



**HAL**  
open science

# Geometry of large random graphs

David Corlin-Marchand

► **To cite this version:**

David Corlin-Marchand. Geometry of large random graphs. Probability [math.PR]. Université Paris-Saclay; Université de Fribourg (Fribourg, Suisse), 2021. English. NNT: 2021UPASM046. tel-03560508

**HAL Id: tel-03560508**

**<https://theses.hal.science/tel-03560508>**

Submitted on 7 Feb 2022

**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.

# Geometry of large random graphs

*Propriétés géométriques à grande échelle  
de graphes aléatoires*

**Thèse de doctorat de l'université Paris-Saclay et de  
l'université de Fribourg**

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

Thèse préparée dans l'unité de recherche **Laboratoire de mathématiques  
d'Orsay (Université Paris-Saclay, CNRS)**,  
sous la direction de **Nicolas CURIEN**, Professeur,  
et la co-direction de **Ioan MANOLESCU**, Professeur

**Thèse soutenue à Paris-Saclay, le 3 décembre 2021, par**

**David CORLIN-MARCHAND**

## Composition du jury

<b>Nathanaël ENRIQUEZ</b> Professeur, Université Paris-Saclay	Président
<b>Thomas DUQUESNE</b> Professeur, Sorbonne Université	Rapporteur & Examineur
<b>Miklós Z. RÁCZ</b> Professeur assistant, HDR, Princeton University	Rapporteur & Examineur
<b>Cécile MAILLER</b> Maîtresse de conférences, University of Bath	Examinatrice
<b>Nicolas CURIEN</b> Professeur, Université Paris-Saclay	Directeur de thèse
<b>Ioan MANOLESCU</b> Professeur, Université de Fribourg	Co-directeur de thèse

**Titre :** Propriétés géométriques à grande échelle de graphes aléatoires

**Mots clés :** Graphes aléatoires, Géométrie, Limites d'échelle, Attachement préférentiel, Percolation

**Résumé :** Dans cette thèse, nous étudions la géométrie à grande échelle de deux modèles de graphes aléatoires.

Dans un premier chapitre, nous nous intéressons aux arbres à attachement préférentiel affine, qui sont des arbres dont le volume croît avec le temps, selon un mécanisme récursif, aléatoire et constant. Partant d'un certain état initial — un arbre fini appelé graine — on ajoute un à un de nouveaux sommets, en reliant chacun d'entre eux par une arête à un sommet présent dans l'arbre courant. Le choix de ce dernier est aléatoire et la probabilité de choisir un sommet en particulier est proportionnelle à une fonction affine et croissante de son degré. Ce que nous montrons dans cette thèse est l'influence asymptotiquement persistante de la graine sur la loi d'un arbre à attachement préférentiel affine. Plus exactement, si  $T_n^S$  et  $T_n^{S'}$  sont deux arbres aléatoires à  $n$  sommets, construits d'après la règle décrite ci-dessus à partir de deux graines  $S$  et  $S'$  distinctes, alors la distance en variation totale entre leur loi respective est uniformément minorée en  $n$  par une constante strictement positive. Autrement dit, la loi d'un arbre à attachement préférentiel affine dépend de la graine dont il est issu, y compris en temps long.

Dans un second chapitre, indépendant du premier, nous étudions cette fois la géométrie de triangulations causales surcritiques, au travers d'un processus de percolation orientée. Les triangulations causales sont des cartes planaires construites, pour faire court, en ajoutant des arêtes « horizontales » entre les sommets d'un arbre situés à même distance du sommet racine. Nous nous intéressons ici plus particulièrement aux triangulations causales obtenues à partir d'un arbre de Bienaymé-Galton-Watson surcritique, conditionné à survivre, de loi de reproduction géométrique. Nous soumettons ces cartes à une percolation orientée de Bernoulli et montrons qu'une transition de phase continue se produit en un point critique  $p_c$ , non trivial, dont l'expression, sous forme de série infinie, est calculée explicitement. Au dessus de ce seuil, une infinité de clusters infinis coexistent dans la carte, ce qui confirme le caractère fondamentalement hyperbolique de sa géométrie. Enfin, nous prouvons qu'au point de percolation critique, un cluster dont le nombre de sommets est conditionné à être grand, correctement renormalisé, converge pour la distance de Gromov-Hausdorff vers l'arbre continu brownien.

**Title:** Geometry of large random graphs

**Keywords:** Random graphs, Geometry, Scaling limits, Preferential attachment, Percolation

**Abstract:** In this thesis, we study the geometry of two random graph models.

In the first chapter, we deal with affine preferential attachment trees. They are trees which recursively grow as follows. Starting with a finite tree—called the seed tree, vertices are attached one by one, each linked by an edge to a random vertex of the current tree, chosen with a probability proportional to an affine function of its degree. We prove the asymptotic influence of the seed on the law of an affine preferential attachment tree. To be more precise, we show that if  $T_n^S$  and  $T_n^{S'}$  are two random trees on  $n$  vertices, both built according to the rule described above but stemming from distinct seed trees  $S$  and  $S'$ , then their laws remain at uniformly positive total variation distance as  $n$  increases.

