}$ from which we deduce the convergence (6.51) for the $\ell^{\infty}_{\downarrow,0}$ topology. To bootstrap this into a convergence for the $\ell^{3/2+\varepsilon}_{\downarrow}$ topology, it suffices to establish tightness in the later (since convergence in $\ell^{\infty}_{\downarrow,0}$ characterizes the limit in $\ell^{3/2+\varepsilon}_{\downarrow}$). For this we claim that it is sufficient to prove that

$$\sup_{n\geqslant 1} \mathbb{E}\left[\sum_{i=0}^{n-m-1}\left(n^{-2/3}\cdot\Delta S_i\right)^{3/2+\varepsilon}\middle| S_{n-m} = n\right] < \infty, \tag{6.52}$$

where S has increments of law $\mu(k)$ given by (6.9). Indeed, for $\xi > 0$ and $\varepsilon > 0$

$$\sum_{i\geqslant 1}\left(n^{-2/3}\cdot\Delta S_i\right)^{3/2+2\varepsilon}\mathbf{1}_{n^{-2/3}\cdot\Delta S_i > \xi} \leqslant \xi^{\varepsilon}\sum_{i=0}^{n-m-1}\left(n^{-2/3}\cdot\Delta S_i\right)^{3/2+\varepsilon},$$

so that using (6.52) the expectation of the right-hand side of the last display can be made arbitrarily small by choosing ξ small. Combining this with the tightness in $\ell^{\infty}_{\downarrow,0}$, it is easy to deduce tightness in $\ell^{3/2+2\varepsilon}_{\downarrow}$ of $(\mathbb{W}_n(\lambda) : n \geqslant 1)$. To prove our claim, notice that by cyclic exchangeability we have

$$\mathbb{E}\left[\sum_{i=0}^{n-m-1}\left(n^{-2/3}\cdot\Delta S_i\right)^{3/2+\varepsilon}\Big|S_{n-m}=n\right] \;=\; (n-m)\sum_{k\geqslant 1}\mu(k)\left|kn^{-2/3}\right|^{3/2+\varepsilon}\frac{\mathbb{P}(S_{n-m-1}=n-k)}{\mathbb{P}(S_{n-m}=n)}.$$

Let us focus first on the k's such that $k \ll n$. In this case, writing $k = yn^{2/3}$ and using (6.9) and (6.12), we deduce that there are constants $C, C' > 0$ (which may depend on our fixed λ) such that the last display is bounded above by

$$Cn^{-2/3}\sum_{k=1}^{\infty}(kn^{-2/3})^{-1+\varepsilon}\cdot\frac{p_1(\lambda-kn^{-2/3})}{p_1(\lambda)} \leqslant C'\int_0^{\infty} y^{-1+\varepsilon}\,\mathrm{d}y < \infty.$$

On the other hand, if k if of order n, rough large deviations estimates show that $\frac{\mathbb{P}(S_{n-m-1}=n-k)}{\mathbb{P}(S_{n-m}=n)}$ is exponentially small (in k and so in n) so that the contribution to the sum is negligible. This finishes the proof of (6.52) and of the proposition.

□

As an application of this methodology, let us revisit a few of the results of [135] discussed in [138, Section 1.4.3]. Consider a slightly supercritical random forest $W(n,m)$ with $m = \frac{n}{2} + \frac{s}{2}$ with $n^{2/3} \ll s \ll n$ whose component sizes are coded by $(\Delta\tilde{S}_i + 2 : 0 \leqslant n-m-1)$ conditioned on $\tilde{S}_{n-m} = s$. According to standard "big-jump" principles, since μ is a subexponential distribution such a random walk has a unique "big-jump" of height of order s and once this jump has been removed, the remaining random walk is close in total variation distance to an unconditioned μ-random walk, see [18, Theorem 1]. This gives another way to prove that the largest cluster in $W(n,m)$ is of size $(1+o(1))s$ and the remaining components converge after normalization by $n^{2/3}$ to the jumps of the *unconditioned* Lévy process $(\mathscr{S}_t : 0 \leqslant t \leqslant 1)$.

III Comments and perspectives

We end this paper by presenting several research directions and connections of our work. This part is informal and we do not claim any mathematical statement. We first draw a parallel between the enumeration of (strongly, fully or nearly) parked trees and random planar maps which gives another support for Conjecture 1. We then present a few fallouts of the study of (generalized) frozen process $F(n,\cdot)$ on the Erdős–Rényi random graph $G(n,\cdot)$. We end with extensions of our work concerning the parking process on random trees.

6.8 Links with planar maps and growth-fragmentation trees

We shall consider strongly parked trees with *outgoing flux*. More precisely, for $n, p \geqslant 0$ we denote by $\mathrm{SP}(n, n+p)$ the number of labeled rooted Cayley trees of size n with $n+p$ labeled cars so that after

parking, all edges have a positive flux and exactly p cars exit the tree. We encode these numbers into the generating function

$$\mathbf{S}(x,y) = \sum_{n\geqslant 1, p\geqslant 0} \frac{\mathrm{SP}(n,n+p)}{n!(n+p)!} x^n y^p,$$

which replaces the univariate generating function $\mathbf{S}(x)$ which we considered in Section 6.4. In particular since King & Yan [119] computed $\mathrm{SP}(n,n) = (2n-2)!$ we have (Proposition 6.10) that

$$\mathbf{S}(x,0) = 1 - \ln(2) - \sqrt{1-4x} + \ln\left(1+\sqrt{1-4x}\right), \quad \text{for } 0 \leqslant x \leqslant x_c = \frac{1}{4}. \tag{6.53}$$

6.8.1 Tutte's equation

To get a functional equation on $\mathbf{S}$ one considers the decomposition of strongly parked trees at the root vertex (see Figure 6.23 left) which shows that $\mathrm{SP}(n,n+p)$ is equal to

$$\sum_{a\geqslant 0}\sum_{k\geqslant 0}\sum_{\substack{n_1,\dots,n_k\geqslant 1\\ p_1,\dots,p_k\geqslant 1}} \frac{1}{k!}\binom{n}{1,n_1,n_2,\dots,n_k}\binom{n+p}{a,n_1+p_1,\dots,n_k+p_k}\prod_{i=1}^{k}\mathrm{SP}(n_i,n_i+p_i)\mathbf{1}_{\substack{n_1+\cdots+n_k=n-1\\ a+p_1+\cdots+p_k=p+1.}}$$

Indeed, the integer a counts the number of cars arriving at the root, the integer k is the number of children of the root vertex and n_i, p_i are the characteristics (number of vertices and outgoing flux) of the subtrees above it. This equation translates into the following equation on $\mathbf{S}$

$$\mathbf{S}(x,y) = \frac{x}{y}\left(\mathrm{e}^{y}\mathrm{e}^{\mathbf{S}(x,y)-\mathbf{S}(x,0)} - 1\right). \tag{6.54}$$

At first sight, one may think that the series $\mathbf{S}(x,0)$ is a necessary input (which we do have) to solve the equation, but a close inspection shows that the equation actually determines the coefficients of $\mathbf{S}$ by induction on $n+p$.

This type of equation is very common in the map enumeration literature where they are called "Tutte" equations, see [43] for a comprehensive survey. More precisely, recall that a map is a planar graph properly embedded in the plane given with one distinguished oriented edge. Following Tutte, when enumerating (various classes of) planar maps by their size n, it is convenient to introduce an external parameter p, the perimeter of the external face (lying on the right of the root edge). When performing the root erasure, certain situations yield a splitting of a map of size n and perimeter p into two components of size n_1 and n_2 having perimeter p_1 and p_2 so that we have (on a high level) $n_1+n_2 \approx n$ and $p_1+p_2 \approx p$ which is similar to penultimate equation above, see Figure 6.23 right. Similar equations arose in [68, 79].

In our case, Equation (6.54) can be solved using the Lambert function. Fix (x,y) and observe that $\mathbf{S} = \mathbf{S}(x,y)$ is solution to an equation of the form $a\mathrm{e}^{\mathbf{S}} + b\mathbf{S} + c = 0$. If we put

$$\Delta = -\exp\left(y - \mathbf{S}(x,0) - \frac{x}{y}\right)\frac{x}{y} \leqslant 0$$

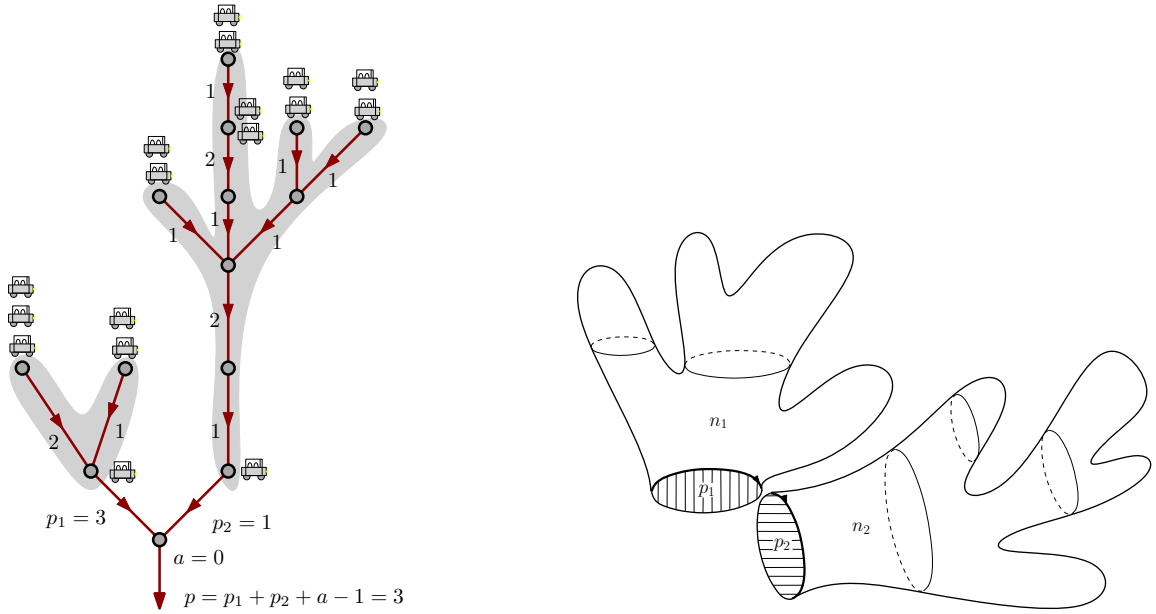


Figure 6.23: Left: Illustration of the recursive decomposition at the root of strongly parked trees to get a functional equation on $\mathbf{S}$. Right: heuristic representation of Tutte's equation in the theory of planar map enumeration.

when $x \geqslant 0$ and $y > 0$, then the above equation has solutions if $\Delta \geqslant -\mathrm{e}^{-1}$ which are

$$-W_i(\Delta) - \frac{x}{y}$$

where W_i is the ith branch of the Lambert function. There is actually a singularity and we need to change branch (see Figure 6.24), more precisely, when $x < x_c = 1/4$ we have

$$\mathbf{S}(x,y) = \begin{cases} -W_{-1}(\Delta) - \frac{x}{y} & \text{if } y \leqslant \frac{1}{2}\left(1 - \sqrt{1-4x}\right) \\ -W_0(\Delta) - \frac{x}{y} & \text{if } \left(1 - \sqrt{1-4x}\right) \leqslant y \leqslant y_c(x), \end{cases} \tag{6.55}$$

where $y_c(x)$ is the radius of convergence of the series in y when x is fixed which is the maximal solution of $\Delta = -\mathrm{e}^{-1}$. At $x = x_c = 1/4$, then $\Delta + \mathrm{e}^{-1}$ vanishes at $y_c = y_c(1/4) = 1/2$ and yields a singularity of type $(y - y_c)^{3/2}$.

6.8.2 Lackner & Panholzer's decomposition and the KP hierarchy?

Actually, there is another completely different way to get a functional equation on $\mathbf{S}$. Adapting an idea of [128] (see in particular Equation (4) there) one can decompose a strongly parked tree according to the travel of the car labeled $n + p$ (the last car) in a sequence of strongly parked trees each given with a distinguished point, see Figure 6.25. This last car decomposition yields to the following equation on $\mathbf{S}(x,y)$:

$$y\partial_y \mathbf{S}(x,y) + x\partial_x \mathbf{S}(x,y) - \mathbf{S}^\bullet(x,0) = \frac{xy\partial_x \mathbf{S}(x,y)}{1 - \mathbf{S}^\bullet(x,0)}, \tag{6.56}$$

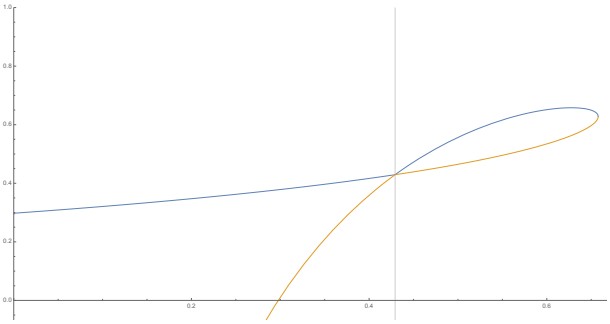


Figure 6.24: Plot of the function $y \mapsto \mathbf{S}(x,y)$ for $x = 0.245$. To get an analytic function, $\mathbf{S}$ changes from the blue to the orange branch at $y \approx 0.41$.

where $\mathbf{S}^\bullet(x,y) = x\partial_x \mathbf{S}(x,y)$ is the generating series of strongly parked trees with an additional distinguished vertex. It should be possible to solve the above equation using the method of characteristics to recover (6.55) but we have not been able to carry the calculations. Also, applying the last car

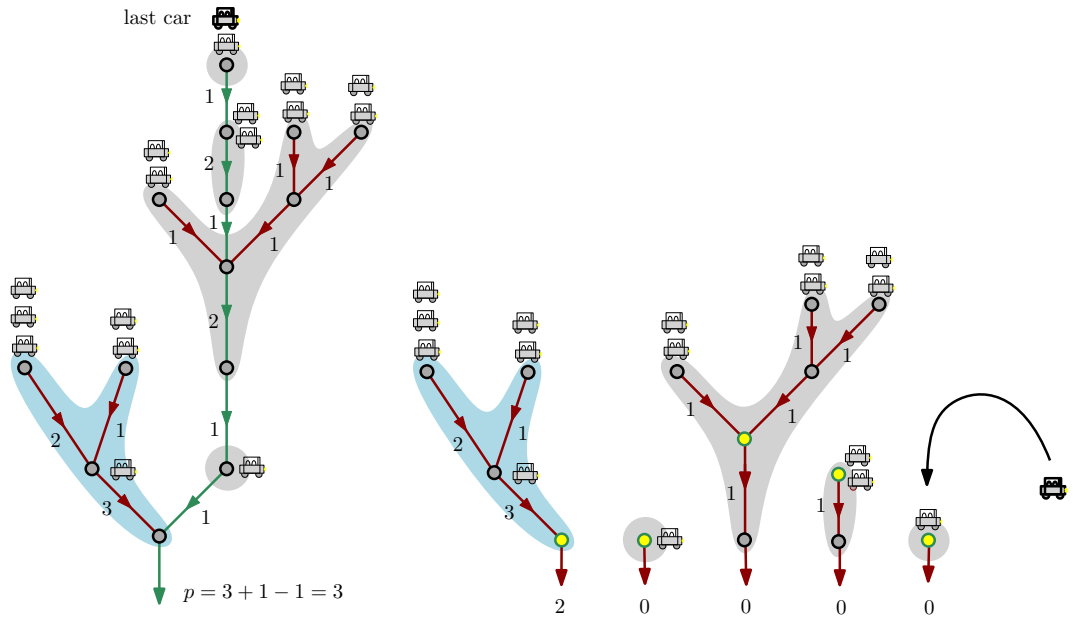


Figure 6.25: Illustration of the decomposition of a strongly parked tree according to the ride of the last car. If we remove the edges through which the last car had to go, then we end up with a sequence of strongly parked trees with distinguished points, where the last of those may have a positive flux at the root.

decomposition to the case $y = 0$ (no flux) one finds the equation

$$\mathbf{S}^\bullet(x,0) = \frac{x}{1 - \mathbf{S}^\bullet(x,0)}$$

involving $\mathbf{S}(x,0)$ only and enables us to recover (6.53) very easily. In the theory of planar maps, there are similar inductive decompositions of planar maps of size n involving planar maps of size n_1 and n_2 with $n_1+n_2 \approx n$ *without* boundary. Those decompositions are obtained via the KP hierarchy, see [101] or [134, Corollary 1] for details. We plan on adapting the "last car decomposition" of Lackner & Panholzer to the enumeration of random planar maps.

We expect that the information on the bivariate generating function $\mathbf{S}(x,y)$ will enable us to perform the asymptotic enumeration of strongly parked trees with flux and prove similar results as in the planar map setting, i.e.

$$\frac{\mathrm{SP}(n,n+p)}{n!(n+p)!} \quad \sim \quad c_1 \cdot 4^n \cdot 2^p \cdot n^{-5/2} \cdot p^{1/2} \exp\left(-c_2 \frac{p^2}{n}\right) \tag{6.57}$$

for some constants $c_1, c_2 > 0$ as long as p^2/n stays in a fixed compact interval of $(0,\infty)$. Those asymptotics are necessary to progress towards Conjecture 1 but also would be a crucial input to compute, for fixed $\lambda \in \mathbb{R}$, the exact distribution of $(\|\mathscr{FM}(\lambda)\|_\bullet, \mathscr{D}(\lambda))$. The behavior (6.57) has been observed in a great generality for a related model [59] which we now describe.

Chen's and Panholzer's generalizations of fully parked trees. Panholzer [148] studies the model of fully parked trees (with no flux) when the underlying Cayley tree is replaced by a combinatorial model such as d-ary trees, ordered trees... and he obtains remarkable explicit formulas. In particular, he finds connections with models of planar maps (OEIS A000139, or OEIS A000260 via OEIS A294084), see Remark 2 in [148]. It is natural to extend the above discussion to those models. In [59], Chen considers the enumeration of *plane* trees (rather than Cayley trees) decorated with i.i.d. (not necessarily Poisson) car arrivals conditioned to be fully parked and with a possible flux at the root. He proves a phase transition for the enumeration of such structures appearing at the same location as the phase transition for the parking process [75]. In the case of bounded car arrivals, he proves in [59, Theorem 3] an asymptotic enumeration of plane fully parked trees with flux of the form (6.57). This supports the belief that the parking processes are in the same universality classes, see below.

6.8.3 Growth-fragmentation trees and conjectural scaling limits.

Let us now give some background for Conjecture 1. By the discussion in the last paragraph, the generating series of strongly parked trees is finite at $x = x_c = \frac{1}{4}$. Hence, for each $p \geqslant 0$, we can define as in Section 6.5.1 a random strongly parked tree S_p with flux p at the root under Boltzmann critical distribution whose law is simply

$$\mathbb{P}(S_p = \mathfrak{s}_p) = \frac{1}{[y^p]\mathbf{S}(x_c,y)} \frac{1}{\|\mathfrak{s}_p\|_\bullet!(\|\mathfrak{s}_p\|_\bullet + p)!} \left(\frac{1}{4}\right)^{\|\mathfrak{s}_p\|_\bullet},$$

for each strongly parked tree $\mathfrak{s}_p$ with flux p at the root. Let us forget the labels of the vertices and the cars and see such a tree as a rooted unordered tree where each vertex is labeled by the flux of car emanating from it (so that the root has label p). It is easy to see that such trees are *Markov branching*

trees, that is, have the same law as the family tree of a system of particles evolving independently of each other. Each particle carries a non-negative integer label (the emanating flux of car from that vertex) and at each step, a particle of label p "splits" into k particles of labels $p_1, p_2, \dots, p_k$ (ordered uniformly at random) with probability

$$\frac{1}{[y^p]\mathbf{S}(x_c, y)} \frac{1}{(p - (p_1 + \cdots + p_k) + 1)!} \frac{1}{k!} \prod_{i=1}^{k} [x^{p_i}]\mathbf{S}(x_c, y).$$

When the scaling limit of the labels along a branch[9] is given by a positive self-similar Markov process, the scaling limit of those trees are described by the growth-fragmentation trees[10] of Bertoin [30]. In our case of random strongly parked trees, the labels evolve in the scaling limit as (versions of) the 3/2-stable Lévy process, exactly as for the Markov branching trees appearing the peeling exploration of random planar maps, see [32] and [31, Section 6]. To be more specific, the growth-fragmentation mechanism involved in Conjecture 1 is the one "canonically associated" to the spectrally positive 3/2-stable Lévy process i.e. with the cumulant function

$$\kappa(q) = \frac{\Gamma(q - \frac{3}{2})}{\Gamma(q - 3)}$$

for $q > 3/2$ and self-similarity index $\alpha = -3/2$, see [31, Section 5]. Our Conjecture 1 concerns *conditioned version* of those Markov branching trees, see the forthcoming work [34] for details.

6.9 Back to Erdős–Rényi

Let us now formulate a few possible consequences of our work on the classical Erdős–Rényi random graph. For this we need to generalize a little the frozen process by introducing a parameter $p \in [0, 1]$.

6.9.1 Generalized frozen process

Given the sequence of *unoriented* edges $(E_i : i \geqslant 1)$ and an independent sequence of uniform random variables $(U_i : i \geqslant 1)$ we construct a generalization of the frozen Erdős–Rényi process as follows. Fix a parameter $p \in [0, 1]$ and define a growing graph process $F_p(n, m)$ with two colors, white and blue, in a way very similar to $F(n, m)$: Initially $F_p(n, 0)$ is made of the n labeled white vertices $\{1, 2, \dots, n\}$ and for $m \geqslant 1$

- if both endpoints of E_m are white vertices then the edge E_m is added to $F_p(n, m-1)$ to form $F_p(n, m)$. If this addition creates a cycle in the graph then the vertices of its component are declared frozen and colored in blue.
- if both endpoints of E_m are blue (frozen vertices), then E_m is discarded.

[9] to be specific, one define a branch by following the locally largest label at each splitting

[10] to be precise, Bertoin defines a growth-fragmentation process from to which we can associate a continuum random tree by [154, Corollary 4.2]

- if E_m connects a white and a blue vertex, then E_m is discarded *if* $U_m > p$ and kept otherwise, in which case the new connected component is declared frozen and colored in blue.

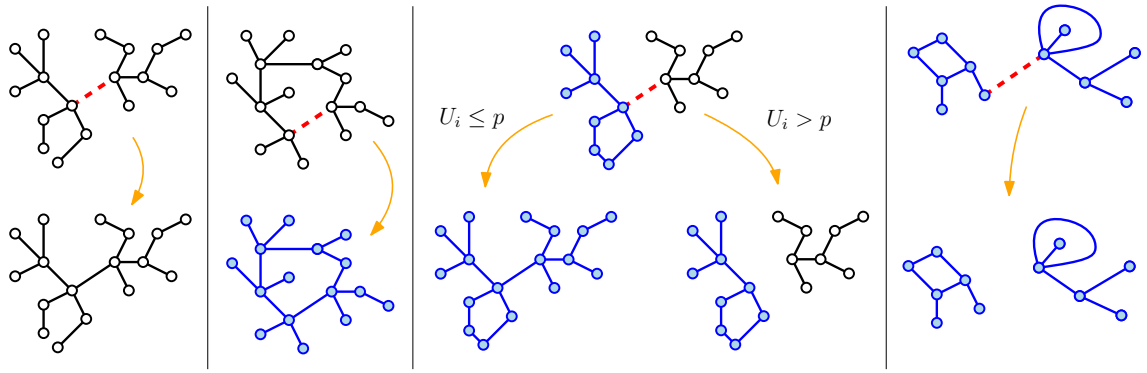


Figure 6.26: Transitions in the frozen Erdős-Rényi with parameter p.

Obviously, in the case $p = \frac{1}{2}$, the process $F_{1/2}$ has the same law as the frozen Erdős–Rényi that we used in this paper[11]. For $p = 0$, the process corresponds to completely stopping the connected components once they create a cycle. In the case $p = 1$, the process is obtained from $G(n, m)$ by discarding the edges which would create a surplus of 2. In particular, we have

$$[F_1(n, m)]_\circ = [F_1(n, m)]_{\text{tree}} = [G(n, m)]_{\text{tree}}, \quad \text{for all } m \geqslant 0. \tag{6.58}$$

Notice however, that the obvious coupling of F_p for all $p \in [0, 1]$ is not monotonic in p.

It should be easy to extend our analysis to the frozen Erdős–Rényi processes with parameter p and in particular Theorem 6.2, Propositions 6.8 and 6.15 and Corollary 2 should hold with the proper changes. E.g., the scaling limit of the rescaled total size of the frozen components in $\mathrm{F}_{n,p}(\lambda)$ should be a pure-jump Feller process $\mathscr{X}_p(\lambda)$ with jump kernel given by

$$\frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{\mathrm{d}y}{y^{3/2}}(y + 2p \cdot x)\frac{p_1(\lambda - x - y)}{p_1(\lambda - x)}. \tag{6.59}$$

Specifying those results for $p = 1$ and using (6.58) we deduce that the process of the total mass of the particles with surplus in the multiplicative coalescent $\mathscr{M}(\lambda)$ has law $\mathscr{X}_1(\lambda)$ and that conditionally on it, the remaining components are distributed as the jump of the conditioned Lévy process $\mathscr{S}^{\lambda - \mathscr{X}_p(\lambda)}$. We were not aware of such a description prior to this work.