In the second chapter, we turn our attention to another random graph model, namely supercriti-

cal causal triangulations, and look at them through the prism of oriented percolation. Causal triangulations are planar graphs obtained roughly speaking by adding horizontal connections between vertices of an infinite tree. Here we consider the case where they are derived from a supercritical Galton-Watson tree conditioned to survive with a geometric offspring distribution. When such maps are subject to a Bernoulli oriented percolation process, we prove that a continuous phase transition occurs at a non trivial critical point  $p_c$  and we compute an explicit expression of the latter as an infinite series. We establish that strictly above the threshold, infinitely many infinite clusters coexist in the map. This is a typical percolation result for graphs with a hyperbolic flavour. Finally, we demonstrate that large critical oriented percolation clusters converge after rescaling towards the Brownian continuum random tree.

université  
PARIS-SACLAY

FACULTÉ  
DES SCIENCES  
D'ORSAY





# Remerciements

Au travers de ces quelques lignes, je tiens à citer et remercier celles et ceux qui ont contribué, d'une façon ou d'une autre, à l'écriture de cette thèse.

Je veux bien sûr commencer par exprimer ma plus sincère reconnaissance à mes deux directeurs de thèse, Nicolas Curien et Ioan Manolescu, pour m'avoir fait travailler sur des sujets riches, passionnants et variés. Je les remercie tous deux d'avoir longuement réfléchi et conçu un projet de thèse, entièrement organisé de façon à respecter mes contraintes personnelles.

Je suis tout particulièrement reconnaissant envers Ioan pour m'avoir offert l'opportunité de réaliser un stage de recherche avec lui à Fribourg à l'automne 2016, ce qui constitua le vrai point de départ de cette thèse. Merci à lui pour son accueil, pour son énergie, pour m'avoir intégré à la vie du département de mathématiques de l'université de Fribourg, et enfin, pour m'avoir fait découvrir quelques bons restaurants fribourgeois! Merci au co-auteur de mon premier article de recherche publié [106] — qui correspond au deuxième chapitre de ce manuscrit — pour les nombreuses et fructueuses discussions mathématiques que nous avons eues dans son grand et confortable bureau, et pour m'avoir fait bénéficier de son expérience en matière de rédaction. Merci enfin pour m'avoir permis de présenter le résultat de nos travaux lors d'un séminaire à Fribourg en octobre 2019.

Avant que Nicolas ne devienne mon directeur de thèse, j'ai eu la chance et le plaisir de le découvrir comme professeur lors de mon année de master recherche à Orsay. Si je pense avoir su mesurer, apprécier et profiter pleinement de ses indéniables talents de pédagogues, je n'ai malheureusement pas — en raison de ma nature bien trop réservée — souvent pu goûter à ses célèbres récompenses chocolatées, destinées aux étudiants participant activement à sa classe. Directeur, Nicolas a toujours été d'une infinie patience, d'une extrême disponibilité et d'une réactivité sans pareil pour relire et corriger les multiples versions du troisième et dernier chapitre de ce manuscrit. Après chacun de nos rendez-vous dans son bureau, mes blocages mathématiques disparaissaient, et je retrouvais systématiquement énergie et motivation. Un remède dont il aurait été sage que j'use plus fréquemment. Par ailleurs, je ne puis exprimer combien ses différents polycopiés — particulièrement celui de son cours à l'école d'été de Saint-Flour — m'ont été utiles au cours de la rédaction de ma thèse. Pour toutes les raisons que je viens d'énumérer, j'adresse ma plus profonde gratitude à Nicolas.

Une troisième personne a joué un rôle décisif dans la définition de mon sujet de thèse, ainsi que son organisation. Il s'agit de Jean-François Le Gall, que je veux remercier pour la confiance qu'il m'a accordée en acceptant tout d'abord d'encadrer mon mémoire de master en 2016, puis en me recommandant auprès de mon futur directeur de thèse Ioan Manolescu, afin d'approfondir le travail que j'avais réalisé sous sa propre direction. Je le remercie également pour son implication lors des discussions ultérieures, qui ont abouti au choix d'une thèse co-encadrée par Ioan et Nicolas.

Mes remerciements suivants vont aux membres de mon jury de soutenance.