6.9.2 Asymptotics when $\lambda \to \infty$

If the above description of the scaling limit $\mathscr{X}_p(\lambda)$ of $n^{-2/3} \cdot \|\mathrm{F}_{n,p}(\lambda)\|_\bullet$ is granted, then one can perform the analysis in the near supercritical regime $\lambda \to \infty$. It is easy to see that $\mathscr{X}_p(\lambda)$ tends to

[11] in the paper we used the orientation of the edges $\vec{E}_i$ and did not require the additional randomness of the U_i.

∞ and it is not hard to see that it is asymptotically larger than λ. Using (6.11) and the definition of the jump kernel (6.59) we can compute formally

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda}\mathbb{E}[\mathscr{X}_p(\lambda)] &= \mathbb{E}\left[\frac{1}{2}\int_0^\infty \frac{\mathrm{d}y}{\sqrt{2\pi}y^{3/2}}\left(2p\mathscr{X}_p(\lambda)+y\right)\cdot y\cdot \frac{p_1(\lambda-\mathscr{X}_p(\lambda)-y)}{p_1(\lambda-\mathscr{X}_p(\lambda))}\right]\\ &\underset{\substack{\lambda\to\infty\\ \mathscr{X}_p(\lambda)-\lambda\to\infty}}{\sim} \mathbb{E}\left[\frac{p\mathscr{X}_p(\lambda)}{\sqrt{2\pi}}\int_0^\infty \frac{\mathrm{d}y}{y^{3/2}}\cdot y\cdot \mathrm{e}^{-y(\lambda-\mathscr{X}_p(\lambda))^2/2}\right] = \mathbb{E}\left[\frac{p\mathscr{X}_p(\lambda)}{\mathscr{X}_p(\lambda)-\lambda}\right].\end{aligned}$$

From this we conjecture the asymptotic rate of growth of the process $\lambda \mapsto \mathscr{X}_p(\lambda)$ for $p \in [0,1]$

$$\frac{\mathscr{X}_p(\lambda)}{\lambda} \xrightarrow[\lambda\to\infty]{(\mathbb{P})} 1+p. \tag{6.60}$$

In particular, when $p = 0$ i.e. when we stop the cluster growth when they have a positive surplus, we believe that the total mass of the frozen part is of order λ and furthermore that $\mathscr{X}_0(\lambda) - \lambda$ converge in distribution as $\lambda \to \infty$ towards a stationary law (an example of self-organized criticality).

In the case $p = 1$ we should have $\mathscr{X}_1(\lambda) \approx 2\lambda$. This is coherent with the result of Luczak [136] saying that the largest cluster in $\mathrm{G}_n(\lambda)$ is of order $2\lambda n^{2/3}$ when $\lambda \to \infty$ (this cluster is likely to be formed by the majority of the connected components of the frozen part).

6.9.3 Process construction

In the case of the multiplicative coalescent, there has been a substantial amount of work describing the process in terms of collections of excursion lengths of evolving random functions [17, 50, 133, 139, 169]. We do not know whether such a construction is doable for our frozen processes.

6.10 Extension of parking process

In this work we used a coupling between the Erdős–Rényi random graph and uniform parking on Cayley trees to study the later. Our results obviously call for generalizations for other models of random trees and other arrival distributions of cars. The ideas of this paper can indeed be extended to cover the model of [64] and this is the subject of a forthcoming work of the first author [66]. In particular, although the precise location of the phase transition depends on the combinatorial details, we believe that the scaling limits unraveled in this paper are common to a large class of models as long as the degree distribution and the car arrivals have a sufficiently light tail. In the presence of group arrival of cars with heavy tail, new scaling limits should occur related to the different universality classes observed for component sizes in configuration models [37, 49, 62, 113].

Once Conjecture 1 has been addressed, we hope to describe a dynamical scaling limit for the geometry of the parked components as well as the flux of cars (and ideally passing our coupling construction to the scaling limit). This should involve spiraling frozen fractals all around [78]... We will try to address those questions in future works.

Table of notation

General notation

$\mathrm{X}_n(\lambda)$	$\mathrm{X}_n(\lambda) = X\left(n, \lfloor \frac{n}{2} + \frac{\lambda}{2} n^{2/3} \rfloor \vee 0\right)$ shorthand notation for a process $X(n,m)$
$\Delta X(n,m)$	$\Delta X(n,m) = X(n,m+1) - X(n,m)$ shorthand notation for the increments of a process $X(n,m)$
T_n	uniform rooted Cayley tree over $\{1,2,\dots,n\}$
$T_\star(n,m)$	for $\star \in \{\mathsf{near}, \mathsf{full}, \mathsf{strong}\}$ different types of components in T_n after parking m cars, see Figure 6.7.
$X_1, Y_1, X_2, Y_2 \dots$	independent uniform numbers over $\{1,2,\dots,n\}$ the X_i's are seen as car arrivals and are independent of T_n while the Y_i's are coupled non-trivially with T_n
$\vec{E}_i = (X_i, Y_i)$	ith oriented edge
$G(n,m)$	Erdős–Rényi random graph built by adding the first m unoriented edges
$F(n,m)$	frozen Erdős–Rényi random graph built from the first m edges
$D(n,m)$	number of discarded edges in the construction of $F(n,m)$ or equivalently of cars that did not manage to park on T_n
$W(n,m)$	uniform unrooted labeled forest with n vertices and m edges
$\mathbb{W}_n(\cdot)$	sequence of renormalized sizes of components in $\mathrm{W}_n(\cdot)$.
$\mathfrak{g}, [\mathfrak{g}]_\circ, [\mathfrak{g}]_\bullet, [\mathfrak{g}]_{\mathrm{tree}}$	a multigraph, its subgraph made of white/blue vertices, and its forest part
$\|\mathfrak{g}\|_\bullet, \|\mathfrak{g}\|_\circ, \|\mathfrak{g}\|_\bullet, \|\mathfrak{g}\|_{\bullet\text{-}\bullet}$	number of vertices, white vertices, blue vertices and edges of $\mathfrak{g}$
$\mathfrak{F}(n,m)$	unrooted forests over $\{1,2,\dots,n\}$ with m edges
$\#\mathfrak{F}(n,m)$	number of unrooted forests over $\{1,2,\dots,n\}$ with m edges
$\mu(k) = 2\mathrm{e}^{-k}\frac{k^{k-2}}{k!}$	step distribution in the random walk S coding the forests

Continuous processes notation

The random variables in the "continuous world" are usually denoted with a mathscr font.

$\mathscr{S}$	3/2-stable spectrally positive Lévy process with Lévy measure $\frac{\mathrm{d}y}{\sqrt{2\pi y^{5/2}}}$
$\mathscr{S}^u$	version of $\mathscr{S}$ conditioned on $\mathscr{S}_1 = u$
p_1, p_s	density of $\mathscr{S}$ at time 1 (Airy distribution) resp. $s \geqslant 0$
$\mathbf{n}_\lambda(x, \mathrm{d}y), g_{x,\lambda}(y)$	jump kernel, see Section 6.7.3 and (6.13)
$(\mathscr{M}(\lambda) : \lambda \in \mathbb{R})$	(standard) multiplicative coalescent
$(\mathscr{FM}(\lambda) : \lambda \in \mathbb{R})$	frozen multiplicative coalescent
$[\mathscr{FM}(\lambda)]_\bullet, \|\mathscr{FM}(\lambda)\|_\bullet$	ℓ^1- part of $\mathscr{FM}(\lambda)$ and its total mass
$[\mathscr{FM}(\lambda)]_\circ$	ℓ^2- part of $\mathscr{FM}(\lambda)$

Generating functions and counting functions

Generating functions are denoted by a mathbf symbol.

$\mathrm{PF}(n,m)$	for $0 \leqslant m \leqslant n$ number of parking functions,
	i.e. of Cayley trees of size n and m cars so that all cars park
$\mathrm{PF}_{\mathrm{root}}(n,m)$	for $0 \leqslant m \leqslant n$ number of parking functions with empty root,
	i.e. of Cayley trees of size n and m cars so that all cars park and the root stays void
$\mathrm{FP}(n,n+p)$	for $n, p \geqslant 0$ number of fully parked trees with flux p
	i.e. of Cayley tree of size n and $n+p$ cars so that exactly p cars do not park
$\mathrm{SP}(n,n+p)$	for $n, p \geqslant 0$ number of strongly parked trees with flux p
	i.e. of Cayley trees of size n and $n+p$ cars so that exactly p cars do not park
	and so that all edges have positive flux
$\mathbf{T}(x)$	Exponential GF $\sum_{n\geqslant 1} \frac{x^n n^{n-2}}{n!}$ for unrooted Cayley trees
$\mathbf{N}(x)$	Exponential GF $\sum_{n\geqslant 1} \frac{\mathrm{PF}(n,n-1)}{n!(n-1)!} x^n$ for nearly parked trees
	which is equal to $\frac{1}{2}\mathbf{T}(2x)$ by Proposition 6.9
$\mathbf{F}(x)$	Exponential GF $\sum_{n\geqslant 1} \frac{\mathrm{FP}(n,n)}{(n!)^2} x^n$ for fully parked trees
$\mathbf{S}(x)$	Exponential GF $\sum_{n\geqslant 1} \frac{\mathrm{SP}(n,n)}{(n!)^2} x^n$ for strongly parked trees
	which is equal to $1 - \ln(2) - \sqrt{1-4x} + \ln\left(1+\sqrt{1-4x}\right)$ by Proposition 6.10
$\mathbf{S}(x,y)$	Exponential GF $\sum_{n\geqslant 1} \frac{\mathrm{SP}(n,n+p)}{n!(n+p)!} x^n y^p$ for strongly parked trees with flux

Notation for Section 6.6

$\mathbf{Cadlag}(I, \mathrm{Pol})$	Space of càdlàg function from an interval $I \subset \mathbb{R}$ to some Polish space Pol
$(\mathscr{M}(\lambda) : \lambda \in \mathbb{R})$	(standard) multiplicative coalescent
$\mathcal{E} = \ell^1_{\downarrow} \times \ell^2_{\downarrow}, \quad \mathrm{d}_{\mathcal{E}}$	state space of the frozen multiplicative coalescent and its metric
$\mathcal{E}_0 = \ell^\infty_{\downarrow,0} \times \ell^\infty_{\downarrow,0}, \quad \mathrm{d}_{\mathrm{sup}}$	proxy state space of pair of decreasing sequences tending to 0 and its metric
$(\mathscr{FM}(\lambda) : \lambda \in \mathbb{R})$	frozen multiplicative coalescent
$[\mathscr{FM}(\lambda)]_\bullet, \|\mathscr{FM}(\lambda)\|_\bullet$	ℓ^1- part of $\mathscr{FM}(\lambda)$ and its total mass
$\mathscr{X}(\lambda) = \|\mathscr{FM}(\lambda)\|_\bullet$	shorthand notation for the total mass of the frozen particles
$\mathrm{O}_n(\lambda_0)$	number of vertices of $\mathrm{G}_n(0)$ whose components
	carry surplus appeared before time λ_0
$\mathrm{F}_n^{[\lambda_0]}, \mathrm{G}_n^{[\lambda_0]}$	frozen (resp standard Erdős–Rényi process) process started from time λ_0
	by removing the components with surplus at time λ_0
$\mathrm{F}_n^{[\lambda_0,\delta]}, \mathrm{G}_n^{[\lambda_0,\delta]}$	same process as above restricted to the specks
	(i.e. components of $[\mathrm{G}_n(\lambda_0)]_{\mathrm{tree}}$) of size at least $\delta n^{2/3}$
$\mathbb{F}_n, \mathbb{F}_n^{[\lambda_0]}, \mathbb{F}_n^{[\lambda_0,\delta]}$	Sequences of renormalized sizes of components (frozen followed by standard)
η-skeleton	graph spanned by the specks of $\mathrm{G}_n(0)$ of mass at least η
	or belonging to a cycle of $\mathrm{G}_n^{[\lambda_0]}(0)$
$\gamma = \gamma_n(\lambda_0, \eta)$	minimal mass of a speck of the η-skeleton

Chapitre 7 :

Parking on the infinite binary tree

Les résultats de ce chapitre sont issus de l'article [12], écrit en collaboration avec David Aldous, Nicolas Curien et Olivier Hénard et publié dans Probability Theory and Related Fields.

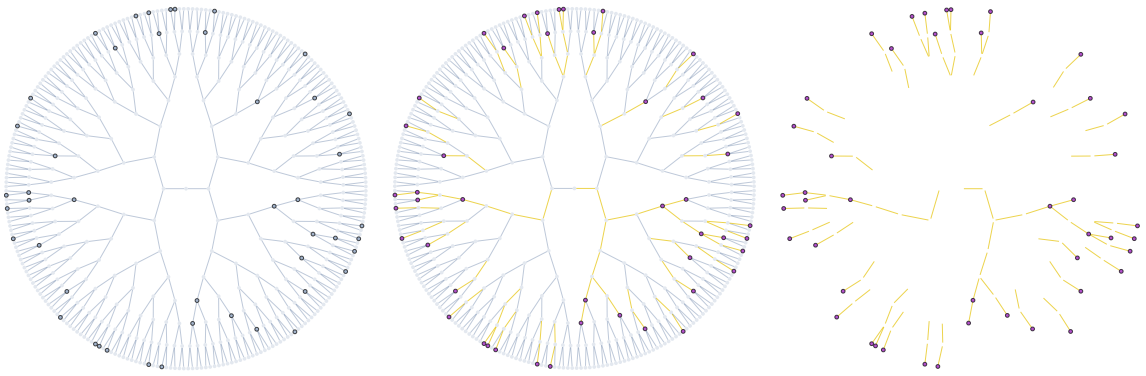


Figure 7.1: Simulation of the parking process on the first 9 levels of the full binary tree (the edges are oriented towards the origin of the tree which is at the center of the figures). The dotted vertices on the left figure initially contain 2 cars whereas the others are void. In the middle and right figures, the highlighted edges have seen a positive flux of car.

Let $(A_u : u \in \mathbb{B})$ be i.i.d. non-negative integers that we interpret as car arrivals on the vertices of the full binary tree $\mathbb{B}$. Each car tries to park on its arrival node, but if it is already occupied, it drives towards the root and parks on the first available spot. It is known [100, 22] that the parking process on $\mathbb{B}$ exhibits a phase transition in the sense that either a finite number of cars do not manage to park in expectation (subcritical regime) or all vertices of the tree contain a car and infinitely many cars do not manage to park (supercritical regime). We characterize those regimes in terms of the law of A in an explicit way. We also study in detail the critical regime as well as the phase transition which turns out to be "discontinuous".

Contents

7.1 Introduction

The parking process is a central algorithm in combinatorics and probability. When the underlying graph is an oriented line, it was first studied by Konheim & Weiss [122] in relation with hash tables and it has led to many developments in probability notably via connections with the Brownian continuum random tree and the additive coalescent [57]. Recently, Lackner & Panholzer [128] started the systematic study of the parking functions on finite rooted trees. This triggered an intense activity on the model of parking on a random *critical* Galton–Watson tree. In particular, a phase transition was proved to occur and the threshold was located in an increasing level of generality [100, 75, 64]. Furthermore a surprising connection with the Erdös–Rényi random graph and the multiplicative coalescent was unraveled in [67].

However, much less is known about the parking scheme on **supercritical** Galton–Watson trees, apart from the existence of a phase transition [100, 22] and despite an intense activity on the closely related Derrida–Retaux model [61]. The goal of this paper is to close this gap and locate and study the phase transition in the case of the parking process on the infinite binary tree (see Section 7.6 for extensions).

Parking on the infinite binary tree. Consider the full planar rooted binary tree. Its vertices can be conveniently represented by the finite words on two letters $\mathbb{B} = \cup_{n\geqslant 0}\{0,1\}^n$, with $\{0,1\}^0 = \varnothing$ being the root of the tree and with edges between the words u and $u0$ and the words u and $u1$. Those vertices will be interpreted as free parking spots, each spot accommodating at most 1 car. On top of that tree, we consider a non-negative integer labeling $(A_u : u \in \mathbb{B})$ representing the number of cars arriving on each vertex $u \in \mathbb{B}$. Each car tries to park on its arrival vertex, and if the spot is occupied, it travels downwards in direction of the root of the tree until it finds an empty vertex to park. If there is no such vertex on the path towards the root $\varnothing$, the car exits the tree, contributing to the *flux* of cars at the root. If we introduce the random variable

$$X := \text{ number of cars which have visited } \varnothing,$$

the outgoing flux of cars is then simply $\mathrm{F} = (X-1)_+ = \max(X-1, 0)$. As we will see in Section 7.2.1, the final configuration (flux and status void/occupied for the vertices), and in particular the value of variables X and F, does not depend upon the order chosen to park the cars.

In the remainder of this paper we shall suppose that the car arrivals $(A_u : u \in \mathbb{B})$ are i.i.d. with a given distribution $\mu = (\mu_k : k \geqslant 0)$ on $\{0,1,2,3,\ldots\}$. To avoid trivialities, we always suppose that $\mu(\{0,1\}) < 1$ for otherwise the cars would always park on their arrival node. We let

$$G(x) = \sum_{k\geqslant 0} \mu_k x^k$$

be the generating function of the law μ. One can then establish a dichotomy (see Lemma 7.1 and also [100, 22] as well as Proposition 7.1 below):

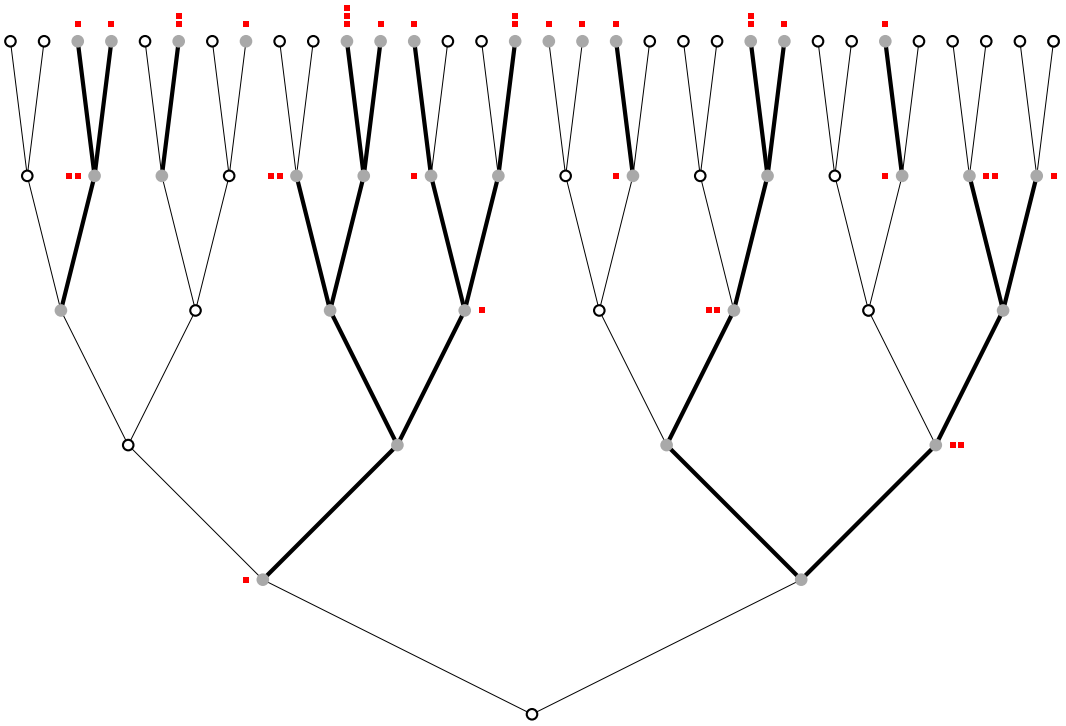


Figure 7.2: Illustration of the parking process in the first 5 levels of the full binary tree. The car arrivals are represented by red squares (including the cars that may come from higher levels on top of the tree). After the parking process, the vertices accommodating a car are displayed in gray, whereas the free spots are displayed in white. The connected components of parked cars are drawn with thick lines, they are *fully parked trees*.

- Either the number X of cars that visited the root $\varnothing$ has a finite mean and all clusters of parked vertices are finite almost surely, we call this phase the **subcritical** regime,
- Or almost surely $X = \infty$ and actually, all vertices of $\mathbb{B}$ are occupied after the parking process, we call this phase the **supercritical** regime.

We shall furthermore distinguish the **critical** regime, when it is not possible to stochastically increase μ and stay subcritical. A first trivial remark is that when $\mathbb{E}[A] > 1$ the process is necessarily supercritical (since there are more cars than parking spots on average). Our main result is then a characterization of those regimes explicitly in terms of the generating function G of μ:

Theorem 7.1 (*Location of the phase transition*)

Suppose that there exists $t_c \in (0, \infty)$ such that

$$t_c = \min\{t \geqslant 0 : \ 2(G(t) - tG'(t))^2 = t^2 G(t) G''(t)\} \qquad (\star).$$

Then the parking process is subcritical if and only if

$$(t_c - 2)G(t_c) \geqslant t_c(t_c - 1)G'(t_c). \tag{7.1}$$

The condition $(\star)$ on the existence of t_c is mild and is for example verified for all generating functions with infinite radius of convergence (see Remark 7.4.2). When $(\star)$ is not verified, we provide a method to check if we are in the subcritical phase, see Section 7.5.3. Checking the signs of the two sides of the inequality (7.1), see Remark 7.5.1, we see in particular that the generating function G must have a radius of convergence at least 2 to be in the subcritical phase. This can easily be explained by probabilistic arguments: otherwise, the maximum of 2^n independent copies of a random variable with law μ is larger than n with high probability, so that the car arrivals at the single level n of $\mathbb{B}$ suffice to guarantee that the root $\varnothing$ is occupied, see Lemma 7.1. The same argument actually even proves that the radius of convergence of X (which is stochastically larger than A) must stay above 2 in the subcritical regime. Notice that deciding whether μ is subcritical for parking depends in a subtle way on the distribution as opposed to the case of critical Galton–Watson trees [100, 75, 64] where its depends only on the first two moments.

Let us give a couple of examples of application of our theorem in the case of a car arrival distribution that is parametrized by a family $(\mu_\alpha : 0 \leqslant \alpha)$ which is stochastically increasing with mean $\alpha \geqslant 0$: in this case the parking is subcritical if $\alpha \leqslant \alpha_c$ and supercritical if $\alpha > \alpha_c$, for some threshold α_c depending on the family of laws:

- *Binary$_{0/2}$ arrivals.* If $\mu_\alpha = (1 - \frac{\alpha}{2})\delta_0 + \frac{\alpha}{2}\delta_2$, then the critical threshold is
$$\alpha_c(\text{Binary}_{0/2}) = \frac{1}{14}.$$
This settles an open problem of Bahl, Barnet & Junge [22]. Obviously the value of α_c is in agreement with the bounds $\frac{1}{32} \leqslant \alpha_c \leqslant \frac{1}{2}$ of [100, Proposition 3.5] and improved to $0.03175 \leqslant \alpha_c \leqslant 0.08698$ in [22, Proposition 4].

- *Binary$_{0/k}$ arrivals.* More generally if $\mu_\alpha = (1 - \frac{\alpha}{k})\delta_0 + \frac{\alpha}{k}\delta_k$ for some $k \in \{2, 3, 4, \ldots\}$, then the threshold is
$$\alpha_c(\text{Binary}_{0/k}) = \frac{k}{1 + 2^{-k-2}\left(3 + \sqrt{\frac{k+7}{k-1}}\right)^k \left((k-1)(k+4) + k\sqrt{(k+7)(k-1)}\right)}.$$

- *Poisson arrivals.* If μ_α is Poisson with mean α, then
$$\alpha_c(\text{Poisson}) = 3 - 2\sqrt{2}.$$

- *Geometric arrivals.* If $\mu_\alpha = p^k(1-p)$ for $k \geqslant 0$ is a geometric law with mean $\alpha = \frac{p}{1-p}$ then $p_c(\text{Geometric}) = 1/9$ and
$$\alpha_c(\text{Geometric}) = \frac{1}{8}.$$

The critical regime. Let us now focus more precisely on the critical regime : we assume that μ is subcritical (and $(\star)$ holds) and that it is not possible to stochastically increase μ while remaining subcritical. As we shall see in Section 7.5, this means that the inequality in (7.1) is actually an

equality. Recall that we denoted by X the number of cars that *visited* the root $\varnothing$ of $\mathbb{B}$ during the parking process. We set for $k \geqslant 0$, $p_k = \mathbb{P}(X = k)$ and will use the shorthands $p_\circ = p_0$ and $p_\bullet = p_1$ for respectively the probability that the root is void and the probability that the root is at the bottom of a parked cluster without flux, after parking. In the following, we shall call *white* the clusters of void vertices, and *black* the clusters of parked vertices.