*First, I wish to warmly thank Thomas Duquesne and Miklós Z. Rácz for having accepted to be the rapporteurs of my thesis. It was a great honor for me to get the following lines reviewed by such experienced mathematicians. Both co-write papers [39, 54] which can be regarded without exaggeration as the seeds — the term looks particularly appropriate — from which the second chapter of the present manuscript sprang. I am also grateful to them for having attended my thesis defense as jury members, either physically or via video-conference.*

Je remercie chaleureusement Nathanaël Enriquez d'avoir accepté de présider mon jury de soutenance, et Cécile Mailler pour son rôle d'examinatrice. Leur présence m'honore. Merci tout par-

ticulièrement à Cécile Mailler d'avoir prolongé son séjour à Paris pour l'occasion.

Je souhaite désormais remercier les différentes institutions, et ceux qui y travaillent, pour m'avoir offert un environnement propice à la réalisation de ma thèse. Je remercie tout d'abord l'université Paris-Saclay et l'université de Fribourg. Je remercie l'Ecole Doctorale Mathématiques Hadamard, et son directeur, Stéphane Nonnenmacher, pour son suivi, sa disponibilité et son écoute. Merci également à Séverine Simon pour avoir répondu avec patience à toutes mes questions d'ordre administratif, notamment lors du fastidieux rituel annuel de la réinscription. J'adresse toute ma gratitude aux personnels du Laboratoire de Mathématiques d'Orsay, particulièrement à Catherine Ardin et Marie-Christine Myoupo du pôle gestion, pour avoir toujours traité avec rapidité mes demandes de remboursement de frais de mission, à Mathilde Rousseau et tout le pôle informatique pour leur assistance régulière. Je veux remercier les organisateurs du séminaire hebdomadaire de probabilités, pour la qualité constante de la programmation. Enfin, merci au département de mathématiques de Fribourg de m'avoir accueilli à plusieurs reprises, que ce soit pour de courts séjours d'une semaine ou lors de mon plus long stage de recherche à l'automne 2016. Je suis infiniment reconnaissant à Isabella Schmutz d'avoir toujours veillé à ce que mon travail puisse s'effectuer dans les meilleures conditions matérielles possibles.

Pendant mes quatre années de thèse, l'opportunité m'a été offerte d'enseigner toutes sortes de mathématiques — algèbre linéaire, analyse, probabilités et statistiques, graphes, etc. — à plus d'une centaine d'élèves aux profils variés. J'ai eu tout d'abord le plaisir d'exercer pendant trois ans comme chargé de TD au département informatique de l'IUT d'Orsay. Mon public était constitué d'étudiants en première et deuxième année de DUT. Une formation technique donc, où les mathématiques pouvaient parfois apparaître, aux yeux des élèves, comme une discipline aride et abstraite. Capter leur attention et leur donner envie de travailler a été pour moi un véritable défi, que j'ai aimé relever. Je remercie infiniment la regrettée Caroline Fabre pour son accueil chaleureux au sein de la petite équipe d'enseignants en mathématiques de l'IUT d'Orsay. Je remercie également Banafsheh Farang-Hariri, Odile Guiot et Elodie Leducq, pour leur encadrement bienveillant. Grâce à elles et à toute l'aide qu'elles m'ont apportée, je me suis épanoui lors de cette première expérience du métier d'enseignant.

Après trois années passées à l'IUT d'Orsay, j'ai occupé durant un an un poste d'ATER à temps plein à la Faculté des Sciences d'Orsay. La préparation de mes cours est par conséquent devenue une activité nettement plus chronophage. Je tiens à remercier Sophie Lemaire pour le souci qu'elle a eu d'aménager, autant que possible, mon service d'enseignement, de façon à le rendre compatible avec les ultimes étapes de rédaction de ma thèse. À la Faculté des Sciences, j'ai rencontré un public d'étudiants différent de celui de l'IUT. Des élèves en effet inscrits en première, deuxième, voire troisième année de licences scientifiques, pour qui les mathématiques occupaient une place tout à fait centrale dans leur formation. Je suis heureux d'avoir ainsi pu diversifier mon expérience de l'enseignement. En outre, j'ai eu la chance de travailler avec des professeurs et maîtres de conférence accordant une large autonomie aux chargés de TD placés sous leur responsabilité, attentifs à leurs remarques et retours d'expérience, les impliquant parfois très activement dans l'organisation du cours et des examens. Je tiens pour ces raisons à remercier vivement Guy David, Stéphane Fischler, Elisabeth Gassiat, Hoang-Chinh Lu, Edouard Maurel-Segala et Hans Henrik Rugh. Je remercie également mes collègues et camarades de doctorat Gabriella Bettonte, Baptiste Cercle, Jacques de Catelan, Brice Hannebicque, Perrine Lacroix et Béwentaoré Sawadogo, avec qui j'ai partagé les charges de surveillances d'examens, de corrections de copies, mais aussi de relecture et rédaction de différents documents liés aux cours où nous intervenions.

J'ai été marqué, tout au long de ma scolarité, par la rencontre de professeurs et chercheurs brillants. Cette thèse leur doit incontestablement beaucoup.

Je veux exprimer toute ma gratitude à Roger Mansuy, mon professeur de mathématiques en première année de classes préparatoires au lycée Louis Le Grand. Remarquable pédagogue, il m'a appris la rigueur, l'ordre et le soin que nécessite une bonne démonstration. Attentif aux parcours de ses élèves, j'ai pu compter sur lui lorsque j'ai eu besoin de conseils.

Mon entrée à l'Ecole Normale Supérieure de Cachan en 2013 n'aurait peut-être pas été possible sans Dominique Guest, mon professeur de physique-chimie et professeur principal en deuxième de

classes préparatoires. Merci à elle pour son travail acharné et ses constants encouragements.

Entré à l'ENS par le concours PSI (pour Physique-Sciences de l'Ingénieur), j'ai souhaité me réorienter en intégrant le département de mathématiques de l'école. Je veux exprimer toute ma reconnaissance à Frédéric Pascal pour avoir accepté de m'y accueillir. Je le remercie également pour son suivi minutieux de ma scolarité à l'ENS, son soutien lorsque j'ai eu besoin d'une lettre de recommandation, et enfin, d'avoir consenti à lourdement aménager mon année de préparation à l'agrégation, afin que je puisse effectuer mon stage de recherche à Fribourg entre septembre et novembre 2016. Un stage qui, comme je l'ai déjà dit, fut la première étape vers la réalisation de cette thèse.

J'adresse tous mes remerciements, pour l'excellence des formations dispensées, aux équipes pédagogiques du département de mathématiques de l'ENS Cachan, de l'université Paris-Diderot — où j'ai suivi et validé plusieurs cours lors de ma première année de master, et enfin, bien sûr, du master 2 « Mathématiques de l'aléatoire » (2015/2016) de l'Université Paris-Sud. J'adresse un remerciement particulier à Francis Comets et Justin Salez, alors respectivement professeur et chargé de TD, pour la qualité de leur cours de M1 « Probability and stochastic processes » (2014) et de M2 « Calcul stochastique et modèles de diffusion » (2014/2015) — tous deux donnés à l'université Paris-Diderot. Ils m'ont définitivement décidé à m'orienter vers l'étude approfondie des probabilités.

Deux stages d'initiation à la recherche réalisés pendant ma scolarité à l'ENS ont largement déterminé mon choix ultérieur de m'engager dans cette thèse. Je remercie Argyris Kalogeratos, Kevin Scaman et Nicolas Vayatis pour l'encadrement de mon stage, avec mon camarade de promotion Patrick Saux, au centre Borelli (ex-CMLA) en 2014. J'exprime aussi toute ma reconnaissance à Julien Berestycki pour m'avoir offert en 2015 l'exceptionnelle opportunité de venir travailler pendant cinq mois au département de statistiques de l'université d'Oxford, et de m'avoir si bien accueilli.

Début janvier, je partirai à Lille pour effectuer un post-doctorat sous la direction de David Coupier et Benoît Henry. Je les remercie de m'avoir accordé leur confiance. Leur décision m'a indéniablement aidé à trouver la motivation nécessaire à la rédaction des ultimes pages de ce manuscrit.

Je voudrais maintenant conclure ce texte sur une note plus intime, en remerciant mes amis et ma famille pour leur soutien, leur affection et l'épanouissement que je trouve à leurs côtés.

Je souhaite pour commencer adresser des remerciements particuliers à mon très cher ami Julien, qui me fait l'honneur de venir assister à ma soutenance, et avec qui j'ai plaisir à refaire quotidiennement le monde — à défaut de le changer... pour le moment ! — depuis (presque) un quinquennat. J'ignore si le volume total de nos échanges écrits excède celui de ce manuscrit. Il faudrait mesurer. Je te remercie en tout cas de partager avec moi tes réflexions et d'entretenir les miennes. À toi maintenant, et à Alexandra, d'achever vos thèses respectives !

Merci à mon ami dijonnais Antoine, rencontré en prépa il y a dix ans, grâce à qui la montée à Paris a assurément été moins difficile à vivre, et que je suis toujours heureux de revoir, que ce soit pour discuter de politique autour d'un bon plat, ou pour visionner ensemble un film russe — peut-être un jour ouzbek ? — de qualité.

Merci enfin à tous mes amis palois de très longue date, Estelle, Jérémy, Lilian, Lucas, Ludovic, Mélanie et Quentin, avec qui, j'en suis heureux, l'amitié ne s'est pas arrêtée aux années lycée, et ce malgré la distance.