Theorem 7.2 (*Critical computations*)

Suppose that $(\star)$ *holds and that* (7.1) *is an equality. Then almost surely the root* $\varnothing$ *is void or it belongs to a finite black cluster, and we have*

$$p_\circ = \frac{t_c^2}{4(t_c-1)G(t_c)} \quad \text{and} \quad p_\bullet = \sqrt{\frac{p_\circ}{\mu_0}} - p_\circ.$$

These calculations have a few surprising consequences:

- $\mathbb{E}[X] < \infty$. The fact that the expectation of the flux of cars is finite in the whole subcritical regime (including at criticality) may be surprising at first, but this can actually be seen from the recursive distributional equation satisfied by X by splitting at the root of $\mathbb{B}$

$$X \overset{(d)}{=} (X_1 - 1)_+ + (X_2 - 1)_+ + A, \tag{7.2}$$

  where X_1, X_2 are two copies of law X independent of the car arrivals A of law μ. Indeed, the RHS has expectation at least $2\mathbb{E}[X] - 2$ which is strictly larger than $\mathbb{E}[X]$ as soon as $\mathbb{E}[X] > 2$. Iterating the argument, one sees that there is no a.s. finite solution to the above recursive distributional equation which has a mean > 2, see [22, Theorem 1.1] for details. Actually, as we already mentioned the variable X must have a radius of convergence larger than 2, even at the critical point, see the forthcoming Lemma 7.1. Also, plugging the value of $p_\circ = \mathbb{P}(X = 0)$ into (7.2) we deduce that

$$\mathbb{E}[X] + \mathbb{E}[A] = 2(1 - p_\circ), \qquad \text{or equivalently} \qquad \mathbb{E}[\mathrm{F}] + \mathbb{E}[A] = 1 - p_\circ$$

  Now, by Remark 7.5.2, on the subcritical regime we have $p_\circ > \frac{1}{2}$, so that the LHS of the left identity is bounded by 1 and the LHS of the right identity by $1/2$; one can also show that these bounds are sharp.

- It will follow from our combinatorial decomposition that the clusters of void vertices are actually Bienaymé–Galton–Watson trees with offspring distribution ξ given by

$$\mathbb{P}(\xi = 0) = \frac{p_\bullet^2}{(p_\circ + p_\bullet)^2}, \quad \mathbb{P}(\xi = 1) = \frac{2p_\bullet p_\circ}{(p_\circ + p_\bullet)^2}, \quad \mathbb{P}(\xi = 2) = \frac{p_\circ^2}{(p_\circ + p_\bullet)^2}.$$

  Again, since by Remark 7.5.2 we have $p_\circ > \frac{1}{2}$, those trees are **supercritical**, implying that at criticality there are (infinitely many) infinite white clusters. On the contrary, we shall see in Proposition 7.1 that in the subcritical regime (including the critical case), there are no infinite black clusters.

Those phenomena underline once again the fact that the phase transition in the parking process is *discontinuous*, in contrast e.g. with the case of Bernoulli percolation on the same tree[1].

Fully parked trees and their enumeration. The proofs of our main results rely on a simple combinatorial decomposition into clusters of parked vertices and the enumeration thereof. More precisely, a fully parked tree $\mathbf{f}$ is a subtree of $\mathbb{B}$ containing the root, decorated with car arrivals, so that all those cars manage to park on $\mathbf{f}$ and that reciprocally all vertices of $\mathbf{f}$ are parked. If $F \equiv F_\mu(x)$ is the generating function of fully parked trees counted with a weight x per vertex and incorporating the μ-weight of car arrivals, see Section 7.3.1 for the precise definition, then the high-level idea of the proof is to write the fixed point equations for $p_\circ$ and $p_\bullet$, which are

$$p_\circ = \mu_0(p_\circ + p_\bullet)^2 \quad \text{and} \quad p_\bullet = p_\circ F_\mu(p_\circ), \tag{7.3}$$

and translate the idea that we decompose the structure into the (finite) clusters of parked vertices. Theorem 7.1 boils down to deciding whether we have a non trivial solution $p_\circ \neq 0$ to these equations (otherwise we are in the supercritical regime). The critical regime corresponds to the case where $p_\circ$ is exactly the radius of convergence of F. Thus the main ingredient in the proof is the "computation" of the generating function F. The enumeration of fully parked trees has already been considered in the combinatorics literature [128, 59, 119, 67, 63] and it shares many similarities with the enumeration of planar maps. The idea is to enumerate a more complicated structure, namely fully parked trees with a possible flux of cars at the origin. Those are defined as fully parked trees, except that now the number of cars may be larger than the number of vertices of the tree so that the number of cars X visiting the root of the tree may be strictly larger than 1. If $F \equiv F_\mu(x, y)$ is the generating function of fully parked trees with weight x per vertex and y per outgoing car, then writing a recursive decomposition at the root vertex we obtain

$$F(x, y) = \frac{x}{y}\left(\big(1 + F(x, y)\big)^2 G(y) - \big(1 + F(x, 0)\big)^2 G(0)\right). \tag{7.4}$$

These equations are reminiscent of Tutte's equation [166] in the realm of planar maps where the perimeter of the external face plays the role of our outgoing flux of cars. In this equation, the variable y is called the *catalytic variable* since its role is to disappear to recover $F(x, 0) \equiv F(x)$, the generating function of fully parked trees with no flux. We apply the standard kernel method [44] to solve those equations, see Section 7.4 for details.

Once we have sufficient information on F, the proofs of our main results are rather straightforward. Deciding whether $p_\circ$ is non trivial boils down to an inequality on $F(x_c, 0)$ at its radius of convergence x_c, see Proposition 7.2. Under the assumption $(\star)$, this inequality is equivalent to (7.1) and the critical case corresponds to the case when $p_\circ$ coincides with the radius of convergence of $F(\cdot, 0)$. Furthermore, in the subcritical case the generating function of the outgoing flux of cars is given by $p_\circ F(p_\circ, y)$ (see (7.6)).

[1] in this model, the probability that the root of the full binary tree belongs to an infinite component is a continuous function of the percolation probability

Growth-fragmentation trees. It will follow from our decomposition that conditionally on $X = 1$, i.e. on $\varnothing$ being the root of a fully parked tree with no flux, then the cluster of parked cars above $\varnothing$ is a random fully parked tree whose size has generating function $F(p_\circ z)/F(p_\circ)$. In the critical regime, since $p_\circ$ corresponds to the radius of convergence of F, the tail of the cluster size has a subexponential decay and in the generic situation (e.g. when the car arrivals have bounded support), we actually have

$$\mathbb{P}(\varnothing \text{ is a the root of a parked cluster of size } n) \sim \mathsf{cst} \cdot n^{-5/2}, \tag{7.5}$$

the exponent $5/2$ being common in the theory of map enumeration. Furthermore, we also believe that in the generic situation, rescaled large fully parked trees converge after normalization towards the growth-fragmentation trees that already appeared in the study of scaling limits of random planar maps and the Brownian sphere, see [31, 32, 130] or [72, Chapter 14.3.2]. We already made a similar conjecture for the scaling limits of parked components in the parking process on large uniform Cayley trees [67, Conjecture 1]. It is interesting to notice that although the phase transition in the parking on $\mathbb{B}$ is of a different flavor (the phase transition in the case of critical Galton–Watson trees is "continuous"), the large scale geometry of the critical components should be the same. However, there are non-generic situations (with specific car arrivals distributions having heavy-tail) where (7.5) does not hold and where we expect different scaling limits. See Section 7.6 and [59] for a similar phenomenon in the case of enumeration of non-binary plane fully parked trees. We plan to address those questions in following works.

Relationship with the literature. In deriving soft arguments about the existence of a phase transition (section 7.2.2), we do reuse (with acknowledgment) some ideas of [22] and [100], mostly to offer a self-contained account; yet the meat of this work is the derivation of explicit formulae (meaning expressions in term of μ) for the probability that the root of the tree is parked, hence for the localisation of the phase transition: more precisely, the key equation (7.3) is derived in Section 7.3.2 and solved in Section 7.4.1 using the explicit solution of the Tutte equation (7.4) satisfied by fully parked trees with an additional flux. Those trees are interesting on their own and have been investigated from a pure analytic combinatorics viewpoint in the work [59]. In fact, [59] was motivated by [75], and [59] together with this paper were conceived following discussions between their authors. In a sense, the contribution of this paper is to bridge the gap between [75] an [22], using methods from analytic combinatorics to give insight on more probabilistic quantities.

Acknowledgments. A.C and N.C. acknowledge the support from ERC 740943 GeoBrown. Part of this work was initiated during a conference in CIRM and we thank our host for its hospitality.

7.2 Background

In this section we formally present the parking process on $\mathbb{B}$ and gather a few "rough" probabilistic results (mostly adapted from [100, 22]).

7.2.1 Parking on infinite trees

Let τ be a rooted locally finite (plane) tree decorated with car arrivals $(a_u : u \in \tau)$. As described in the introduction, cars try to park on their arrival node, and if the spot is taken they travel downwards in search of the first empty spot and, in case there is no such spot, exit at the root. In the case τ is finite, an easy Abelian property shows that the number of cars visiting each vertex of the tree does not depend on the order in which we park the cars, see Section 2.1 of [128].

On infinite trees, to prevent cumbersome issues, we shall stick to a given parking procedure: *park the lowest cars first*. More precisely, for each $n \geqslant 0$ let us consider the finite tree $[\tau]_n$ made of the first n generations above the root $\varnothing$ (recall that τ is supposed locally finite) together with the restriction of the car arrivals on these vertices. We can then perform the parking on $[\tau]_n$ and construct variables

$$x_n(u) \quad u \in [\tau]_n,$$

representing the number of cars that visited the vertices of $[\tau]_n$ in the parking process (recall that those variables do not depend on the order in which we parked the cars on $[\tau]_n$). Notice that for a given vertex $u \in \tau$, the function $n \mapsto x_n(u)$ is non-decreasing (it is defined for n larger than the height of u) so that we can let $n \to \infty$ and define

$$x(u) = \lim_{n\to\infty} x_n(u),$$

as the limiting number of cars visiting u in the parking process on τ. This morally corresponds to parking the lowest cars first[2]. In particular we say that u is void if $x(u) = 0$, that u is occupied if $x(u) \geqslant 1$ and the flux of outgoing cars at u is $f(u) = (x(u) - 1)_+$.

7.2.2 Rough phase transition

We now focus on the case of the binary tree $\mathbb{B}$ with i.i.d. car arrivals $(A_u : u \in \mathbb{B})$ with law μ satisfying $\mu(\{0,1\}) < 1$. We denote by $X(u)$ the number of cars that visited vertex $u \in \mathbb{B}$ as defined in the preceding section and will use the shorthand notation $X = X(\varnothing)$. We first establish a dichotomy on X in the next lemma, which we then interpret in more geometric terms by proving that there cannot be infinite black clusters with a finite flux.

Lemma 7.1 (Dichotomy subcritical/supercritical)**.** *We have the following dichotomy:*
Subcritical case. *Either the sequence* $(2^n\mathbb{P}(X > n) : n \geqslant 0)$ *is bounded.*
Supercritical case. *Or* $X = \infty$ *a.s, in which case all vertices are parked a.s.*

Proof. Assume that $(2^n\mathbb{P}(X > n) : n \geqslant 0)$ is not bounded, and observe that the same is then true of the sequence $(2^n\mathbb{P}(X > n + k) : n \geqslant 0)$ for any integer k. Then consider the collection of the 2^n i.i.d. variables $X(u)$ attached to the vertices u of $\mathbb{B}$ at height n. We have the upper bound

$$\mathbb{P}\Big(\bigcap_{u:|u|=n} \{X(u) \leq n + k\}\Big) \leq \mathrm{e}^{-2^n\mathbb{P}(X>n+k))},$$

[2]In fact, we could equivalently fix an exhaustion of τ by finite trees $\tau_1 \subset \tau_2 \subset \cdots$ and define the parking on τ as the limit of the parking procedure over the τ_n's.

with the right-hand side going to 0 along a subsequence. But on the complement of the event on the left-hand side, one of the variables $X(u)$ is strictly larger $n+k$ and this contribution only suffices to imply that $X = X(\varnothing) \geqslant k$. Combined with our assumption, this implies that, almost surely, $X \geqslant k$, hence, k being arbitrary, $X = \infty$. This means that the root of $\mathbb{B}$ almost surely contains a car, and it is the same for any other vertex.

□

The next lemma says that the above dichotomy is equivalent to the existence of infinite black clusters. In particular, it rules out the possibility of having an infinite black cluster and a finite flux.

Proposition 7.1. *In the subcritical regime, there is no infinite black cluster.*

Proof. Suppose that μ is subcritical, so that all variables $X(u)$ are finite after the parking process. It suffices to prove that the probability that the cluster of the origin $\mathcal{C}(\varnothing)$ is infinite is 0. Fix $p \geqslant 0$ and let us consider the event $\mathcal{E} = \{X(\varnothing) = p \text{ and } \mathcal{C}(\varnothing) \text{ is infinite}\}$. We shall explore the process by parking on the first n levels of $\mathbb{B}$ as in the preceding section. More precisely, let $\mathcal{F}_n$ be the sigma field generated by the variables $X_n(u)$ for $X_n(u)$ the number of cars visiting the vertex u when restricting the parking on $[\mathbb{B}]_n$. We then construct a sequence of stopping times $\theta_1 < \theta_2 < \cdots$ obtained as follows: $\theta_1 = \inf\{n \geqslant 0 : X_n(\varnothing) = p\}$ and then by induction $\theta_{i+1} = \inf\{n > \theta_i : \varnothing \leftrightarrow \partial[\mathbb{B}]_n\}$ where $\varnothing \leftrightarrow \partial[\mathbb{B}]_n$ means that $\varnothing$ is connected to the level n by a path whose vertices satisfy $X_n(u) \geqslant 1$. A moment's though shows that on the event $\mathcal{E}$ all these stopping times are finite for otherwise the black cluster of the origin would not be infinite. For $n \geqslant 2$, on the event $\{\theta_n < \infty\}$, let v_n be the (first, for definiteness) vertex of $\partial[\mathbb{B}]_{\theta_n}$ to be connected to the root when parking on $[\mathbb{B}]_{\theta_n}$. Set $\mathcal{D}_n = \{A_{v_n 0}, A_{v_n 1} \in \{0,1\}\}$ the event that the two children of v_n have car arrivals $\leqslant 1$. Plainly, $\mathcal{E} \subset \mathcal{D}_n \cap \{\theta_n < \infty\}$ since otherwise, the flux coming from these two vertices would go down all the way to $\varnothing$ and we would have $X(\varnothing) > p$. In particular we have

$$\mathbb{P}(\mathcal{D}_n \mathbf{1}_{\theta_n < \infty} \mid \mathcal{F}_{\theta_n}) = (\mu_0 + \mu_1)^2 < 1.$$

Notice then that $\mathcal{D}_1, ..., \mathcal{D}_{n-1}$ are $\mathcal{F}_{\theta_n}$-measurable so that by induction we have

$$\mathbb{P}(\mathcal{E}) \leqslant \mathbb{P}(\theta_n < \infty) \prod_{k=2}^{n} \mathbb{P}(\mathcal{D}_k | \theta_n < \infty) \leqslant (\mu_0 + \mu_1)^{2(n-1)},$$

which implies $\mathbb{P}(\mathcal{E}) = 0$ since n is arbitrary and we assumed the distribution μ satisfies $\mu_0 + \mu_1 < 1$.

□

As a consequence of the (proof of) Lemma 7.1, there is no lower bound for $\mathbb{E}[A]$ for supercritical parking, since one may cook up distributions μ with arbitrarily small expectation but $\mathbb{E}[2^A] = \infty$. However, if the car arrival distribution is bounded, one can obtain a lower bound for the expectation $\mathbb{E}[A]$ of the car arrival distribution for supercritical parking using a first moment method, see [100, Proposition 3.5] and [22] for details.

To speak of a phase transition, one may imagine a family $(\mu^\alpha)_{\alpha \geqslant 0}$ of car arrival distributions that is stochastically increasing in the mean α. In this case, the subcritical phase is identified with

a *closed set* $\alpha \in [0, \alpha_c]$, and the supercritical phase with the set $]\alpha_c, 1]$. The fact that α_c is actually subcritical (i.e. satisfies the first alternative of the dichotomy) can be seen by monotone convergence since the expectation of the flux is bounded above by 2 in the whole subcritical phase as recalled in the introduction (see [22] for details).

7.3 Decomposition into fully-parked components

In this section we present our combinatorial decomposition which underlies our main results. The idea is very simple: we decompose the final configuration on $\mathbb{B}$ into the black clusters of parked vertices and the white empty vertices. This shows that we can decompose the final configuration as a two-type Bienaymé–Galton–Watson tree whose offspring distribution is related to the generating function F of fully parked trees studied in detail in the next section.

7.3.1 Fully parked trees

Suppose that we performed the parking process on $\mathbb{B}$, and recall that the black vertices are those $u \in \mathbb{B}$ satisfying $X(u) \geqslant 1$, the other ones being the empty or white vertices. The *finite* black connected components are *fully parked trees* $\mathbf{t}$, i.e. connected subsets of the binary tree decorated by car arrivals $(a_u)_{u \in \mathbf{t}}$ such that after parking all vertices are occupied. If such a tree appears as the black component of the root $\varnothing$, then the fully parked tree may have an outgoing flux at the root (i.e. containing more cars than vertices), otherwise it contains as many cars as vertices. See Figure 7.3.

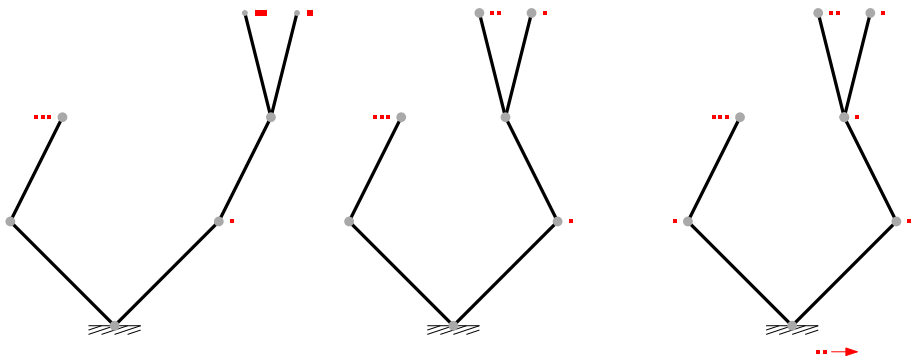


Figure 7.3: Three examples of fully parked trees. Notice that the first two have no outgoing flux and represent the same plane tree but their embeddings in $\mathbb{B}$ is different. The last fully parked tree on the right has an outgoing flux of 2 cars.

For the enumeration of the fully parked trees we shall always consider that their bottom vertex is $\varnothing$. Each plane rooted structure of a fully parked tree with m vertices with 1 child actually corresponds to 2^m different embeddings as a subset of $\mathbb{B}$ (with $\varnothing$ as the root): for this reason, later in the decomposition we shall put a weight of 2 for vertices with outdegree 1. Let us denote by $\mathbb{T}_n^{(p)}$

the set of all fully parked trees with root $\varnothing$, with n vertices and having outgoing flux $p \geqslant 0$ (i.e. $p+1$ cars have visited the root vertex). The weight $w(\mathbf{t})$ of a fully parked tree $\mathbf{t} \equiv (\mathbf{t} : (a_u : u \in \mathbf{t}))$ is the weight of its car decoration, that is

$$w(\mathbf{t}) = \prod_{u \in \mathbf{t}} \mu_{a_u}.$$

We can then form the bivariate generating series of fully parked trees (with flux) as

$$\begin{aligned} F(x,y) \equiv F_\mu(x,y) &:= \sum_{n \geqslant 1} \sum_{p \geqslant 0} \sum_{\mathbf{t} \in \mathbb{T}_n^{(p)}} x^n y^p w(\mathbf{t}) \\ &= x\big(\mu_1 + \mu_2 y + \mu_3 y^2 + \cdots\big) + x^2\big(2(\mu_2 + \mu_1^2) + 2y(\mu_3 + 2\mu_1\mu_2) + \cdots\big) + \cdots \end{aligned}$$

Section 7.4 is devoted to the study of F via a functional equation obtained by splitting a fully parked tree at the root, see (7.11). But before doing so, let us present the combinatorial decomposition and the characterization of subcriticality in terms of F. It turns out that most equations simplify if one introduces

$$\mathbf{F}(x,y) := 1 + F(x,y) \quad \text{and} \quad \mathbf{F}_0(x) := \mathbf{F}(x,0) := 1 + F(x,0).$$

7.3.2 Decomposition

Recall that the law of X is $(p_k : k \geqslant 0)$ and that we gave a short-hand notation $p_\circ = p_0$ for the probability that the root vertex is empty. We write $\mathcal{C}(\varnothing)$ for the monochromatic cluster of the origin in $\mathbb{B}$ after parking. Notice that the number of vertices adjacent from above to a fully parked tree $\mathbf{t} \subset \mathbb{B}$ with n vertices is $n+1$, regardless of the shape of $\mathbf{t}$. Recalling Proposition 7.1 we have for $k \geqslant 0$

$$\begin{aligned} \mathbb{P}(X(\varnothing) = k+1) &\underset{\text{Prop.7.1}}{=} \mathbb{P}(X(\varnothing) = k+1 \text{ and } \#\mathcal{C}(\varnothing) < \infty) \\ &= \sum_{\mathbf{t} \in \mathbb{T}_n^{(k)}} \mathbb{P}(\mathcal{C}(\varnothing) = \mathbf{t}) \\ &= \sum_{n \geqslant 1} \sum_{\mathbf{t} \in \mathbb{T}_n^{(k)}} w(\mathbf{t}) p_\circ^{n+1} = p_\circ [y^k] F(p_\circ, y). \end{aligned} \tag{7.6}$$

The other fundamental equation is obtained by noticing that the event $\{X(\varnothing) = 0\}$ occurs if and only if $A_\varnothing = 0$ and $\{X(u) \in \{0,1\} : \text{ for } u \in \{0,1\}\}$ which turns into

$$p_\circ = \mu_0 (p_\circ + p_\bullet)^2. \tag{7.7}$$

Specializing (7.6) to $k = 0$, we recover together with the previous display the fixed point equation $p_\bullet = p_\circ F(p_\circ, 0)$, mentioned in the introduction. In particular, re-injecting in (7.7) we obtain

$$p_\circ = \mu_0 p_\circ^2 \big(\mathbf{F}_0(p_\circ)\big)^2, \quad \text{where we recall that } \mathbf{F}_0(x) = 1 + F(x,0).$$

Notice that the function $x \mapsto \mu_0 x^2 (\mathbf{F}_0(x))^2$ is strictly convex and that $p_\circ = 0$ is a trivial solution to the above equation, so there is **at most one positive solution** $p_\circ$. Under the same hypothesis, splitting according to the values of $X(\varnothing)$ we also obtain thanks to (7.6)

$$1 = p_\circ + p_1 + p_2 + \cdots \underset{(7.6)}{=} p_\circ \mathbf{F}(p_\circ, 1). \tag{7.8}$$

Proposition 7.2 (F-characterization of subcriticality)**.** *The law μ is subcritical if and only if there is a positive solution to the equation*

$$1 = \mu_0 x(\mathbf{F}_0(x))^2. \tag{7.9}$$

Proof. Let μ be a subcritical law for the parking on $\mathbb{B}$. Since $p_\circ \neq 0$, the above calculations show that $p_\circ$ is indeed a solution to the equation (7.9).

Conversely, suppose that there is a positive solution $x_\circ$ to (7.9). As a special case of equation (7.11) below for $\mathbf{F}(x,y)$, we know that the series $\mathfrak{f}(y) = \mathbf{F}(x_\circ, y)$ is a solution to $y + \mathfrak{f}^2 x_\circ G(y) - y\mathfrak{f} - 1 = 0$. Solving the quadratic equation and taking the combinatorial solution we have

$$\mathfrak{f}(y) = \frac{y + \sqrt{y^2 + 4x_\circ(1-y)G(y)}}{2x_\circ G(y)}.$$

At first, the above equality holds only as a formal power series in y. But notice that the function $y^2 + 4x_\circ(1-y)G(y)$ inside the square-root does not vanish over $y \in [0,1]$ so that the solution above is analytic over $[0,1]$. By Pringsheim's theorem [92, Theorem IV.6 p.240], the function $\mathfrak{f}$ has radius of convergence at least 1 and we have $\mathfrak{f}(1) = \frac{1}{x_\circ}$ which is $x_\circ \mathbf{F}(x_\circ, 1) = 1$. This in turn ensures that there exists a random variable Z (the outgoing flux of cars) whose generating function is

$$\mathbf{Z}(y) = x_\circ \mathbf{F}(x_\circ, y).$$

We then compute, using Tutte's equation (7.4) (see (7.11) below) as well as (7.9):

$$\begin{aligned}
&\frac{1}{y}\left(\mathbf{Z}(y)^2 G(y) - \mathbf{Z}(0)^2 G(0)\right) + \mathbf{Z}(0)^2 G(0)\\
=&x_\circ \left(\frac{x_\circ}{y}\left(\mathbf{F}(x_\circ, y)^2 G(y) - \mathbf{F}(x_\circ, 0)^2 G(0)\right)\right) + x_\circ^2 \mathbf{F}(x_\circ, 0)^2 G(0)\\
=&x_\circ F(x_\circ, y) + x_\circ = \mathbf{Z}(y),
\end{aligned}$$

but this identity is equivalent to the following recursive distributional equation for Z:

$$Z \overset{(d)}{=} (Z_1 + Z_2 + A - 1)_+,$$

where on the right-hand side the variables are independent and Z_1, Z_2 are two copies of law Z.

This recursive distributional equation enables us to decorate the vertices of $\mathbb{B}$ by i.i.d. variables A_u in such a way that for every n, the parking on $[\mathbb{B}]_n$ together with i.i.d. fluxes on $\partial[\mathbb{B}]_n$ yields a flux of law Z at the root (in a coherent manner). Replacing the i.i.d. fluxes on $\partial[\mathbb{B}]_n$ by null fluxes on $\partial[\mathbb{B}]_n$, and writing $X_n = X([\mathbb{B}]_n)$ for the number of cars visiting the root for the parking on $[\mathbb{B}]_n$, we deduce by comparison that the flux at the root $\mathrm{F_n} = (X_n - 1)_+$ is dominated by the a.s. finite random variable Z, which implies in particular that the parking on $\mathbb{B}$ with law μ is subcritical. Et voilà.

□

7.4 Enumeration of fully parked trees

This section is the analytic core of the paper. We write the recursive equations (Tutte's equation) for fully parked trees and solve them using the kernel method of Bousquet-Mélou & Jehanne [44]. Combined with Proposition 7.2 this enables us to prove our main results easily. The results are similar to the work of Chen [59] which considered plane fully-parked trees (as opposed to our binary case). Notice also that the technical part of [59] consists in obtaining asymptotics for the coefficients, a goal that we did not pursue in these pages.

7.4.1 Solving Tutte's equation

Recall that $F(x,y)$ is the bivariate generating function of the fully parked trees where x encodes the number of vertices of the tree and y the **flux** of cars and G is the generating function of the car arrivals. To enumerate fully parked trees, we decompose them at their root vertices. Take a fully parked tree with $n \geqslant 1$ vertices and $n+p$ cars in total (the flux of cars is p). Then

- either $n = 1$ which means that the root vertex has no vertex above it. In this case, at least one car arrives on this vertex (since the root vertex should contain a parked car) and the number of cars arriving on this vertex is $1+p$. Summing over p gives the term $x(G(y)-G(0))/y$.
- Another possibility is that the root vertex has a unique child in the fully parked tree, which can be the left or right neighbor in $\mathbb{B}$. In that case, the subtree above this child is a fully parked tree with $n-1$ vertices and a flux of cars p_1 where $p_1+\ell-1=p$ if there are ℓ cars arriving on the initial root vertex. Notice that the case $p_1=\ell=0$ is excluded since otherwise the root vertex is not parked. Summing over p yields the term $2x(F(x,y)G(y)-F(x,0)G(0))/y$.
- The last case is when the root vertex has two parked children, each carrying a fully parked tree above it with respective sizes $k \geqslant 1$ and $n-k-1 \geqslant 1$ and flux of cars p_1 and p_2. To obtain a flux of cars p at the root, one must have $p_1+p_2+\ell-1=p$ where ℓ is the number of cars arriving at the root vertex. Again the case $p_1=p_2=\ell=0$ is excluded. We thus obtain a term $x(F(x,y)^2G(y)-F(x,0)^2G(0))/y$.

Summing these three terms, we obtain the following recursive equation for F:

$$yF(x,y) = x(G(y)-G(0)) + 2x(F(x,y)G(y)-F(x,0)G(0)) + x(F(x,y)^2G(y)-F(x,0)^2G(0)) \tag{7.10}$$

With our notation $\mathbf{F} = F+1$ and $\mathbf{F}_0(x) = \mathbf{F}(x,0) = F(x,0)+1$, this equation simplifies to

$$P(\mathbf{F}(x,y),\mathbf{F}_0(x),x,y)=0 \quad \text{where} \quad P(f,f_0,x,y) = y + f^2xG(y) - fy - f_0^2xG(0). \tag{7.11}$$

To solve this equation, we apply the kernel method of Bousquet-Mélou and Jehanne [44] and look for a (formal) power series $Y = Y(x)$ such that $\partial_f P(\mathbf{F}(x,Y(x)),\mathbf{F}_0(x),x,Y(x)) = 0$ so that combined with (7.11) we also find automatically $\partial_y P(\mathbf{F}(x,Y(x)),\mathbf{F}_0(x),x,Y(x)) = 0$. This introduction may

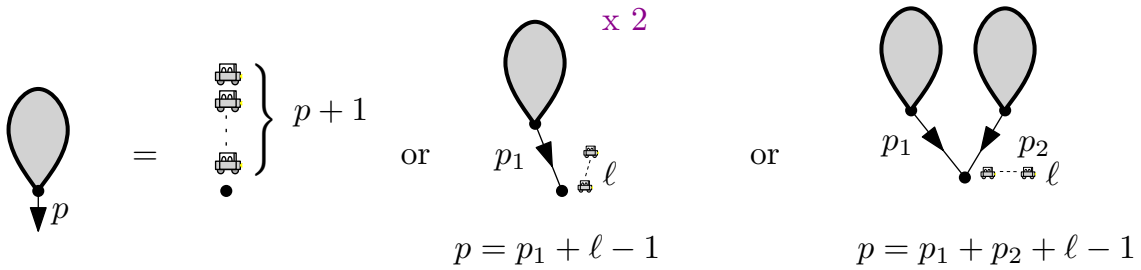


Figure 7.4: Illustration of Tutte's recursive decomposition at the root vertex. On the left a fully parked tree with flux p. If $n = 1$, then the tree is just a vertex with $p+1$ cars arriving on it. Otherwise, it has one or two children which are the root of smaller fully parked tree.

seem ad-hoc, but it enables us to find a system of three equations on the three unknowns $\mathbf{F}, \mathbf{F}_0$ and Y, so that with a little luck we will find an "expression" for those. Actually, as we will see below $x \mapsto Y(x)$ is a convenient change of variable which simplifies our expressions. To summarize, we are looking for a solution $Y \equiv Y(x)$ to the following system:

$$\begin{cases} Y - 2x\mathbf{F}G(Y) = 0, \\ 1 + xG'(Y)\mathbf{F}^2 = \mathbf{F}, \\ Y + xG(Y)\mathbf{F}^2 = Y\mathbf{F} + xG(0)\mathbf{F}_0^2. \end{cases} \tag{7.12}$$

Thanks to the first equation, we know that $\mathbf{F} = Y/(2xG(Y))$. Replacing $\mathbf{F}$ by this quantity in the second equation, we obtain

$$1 + \frac{Y^2 G'(Y)}{4xG(Y)^2} = \frac{Y}{2xG(Y)} \quad \text{that is} \quad Y = x\left(\frac{4G(Y)^2}{2G(Y) - YG'(Y)}\right), \tag{7.13}$$

which makes it clear that $Y \equiv Y(x)$ exists as a power series (and even with a positive radius of convergence in a neighborhood of 0). Once the existence of Y is granted, we use again the system of equations (7.12) to obtain an equation that only involves $\mathbf{F}_0(x)$ and $Y(x)$. If we replace in the third equation $\mathbf{F}$ by $Y/(2xG(Y))$ (which is a consequence of the first equation) and x by $\frac{Y(2G(Y)-YG'(Y))}{4G(Y)^2}$, we obtain

$$\frac{4YG(Y)}{2G(Y) - YG'(Y)} + \frac{\mathbf{F}_0^2 YG(0)(2G(Y) - YG'(Y))}{G(Y)^2} = 4Y.$$

This equation is quadratic in $\mathbf{F}_0(x)$ and using the fact that it has non negative coefficients we obtain

$$\mathbf{F}_0(x) = \frac{2G(Y)\sqrt{G(Y) - YG'(Y)}}{(2G(Y) - YG'(Y))\sqrt{G(0)}} \quad \text{with} \quad Y \equiv Y(x) \text{ as in (7.13).} \tag{7.14}$$

We have found here an "explicit" solution for $\mathbf{F}_0(x)$ around $x = 0$. Coming back to Tutte's equation, once $\mathbf{F}_0$ is known this equation is quadratic in $\mathbf{F}$ and we can solve it into

$$\mathbf{F}(x, y) = \frac{y \pm \sqrt{y^2 + 4G(y)x(G(0)\mathbf{F}_0(x)^2 x - y)}}{2G(y)x}. \tag{7.15}$$

The sign in front of the square root can actually change since the function inside the square root vanishes when $y = Y(x)$ and we need to change branch to keep an analytic function. But we shall not use the exact expression in what follows.

7.4.2 Radius of convergence

In this section we use the explicit resolution of the functional equation (7.11) to determine the radius of convergence x_c of $\mathbf{F}_0$ and the value of $\mathbf{F}_0(x_c)$. The important fact for our application to parking being that under the condition $(\star)$ on the existence of t_c in Theorem 7.1 we have

$$\mathbf{F}_0(x_c) = \frac{2G(t_c)\sqrt{G(t_c) - t_c G'(t_c)}}{(2G(t_c) - t_c G'(t_c))\sqrt{G(0)}}.$$

Recall from (7.14) that $\mathbf{F}_0(x)$ is an explicit function of $Y(x)$ itself given by the implicit equation (7.13).

Analyticity of Y. We first determine the analytic properties of the change of variable $x \mapsto Y(x)$. Recall from (7.13) that x and $Y = Y(x)$ are linked by the equation

$$x = \frac{Y(2G(Y) - YG'(Y))}{4G(Y)^2}, \quad \text{equivalently} \quad x = \psi(Y(x)) \quad \text{with } \psi(y) = \frac{y(2G(y) - yG'(y))}{4G(y)^2}. \tag{7.16}$$

Note that

$$\partial_y \psi(y) = \frac{2(G(y) - yG'(y))^2 - y^2 G(y)G''(y)}{4G(y)^3}$$

and in particular $\psi(0) = 0$ and $\psi'(0) > 0$ so that by the implicit function theorem, we can define Y in a neighborhood of 0 such that $x = \psi(Y(x))$. Recall the condition $(\star)$ from Theorem 7.1 which says that the function $y \mapsto y^2 G''(y)G(y) - 2(G(y) - yG'(y))^2$ at the numerator of $\partial_y \psi(y)$ reaches 0 at time $t_c \in (0, \infty)$, see Figure 7.5.

Remark. In particular if G has an infinite radius of convergence then $(\star)$ holds. Indeed, the quantity $G(y) - yG'(y) = \sum_{k \geqslant 0} \mu_k (1-k) y^k$ equals $G(0) = \mu_0 > 0$ at $y = 0$ and is bounded from above by $\mu_0 - (1 - \mu_0 - \mu_1)y^2$ for $y > 1$ which goes to $-\infty$ as $y \to +\infty$. Thus, there exists z_c such that $G(z_c) - y_c G'(z_c) = 0$ and the function $y \mapsto y^2 G''(y)G(y) - 2(G(y) - yG'(y))^2$ is positive at $y = z_c$. Since it is negative at $y = 0$, then $(\star)$ holds and $t_c \in (0, z_c)$.
The assumption $(\star)$ is also satisfied when G has a finite radius of convergence y_c and (at least) one of the three quantities $G(y_c)$, $G'(y_c)$, $G''(y_c)$ is infinite. In case $G(y_c) = \infty$, starting from $yG'(y) - G(y) = y \sum_{k \geqslant 0} \mu_k (k-1) y^{k-1}$, and noting that $\mu_k (k-1) y^{k-1} \sim \mu_k k y^{k-1}$, we deduce that $yG'(y) - G(y) \sim yG'(y)$ as $y \to y_c$, hence $G(y) - yG'(y)$ again has limit $-\infty$ as $y \to \infty$, and the same argument as above applies. The last cases are obvious : in case $G'(y_c) = \infty$ but $G(y_c) < \infty$, $G(y) - yG'(y)$ plainly has limit $-\infty$; last, in case $G''(y_c) = \infty$ but $G'(y_c) < \infty$, we directly have $\lim_{y \to y_c} y^2 G''(y)G(y) - 2(G(y) - yG'(y))^2 = +\infty$.

To clarify the reader's spirit and for latter discussion, let us classify the possible scenarios according to the three cases identified in Chen [59, Figure 1], see Figure 7.5:

- the most common case is when t_c exists and is strictly less than the radius of convergence y_c of G. At this point we have

$$\frac{\partial^2}{\partial y \partial y}\psi(y)\Big|_{y=t_c} = \frac{t_c}{4G(t_c)^3}\left(2(t_cG'(t_c) - 2G(t_c))G''(t_c) - t_cG(t_c)G'''(t_c)\right),$$

  and since $G(t_c) \geqslant t_cG'(t_c)$ this second derivative is strictly negative so that the function $y \mapsto \psi(y)$ reaches a local maxima at this point. We call this situation the **generic** situation.

- it could also happen that t_c exists and is equal to the radius of convergence of G. In this case, although $y \mapsto \psi(y)$ reaches a maxima at t_c, the local behavior around the maximum may not be quadratic. We call this situation the **non-generic situation**.

- Otherwise t_c does not exists and in particular the radius y_c of convergence of G is finite and $y \mapsto \psi(y)$ has a finite positive derivative at y_c. This is the **dense** situation which leave aside for the moment.

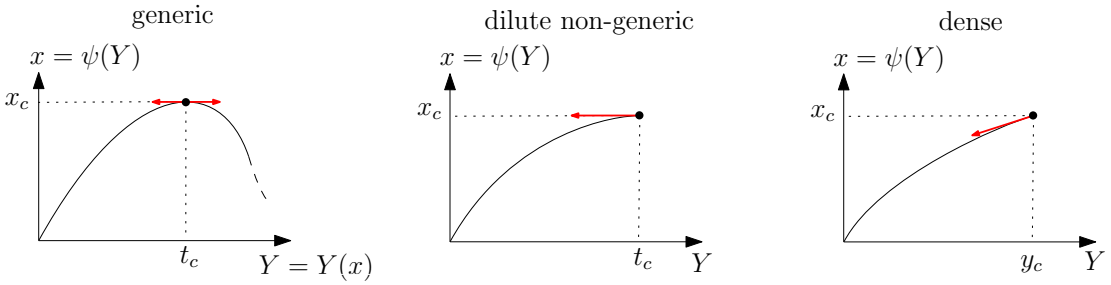


Figure 7.5: Illustration of the three scenarios for the change of variable $x \mapsto Y(x)$. Our standing assumption $(\star)$ holds in the first two cases.

Under the assumption $(\star)$ –i.e. except in the dense situation– we introduce

$$x_c = \psi(t_c) = \frac{t_c(2G(t_c) - t_cG'(t_c))}{4G(t_c)^2}. \tag{7.17}$$

Lemma 7.2. *Under assumption $(\star)$ the function $x \mapsto Y(x)$ is increasing and analytic over $[0, x_c)$ and furthermore*

$$\lim_{x \to x_c} Y'(x) = \infty.$$

Proof. Under the assumption $(\star)$ the function $y \mapsto \psi(y)$ is increasing and analytic over $[0, t_c)$ so that by the analytic version of the implicit function theorem one can define its increasing inverse function $x \mapsto Y(x)$ over $[0, x_c)$. Note that since $\psi'(y) \to 0$ as $y \uparrow y_c$ we have $Y'(x) \to \infty$ as $x \uparrow x_c$.

□

Radius of convergence of $\mathbf{F}_0$. We still suppose $(\star)$. Coming back to $\mathbf{F}_0$, notice that $(G(Y(x)) - Y(x)G'(Y(x)))$ is always positive for $x \in [0, x_c]$, thus by Equation (7.14) the function $\mathbf{F}_0$ is also analytic over $[0, x_c)$. Since $\mathbf{F}_0$ only has positive coefficients, by Pringsheim's Theorem (see for example [92, Theorem IV.6 p.240]), its radius of convergence is *at least* x_c. The following lemma shows that it actually coincides with it

Lemma 7.3. *Suppose $(\star)$ then we have*

$$\lim_{x \to x_c} \mathbf{F}_0''(x) = \infty,$$

in particular the radius of convergence of $\mathbf{F}_0$ is equal to x_c.

Proof. We use our explicit computations of $\mathbf{F}_0$ and Y to derive formulas for the first two derivatives of x. Even if we don't need it for this lemma, we start with the first derivative. We take the expression of $\mathbf{F}_0$ given by Equation (7.14), differentiate it with respect to x and replace the occurence of $Y'(x)$ by $1/(\partial_y \psi(Y(x)))$ thanks to Equation (7.16). We obtain

$$\mathbf{F}_0'(x) = \frac{4G(Y(x))^3 G'(Y(x))}{\sqrt{G(Y(x)) - Y(x)G'(Y(x))}(2G(Y(x)) - Y(x)G'(Y(x)))^2}.$$

This quantity has a finite limit when $Y(x)$ converges to t_c. We thus need to compute the second derivative of $\mathbf{F}_0$. To do so, we differentiate the above expression of $\mathbf{F}_0'$ and again replace the occurence of $Y'(x)$ by $1/(\partial_y \psi(Y(x)))$. We then obtain a fraction involving the derivatives of G at $Y(x)$. Using the definition of t_c under assumption $(\star)$, we can show that

$$\lim_{x \to x_c} \mathbf{F}_0''(x) = \lim_{t \to t_c} \frac{16G(t)^6(2G(t) - tG'(t))^3(G(t) - tG'(t))(2G(t) - tG'(t))^2}{t^2(G(t) - tG'(t))^{3/2} \cdot (2(G(t) - tG'(t))^2 - t^2 G(t)G''(t))} = +\infty,$$

since all but the factor $(2(G(t) - tG'(t))^2 - t^2 G(t)G''(t)$ are positive for $t < t_c$ and have a positive limit as $t \to t_c$.

□

7.5 Probabilistic consequences

Armed with our enumeration results and the criterion of Proposition 7.2, we can now proceed to the proof of our main results.

7.5.1 Theorem 7.1: Location of the threshold

Recall that by Proposition 7.2, the parking process is subcritical if and only if there exists a positive solution to (7.9). When $(\star)$ holds, since the function $x \mapsto G(0)x\mathbf{F}_0(x)$ is strictly increasing, Equation (7.9) has a solution if and only if $G(0)x_c\mathbf{F}_0(x_c)^2 \geqslant 1$ where x_c is the radius of convergence of $\mathbf{F}_0$ found in the previous paragraph. Now, since $t_c = Y(x_c)$ and plugging the value of $\mathbf{F}_0(x_c)$ given by Equation (7.14) in $G(0)x_c\mathbf{F}_0(x_c)^2 \geqslant 1$, we obtain $t_c + \frac{t_c G(t_c)}{t_c G'(t_c) - 2G(t_c)} \geqslant 1$.

By definition of t_c under Assumption $(\star)$, the quantity $t_cG'(t_c) - 2G(t_c) < t_cG'(t_c) - G(t_c)$ is always negative. Hence there is a solution x to (7.9) if and only if t_c satisfies

$$(t_c - 2)G(t_c) \geqslant t_c(t_c - 1)G'(t_c),$$

which together with Proposition 7.2 concludes the proof of Theorem 7.1.

Remark. Notice that if t_c satisfies the condition of Theorem 7.1, then it is greater than 2. This implies that if the parking process is subcritical, then the radius of convergence of the generating function G of the car arrivals is at least 2. To see it, first note that the inequality $(t-2)G(t) \geqslant t(t-1)G'(t)$ can not be satisfied for $t \in [1,2]$ since the left-hand side is non-positive and equals 0 only for $t = 2$, whereas the right-hand side is non-negative and equals 0 for $t = 0$ and $t = 1$ only. Neither can this inequality be satisfied for $t \in (0,1)$, because for such t, we can bound from above the quantity $t(t-1)G'(t)/(t-2)$ by $(3-2\sqrt{2})G'(1) \leqslant (3-2\sqrt{2})G'(1) \leqslant 1/5$ since $G'(1) = \mathbb{E}[A] \leqslant 1$ in the subcritical case. On the other side, the quantity $G(t)$ is bounded from below by $G(0) \geqslant 1/2$.

7.5.2 Theorem 7.2: critical computations

Before moving on to the critical computation (Theorem 7.2) let us prove that the critical case characterized by the equality in the second display of Theorem 7.1 actually corresponds to the natural fact that one "cannot increase the number of cars" and stay subcritical:

Lemma 7.4 (Criticality). *Suppose $(\star)$. Then we have equality in (7.1) iff for any $\varepsilon > 0$ the law with generating function $G_\varepsilon(t) = G(t) + \varepsilon t - \varepsilon$ is supercritical.*

Proof. Suppose that μ is subcritical in the sense of Theorem 7.1 and look at the probability measure μ_ε such that its generating function is given by $G_\varepsilon(t) = G(t) + \varepsilon t - \varepsilon$ for some $\varepsilon > 0$. First notice that μ_ε satisfies Assumption $(\star)$ for ε small enough. Indeed the radius of convergence of G_ε is that of G and the quantity

$$\begin{aligned} 2(G_\varepsilon(t) - tG'_\varepsilon(t))^2 - t^2G_\varepsilon(t)G''_\varepsilon(t) &= (2(G(t) - tG'(t) - \varepsilon)^2 - t^2(G(t) + \varepsilon(t-1))G''(t) \\ &= 2(G(t) - tG'(t))^2 - t^2G(t)G''(t) \\ &\quad + 2\varepsilon^2 - 2\varepsilon(G(t) - tG'(t)) - \varepsilon t^2(t-1)G''(t) \end{aligned}$$

is negative at t_c when ε is small enough, so that $t_c^\varepsilon := \min\{t \geqslant 0,\ 2(G_\varepsilon(t) - tG'_\varepsilon(t))^2 = t^2G_\varepsilon(t)G''_\varepsilon(t)\} < t_c$ and the function $\varepsilon \mapsto t_c^\varepsilon$ is continuous in a positive neighborhood of 0. To determine whether μ_ε is subcritical or not, we then need to determine the sign of

$$(t_c^\varepsilon - 2)G_\varepsilon(t_c^\varepsilon) - t_c^\varepsilon(t_c^\varepsilon - 1)G'_\varepsilon(t_c^\varepsilon),$$

which is then continuous is ε. Thus if Equation (7.1) is not an equality, we can increase μ and remain subcritical.

Suppose now that (7.1) is an equality. We can show that

$$\partial_y\varphi_\varepsilon(t_c - \delta) = \frac{6G(t_c) + t_c^2G^{(3)}(t_c)(t_c-1)^3}{4G(t_c)^2(t_c-1)^3}\delta - \frac{3}{2(t_c-1)G(t_c)^2}\varepsilon + o(\varepsilon) + o(\delta).$$

When $t_c - \delta = t_c^\varepsilon$, then the left hand side is zero so that the main asymptotic on the right hand side should be 0. Hence $\delta = t_c - t_c^\varepsilon$ is of order ε. Moreover,

$$(t_c^\varepsilon - 2)G_\varepsilon(t_c^\varepsilon) - t_c^\varepsilon(t_c^\varepsilon - 1)G'_\varepsilon(t_c^\varepsilon) = -2\varepsilon(t_c^\varepsilon - 1) + O((t_c^\varepsilon - t_c)^2).$$

is negative when ε small enough and μ_ε is supercritical by Theorem 7.1.

□

Proof of Theorem 7.2. Let us focus now on the case when (7.1) is an equality. In that case, since (7.9) is an equality for $x = x_c$, then $x_c = p_\circ$. If x satisfies Equation (7.16) and the parking is critical, so that $y = t_c$ is solution of $(t-2)G(t) = t(t-1)G'(t)$, then

$$x = x_c = \frac{t_c^2}{4(t_c - 1)G(t_c)} \qquad \text{and} \qquad \mathbf{F}_0(x_c)^2 = \frac{1}{p_\circ G(0)} = \frac{1}{p_\circ \mu_0}. \tag{7.18}$$

Moreover, using Equation (7.7), we obtain

$$p_\bullet = \sqrt{\frac{p_\circ}{\mu_0}} - p_\circ,$$

and this concludes the proof of Theorem 7.2.

□

Remark. The fact that $p_\circ > 1/2$ is a consequence of the equation for t given by Theorem 7.1. Indeed, in the critical case, then $p_\circ = x_c = t_c^2/(4(t_c-1)G(t_c))$ where t_c is given by Theorem 7.1. But since $t_c > 1$, we have $G(t_c) - G(1) \leqslant (t_c - 1)G'(t_c) = (t_c - 2)G(t_c)/t_c$. Thus $G(t_c) \leqslant t_c/2$ and

$$p_\circ \geqslant \frac{2t_c}{t_c - 1} > 1/2.$$

7.5.3 Examples

Let us proceed to the computation of the critical threshold for parking for various families of stochastically increasing laws. In the first four cases below, it is easy to check that condition $(\star)$ holds so that we can just apply the general formulas. In the last example we explain how our techniques can be applied even if $(\star)$ does not hold.

Binary$_{0/2}$ car arrivals. As a first example, we can imagine that either 0 or 2 cars arrive at each spot, i.e. the law of the car arrivals is $\mu = (1 - \frac{\alpha}{2})\delta_0 + \frac{\alpha}{2}\delta_2$, so that $G(t) = (1 - \frac{\alpha}{2}) + \frac{\alpha}{2}t^2$. This is the example considered in [100, Proposition 3.5] and [22, Proposition 4]. Explicit computations show in this case that

$$t_c = Y(x_c) = \sqrt{\frac{2-\alpha}{3\alpha}}, \qquad x_c = \frac{3\sqrt{3}}{16\sqrt{\alpha(2-\alpha)}} \qquad \text{and} \qquad \mathbf{F}_0(x_c) = \frac{4\sqrt{6}}{9}.$$

Note that $\mathbf{F}_0(x_c)$ does not depend on α, see the remark below. We then see that for $t = t_c$, the Inequality (7.1) is quadratic in α and is satisfied as soon as $\alpha < \alpha_c = 1/14$.