Je veux exprimer mon infinie gratitude à la famille des Cazalet, ma famille de cœur, pour m'avoir ouvert leur foyer il y a dix ans et accueilli comme l'un des leurs. Merci à mon beau-père Jean-Michel, à ma belle-mère Christine et à mon beau-frère Guillaume.

Il est impossible d'exprimer tout ce que je dois à ma mère. Merci à elle pour son amour et pour avoir si bien veillé à mon éducation. J'espère qu'elle sera fière de mon travail.

Enfin, j'adresse mes ultimes et plus chaleureux remerciements à ma chère et tendre Julie, dont j'ai le bonheur de partager la vie. Par ses encouragements permanents et son amour, chaque ligne écrite ici lui doit beaucoup.





# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Graphes, arbres et cartes . . . . .	8
1.2	De l'aléatoire chez les graphes, arbres et cartes . . . . .	12
1.3	De la géométrie des graphes, arbres et cartes . . . . .	17
1.4	Reconnaissance de la graine dans un arbre à attachement préférentiel <i>affine</i> . . . . .	24
1.5	Percolation orientée sur triangulations causales surcritiques . . . . .	38
1.6	A brief introduction for non-French speakers . . . . .	58
<b>2</b>	<b>Influence of the seed in affine preferential attachment trees</b>	<b>61</b>
2.1	Introduction . . . . .	62
2.2	Planar affine preferential attachment model . . . . .	65
2.3	First and second moment of a class of observables . . . . .	68
2.4	Observables and their difference around the seed tree . . . . .	81
2.5	Open problems and future research . . . . .	89
<b>3</b>	<b>Percolation on supercritical causal triangulations</b>	<b>91</b>
3.1	Introduction and main results . . . . .	92
3.2	The supercritical causal triangulation of the half-plane . . . . .	95
3.3	Percolation on SCTHP . . . . .	99
3.4	Percolation on SCT . . . . .	117



# Chapitre 1

## Introduction

Dans ce chapitre d'introduction, nous exposons tout d'abord le cadre mathématique minimal, commun aux différents travaux présentés dans ce manuscrit. Dans la Section 1.1, nous rappelons la définition des graphes, arbres et cartes, et motivons leur introduction. Nous donnons dans la Section 1.2 des exemples de modèles *aléatoires* impliquant de tels objets. Enfin, dans la Section 1.3, nous expliquons ce que recouvre pour eux la notion de *géométrie*, et particulièrement ce que nous entendons par *géométrie à grande échelle* de graphes aléatoires.

Dans les sections 1.4 et 1.5, nous présentons les principaux résultats mathématiques originaux, qui ont été obtenus au cours de cette thèse, ainsi que le contexte dans lequel ils s'inscrivent.

La Section 1.4 introduit au Chapitre 2 de ce manuscrit, qui reproduit un travail [106] mené en commun avec Ioan Manolescu sur la question de l'influence de la *graine* dans les arbres à attachement préférentiel *affine*.

La Section 1.5 introduit, quant à elle, au Chapitre 3, fruit d'un travail autonome [105], où nous analysons un processus de percolation orientée sur un modèle de triangulations causales surcritiques.

**Les contributions originales de cette thèse sont les résultats grisés et encadrés.**

*In section 1.6, the non-French speaker will find a brief presentation of the mathematical problems addressed in this dissertation.*

### Sommaire

---

1.1	Graphes, arbres et cartes . . . . .	8
1.2	De l'aléatoire chez les graphes, arbres et cartes . . . . .	12
1.3	De la géométrie des graphes, arbres et cartes . . . . .	17
1.4	Reconnaissance de la graine dans un arbre à attachement préférentiel <i>affine</i> . . . . .	24
1.5	Percolation orientée sur triangulations causales surcritiques . . . . .	38
1.6	A brief introduction for non-French speakers . . . . .	58

---

## 1.1 Graphes, arbres et cartes

Un graphe  $G$  est la donnée de deux ensembles :

- un ensemble  $V$  de *sommets* ;
- un ensemble  $E \subset V^2$  d'*arêtes*.

Les arêtes déterminent les relations d'*adjacence* (ou de *voisinage*) entre sommets du graphe. On se figure en effet généralement un graphe comme un ensemble de points (les sommets) et de segments (les arêtes), où un segment relie un sommet  $x \in V$  au sommet  $y \in V$  si  $(x, y) \in E$ .