Binary$_{0/k}$ car arrivals. In the case when $\mu = (1-\frac{\alpha}{k})\delta_0 + \frac{\alpha}{k}\delta_k$, so that $G(t) = (1-\frac{\alpha}{k}) + \frac{\alpha}{k}t^k$ with $k \geqslant 3$, we have

$$Y(x_c) = \left(\frac{(k-a)(-4+k(3+k-\sqrt{(k-1)(k+7)}))}{2a(k-2)(k-1)}\right)^{1/k},$$

$$\mathbf{F}_0(x_c) = \frac{\sqrt{2}(3k-\sqrt{(k-1)(k+7)}-3)}{(k-2)(k-1)\sqrt{\frac{5-k+\sqrt{(k-1)(k+7)}}{(k-1)k}}}.$$

so that the model is critical at

$$\alpha_c(\text{Binary}_{0/k}) = \frac{k}{1+2^{-k-2}\left(3+\sqrt{\frac{k+7}{k-1}}\right)^k\left((k-1)(k+4)+k\sqrt{(k+7)(k-1)}\right)}.$$

Poisson. Suppose the law of the car arrivals is Poisson with mean $\alpha > 0$, so that in this case $G(t) = e^{\alpha(t-1)}$. Again, explicit computation show that

$$Y(x_c) = \frac{2-\sqrt{2}}{\alpha}, \qquad x_c = \frac{(\sqrt{2}-1)e^{\alpha-2+\sqrt{2}}}{2\alpha} \qquad \text{and} \qquad \mathbf{F}_0(x_c) = \sqrt{2(\sqrt{2}-1)}e^{1-1/\sqrt{2}},$$

so that the model is subcritical for parking as long as $\alpha \leqslant \alpha_c$ with

$$\alpha_c = 3-2\sqrt{2}.$$

Geometric. Consider here the case when $G(t) = 1/(1+\alpha-\alpha t)$. Then

$$Y(x_c) = \frac{1+\alpha}{3\alpha}, \qquad x_c = \frac{(1+\alpha)^2}{12\alpha} \qquad \text{and} \qquad \mathbf{F}_0(x_c) = \frac{2}{\sqrt{3}}.$$

so that the model is subcritical for parking as long as $\alpha \leqslant \alpha_c$ with $\alpha_c = 1/8$.

Remark (Combinatorial counting). The reader may be puzzled by the fact that in the last four cases the value $\mathbf{F}_0(x_c)$ does not depend on the parameter α. This is because in each case, the dependence on α of the μ_α-weight of a fully parked tree of size n is of the form $(c_\alpha)^n$, for a constant c_α depending on α that cancels out at criticality. To wit, consider fully-parked trees associated with $0/k$ arrivals : in this case, for n a multiple of k, fully parked trees of size n have weight $\prod_v \mu_{a(v)} = \mu_0^{n/k}\mu_k^{(1-1/k)n} = (\mu_0^{1/k}\mu_k^{(1-1/k)})^n$.

Without ($\star$). When hypothesis ($\star$) is not satisfied we can still apply our method: If the generating series G has a radius of convergence y_c then the value t_c is then replaced by

$$\tilde{t}_c = \min\left\{y_c, \min\{t \geqslant 0 :\ 2(G(t)-tG'(t))^2 = t^2G(t)G''(t)\}\right\},$$

and we need to check that $\tilde{x}_c$ defined analogously by (7.17) is again the radius of convergence of the series $\mathbf{F}_0$ (we did not try to prove such a general statement here). Then using Proposition 7.2,

the subcriticality of the parking is equivalent to the fact that $\mu_0 \tilde{x}_c (\mathbf{F}_0(\tilde{x}_c))^2 \geqslant 1$, the only different point is that we cannot use the equality $2(G(\tilde{t}_c) - \tilde{t}_c G'(\tilde{t}_c))^2 = \tilde{t}_c^2 G(\tilde{t}_c) G''(\tilde{t}_c)$ to further simplify the expression. As an example of such law, consider the generating function

$$G(t) = 1 + \frac{1+t^2}{26} - \frac{1}{13}\left(\frac{3-t}{2}\right)^{7/3}. \tag{7.19}$$

The radius of convergence of G is equal to 3 but $G'(3)$ and $G''(3)$ exist. An explicit computation shows that $(\star)$ holds with $t_c = y_c = \tilde{t}_c = 3$ for G, and furthermore G is critical for the parking process. However if one considers $\tilde{G} = 0.9 + 0.1G$ then $\tilde{t}_c = 3$ and $(\star)$ does not hold but still $\tilde{G}$ is subcritical for the parking process.

7.6 Extensions and comments

In this work, we voluntarily stick to the simplest case of the binary tree with i.i.d. arrivals without specific conditions to keep the paper accessible to a wide audience. Let us mention a few perspectives that our approach opens:

7.6.1 Non-generic and dense case

In this work, we focused on the localization of the threshold and on the computation of some critical quantities. One could also try to get scaling limits of critical components and compute several critical or near-critical exponents. As mentioned in the introduction, we expect that a large family of critical car arrivals (say, with bounded support) belong to a common universality class where we expect that the cluster size of the root has a tail that decays as $n^{-5/2}$ as $n \to \infty$ and where the scaling limits of the critical components are given by 3/2-stable Growth-Fragmentation trees. But when the car arrivals have a heavy tail (and when the parameters are fine-tuned so that the law is critical), we hope to see different universality classes. Actually, as seen in Section 7.4.2, the singular behavior of $\mathbf{F}_0$ near its radius of convergence is linked to the behavior of the $Y(x)$ near near x_c, see Figure 7.5. For instance, in the example (7.19) an explicit computation shows that the singular behavior of $\mathbf{F}_0(3-x) - \mathbf{F}_0(3)$ is of the form $Cx^{4/3}$ which indicates a polynomial decay of the critical cluster with exponent $n^{-7/3}$. This is very similar to the scenarios that happened in the enumeration of plane fully parked trees in Chen [59], with the notable difference that in our model the dense case can be critical for the parking process. See also [61, Theorem 1.2] for the Derrida-Retaux model with heavy-tailed distributions where the free-energy has a peculiar behavior. We plan on studying those different behaviors in forthcoming works.

7.6.2 General case of d-ary tree and GW trees

Our work may be extended to parking on more general trees such as d-ary trees and perhaps super-critical Bienaymé–Galton–Watson trees. The crux is of course the enumeration of fully parked trees. In the case of Bienaymé–Galton–Watson trees with a geometric offspring distribution, we can use

the enumeration of plane fully parked trees already performed by Chen [59]. In the case of general supercritical Bienaymé–Galton–Watson trees, one would probably need the addition of another variable z counting the number of adjacent vertices of the fully parked tree inside the global tree (in our case, we had a fixed number $n+1$ of vertices adjacent from above to a fully parked tree of $\mathbb{B}$ of size n). We wonder whether the randomness of the underlying tree may yield to different universality classes compared to the case of d-ary trees and plan on investigating this in forthcoming works.

7.6.3 Links with Derrida-Retaux model

As mentioned several times in the paper, the Derrida-Retaux model [61] is closely related to the parking process on $\mathbb{B}$. We wonder whether a firm connection can be made between the two models.

Chapitre 8 :

Last Car Decomposition of Planar Maps

Les résultats de ce chapitre sont issus de l'article [63] et ont été soumis pour publication.

We give new equations which characterize the generating functions of planar quadrangulations and planar triangulations, with zero, one or two boundaries. The proof is inspired by the Lackner–Panholzer last car decomposition of parking trees [128] and consists in applying a similar decomposition to the peeling trees of planar maps.

Contents

8.1 Introduction

Since their introduction by Tutte in his series of "census" papers [165, 166, 167, 168], maps are fundamental objects which have been extensively studied especially in combinatorics and in probability. The purpose of this work is to establish new recursive equations to enumerate particular types of maps which are inspired both by the well-known peeling procedure for maps [72] and by the last car decomposition of fully parked trees introduced first by Lackner and Panholzer in [128].

Enumeration of planar maps. In this paper, we will focus on *planar maps*, which are finite connected graphs (possibly with loops and multiple edges) properly embedded in the two-dimensional sphere, seen up to continuous deformation. To avoid symmetries, all planar maps will be rooted at a distinguished oriented edge. The number of edges incident to a face, i.e. that of its underlying polygon, is called the *degree* (or sometimes the *perimeter* or *length*) of this face. We will particularly study the case of planar *quadrangulations* and *triangulations* which are maps where all faces are quadrangles (resp. triangles). Sometimes our maps will have a boundary, i.e. a distinguished face which may have a different degree (but this degree has to be even in the case of quadrangulations), in which case the root edge lies on the boundary with the distinguished face lying to its right. This face will be called the external face.
One possible way to enumerate quadrangulations or triangulations is to use Tutte's method based on a recursive decomposition which is obtained by removing the root edge. The point is that when removing an edge from a quadrangulation, it may not be a quadrangulation anymore but a quadrangulation with a boundary. More precisely, when erasing the root edge of a quadrangulation with a boundary with degree $2p$ with $p \geqslant 1$, one of these two possible events occurs (see Figure 8.1):

- either the quadrangulation stays connected, which means that one discovers a new face of the initial quadrangulation, and one re-roots the quadrangulation at the edge adjacent to the left of the oriented peeled edge.
- or the deletion of the root edge disconnects the quadrangulation and one gets two quadrangulations with half-perimeter $p_1 \geqslant 0$ and $p_2 \geqslant 0$ such that $p_1 + p_2 = p - 1$, that one can re-root using the two endpoints of the removed edge.

Note that we considered here the half-perimeter since the quadrangulations are bipartite and we only encounter even boundaries during the exploration. In the case $p = 0$, the map is just a vertex map, i.e. the map composed of one single vertex and no edge. One can iterate Tutte's decomposition in each quadrangulation with a boundary until one obtains a collection of vertex maps. This edge-by-edge peeling exploration of a planar quadrangulation can be encoded in a so-called *peeling tree*[1] by recursively labeling the vertices with the half-perimeters of the boudaries, see Figure 8.1. Conversely,

[1] In random map theory, we often consider different ways to re-root the components, or *peeling algorithms*, which may yield different peeling trees, see [72]. For the sake of simplicity, we shall stick here to the rules presented above and do not use any other peeling algorithm.

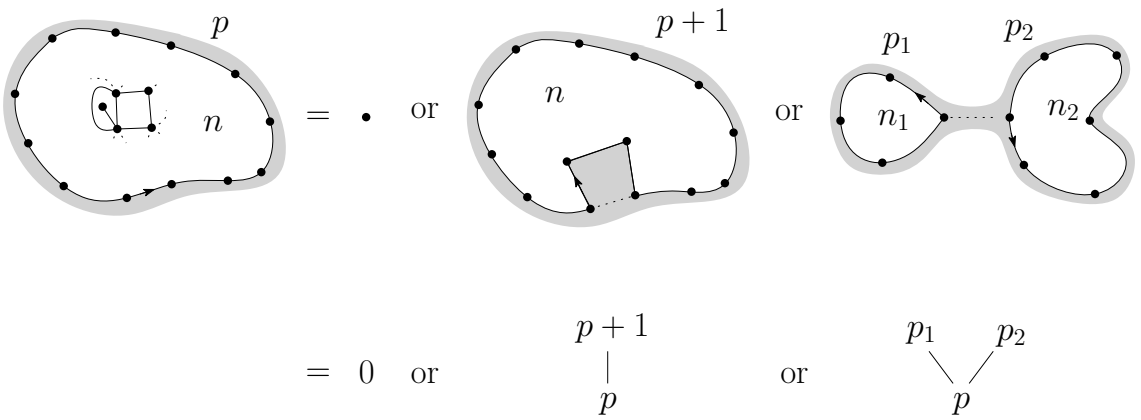


Figure 8.1: Tutte's recursive decomposition for quadrangulations and the local correspondence in the peeling tree. In the case when the removal of the root edge splits the quadrangulation (right), we put the component attached to the origin of the root edge on the left of the peeling tree.

given a labeled plane tree whose labeling follows the appropriate rules described above, one can recover the initial quadrangulation. Note that the vertices of the peeling tree with label 0 are the leaves and correspond to the vertices in the quadrangulation, whereas the other vertices correspond to the edges of the map, see Figure 8.2.

Tutte then obtained recursive equations for the number of quadrangulations using the boundary length as a catalytic variable, which can be summarize into the following equation on $\mathbf{Q}$ the bivariate generating function of quadrangulations with a boundary, where the variable x counts the number of vertices and y counts the half-perimeter of the boundary

$$\mathbf{Q} = x + y\mathbf{Q}^2 + \frac{1}{y}\left(\mathbf{Q} - x - y\mathfrak{Q}\right), \tag{8.1}$$

where $\mathfrak{Q} = [y^1]\mathbf{Q}$ is the (univariate) generating function of quadrangulations with a boundary of length 2. On the right, the term x stands for the vertex map, and the term $y\mathbf{Q}^2$ encompasses the case where the removal of the root edge splits the quadrangulation. The remaining term corresponds to the case where the quadrangulation stays connected, in which case the new quadrangulation has at least a boundary of length 4. This equation characterizes the power series $\mathbf{Q}$ and has been solved explicitly using the so-called "quadratic method" and then its generalization introduced by Bousquet-Mélou and Jehanne [44].

Since then, other methods have been developed to enumerate maps: via matrix integrals [45, 164], bijections with other labels trees "à la Schaeffer" [69, 160, 159] or correspondence with the KP hierarchy [56, 101]. Our work concentrates on peeling trees but uses another method to enumerate them which is based on Lackner–Panholzer last car decomposition of parking trees [128]. This link between parking models and maps was already suggested by Panholzer in [147, Remark 2] and by

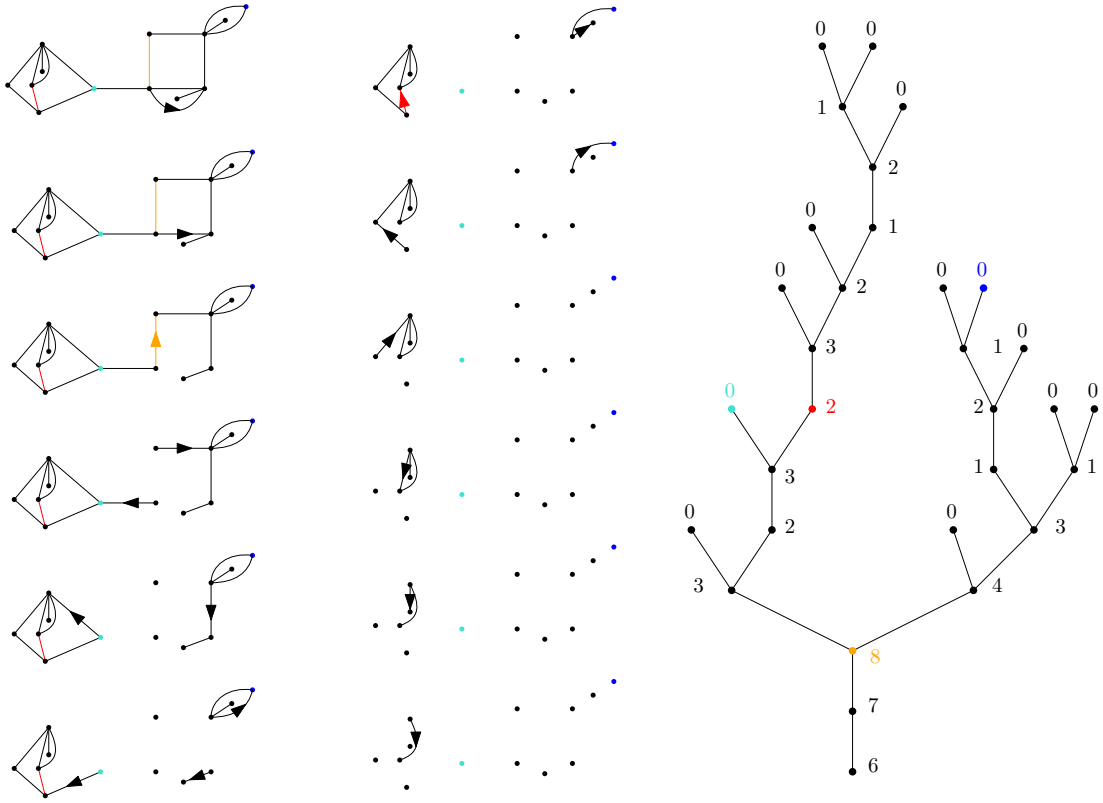


Figure 8.2: Step-by-step example of the peeling process on a quadrangulation with $n = 12$ vertices and a boundary of degree $2p$ with $p = 6$ and its corresponding peeling tree. In blue, two vertices of the quadrangulation and the two corresponding leaves with label 0 in the peeling tree. In orange and rouge, two edges and their corresponding inner vertices in the peeling tree.

the author and Curien in [67, Section 8]. Panholzer found remarkable explicit enumeration formulas of parked trees for a large class of combinatorial models, one of which is linked to the enumeration of non-decomposable maps. The non-decomposable maps also have links with description trees [68] of Cori and Schaeffer whose construction shares similarities with that of parking trees which we explain now. The link between bipartite planar maps and "degree trees" pointed out by Fang in [86, 87] also supports this strong link between parking trees and map models.

Parking on trees. Our decomposition uses an idea introduced by Lackner and Panholzer in [128] in the context of parking trees. Let us recall this model for the readers' convenience, although we shall not use it in this paper. Let $\mathbf{t}$ be a finite (plane) rooted tree which is our parking lot. Each vertex represents a parking spot which can accommodate at most one car. We then let cars arrive on the vertices of $\mathbf{t}$. Each car tries to park on its arrival node, but if the spot is already occupied, it drives towards the root and parks as soon as possible. An important property of this model is its Abelian

property: the final configuration, the flux of cars which go through a given edge and the outgoing flux of cars do not depend upon the order chosen to park the cars. In particular, one can recover the initial configuration of cars from the final configuration of parked cars and flux on the edges. This model undergoes a phase transition and was first studied on a directed line [122] and raised recently a growing interest, especially on random tree models with an increasing level of generality [128, 100, 60, 75, 64].

We will focus here on the connected components of parked vertices in the final configuration in this model, which are called *fully parked trees*. They consist of plane trees with a decoration of cars so that all vertices accommodate a car and some cars possibly contribute to the outgoing flux. We can decorate the edges of those trees with the flux of cars and in a specific case of trees and car arrivals, the labeled trees obtained that way (pushing up the labeling from the edges to the vertices above) are very similar to our peeling trees since the rules for the labels are identical, see Figure 8.3. Specifically, consider a fully parked tree where each vertex has 0, 1 or 2 children and where exactly one car arrives on each leaf (vertex with 0 child), no car arrives on vertices with one child and two cars arrive on vertices with two children. Then the leaves will all get label 0 since one car arrives and parks on each leaf and no car can come from above. If a vertex has one child with label $\ell \geqslant 1$ (and no car arriving on it), then one of the ℓ cars arriving from above parks and the vertex will get label $\ell - 1$. Lastly, a vertex with two children with label ℓ_1 and $\ell_2 \geqslant 0$ has two cars arriving on it and $\ell_1 + \ell_2$ coming from above. One of them parks and it remains $\ell_1 + \ell_2 + 1$ cars contributing to the flux on the edge below. Those local rules are exactly the same as that of peeling trees of quadrangulations and we can thus match bijectively these fully parked trees (with prescribed car arrivals) with quadrangulations with a boundary.

To enumerate such general plane trees with parking, Chen [59] uses a method which is similar to Tutte's since he decomposes the fully parked trees at their root using the outgoing flux of cars as a catalytic variable.

The technique introduced by Lackner & Panholzer [128] and deepened by Panholzer in [147] is different: it consists in a decomposition of the initial fully parked tree according to the parking spot of a distinguished car seen as the "last" car. Thanks to the Abelian property, we can imagine that we first park all the cars but this distinguished car and that it arrives at last and parks on a vertex, which was empty before the last car arrived. And when removing this last car, the trees which are attached to this free spot are also fully parked trees. See also [67, Section 8.2] for the decomposition of fully parked trees with a different notion of components and with possibly a flux of outgoing cars.

The heart of this work is to adapt the "last car decomposition" to the enumeration of planar maps, more precisely of the peeling trees of planar maps. This new decomposition is explained in Section 8.2 in the case of quadrangulations and in Section 8.3 for triangulations with zero, one or two boundaries. In both cases, this decomposition enables us to establish new equations on the corresponding generating functions which we state now.

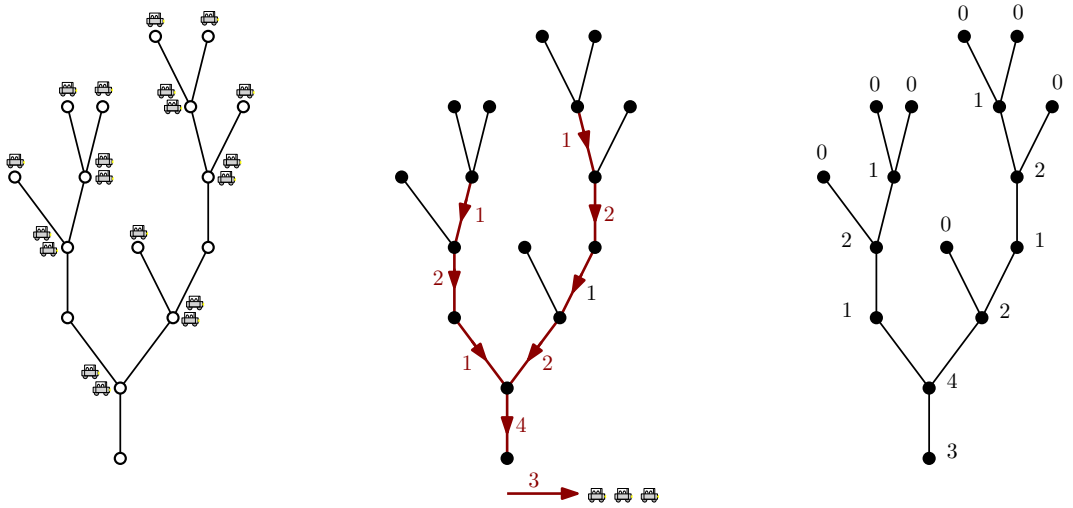


Figure 8.3: Left, an example of a fully parked tree with 16 vertices and 19 cars, three of which can not park. We put exactly one car arriving on every leaf, two cars on every vertex with two children, and no car on vertices with one child. In the middle, the corresponding final configuration where the edges are decorated with the flux of cars (when non-zero). When pushing up the labels to the above vertices, one obtains a labeled tree which follows the same local rules as the peeling tree of quadrangulations (right).

Quadrangulations. We introduce Q_n to be the number of rooted quadrangulations (without boundary) with n vertices, and we denote by $\mathfrak{Q}$ its corresponding generating function

$$\mathfrak{Q}(x) := \sum_{n\geqslant 2} Q_n x^n = x^2 + 2x^3 + 9x^4 + 54x^5 + 378x^6 + \cdots .$$

We already encountered $\mathfrak{Q}$ as the generating function of quadrangulations with a boundary of perimeter 2. Indeed we will see later that we can transform the root edge of a quadrangulation without boundary to get (bijectively) a quadrangulation with boundary of degree 2. By convention, the map with two vertices linked by an edge is considered as a quadrangulation (without boundary) with 0 face, which explains the term x^2.

Theorem 8.1

Writing $\mathfrak{Q}^\bullet = x\mathfrak{Q}'(x)$, the "last car decomposition" translates into the equation

$$\mathfrak{Q}^\bullet = 2x^2 + 6x\left(\frac{\mathfrak{Q}^\bullet - \mathfrak{Q}}{1 - \mathfrak{Q}^\bullet/x}\right), \tag{8.2}$$

which characterizes $\mathfrak{Q}$ and which is equivalent to the following recursive equation: $Q_2 = 1$ and for $n \geqslant 3$,

$$nQ_n = \sum_{k=2}^{n-1} k(n+1-k)Q_k Q_{n+1-k} + (4n-10)Q_{n-1}. \tag{8.3}$$

Notice that Equation (8.2) or equivalently (8.3) does not make use of quadrangulations with a boundary. They are fundamentally different from those of Tutte (8.1) which necessitated the introduction of a *catalytic variable* y to write the equation on the bivariate generating function of quadrangulations with a boundary (see $\mathbf{Q}$ bellow) and eventually characterize $\mathfrak{Q}$ alone. The form of the equations above may remind the knowledgeable readers the decomposition obtained via the KP hierarchy, see for example [134, Corollary 2] or [56]. But it seems that those equations can not be deduced one from another. Indeed with our notation, the equation coming from the KP hierarchy is

$$\mathfrak{Q}^\bullet - \mathfrak{Q} = 4x(2\mathfrak{Q}^\bullet - 3\mathfrak{Q}) + 3(2\mathfrak{Q}^\bullet - 3\mathfrak{Q})^2 + x^2. \tag{8.4}$$

Our decomposition also allows us to deduce a recursive decomposition of quadrangulations with a boundary. Let $Q_n^{(p)}$ be the number of rooted quadrangulations with a boundary of length $2p$ and n vertices, and we denote by $\mathbf{Q}$ its corresponding bivariate generating function

$$\mathbf{Q}(x,y) := \sum_{n\geqslant 1}\sum_{p=0}^{n} Q_n^{(p)} x^n y^p = x + y\left(x^2 + 2x^3 + 9x^4 + \cdots\right) + y^2\left(2x^3 + 9x^4 + 54x^5 + \cdots\right) + \cdots.$$

By convention, we consider the single isolated vertex as a planar quadrangulation with a boundary of perimeter 0 (and 0 face), which explains the term x. The last car decomposition gives the following differential equation on $\mathbf{Q}$

$$\mathbf{Q}^\bullet = x + 6y\mathbf{Q}^\bullet\left(\frac{\mathfrak{Q}^\bullet - \mathfrak{Q}}{1 - \mathfrak{Q}^\bullet/x}\right) + 2xy\left(3\mathbf{Q}^\bullet - 2\mathbf{Q} - y\partial_y\mathbf{Q}\right), \tag{8.5}$$

where $\mathbf{Q}^\bullet = x\partial_x\mathbf{Q}$. This equation is not as "simple" as Tutte's Equation (8.1) since it also involves the partial derivatives of the generating function $\mathbf{Q}$.

Lastly we can also consider quadrangulations with multiple boundaries that are enumerated by Tutte's slicing formula, see [72, Theorem 3.4]. We will only deal with the case of two boundaries at the end of Section 8.2 but the general case can be obtained by stacking and gluing the appropriate number of peeling trees with the same procedure.

Triangulations. An adaptation of the "last car" technique gives similar results in the case of triangulations. As above we start by the case without boundary, and the map with two vertices linked by an edge is considered as a triangulation with 0 face. We denote by T_n the number of rooted triangulations with n vertices without boundary and we denote its generating series by

$$\mathfrak{T}(x) = \sum_{n\geqslant 2} T_n x^n = x^2 + 4x^3 + 32x^4 + 336x^5 + \dots$$

Theorem 8.2

The last car decomposition yields the following equation

$$3\mathfrak{T}^\bullet - 4\mathfrak{T} = 2\left(\frac{\mathfrak{T}^\bullet - \mathfrak{T}}{1 - \mathfrak{T}^\bullet/x}\right), \tag{8.6}$$

where $\mathfrak{T}^\bullet = x\mathfrak{T}'(x)$. *Assuming* $T_2 = 1$, *this equation characterizes* $\mathfrak{T}$ *and is equivalent to the recursive equation which is similar to* (8.3)

$$T_n := \frac{1}{n-2}\sum_{k=2}^{n-1}(3k-4)(n+1-k)T_kT_{n+1-k}. \tag{8.7}$$

As above, this equation may remind the reader of the following equation coming from the KP hierarchy:

$$\mathfrak{T}^\bullet - \mathfrak{T} = (6\mathfrak{T}^\bullet - 8\mathfrak{T} + x)^2, \tag{8.8}$$

see [101, Theorem 5.4 and Equation 45]. However, we were not able to deduce one from another by simple computation. We now let $T_n^{(p)}$ be the number of rooted triangulations with a boundary of length p and n vertices in total, and we denote by $\mathbf{T}$ its corresponding generating function

$$\mathbf{T}(x,y) = \sum_{n\geqslant 1, p\geqslant 0} T_n^{(p)} x^n y^p = x + y(x^2 + 4x^3 + 32x^4 + \dots) + \dots.$$

Then our last car decomposition shows that $\mathbf{T}$ satisfies the following differential equation

$$6\mathbf{T}^\bullet - 2y\partial_y\mathbf{T} - 6\mathbf{T} + y\partial_y(y\mathbf{T}) = \frac{4y}{x}\mathbf{T}^\bullet\left(\frac{\mathfrak{T}^\bullet - \mathfrak{T}}{1-\mathfrak{T}^\bullet/x}\right) + y\left(4\mathbf{T}^\bullet - 3\mathbf{T} - y\partial_y\mathbf{T}\right). \tag{8.9}$$

where $\mathbf{T}^\bullet = x\partial_x\mathbf{T}$, which is the analogue of Equation (8.5). We will see later that a small local transformation on the triangulations shows that $\mathfrak{T} = [y^1]\mathbf{T} = [y^2]\mathbf{T} = [y^3]\mathbf{T}$. We can also adapt our decomposition in the case of triangulations with two boundaries, see the end of Section 8.3.

Acknowledgements. We acknowledge support from ERC 740943 *GeoBrown*. We thank Thomas Budzinski and Baptiste Louf for stimulating discussions. I am also grateful to Nicolas Curien for his precious suggestions.

8.2 "Last Car" decomposition of quadrangulations

We have given above an insight into the peeling exploration technique on quadrangulations with a boundary. Let us precisely explain the correspondence between quadrangulations with a boundary and their peeling tree. Recall that the removal of the root edge in a quadrangulation can produce two possible events: either it stays connected and the half-perimeter rises by 1, or it splits in two parts and the sum of the two half-perimeter is prescribed, see Figure 8.1.

Peeling tree of quadrangulations with one boundary. With each rooted quadrangulation with n vertices and with a boundary of perimeter $2p$, we can match bijectively its peeling tree, which is a labeled plane tree whose vertices have between zero and two children including n leaves. The labeling satisfies the following rules:

- The root vertex has label p.
- The leaves have label 0.
- ⋆ *(local rules)* All inner vertices have a label $\ell \geqslant 1$ and either one child with label $\ell + 1$ or two children with labels $\ell_1, \ell_2 \geqslant 0$ such that $\ell_1 + \ell_2 - 1 = \ell$. See Figure 8.2.

Note that the labels of the vertices can be seen as the half-perimeter of the successive boundaries during the exploration, and that the local rules imply that it has $2n - p - 2$ inner vertices, which is also the number of edges of the quadrangulation.

We will now concentrate on such peeling trees and (almost) forget about their interpretation in terms of maps. We want to apply a "last car" decomposition to such a tree. Let us first explain the influence of removing a car from a parking tree or fully parked tree. Recall that a fully parked tree is a rooted tree together with a car configuration where all parking spots are occupied (eventually with an outgoing flux) and that we can label each vertex by the flux of cars that go through the edge just below (or the outgoing flux for the root vertex). Imagine now that one removes a distinguished car from this tree, or more precisely that we first park all the cars but this one and then try to park this distinguished car so that we can easily remove it. Take for example a car which contributed to the outgoing flux. Then the effect of this removal is that the flux in the edges of the branch between the arriving spot of this car and the root vertex has decreased by 1, see Figure 8.4.

In the next sections, we will try to subtract 1 from the labels in a branch of a peeling tree. We will see that, for the inner vertices of the branch, this transformation preserves the local rules.

8.2.1 Quadrangulations without boundary or with a boundary of length 2.

To enumerate quadrangulations without boundary we actually first transform them into quadrangulations with a boundary of degree 2. To this end, the standard trick is to cut along the root edge and "open" it, see Figure 8.5. This transformation does not affect the number of vertices of the initial quadrangulation and is a bijection between quadrangulations without boundary and with a boundary of perimeter 2 with the same number of vertices. Since we only consider planar maps, Euler's formula implies that a quadrangulation without boundary has $2n - 4$ edges (and $n - 2$ faces), and applying the root transform only increases the number of edges and faces by 1.

Thanks to this trick, we now only need to concentrate on peeling trees of quadrangulations with a boundary of length 2, i.e. those whose root has label 1.
Inspired by removing the last car in fully parked trees [128], we want to remove 1 from all labels on a

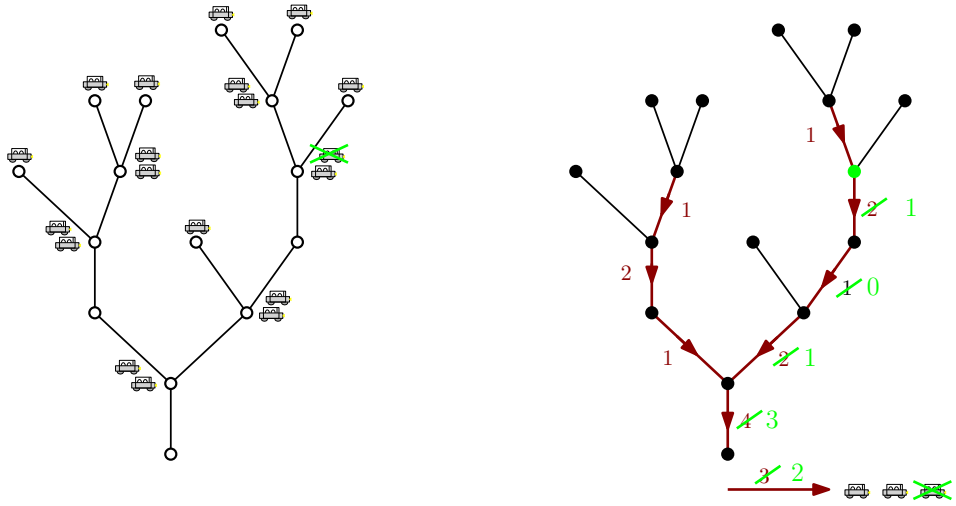


Figure 8.4: The effect of removing a car (seen as the last car) in a fully parked tree. We remove here the car in green on the left and imagine it parked last. On the right, it decreases by 1 the flux of the edges between the green point (location of the car arrival) and the root edge.

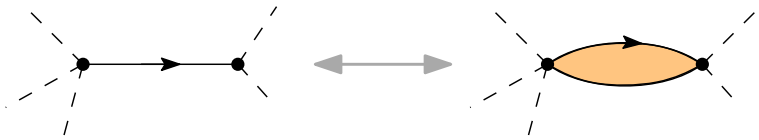


Figure 8.5: Root transform for quadrangulations (or more generally for bipartite maps). On the left the initial root edge and on the right the distinguished face of degree 2 together with the new root edge.

"branch" of the peeling tree. To do this, we need to distinguish a leaf of the tree, which corresponds to a vertex in the initial quadrangulation and we consider the branch between this leaf and the root vertex. The key observation is that subtracting 1 to all labels of this branch preserves the local rules of the tree which we described in ($\star$), but two issues may appear.

- First, the initial distinguished leaf cannot get a label -1, but it has a parent with some label $k \geqslant 1$ and a sibling with label $k-1$, so that we can just contract these three vertices into a vertex with label $k-1$. We distinguish it to remember where we removed the distinguished leaf. We also need to remember whether the leaf was on the left or on the right.

- The other possible issue is that the vertices with label 1 will get label 0 and therefore have to be leaves. This is also easily solved by cutting the edges just above the new 0's and distinguishing them to remind the location of the cuts, see Figure 8.6.

Lastly, when $n \geqslant 3$, the root vertex of the tree has always a single child with label 2 so that we

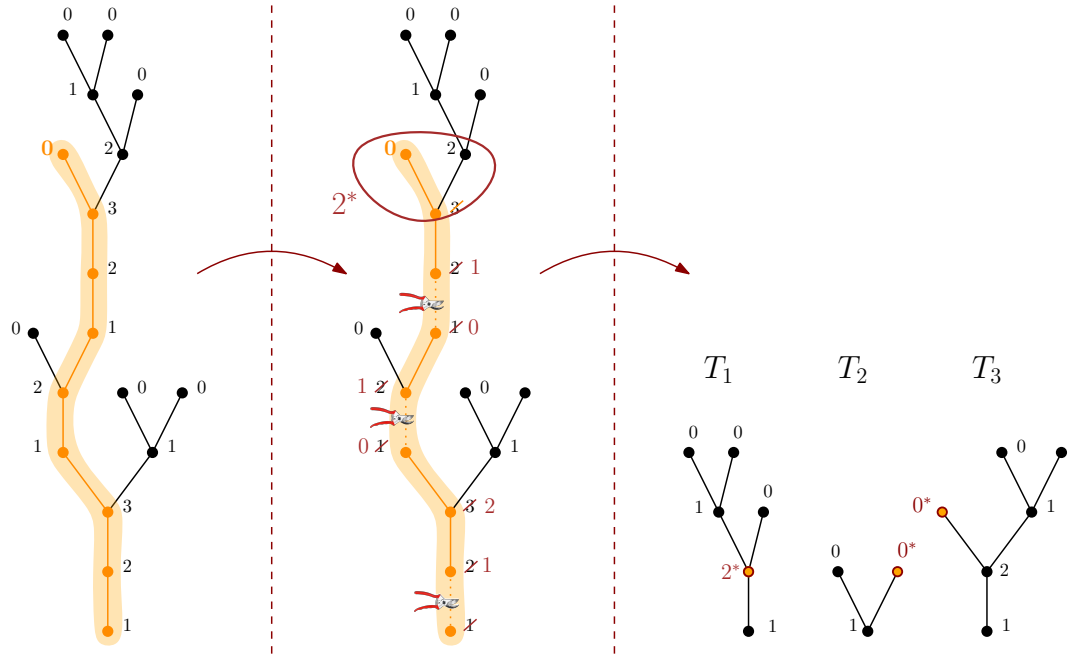


Figure 8.6: Example of the last car decomposition on a peeling tree of quadrangulation with $n = 7$ leaves into $k = 3$ trees with $n + k - 2 = 8$ leaves in total. On the left, the initial peeling tree where the distinguished leaf is displayed in orange as well as the branch between this leaf and the root. In the middle, we point out the needed transformations when removing one in this branch to preserve the local rules: the father and sibling of the distinguished leaf are contracted into a (marked) vertex labeled 2, we remove the initial root (or cut just above), and we cut above the two vertices which get label 0. On the right, the resulting sequence of three trees.

just remove the initial root and root the new tree at this child which gets label 1.

To summarize, our transformation converts a peeling tree starting from 1 with $n \geqslant 3$ leaves, marked at one leaf, into a sequence of k trees for some $k \geqslant 1$, with root label 1, with $n + k - 2$ leaves in total and such that the first tree has a distinguished vertex (leaf or inner vertex) and the eventual following trees have a marked leaf. Conversely, given such a sequence, we can recover the initial peeling tree by gluing successively the root of a tree with the distinguished leaf of the next tree of the sequence and add 1 to the labels in the appropriate branch. Noting that a quadrangulation with a boundary of perimeter 2 and n vertices has $2n - 3$ edges hence its peeling tree has $3n - 3$ vertices in total, we obtain the following differential equation on $\mathfrak{Q}$:

$$\mathfrak{Q}^\bullet = 2x^2 + 2x\left(\frac{3\mathfrak{Q}^\bullet - 3\mathfrak{Q}}{1 - \mathfrak{Q}^\bullet/x}\right),$$

which is Equation (8.2). Let us explain this in more details:

- The term $2x^2$ stands for the map with two vertices connected by the root edge, which is a

quadrangulation by convention.

- The factor $2x$ comes from the fact that there are two possible ways to "reglue" the initial marked leaf with label 0 on the first tree of the transformed sequence (left or right).
- The factor $3\mathfrak{Q}^\bullet - 3\mathfrak{Q}$ comes from the first tree in the sequence, which has a distinguished vertex, and is divided by $1 - \mathfrak{Q}^\bullet/x$ for all possible cuts, see SEQ construction, for example in [92, Theorem III.1 p.166].

By identifying the coefficient in the equation above, we obtain Equation (8.3) which concludes the proof of Theorem 8.1. This is a new equation which characterizes A000168 in Sloane online encyclopedia for integer sequences.

8.2.2 Quadrangulations with a boundary of degree $2p \geqslant 4$

The transformation which we described above also works for quadrangulations with a boundary of perimeter $2p \geqslant 4$. The only difference is this in that case, we do not remove the initial root of the tree. Starting from a peeling tree with $n \geqslant 3$ leaves, a root labeled $p \geqslant 2$ and a distinguished leaf, we apply our last car decomposition in the branch between the distinguished leaf and the root of the tree. We obtain a sequence of $k \geqslant 1$ peeling trees for some $k \geqslant 1$ such that:

- the first tree is marked at a vertex (leaf or not),
- all eventual following trees have a marked leaf,
- all but the last tree have a root label 1 and the last tree has a root labeled $p - 1 \geqslant 1$,
- the trees have $n + k - 2$ leaves with label 0 in total.

See Figure 8.7. Note that the first and the last tree can be confounded. We then get the differential Equation (8.5) on $\mathbf{Q}$ that we recall here:

$$\mathbf{Q}^\bullet = x + 6y\mathbf{Q}^\bullet \left(\frac{\mathfrak{Q}^\bullet - \mathfrak{Q}}{1 - \mathfrak{Q}^\bullet/x} \right) + 2xy \left(3\mathbf{Q}^\bullet - 2\mathbf{Q} - y\partial_y \mathbf{Q} \right),$$

where $\mathbf{Q}^\bullet = x\partial_x \mathbf{Q}$ and $\mathfrak{Q} = [y^1]\mathbf{Q}$, (resp. $\mathfrak{Q}^\bullet = [y^1]\mathbf{Q}^\bullet$). Indeed,

- The term x corresponds to the map composed of a single vertex and no edge whose peeling tree is simply the tree with one vertex labeled 0.
- The second term corresponds to the case where our transformation gives more than one tree (and in particular, the first and the last trees are not the same), including the case of quadrangulations with a boundary of perimeter 2 where the last tree after transformation is the tree with one vertex labeled 0. In that case, the last tree is a tree with root label $p - 1$ and a marked leaf, which corresponds to the factor $y\mathbf{Q}^\bullet$. The first tree has a distinguished vertex and its root has label 1, hence the factor $3\mathfrak{Q}^\bullet - 3\mathfrak{Q}$ as in the case of a boundary of perimeter 2. The

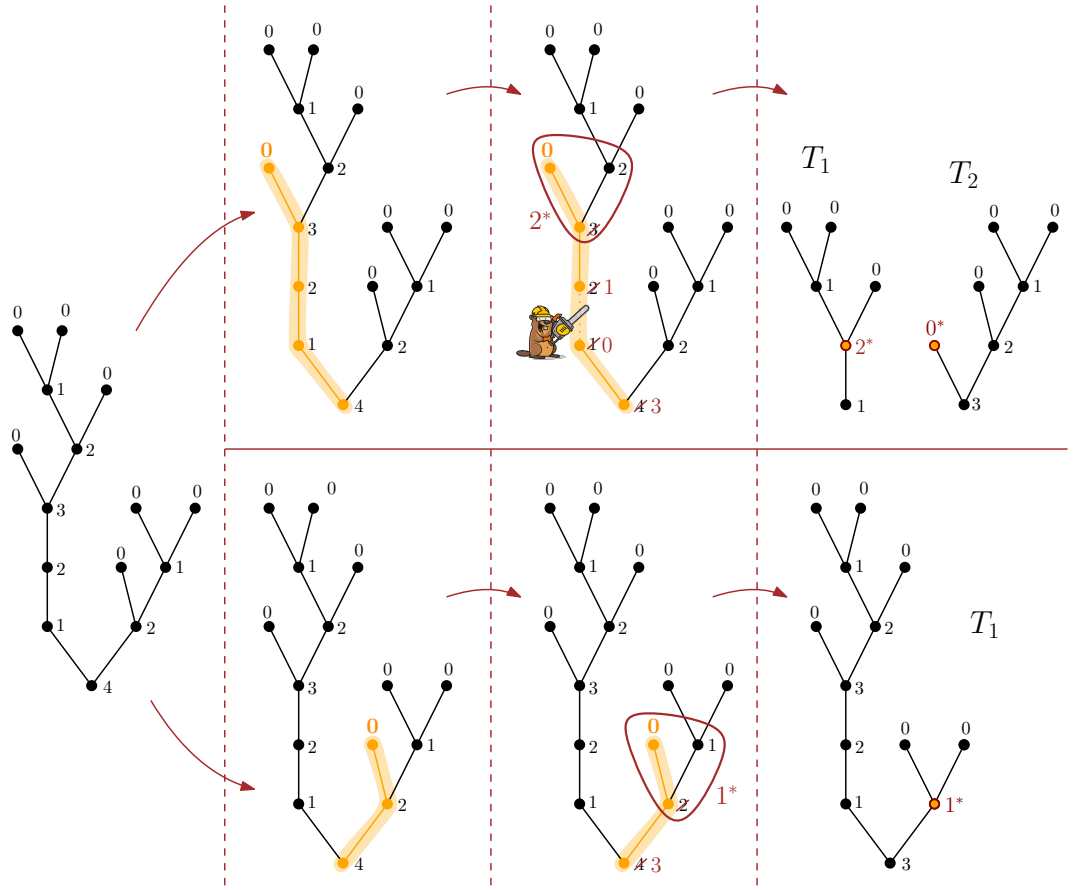


Figure 8.7: Two examples of our transformation on the same tree with root labeled $p = 4$ and $n = 7$ leaves with two different distinguished leaves. On the top, the transformation produces $k = 2$ trees and have $n - k - 2 = 7$ leaves in total. The first one has a distinguished vertex and root labeled 1, and the second one has a distinguished leaf and root labeled $p - 1 = 3$. At the bottom, we get $k = 1$ tree with $n - k - 1 = 6$ leaves, a distinguished vertex and a root labeled $p - 1 = 3$. In particular, the first and last tree are the same.

factor $1/(1 - \mathfrak{Q}^\bullet/x)$ stands for the other possible cuts. There is also a factor 2 which comes from the fact that there are two possible ways to reglue the initial distinguished leaf labeled 0 on the first tree.

- The last term corresponds to the quadrangulations with $n \geqslant 2$ vertices where we get only one tree when applying our transformation on its peeling tree. In that case the resulting tree has $n - 1$ leaves, a root labeled $p - 1$ hence $2(n-1) - (p-1) - 2 =$ inner vertices (edges in the quadrangulation) and $3(n-1) - 2 - (p-1)$ vertices in total, one of them is distinguished, explaining the factor $3\mathbf{Q}^\bullet - 2\mathbf{Q} - y\partial_y\mathbf{Q}$.

8.2.3 Quadrangulations with two boundaries

We can also give a recursive equation for the number of quadrangulations with two boundaries. We interpret the first boundary as above but the second boundary as a distinguished face of perimeter $2q$, and denote by $Q_n^{(p,q)}$ the number of rooted quadrangulations with a boundary of length $2p$, a distinguished face of perimeter $2q$ and n vertices in total. In the peeling exploration, the discovery of the distinguished face corresponds to the event of seeing a boundary of half-perimeter $r \geqslant 1$ which becomes a boundary of half-perimeter $r+q-1$ when removing an edge. This matches in the peeling tree to a vertex with label $r \geqslant 1$ which gives birth to a vertex labeled $r+q-1$ when discovering this distinguished face. The rest of the local transitions are exactly the same as before, see Figure 8.8. Thus, we introduce $A_n^{(p,r)}$ the number of peeling trees of "quadrangulations" with n leaves labeled 0, a distinguished leaf with label r and a root labeled p.

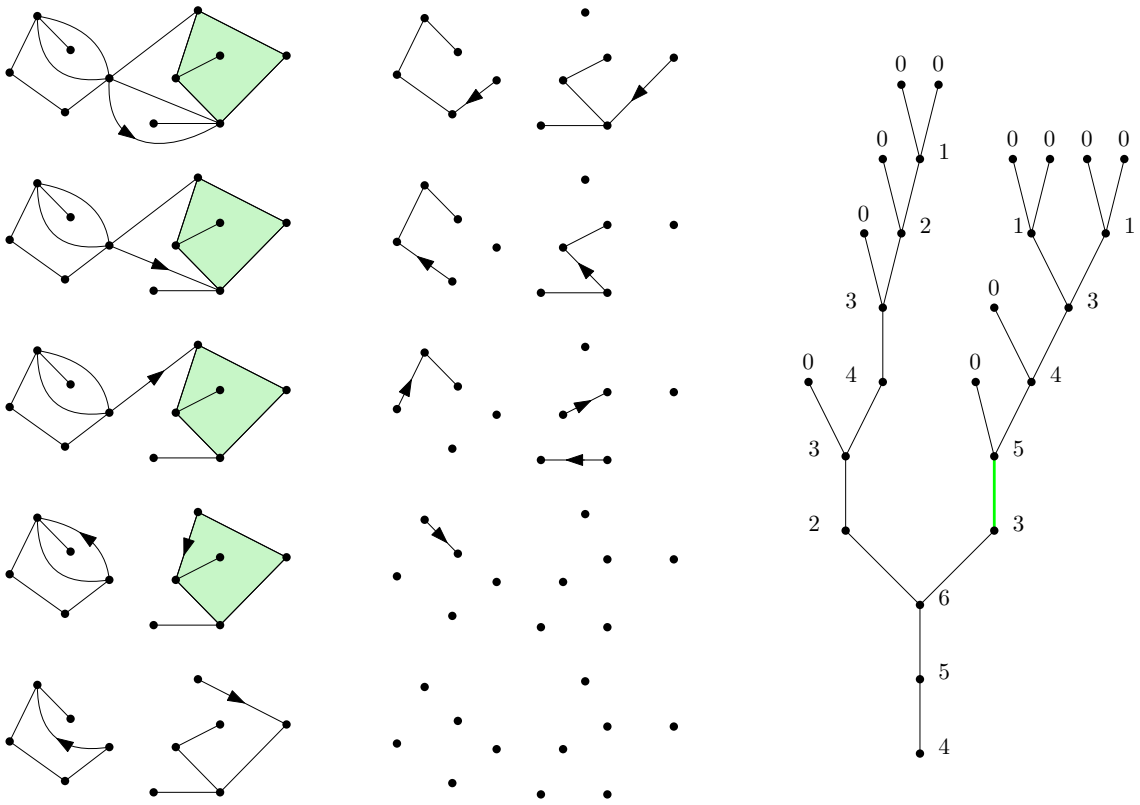


Figure 8.8: Step-by-step peeling exploration of a quadrangulation with a boundary of perimeter $8 = 2 \cdot 4$ and a distinguished hexagon in green. The discovery of this face leads to a transition from a boundary of half-perimeter 3 to one with half-perimeter $3+3-1 = 5$ (green edge in the peeling tree on the right). Removing this green edge in the peeling tree on the right, the above part is a usual peeling tree of a quadrangulation with a boundary of half-perimeter 5 and 6 leaves. The bottom part starts from a label 4, has 5 leaves with label 0 and a distinguished leaf with label 3.