Dans cette thèse, les graphes seront généralement supposés *simples* : une seule arête au maximum connecte une paire de sommets. Les ensembles  $V$  et  $E$  seront en outre toujours **finis** ou **infinis dénombrables**. Enfin, les graphes seront, sauf exception, *localement finis*, au sens où tout sommet n'admet qu'un nombre fini de voisins. Ce nombre est appelé *degré* du sommet. Pour le sommet  $x$  d'un graphe  $G$ , nous noterons cette quantité  $\text{deg}_G(x)$ . Sur la Figure 1.1, nous représentons deux graphes où tous les sommets ont le même degré. À gauche, le graphe dit *complet* à six sommets : chaque sommet est relié à tous les autres sommets et admet donc cinq voisins. À droite, le réseau carré sur  $\mathbb{Z}^2$  : tous les sommets ont exactement quatre voisins.

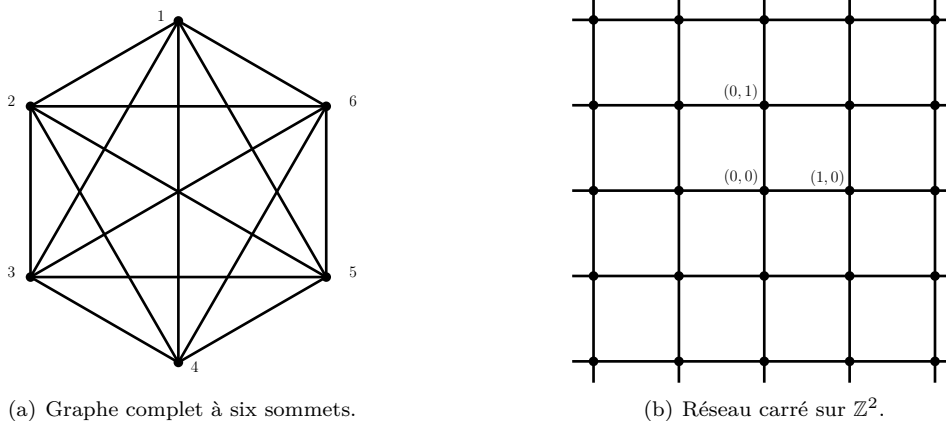


FIGURE 1.1 – Deux exemples de graphes fini (à gauche) ou infini dénombrable (à droite), où tous les sommets ont le même degré.

### Pourquoi des graphes ?

En sciences appliquées [117], on utilise des graphes pour représenter les "relations" existant entre éléments d'un même réseau réel, afin de mieux comprendre la structure ou "géométrie" de ce dernier. Ces relations peuvent être de différentes natures : liens d'amitié entre individus sur un célèbre réseau social, citations entre articles de recherche, voies ferroviaires connectant un réseau de gares, interactions des neurones du cerveau via les axones, etc.

Les graphes s'avèrent aussi utiles pour résoudre des problèmes pratiques d'optimisation discrète de type maximisation de flot/minimisation de coût [93] : planification des horaires de train, gestion de la puissance électrique délivrée par un réacteur nucléaire, etc. Ils sont un outil pour discrétiser des problèmes physiques continus. Pensons par exemple à la trajectoire d'une particule dans l'espace, qu'il peut être commode d'approcher par ses déplacements successifs sur un réseau cubique  $\delta \cdot \mathbb{Z}^3$ , avec un pas  $\delta$  tendant vers 0.

En mathématiques, l'étude des graphes constitue un domaine de recherche en soi, mais interagissant largement avec d'autres. Par exemple, à travers la notion de graphe de Cayley, la théorie des graphes a des applications en combinatoire et géométrie des groupes [60].

### Les arbres

On appelle *arbre* tout graphe connexe sans cycle : tout sommet du graphe est atteignable par un chemin d'arêtes depuis n'importe quel autre sommet (*connexité* du graphe), et le chemin pour

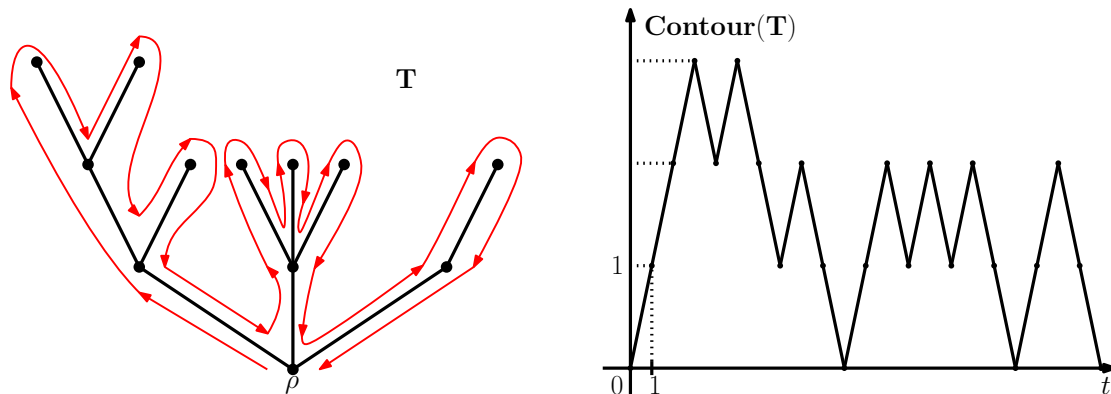
y parvenir est toujours unique. Il est donc impossible de quitter un sommet, puis d'y revenir, sans passer deux fois par une même arête. Un exemple d'arbre est représenté en Figure 1.2(a).

En identifiant un sommet particulier de l'arbre, c'est-à-dire en l'*enracinant*, on peut classer les autres sommets selon le nombre d'arêtes à traverser pour y parvenir depuis le sommet racine. Comme le chemin d'arête menant de la racine à un sommet précis est unique, une structure "généalogique" apparaît : un sommet situé à  $r$  arêtes de la racine est voisin et "enfant" d'un et d'un seul sommet situé à  $r - 1$  arêtes de la racine, qui lui-même est voisin et "enfant" d'un et d'un seul sommet situé à  $r - 2$  arêtes de la racine, etc. Un arbre *enraciné*, c'est donc tout comme on se figure un arbre généalogique familial. À partir d'un géniteur initial, on trace les relations de parenté entre cet individu et ses enfants, puis entre ses enfants et les petits-enfants, etc.

### Quoi de propre aux arbres ?

Les arbres se distinguent dans la famille des graphes par de remarquables propriétés combinatoires. Des formules exactes de dénombrement existent par exemple pour les arbres enracinés et étiquetés à  $n \geq 1$  sommets — "étiquetés" signifie que l'ensemble de sommets de l'arbre est  $\{1, \dots, n\}$ , pour les arbres plans (voir la définition ci-dessous), etc. C'est rarement le cas des graphes avec cycles. La raison tient essentiellement au fait que les arbres sont souvent codables par des processus unidimensionnels, à l'instar des arbres plans de taille finie, que l'on peut mettre en bijection avec des suites d'entiers positifs.

On appelle *arbre plan* un arbre enraciné muni d'un système d'orientations : pour chaque sommet  $v$  de l'arbre, on numérote ses enfants, de sorte que si on représente l'arbre dans le plan, l'enfant n° 1 de  $v$  sera à la gauche de tous les autres enfants, l'enfant n° 2 immédiatement à la droite du premier, etc. Voir Figure 1.2(a) pour un exemple. En suivant les contours de l'arbre de la gauche vers la droite, on peut définir une suite d'entiers initialisée en 0, qui effectue des sauts d'amplitude  $+1$  lorsqu'on s'éloigne de la racine, et des sauts d'amplitude  $-1$  lorsqu'on s'en rapproche. Cette suite d'entiers, appelé *processus de contour*, caractérise l'arbre lorsqu'il est de taille finie, au sens où la donnée du processus suffit à le reconstituer. D'autres processus existent, comme le *processus de hauteur*, dont nous parlerons dans la suite, mais ce ne sont généralement que des variantes du processus de contour. Sur la Figure 1.2(b), nous avons tracé le processus de contour associé à l'arbre plan de la Figure 1.2(a).



(a) Arbre plan  $T$  enraciné en  $\rho$  (en noir) et parcours de contournement de l'arbre (en rouge).

(b) Processus de contour associé.

FIGURE 1.2 – Un arbre plan est caractérisé par son processus de contour.

La spécificité des arbres tient aussi à leur caractère "minimaliste". En effet, parmi les graphes connexes à  $n \geq 1$  sommets, les arbres sont exactement ceux minimisant la quantité totale d'arêtes, à savoir  $n - 1$ , requise pour garantir la connexité. Il existe un algorithme — celui de l'arbre couvrant minimal — qui dégarnit progressivement un graphe connexe à  $n$  sommets — c'est-à-dire en lui retirant des arêtes — jusqu'à obtenir un arbre connectant le même ensemble de sommets. C'est un problème d'optimisation aux applications tout à fait concrètes : minimiser par exemple le nombre

d'infrastructures nécessaires (lignes électriques, voies ferrées, etc) pour connecter les éléments d'un même réseau (foyers, gares, etc).

### Graphes et cartes planaires

Un graphe est dit *planaire* s'il peut être représenté dans le plan sans qu'aucune de ses arêtes n'en croise une autre. Le réseau carré sur  $\mathbb{Z}^2$  est évidemment planaire (voir Figure 1.1(b)). Un arbre est, lui, toujours un graphe planaire. En revanche, le graphe complet à  $n \geq 5$  sommets — chaque sommet est connecté à tous les autres — n'est pas planaire. Le lecteur pourra s'en convaincre avec le cas  $n = 6$  représenté en Figure 1.1(a).