Indeed, if we decompose the peeling tree of a quadrangulation with two boundaries of length $2p$ and $2q$ according to this different transition, then for some $r \geqslant 1$, the above part is the usual peeling tree of a quadrangulation with one boundary of length $r+q-1$ and the other one has root label p and a distinguished leaf with label r. And the total number of leaves with label 0 should be n as the quadrangulation had n vertices. With this decomposition, we obtain

$$Q_n^{(p,q)} = \sum_{r=1}^{n} \sum_{k=0}^{n-r-q} A_k^{(p,r)} Q_{n-k}^{(r+q-1)}. \tag{8.10}$$

Knowing the $A_n^{(p,q)}$'s, this equation allows us to deduce the $Q_n^{(p,q)}$. Note that the sum on r when $k=0$ encompasses the case where the bottom part has only one leaf with label r and is just a straight vertical line (thus $A_0^{(p,r)}$ equals 1 if $r \geqslant p$ and 0 otherwise).

It only remains to enumerate peeling trees with a distinguished leaf with label $r \geqslant 1$. For this, we can use our last car decomposition and subtract 1 along the branch to the distinguished leaf. In this case, we do not need to transform the father and the sibling of the distinguished leaf. When $r = 1$, then removing 1 in the branch between this leaf and the root of the tree creates a leaf with label 0 and thus a usual peeling tree of a quadrangulation with one boundary. Hence we have $A_n^{(p,0)} = (n+1)Q_{n+1}^{(p)}$ when $n \geqslant \mathsf{max}(1, p-1)$.
When $r \geqslant 2$, the distinguished leaf just get label $r-1 \geqslant 1$. We then obtain the following recursive equation, which we explain bellow:

$$\forall r \geqslant 1, \forall p \geqslant 1, \forall n \geqslant 1, \quad A_n^{(p,r)} = A_n^{(\max(p-1,1),r-1)} + \sum_{k=0}^{n-1} A_k^{(1,r-1)} A_{n-k}^{(p,1)}. \tag{8.11}$$

The first term on the right corresponds to the case where there is no 1 in the branch so that there is no cut in the decomposition. When $p = 1$, we remove the initial root whereas we do not remove it when $p \geqslant 2$ so that the root of the new tree has label $p-1$, which explained the index $\mathsf{max}(p-1,1)$. The last term corresponds to the case where there is a cut and decompose the initial tree according to this highest cut. In that case, we only subtract 1 in the top part, since the bottom part is just a smaller peeling tree with a distinguished leaf labeled 1. Notice that when $k=0$, the only possibility to have no leaf labeled 0 is to have a straight-line from label 1 to label $r-1$ so that $A_0^{(1,r-1)} = 1$ when $r \geqslant 2$.

8.3 "Last Car" decomposition of triangulations

We now apply our techniques to the case of triangulations. We first describe the encoding of triangulations by their peeling trees which is similar to the case of quadrangulations.

8.3.1 Peeling triangulations

The peeling technique can be adapted to triangulations. Indeed, we can also remove the edges one-by-one "à la Tutte" to get recursive equations on the number of triangulations with a boundary.

As in the case of quadrangulations, two possible events can occur when erasing the root edge of a triangulation with a boundary $p \geqslant 1$:

- either the triangulation stays connected, which means that one discovered a new face of the initial triangulation, and one re-roots the triangulation at the left-most edge of the new face. The new triangulation has then a boundary of length $p+1$.
- or the deletion of the root edge disconnects the triangulation and one gets two triangulations with perimeter $p_1 \geqslant 0$ and $p_2 \geqslant 0$ such that $p_1 + p_2 + 2 = p$, that one can re-root easily using the two endpoints of the removed edge as in the case of quadrangulations, see Figure 8.9.

Here, we do not consider the half-perimeter of the boundary since the triangulations are not bipartite. As for quadrangulations, we can encode this edge-by-edge peeling exploration in a peeling tree by recursively labeling the vertices with the perimeter of the boundaries, see Figure 8.9. Note in particular that when the removal of the root edge disconnects the triangulation, we put the component attached to the origin of the root to the left in the peeling tree. To fix ideas, the peeling process gives us the following correspondence between triangulations and trees.

Peeling trees of triangulations with one boundary. To each rooted triangulation with n vertices and with a boundary of perimeter p, we can bijectively match its peeling tree which is a plane tree with labeled vertices with zero, one or two children encoding recursively the perimeter of the boundaries in the peeling exploration. The labeling satisfies:

- the tree has n leaves with label 0,
- its root has label p,
- *(local rules)* Each vertex with label $\ell \geqslant 1$ has either a child with label $\ell+1$ or two children with label ℓ_1 and ℓ_2 such that $\ell_1 + \ell_2 + 2 = \ell$. See Figure 8.9.

In particular, the local rules imply that such tree has $3n - p - 3$ inner vertices.

As for the case of quadrangulations, it will be convenient to see triangulations of the sphere as triangulations with a boundary. The most natural idea would be to see a triangulation of the sphere as a triangulation with a boundary of perimeter 2 after unzipping the root edge or 3 if we see the triangle lying on the right of the root edge as the external face. It implies $[y^2]\mathbf{T} = [y^3]\mathbf{T}$. But since the triangulations are not bipartite, the root edge can be a loop and it will be more convenient to use another root transform. We shall actually view triangulations without boundary as triangulations with a boundary of length 1 (i.e. a loop): we cut along the root edge and "open" it to get a double edge, and then insert a loop inside this double edge at the starting point of the initial root edge to obtain a triangle and root the new triangulation on this loop in clockwise direction so that the new triangulation has the 1-gon to its right, see Figure 8.10. We also obtain $[y^1]\mathbf{T} = [y^2]\mathbf{T} = [y^3]\mathbf{T}$.

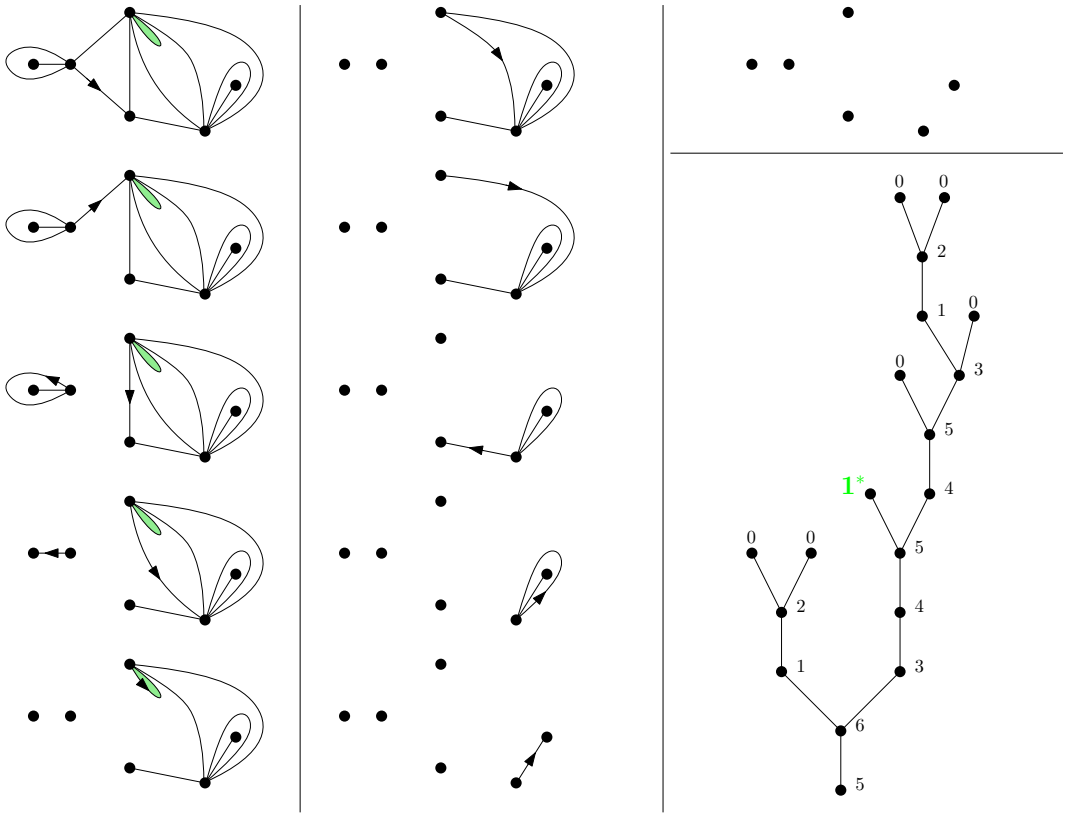


Figure 8.9: Step-by-step peeling exploration of a triangulation with a boundary of perimeter $p = 5$, with $n = 6$ vertices and a distinguished loop. This loop in green matches in the tree with the green leaf labeled 1, which we can put indifferently left or right.

The choice of the clockwise orientation of the new root edge on the new loop is canonical since we imposed that the maps with a boundary have their distinguished face to their right.

In fact, we can apply this root-transformation on any distinguished oriented edge even in triangulations which already have a boundary. We will apply our decomposition to peeling trees of triangulations with a boundary and a distinguished face of degree 1 or *loop*. When the distinguished loop does not lie on the boundary, we can "invert" the root transform and see the distinguished loop as coming from a distinguished oriented edge. To summarize, our decomposition will be based on the following trees.

Peeling trees of triangulations with one boundary and a distinguished loop. To each rooted triangulation with n vertices and with a boundary of perimeter p and a distinguished loop, we can bijectively match its peeling tree which is a plane tree with labeled vertices with between 0 and 2 children encoding recursively the perimeter of the boundaries in the peeling exploration. The labeling satisfies:

- the tree has n leaves with label 0,
- its root has label p,
- one inner vertex with label $\ell \geqslant 1$ has two children, one of which is a leaf with label 1 and the other one hase label $\ell - 1$. For this leaf (and only for this leaf !), the planar ordering does not matter, see Figure 8.9.
- *(local rules)* All other vertices with label $\ell \geqslant 1$ has either a child with label $\ell + 1$ or two children with label ℓ_1 and ℓ_2 such that $\ell_1 + \ell_2 + 2 = \ell$. See Figure 8.9.

In particular, the local rules imply that it has $3n - p - 3$ inner vertices.

It will then be more convenient to apply our last car decomposition in the branch between this leaf with label 1 and the root of the tree.

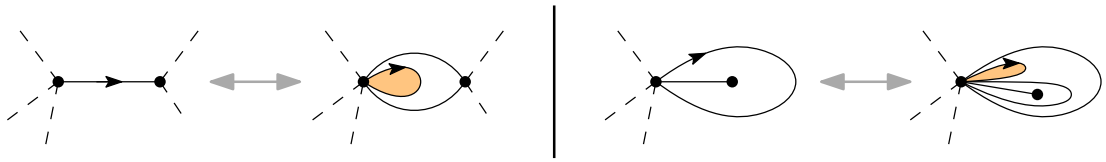


Figure 8.10: Root transform for triangulations: on the right, the case where the initial root edge is a loop and on the left, the case where the two endpoints of the root edge are different. In both cases, the initial triangulation is on the left and on the right, we obtain a triangulation with a boundary of perimeter 1.

8.3.2 Triangulations without boundary

We first want to enumerate triangulations without boundary, and to do this, we apply the above root-transform which gives a bijection between rooted triangulations without boundary and triangulations with a boundary of length 1 and preserves the number of vertices. Instead of distinguishing a vertex to apply our decomposition, we shall this time distinguish an oriented edge, to which we apply another time the above root transform to obtain a triangulation with a boundary of perimeter 1 (obtained from the transformation of the initial root edge) and with a distinguished loop (obtained from the additional distinguished oriented edge). Note in particular that the oriented distinguished

edge can be the root edge, but will be different after applying the root transform for the second time. Therefore the distinguished loop can not lie on the boundary and its edge is part of an internal face so that we can inverse the root transform to recover the initial distinguished oriented edge.

We then build its peeling tree to which we can apply our transformation, i.e. subtract 1 in the whole branch between the distinguished leaf labeled 1 and the root of the tree. The transformation works then as in the case of quadrangulations: the local rules (◦) are preserved but two issues may appear.

- The initial distinguished leaf with label 1 has a parent with label $k \geqslant 1$ and a sibling with label $k-1$, so that we just contract these three vertices into a single marked vertex with label $k-1$.
- if a vertex labeled 1 becomes a vertex with label 0, then we cut the edge just above it; and we remove the initial root which had label 1.

The transformation then takes a tree as described above and maps it into a sequence of k trees with root labeled 1 and such that the first tree has a distinguished vertex (leaf or not), each possible other tree has a distinguished leaf, and the trees have $n+k-1$ leaves in total, see Figure 8.11.

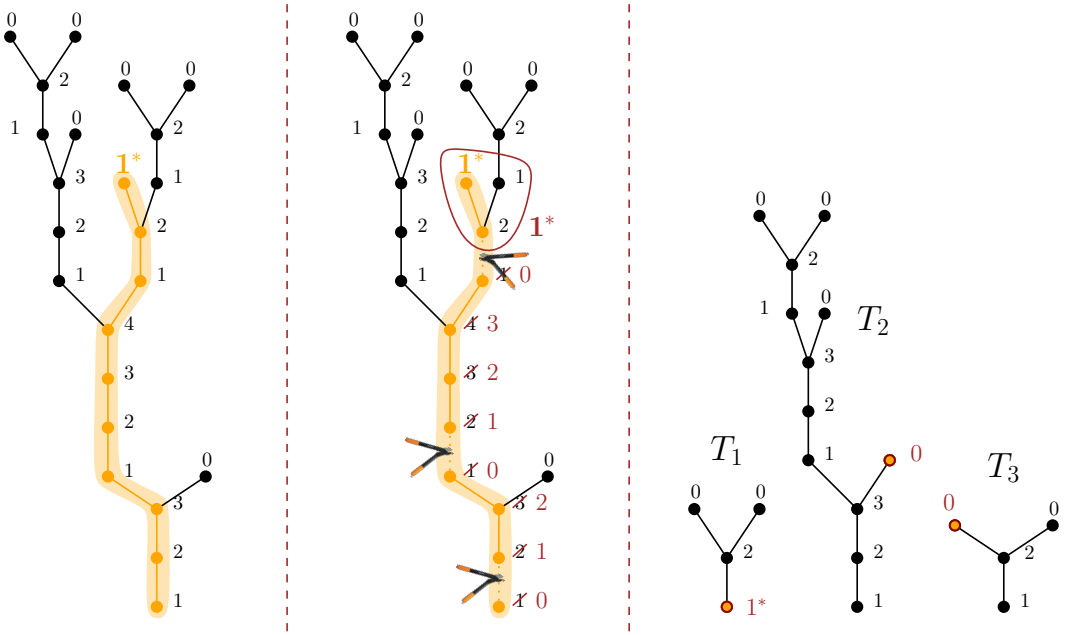


Figure 8.11: Example of our decomposition on a tree with $n = 6$ leaves labeled 0 into $k = 3$ trees with $n+k-1 = 8$ leaves in total. On the left, the initial peeling tree where the distinguished 1-leaf is displayed in orange as well as the branch between this leaf and the root. After removing 1 on all labels of the orange branch (in the middle), we need to remove the initial root (or cut just above), cut above the two vertices which get label 0 and contract the father and sibling of the distinguished leaf into a (marked) vertex labeled 1. On the right, the resulting sequence of three trees.

This gives the following differential equation on $\mathfrak{T}$

$$6\mathfrak{T}^\bullet - 8\mathfrak{T} = 4\left(\frac{\mathfrak{T}^\bullet - \mathfrak{T}}{1 - \mathfrak{T}^\bullet/x}\right)$$

which is equivalent to Equation (8.6). Let us explain this equality:

- On the left, since a triangulation of the 1-gon has $3n-4$ edges, the function $6\mathfrak{T}^\bullet - 8\mathfrak{T}$ is the generating function of triangulations with a boundary of length 1 given with an oriented edge.
- On the right, the first tree of our transformation is a tree which has n_1 leaves for some $n_1 \geqslant 1$ hence $3n_1 - 1 - 3$ inner vertices and $4n_1 - 4$ vertices in total, one of them is distinguished which gives the factor $4\mathfrak{T}^\bullet - 4\mathfrak{T}$.
- The factor $1/(1 - \mathfrak{T}^\bullet/x)$ gives the potential other trees (SEQ construction) which are marked at a leaf.

The equation above can rewritten as $\mathfrak{T}^\bullet - 2\mathfrak{T} = \mathfrak{T}^\bullet(3\mathfrak{T}^\bullet - 4\mathfrak{T}))/x$. By identifying the coefficients, a straightforward computation gives Equation (8.7), which characterizes entry A002005 in Sloane online encyclopedia for integer sequences.

8.3.3 Triangulations with one boundary

As in the case of quadrangulations, the transformation which we described above can be adapted to triangulations with a boundary of perimeter $p \geqslant 2$. More precisely, we consider peeling trees of triangulations with a boundary of length $p \geqslant 2$ and one distinguished loop that we described in Section 8.3.1. Such peeling trees have a distinguished leaf with label 1 coming from the distinguished loop for which we recall its local rule: if its father has label ℓ, its sibling has label $\ell - 1$. We can apply now our "last car"-transformation on the branch between this leaf labeled 1 and the root of the peeling tree. The only difference with that of peeling trees starting from 1 is that we do not remove the root of the trees starting from $p \geqslant 2$. Our last car decomposition of a peeling tree starting from p with n leaves labeled 0 produces then a sequence of k peeling trees of triangulations for some $k \geqslant 0$ such that:

- the first tree is marked at a vertex (leaf or not)
- all eventual following trees have a marked leaf
- all but the last tree have a root label 1 and the last tree has a root labeled $p - 1 \geqslant 1$
- the trees have $n + k - 1$ leaves labeled 0 in total.

Conversely, given the sequence, we can recover the initial tree by stacking and gluing the trees and adding 1 in the appropriate branch so that it is really a bijection, which gives Equation (8.9), which we recall here:

$$6\mathbf{T}^\bullet - 2y\partial_y\mathbf{T} - 6\mathbf{T} + y\partial_y(y\mathbf{T}) = \frac{4y}{x}\mathbf{T}^\bullet\left(\frac{\mathfrak{T}^\bullet - \mathfrak{T}}{1 - \mathfrak{T}^\bullet/x}\right) + y\left(4\mathbf{T}^\bullet - 3\mathbf{T} - y\partial_y\mathbf{T}\right).$$

where $\mathbf{T}^\bullet = x\partial_x\mathbf{T}$. Let us explain this equation in more details:

- The left-hand side enumerates triangulations with a boundary of length $p \geqslant 2$ and one distinguished loop with a weight x per vertex and y per boundary length. If the loop is on the boundary, then we can map the triangulation to a triangulation with perimeter $p-1$ and a distinguished vertex on the boundary where the initial loop is attached. Therefore, this type of triangulation are enumerated by the factor $y\partial_y(y\mathbf{T})$. When the loop is not on the boundary, then this edge is really an edge of an inner triangle so that we can perform the converse of the root transform and obtain a distinguished oriented edge. Since there are $3n-p-3$ (unoriented) edges in a triangulation with n vertices and boundary of length p, these are enumerated by the term $6\mathbf{T}^\bullet - 2y\partial_y\mathbf{T} - 6\mathbf{T}$, which explained the left-hand side term.

- On the right, the first term corresponds to the case where our transformation gives more than a tree (and in particular, the first and the last tree are not the same), including the case of triangulation with boundary 1 where the last tree after transformation is the tree with one vertex labeled 0. The last tree is in that case a tree with root label $p-1$ and a marked leaf, which corresponds to the factor $y\mathbf{T}^\bullet$. The first tree has a distinguished vertex and its root has label 1, hence the factor $4\mathfrak{T}^\bullet - 4\mathfrak{T}$, and the division by $(1-\mathfrak{T}^\bullet/x)$ stands for the other possible cuts. There is no factor 2 in that case since the discovery of the marked loop can be indifferently put left or right in the tree, but a factor $1/x$ to obtain the appropriate total number of leaves.

- The last term corresponds to the triangulations where we get only one tree when applying our transformation on its peeling tree. In that case the resulting tree has $n-1$ leaves, a root labeled $p-1$ hence $3(n-1)-(p-1)-3$ inner vertices in the tree and $4(n-1)-3-(p-1)$ vertices in total, one of them is distinguished explaining the factor $4\mathbf{T}^\bullet - 3\mathbf{T} - y\partial_y\mathbf{T}$.

8.3.4 Triangulations with two boundaries

As in the case of quadrangulations, we are also able to enumerate triangulations with two boundaries. Considering the second boundary as a distinguished face, we can do the same decomposition according to the discovery of this different face. The discovery of this distinguished face corresponds in the peeling tree to a transition from a vertex labeled $r \geqslant 1$ to a vertex labeled $r+q-2$ since we consider here the perimeter (and not the half-perimeter).
We introduce for this purpose $B_n^{(p,r)}$ the number of peeling trees of triangulations with n leaves 0 and a distinguished leaf with label r and boundary of length p. When $r=0$, we set $B_n^{(p,0)} = (n+1)T_n^{(p)}$ when $n \geqslant 1$. The last car decomposition leads to the following recursive equations:

$$\forall r \geqslant 1, \forall p \geqslant 1, \forall n \geqslant 1, \quad B_n^{(p,r)} = B_n^{(\max(p-1,1),r-1)} + \sum_{k=0}^{n-1} B_k^{(1,r-1)} B_{n-k}^{(p,1)}.$$

This equation are very similar to Equation (8.11). We only need to adapt the initial conditions in the $B_n^{(p,0)}$'s. Given the $B_n^{(p,r)}$'s, we are now able to deduce the number $T_n^{(p,q)}$ of rooted triangulations with

n vertices and with a boundary of length p and a distinguished face of length q from the equation

$$T_n^{(p,q)} = \sum_{r=1}^{2n-q} \sum_{k=0}^{n-\lfloor \frac{r+q}{2} \rfloor + 1} B_k^{(p,r)} T_{n-k}^{(r+q-2)},$$

which is the analog of Equation (8.10) for quadrangulations with two boundaries. The only changes are that we replaced the initial conditions and adjusted the bounds of the sums. Notice that we already enumerated the the triangulation with one boundary and a distinguished loop (counted by the $T_n^{(p,1)}$'s) by $6\mathbf{T}^\bullet - 2y\partial_y\mathbf{T} - 6\mathbf{T} + y\partial_y(y\mathbf{T})$ the left-hand side of (8.9).

8.4 Comments and perspectives

We mention here a few possible developments of this work.

Other models of maps. We applied here a "last car decomposition" in the case of quadrangulations and triangulations with zero, one or two boundaries. We believe that this decomposition can be adapted for other models of planar maps such as p-angulations (at least for p even) or (bipartite) maps with Boltzmann weights.

Solving equations. We gave here new equations which characterize the enumeration of a certain type of maps. Some of them were already explicitly enumerated (quadrangulations with zero, one or two boundaries, triangulations without boundary...) so it can be easily checked that their generating functions satisfy our equations. But we may wonder if we can recover those coefficients directly by solving our equations explicitly. This is also a relevant question in the cases when no explicit formula is known: can we extract from these equations explicit formulas for the coefficients?

Decomposition of maps. Here we gave a decomposition of the peeling tree into smaller peeling trees, i.e. peeling trees of "smaller" maps. We do not know if this decomposition can be easily interpreted on the maps. In particular, we choose a specific way to reroot the maps in the peeling exploration, but there are many ways to do it by choosing different *peeling algorithm*. There may be a choice of peeling algorithm for which the transformation on the maps is "natural". We have no hope that it is local but it may have similarities with the cut and slice operation of Louf [134].

Bibliography

[1] R. Abraham and J.-F. Delmas, *Local limits of conditioned galton-watson trees: the condensation case*, Electronic Journal of Probability, 56 (2014), pp. Article–56.

[2] ——, *Local limits of conditioned Galton-Watson trees: the infinite spine case*, Electron. J. Probab., 19 (2014), p. 19 pp.

[3] L. Addario-Berry, *A probabilistic approach to block sizes in random maps*, ALEA Lat. Am. J. Probab. Math. Stat, 16 (2019), pp. 1–13.

[4] L. Addario-Berry, N. Broutin, C. Goldschmidt, and G. Miermont, *The scaling limit of the minimum spanning tree of the complete graph*, (2013).

[5] M. Aizenman and D. J. Barsky, *Sharpness of the phase transition in percolation models*, Communications in Mathematical Physics, 108 (1987), pp. 489–526.

[6] D. Aldous, *The random walk construction of uniform spanning trees and uniform labelled trees*, SIAM J. Discrete Math., 3 (1990), pp. 450–465.

[7] ——, *Asymptotic fringe distributions for general families of random trees*, Ann. Appl. Probab., 1 (1991), pp. 228–266.

[8] ——, *The continuum random tree. I*, Ann. Probab., 19 (1991), pp. 1–28.

[9] ——, *The continuum random tree III*, Ann. Probab., 21 (1993), pp. 248–289.

[10] ——, *Brownian excursions, critical random graphs and the multiplicative coalescent*, The Annals of Probability, (1997), pp. 812–854.

[11] ——, *The percolation process on a tree where infinite clusters are frozen*, in Mathematical Proceedings of the Cambridge Philosophical Society, vol. 128, 2000, pp. 465–477.

[12] D. Aldous, A. Contat, N. Curien, and O. Hénard, *Parking on the infinite binary tree*, Probability Theory and Related Fields, (2023), pp. 1–24.

[13] D. Aldous, G. Miermont, and J. Pitman, *Brownian bridge asymptotics for random p-mappings*, Electronic Journal of Probability, 9 (2004), pp. 37–56.

[14] D. Aldous and J. M. Steele, *The objective method: probabilistic combinatorial optimization and local weak convergence*, in Probability on discrete structures, vol. 110 of Encyclopaedia Math. Sci., Springer, Berlin, 2004, pp. 1–72.

[15] O. Angel, *Growth and percolation on the uniform infinite planar triangulation*, Geom. Funct. Anal., 13 (2003), pp. 935–974.

[16] O. Angel and O. Schramm, *Uniform infinite planar triangulation*, Comm. Math. Phys., 241 (2003), pp. 191–213.

[17] I. Armendáriz, *Dual fragmentation and multiplicative coagulation*, Unpublished preprint, (2005).

[18] I. Armendáriz and M. Loulakis, *Conditional distribution of heavy tailed random variables on large deviations of their sum*, Stochastic processes and their applications, 121 (2011), pp. 1138–1147.

[19] J. Aronson, A. Frieze, and B. G. Pittel, *Maximum matchings in sparse random graphs: Karp–sipser revisited*, Random Structures & Algorithms, 12 (1998), pp. 111–177.

[20] S. Asmussen, S. Foss, and D. Korshunov, *Asymptotics for sums of random variables with local subexponential behaviour*, Journal of Theoretical Probability, 16 (2003), pp. 489–518.

[21] R. Bahl, P. Barnet, T. Johnson, and M. Junge, *Diffusion-limited annihilating systems and the increasing convex order*, Electronic Journal of Probability, 27 (2022), pp. 1–19.

[22] R. Bahl, P. Barnet, and M. Junge, *Parking on supercritical Galton-Watson trees*, ALEA, 18 (2021), pp. 1801–1815.

[23] C. Banderier, P. Flajolet, G. Schaeffer, and M. Soria, *Random maps, coalescing saddles, singularity analysis, and Airy phenomena*, Random Structures & Algorithms, 19 (2001), pp. 194–246.

[24] M. Bauer and O. Golinelli, *Core percolation in random graphs: a critical phenomena analysis*, The European Physical Journal B-Condensed Matter and Complex Systems, 24 (2001), pp. 339–352.

[25] M. Bauer and O. Golinelli, *Random incidence matrices: moments of the spectral density*, Journal of Statistical Physics, 103 (2001), pp. 301–337.

[26] I. Benjamini, O. Schramm, et al., *Percolation beyond $\mathbb{Z}^s$, many questions and a few answers*, Electronic Communications in Probability, 1 (1996), pp. 71–82.

[27] P. Bermolen, M. Jonckheere, and P. Moyal, *The jamming constant of uniform random graphs*, Stochastic Processes and their Applications, 127 (2017), pp. 2138–2178.

[28] J. Bertoin, *Lévy processes*, vol. 121 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 1996.

[29] ——, *Burning cars in a parking lot*, Communications in mathematical physics, 306 (2011), pp. 261–290.

[30] ——, *Markovian growth-fragmentation processes*, Bernoulli, 23 (2017), pp. 1082–1101.

[31] J. Bertoin, T. Budd, N. Curien, and I. Kortchemski, *Martingales in self-similar growth-fragmentations and their connections with random planar maps*, Probability Theory and Related Fields, 172 (2018), pp. 663–724.

[32] J. Bertoin, N. Curien, and I. Kortchemski, *Random planar maps and growth-fragmentations*, Ann. Probab., 46 (2018), pp. 207–260.

[33] ——, *On conditioning a self-similar growth-fragmentation by its intrinsic area*, arXiv preprint arXiv:1908.07830, (2019).

[34] J. Bertoin, N. Curien, and A. Riera, *Scaling limits for branching process with integers types and their conditional versions*, (in preparation).

[35] J. Bertoin and G. Miermont, *Asymptotics in Knuth's parking problem for caravans*, Random Structures & Algorithms, 29 (2006), pp. 38–55.

[36] S. Bhamidi, A. Budhiraja, and X. Wang, *The augmented multiplicative coalescent, bounded size rules and critical dynamics of random graphs*, Probability Theory and Related Fields, 160 (2014), pp. 733–796.

[37] S. Bhamidi, R. van der Hofstad, and S. Sen, *The multiplicative coalescent, inhomogeneous continuum random trees, and new universality classes for critical random graphs*, Probability Theory and Related Fields, 170 (2018), pp. 387–474.

[38] N. Blum, *A new approach to maximum matching in general graphs*, in Automata, Languages and Programming: 17th International Colloquium Warwick University, England, July 16–20, 1990 Proceedings 17, Springer, 1990, pp. 586–597.

[39] T. Bohman and A. Frieze, *Karp–Sipser on random graphs with a fixed degree sequence*, Combinatorics, Probability and Computing, 20 (2011), pp. 721–741.

[40] B. Bollobás, *A probabilistic proof of an asymptotic formula for the number of labelled regular graphs*, European J. Combin., 1 (1980), pp. 311–316.

[41] B. Bollobás and B. Béla, *Random graphs*, no. 73, Cambridge university press, 2001.

[42] B. Böttcher, R. Schilling, and J. Wang, *Lévy matters. iii*, Lecture Notes in Mathematics, 2099 (2013), pp. 71–80.

[43] M. Bousquet-Mélou, *Rational and algebraic series in combinatorial enumeration*, Proceedings of the ICM 2006 (arXiv:0805.0588), (2008).

[44] M. Bousquet-Mélou and A. Jehanne, *Polynomial equations with one catalytic variable, algebraic series and map enumeration*, J. Combin. Theory Ser. B, 96 (2006), pp. 623–672.

[45] E. Brézin, C. Itykson, G. Parisi, and J.-B. Zuber, *Planar diagrams*, Comm. Math. Phys., (1978).

[46] G. Brightwell, S. Janson, and M. Luczak, *The greedy independent set in a random graph with given degrees*, Random Structures & Algorithms, 51 (2017), pp. 565–586.

[47] V. Britikov, *Asymptotic number of forests from unrooted trees*, Mathematical notes of the Academy of Sciences of the USSR, 43 (1988), pp. 387–394.

[48] A. Z. Broder, *Generating random spanning trees*, in FOCS, vol. 89, Citeseer, 1989, pp. 442–447.

[49] N. Broutin, T. Duquesne, and M. Wang, *Limits of multiplicative inhomogeneous random graphs and Lévy trees*, arXiv preprint arXiv:1804.05871, (2018).

[50] N. Broutin and J.-F. Marckert, *A new encoding of coalescent processes: applications to the additive and multiplicative cases*, Probability Theory and Related Fields, 166 (2016), pp. 515–552.

[51] T. Budd, *The peeling process of infinite boltzmann planar maps*, The Electronic Journal of Combinatorics, 23 (2016), pp. P1–28.

[52] T. Budzinski, A. Contat, and N. Curien, *The critical Karp–Sipser core of random graphs*, arXiv preprint arXiv:2212.02463, (2022).

[53] S. Butler, R. Graham, and C. H. Yan, *Parking distributions on trees*, European Journal of Combinatorics, 65 (2017), pp. 168–185.

[54] M. Camarri and J. Pitman, *Limit distributions and random trees derived from the birthday problem with unequal probabilities*, Electronic Journal of Probability, 5 (2000), pp. 1–18.

[55] G. Cannizzaro and M. Hairer, *The brownian castle*, Communications on Pure and Applied Mathematics, (2020).

[56] S. R. Carrell and G. Chapuy, *Simple recurrence formulas to count maps on orientable surfaces*, Journal of Combinatorial Theory, Series A, 133 (2015), pp. 58–75.

[57] P. Chassaing and G. Louchard, *Phase transition for parking blocks, Brownian excursion and coalescence*, Random Structures & Algorithms, 21 (2002), pp. 76–119.

[58] L. Chaumont and R. Liu, *Coding multitype forests: application to the law of the total population of branching forests*, Transactions of the American Mathematical Society, 368 (2016), pp. 2723–2747.

[59] L. Chen, *Enumeration of fully parked trees*, arXiv preprint arXiv:2103.15770, (2021).

[60] Q. Chen and C. Goldschmidt, *Parking on a random rooted plane tree*, Bernoulli, 27 (2021), pp. 93–106.

[61] X. Chen, V. Dagard, B. Derrida, Y. Hu, M. Lifshits, and Z. Shi, *The Derrida–Retaux conjecture on recursive models*, arXiv preprint arXiv:1907.01601, (2019).

[62] G. Conchon-Kerjan and C. Goldschmidt, *The stable graph: the metric space scaling limit of a critical random graph with iid power-law degrees*, arXiv preprint arXiv:2002.04954, (2020).

[63] A. Contat, *Last car decomposition of planar maps*, arXiv preprint arXiv:2205.10285, (2022).

[64] ——, *Sharpness of the phase transition for parking on random trees*, Random Structures & Algorithms, 61 (2022), pp. 84–100.

[65] ——, *Surprising identities for the greedy independent set on Cayley trees*, Journal of Applied Probability, 59 (2022), pp. 1042–1058.

[66] ——, *Parking on random trees via the configuration model*, (In Preparation).

[67] A. Contat and N. Curien, *Parking on Cayley trees & Frozen Erdős-Rényi*, arXiv preprint arXiv:2107.02116, (2021).

[68] R. Cori and G. Schaeffer, *Description trees and Tutte formulas*, Theoretical Computer Science, 292 (2003), pp. 165–183.

[69] R. Cori and B. Vauquelin, *Planar maps are well labeled trees*, Canad. J. Math., 33 (1981), pp. 1023–1042.

[70] S. Coste and J. Salez, *Emergence of extended states at zero in the spectrum of sparse random graphs*, The Annals of Probability, 49 (2021), pp. 2012–2030.

[71] E. Crane, N. Freeman, and B. Tóth, *Cluster growth in the dynamical Erdős-Rényi process with forest fires*, Electronic Journal of Probability, 20 (2015).

[72] N. Curien, *Peeling random planar maps, Saint-Flour course 2019*, available at https://www.imo.universite-paris-saclay.fr/~curien/.

[73] ——, *Random walks and random graphs, a few topics*, available at https://www.imo.universite-paris-saclay.fr/~nicolas.curien/enseignement.html.

[74] ——, *Stationary random graphs*, available at https://www.imo.universite-paris-saclay.fr/~nicolas.curien/enseignement.html.

[75] N. Curien and O. Hénard, *The phase transition for parking on Galton-Watson trees*, arXiv:1912.06012, (2019).

[76] N. Curien and I. Kortchemski, *Random non-crossing plane configurations: a conditioned Galton-Watson tree approach*, Random Structures Algorithms, 45 (2014), pp. 236–260.

[77] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Jones and Bartlett Publishers, Boston, MA, 1993.

[78] Disney, *Let it go (frozen)*.

[79] E. Duchi, V. Guerrini, S. Rinaldi, and G. Schaeffer, *Fighting fish: enumerative properties*, Sém. Lothar. Combin. B, 78 (2017), p. 2017.

[80] H. Duminil-Copin, A. Raoufi, and V. Tassion, *Sharp phase transition for the random-cluster and Potts models via decision trees*, Annals of Mathematics, 189 (2019), pp. 75–99.

[81] M. Dyer, A. Frieze, and B. Pittel, *The average performance of the greedy matching algorithm*, The Annals of Applied Probability, (1993), pp. 526–552.

[82] J. Edmonds, *Paths, trees, and flowers*, Canadian Journal of mathematics, 17 (1965), pp. 449–467.

[83] P. Erdős and A. Rényi, *On random graphs I*, Publ. math. debrecen, 6 (1959), p. 18.

[84] ——, *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci, 5 (1960), pp. 17–60.

[85] S. N. Ethier and T. G. Kurtz, *Markov processes: characterization and convergence*, vol. 282, John Wiley & Sons, 2009.

[86] W. Fang, *Planar triangulations, bridgeless planar maps and tamari intervals*, European Journal of Combinatorics, 70 (2018), pp. 75–91.

[87] ——, *Bijective link between Chapoton's new intervals and bipartite planar maps*, European Journal of Combinatorics, 97 (2021), p. 103382.

[88] L. Federico, R. Van Der Hofstad, F. Den Hollander, and T. Hulshof, *Expansion of percolation critical points for hamming graphs*, Combinatorics, Probability and Computing, 29 (2020), pp. 68–100.

[89] V. Féray and I. Kortchemski, *The geometry of random minimal factorizations of a long cycle via biconditioned bitype random trees*, Annales Henri Lebesgue, 1 (2018), pp. 149–226.

[90] S. R. Finch, *Mathematical constants*, Cambridge university press, 2003.

[91] P. Flajolet, D. E. Knuth, and B. Pittel, *The first cycles in an evolving graph*, Discrete Mathematics, 75 (1989), pp. 167–215.

[92] P. Flajolet and R. Sedgewick, *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009.

[93] P. J. Flory, *Intramolecular reaction between neighboring substituents of vinyl polymers*, Journal of the American Chemical Society, 61 (1939), pp. 1518–1521.

[94] A. Frieze and M. Karoński, *Introduction to random graphs*, Cambridge University Press, 2016.

[95] A. Frieze and C. McDiarmid, *Algorithmic theory of random graphs*, Random Structures & Algorithms, 10 (1997), pp. 5–42.

[96] H. N. Gabow and R. E. Tarjan, *Faster scaling algorithms for general graph matching problems*, Journal of the ACM (JACM), 38 (1991), pp. 815–853.

[97] A. Georgakopoulos and S. Wagner, *Limits of subcritical random graphs and random graphs with excluded minors*, arXiv preprint arXiv:1512.03572, (2015).

[98] E. N. Gilbert, *Random graphs*, The Annals of Mathematical Statistics, 30 (1959), pp. 1141–1144.

[99] C. Goldschmidt and E. Kreačić, *The spread of fire on a random multigraph*, Advances in Applied Probability, 51 (2019), pp. 1–40.

[100] C. Goldschmidt and M. Przykucki, *Parking on a random tree*, Combinatorics, Probability and Computing, 28 (2019), pp. 23–45.

[101] I. P. Goulden and D. M. Jackson, *The KP hierarchy, branched covers, and triangulations*, Adv. Math., 219 (2008), pp. 932–951.

[102] B. Haas and G. Miermont, *Scaling limits of Markov branching trees, with applications to Galton-Watson and random unordered trees*, Ann. of Probab., 40 (2012), pp. 2589–2666.

[103] N. Ikeda and S. Watanabe, *Stochastic differential equations and diffusion processes*, Elsevier, 2014.

[104] J. Jacod and A. N. Shiryaev, *Limit theorems for stochastic processes*, vol. 288 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, second ed., 2003.

[105] S. Janson, *Simply generated trees, conditioned Galton–Watson trees, random allocations and condensation.*, Probability Surveys, 9 (2012), pp. 103–252.

[106] ——, *Asymptotic normality of fringe subtrees and additive functionals in conditioned Galton–Watson trees*, Random Structures & Algorithms, 48 (2016), pp. 57–101.

[107] S. Janson, D. E. Knuth, T. Łuczak, and B. Pittel, *The birth of the giant component*, Random Structures & Algorithms, 4 (1993), pp. 233–358.

[108] S. Janson and M. J. Luczak, *A simple solution to the k-core problem*, Random Structures & Algorithms, 30 (2007), pp. 50–62.

[109] ——, *Susceptibility in subcritical random graphs*, Journal of mathematical physics, 49 (2008), p. 125207.

[110] M. Jonckheere and M. Sáenz, *Asymptotic optimality of degree-greedy discovering of independent sets in configuration model graphs*, Stochastic Processes and their Applications, 131 (2021), pp. 122–150.

[111] O. D. Jones, *Runoff on rooted trees*, 2018.

[112] T. Jonsson and S. Ö. Stefánsson, *Condensation in nongeneric trees*, Journal of Statistical Physics, 142 (2011), pp. 277–313.

[113] A. Joseph, *The component sizes of a critical random graph with given degree sequence*, Annals of Applied Probability, 24 (2014), pp. 2560–2594.

[114] O. Kallenberg, *Foundations of Modern Probability*, Springer, New York, second ed., 2002.

[115] R. M. Karp, *Reducibility among combinatorial problems*, in Complexity of computer computations, Springer, 1972, pp. 85–103.

[116] R. M. Karp and M. Sipser, *Maximum matching in sparse random graphs*, in 22nd Annual Symposium on Foundations of Computer Science (sfcs 1981), IEEE, 1981, pp. 364–375.

[117] H. Kesten, *Subdiffusive behavior of random walk on a random cluster*, Ann. Inst. H. Poincaré Probab. Statist., 22 (1986), pp. 425–487.

[118] W. King and C. Yan, *Parking functions on directed graphs and some directed trees*, arXiv preprint arXiv:1905.12010, (2019).

[119] W. King and C. H. Yan, *Prime parking functions on rooted trees*, Journal of Combinatorial Theory, Series A, 168 (2019), pp. 1–25.

[120] D. Kiss, *Frozen percolation in two dimensions*, Probability Theory and Related Fields, 163 (2015), pp. 713–768.

[121] J. Kollár, *Lectures on resolution of singularities (AM-166)*, Princeton University Press, 2009.

[122] A. G. Konheim and B. Weiss, *An occupancy discipline and applications*, SIAM Journal on Applied Mathematics, 14 (1966), pp. 1266–1274.

[123] I. Kortchemski, *Limit theorems for conditioned non-generic Galton–Watson trees*, in Annales de l'IHP Probabilités et statistiques, vol. 51, 2015, pp. 489–511.

[124] E. Kreačić, *Some problems related to the Karp-Sipser algorithm on random graphs*, PhD thesis, University of Oxford, 2017.

[125] M. Krivelevich, T. Mészáros, P. Michaeli, and C. Shikhelman, *Greedy maximal independent sets via local limits*, in 31st International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, 2020.

[126] T. G. Kurtz, *Solutions of ordinary differential equations as limits of pure jump markov processes*, Journal of applied Probability, 7 (1970), pp. 49–58.

[127] H. J. Kushner, *On the weak convergence of interpolated markov chains to a diffusion*, The annals of Probability, (1974), pp. 40–50.

[128] M.-L. Lackner and A. Panholzer, *Parking functions for mappings*, Journal of Combinatorial Theory, Series A, 142 (2016), pp. 1 – 28.

[129] J.-F. Le Gall, *Random trees and applications*, Probability Surveys, (2005).

[130] J.-F. Le Gall and A. Riera, *Growth-fragmentation processes in Brownian motion indexed by the Brownian tree*, Annals of Probability, 48 (2020), pp. 1742–1784.

[131] T. M. Liggett, *An invariance principle for conditioned sums of independent random variables*, Journal of Mathematics and Mechanics, 18 (1968), pp. 559–570.

[132] V. Limic, *A playful note on spanning and surplus edges*, arXiv preprint arXiv:1703.02574, (2017).

[133] ——, *The eternal multiplicative coalescent encoding via excursions of lévy-type processes*, Bernoulli, 25 (2019), pp. 2479–2507.

[134] B. Louf, *A new family of bijections for planar maps*, Journal of Combinatorial Theory, Series A, 168 (2019), pp. 374–395.

[135] T. Łuczak and B. Pittel, *Components of random forests*, Combinatorics, Probability and Computing, 1 (1992), pp. 35–52.

[136] T. Łuczak, B. Pittel, and J. C. Wierman, *The structure of a random graph at the point of the phase transition*, Transactions of the American Mathematical Society, 341 (1994), pp. 721–748.

[137] R. Lyons and Y. Peres, *Probability on Trees and Networks*, vol. 42 of Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, New York, 2016. Available at https://rdlyons.pages.iu.edu/prbtree/prbtree.html.

[138] J. Martin and D. Yeo, *Critical random forests*, Latin American Journal of Probability and Mathematical Statistics, 15 (2018).

[139] J. B. Martin and B. Ráth, *Rigid representations of the multiplicative coalescent with linear deletion*, Electronic Journal of Probability, 22 (2017).

[140] C. McDiarmid, *Colouring random graphs.*, Annals of Operations Research, 1 (1984).

[141] A. Meir and J. W. Moon, *The expected node-independence number of random trees*, in Indagationes Mathematicae (Proceedings), vol. 76, Elsevier, 1973, pp. 335–341.

[142] M. Molloy and B. Reed, *A critical point for random graphs with a given degree sequence*, Random structures & algorithms, 6 (1995), pp. 161–180.

[143] J. W. Moon, *Counting labelled trees*, (1970).

[144] J. Neveu, *Arbres et processus de Galton-Watson*, Ann. Inst. H. Poincaré Probab. Statist., 22 (1986), pp. 199–207.

[145] E. Page, *The distribution of vacancies on a line*, Journal of the Royal Statistical Society: Series B (Methodological), 21 (1959), pp. 364–374.

[146] I. Palásti, *On some random space filling problems*, Publ. Math. Inst. Hung. Acad. Sci, 5 (1960), pp. 353–359.

[147] A. Panholzer, *A combinatorial approach for discrete car parking on random labelled trees*, Journal of Combinatorial Theory, Series A, 173 (2020), p. 105233.

[148] ——, *Parking function varieties for combinatorial tree models*, Advances in Applied Mathematics, 128 (2021), p. 102191.

[149] J. Pitman, *Coalescent random forests*, Journal of Combinatorial Theory, Series A, 85 (1999), pp. 165–193.

[150] B. Pittel, J. Spencer, and N. Wormald, *Sudden emergence of a giantk-core in a random graph*, Journal of Combinatorial Theory, Series B, 67 (1996), pp. 111–151.

[151] M. D. Plummer and L. Lovász, *Matching theory*, Elsevier, 1986.

[152] B. Ráth, *Mean field frozen percolation*, Journal of Statistical Physics, 137 (2009), pp. 459–499.

[153] B. Ráth and B. Tóth, *Erdős-Rényi random graphs+ forest fires= self-organized criticality*, Electronic Journal of Probability, 14 (2009), pp. 1290–1327.

[154] F. Rembart, M. Winkel, et al., *Recursive construction of continuum random trees*, The Annals of Probability, 46 (2018), pp. 2715–2748.

[155] A. Rényi, *On a one-dimensional problem concerning space-filling*, Publ. Math. Inst. Hungar. Acad. Sci., 3 (1958), pp. 109–127.

[156] ——, *Some remarks on the theory of trees*, Magyar Tud. Akad. Mat. Kutató Int. Közl, 4 (1959), p. 12.

[157] R. Rossignol, *Scaling limit of dynamical percolation on critical Erdős–Rényi random graphs*, The Annals of Probability, 49 (2021), pp. 322–399.

[158] J. Salez, *Some implications of local weak convergence for sparse random graphs*, PhD thesis, Université Pierre et Marie Curie-Paris VI; Ecole Normale Supérieure de Paris . . . , 2011.

[159] G. Schaeffer, *Bijective census and random generation of Eulerian planar maps with prescribed vertex degrees*, Electronic Journal of Combinatorics, 4 (1997), p. 20.

[160] G. Schaeffer, *Conjugaison d'arbres et cartes combinatoires aléatoires. PhD thesis*, (1998).

[161] E. Schertzer, R. Sun, and J. M. Swart, *The Brownian web, the Brownian net, and their universality*, Advances in disordered systems, random processes and some applications, (2017), pp. 270–368.

[162] B. Stufler, *The continuum random tree is the scaling limit of unlabelled unrooted trees*, arXiv preprint arXiv:1412.6333, (2014).

[163] ——, *Local limits of large Galton–Watson trees rerooted at a random vertex*, in Annales de l'Institut Henri Poincaré, Probabilités et Statistiques, vol. 55, Institut Henri Poincaré, 2019, pp. 155–183.

[164] G. 't Hooft, *A planar diagram theory for strong interactions.*, Nuclear Physics B, 72 (1974), pp. 461–473.

[165] W. T. Tutte, *A census of Hamiltonian polygons*, Canad. J. Math., 14 (1962), pp. 402–417.

[166] ——, *A census of planar triangulations*, Canad. J. Math., 14 (1962), pp. 21–38.

[167] ——, *A census of slicings*, Canad. J. Math., 14 (1962), pp. 708–722.

[168] ——, *A census of planar maps*, Canad. J. Math., 15 (1963), pp. 249–271.

[169] G. Uribe Bravo, *Markovian bridges, Brownian excursions, and stochastic fragmentation and coalescence*, PhD thesis, UNAM, (2007).

[170] R. van der Hofstad, *Random Graphs and Complex Networks. Vol. II*, available at https://www.win.tue.nl/~rhofstad/NotesRGCN.html, preliminary version.

[171] V. V. Vazirani, *A theory of alternating paths and blossoms for proving correctness of the $o(\sqrt{V}e)$ general graph maximum matching algorithm*, Combinatorica, 14 (1994), pp. 71–109.

[172] Y. Watabiki, *Construction of non-critical string field theory by transfer matrix formalism in dynamical triangulation*, Nuclear Phys. B, 441 (1995), pp. 119–163.

[173] N. C. Wormald, *Differential equations for random processes and random graphs*, The annals of applied probability, (1995), pp. 1217–1235.

[174] ——, *The differential equation method for random graph processes and greedy algorithms*, Lectures on approximation and randomized algorithms, 73 (1999), pp. 0943–05073.

[175] M. Xiao and H. Nagamochi, *Confining sets and avoiding bottleneck cases: A simple maximum independent set algorithm in degree-3 graphs*, Theoretical Computer Science, 469 (2013), pp. 92–104.

[176] ——, *Exact algorithms for maximum independent set*, Information and Computation, 255 (2017), pp. 126–146.

[177] V. M. Zolotarev, *One-dimensional Stable Distributions*, vol. 65, American Mathematical Society, translations of mathematical monographs ed., 1986.