On définit une *carte planaire finie* comme un graphe planaire (connexe) de taille finie, plongé proprement dans le plan (ou la sphère  $\mathbb{S}^2$ ), le plongement étant considéré à homéomorphisme préservant l'orientation près. Un arbre plan, comme celui de la Figure 1.2, est une carte planaire. Le système d'orientations mentionné dans sa définition caractérise en effet son plongement dans le plan. Dans la vie courante, plan du réseau métropolitain, circuits imprimés... sont tous des cartes planaires. Sur la Figure 1.3, nous présentons deux cartes planaires finies obtenues à partir d'un même graphe mais dont les plongements sur  $\mathbb{S}^2$  ne sont pas homéomorphes.

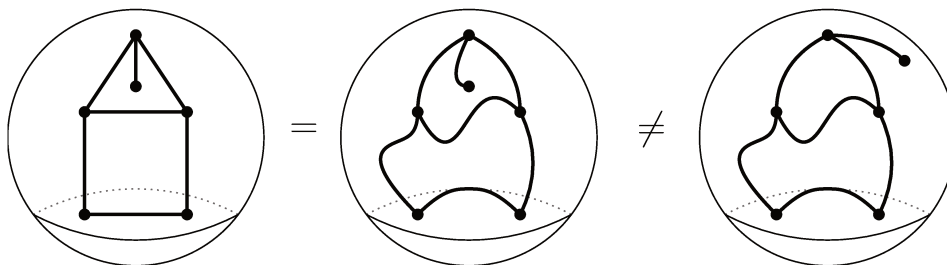


FIGURE 1.3 – Les deux cartes planaires à gauche et au centre — plongées sur  $\mathbb{S}^2$  — sont identiques à homéomorphisme préservant l'orientation près, mais différent de la carte représentée à droite.

Les *faces* d'une carte planaire sont par définition les composantes connexes du complémentaire de son plongement dans le plan. Autrement dit, ce sont les régions du plan bordées par un cycle d'arêtes. Le nombre d'arêtes qui bordent une face est son *degré*. On appelle *triangulations* les cartes planaires dont toutes les faces sont de degré 3, *quadrangulations* celles dont toutes les faces sont de degré 4, etc. Une définition alternative des cartes planaires consiste à les interpréter comme un assemblage de polygones — correspondant aux faces de la carte — collés les uns aux autres, de façon à ce que la variété construite soit une sphère topologique. Une face de degré 3 correspond à un triangle, une face de degré 4 correspond à un carré, etc. Cette définition plus combinatoire, équivalente à la première, présente plusieurs avantages, notamment celui de rendre évident le fait que le nombre de cartes à  $n$  sommets/arêtes/faces est fini. Sur la Figure 1.4, nous illustrons cet autre point de vue sur les cartes.

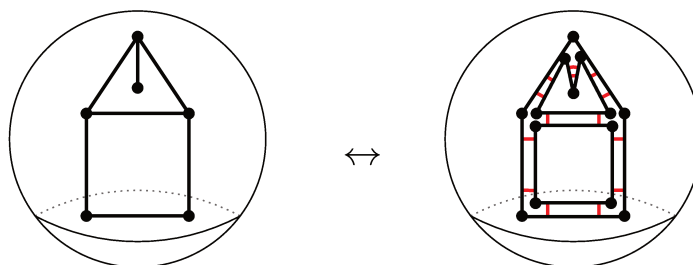


FIGURE 1.4 – Une carte planaire vue comme un assemblage de polygones — un carré et deux pentagones — collés les uns aux autres.

## Quoi de propre aux cartes planaires ?

Les cartes planaires forment une (rare) classe de graphes avec cycles, dont on connaît de nombreuses propriétés combinatoires. Citons par exemple la célèbre formule d'Euler, qui relie le nombre de sommets  $|V|$  aux nombres d'arêtes  $|E|$  et de faces  $|F|$  de n'importe quelle carte planaire finie :

$$|V| + |E| - |F| = 2.$$

Citons aussi le théorème des quatre couleurs, qui a longtemps motivé l'étude des cartes planaires : toutes les faces d'une carte planaire peuvent être coloriées avec seulement quatre couleurs, de telle sorte que deux faces adjacentes — c'est-à-dire partageant une arête en commun — ne seront jamais coloriées de la même façon. Ce résultat difficile fut démontré pour la première fois en 1976 par Kenneth Appel et Wolfgang Haken, avec l'aide de l'outil informatique.

Enfin, grâce à une série de travaux réalisés par Tutte dans les années 60 [133, 134, 135, 136, 137], on a aujourd'hui de nombreux résultats d'énumération de cartes planaires. La méthode imaginée par Tutte repose sur une décomposition récursive des cartes par l'effacement d'une arête préalablement choisie (l'arête *racine*). De cette décomposition, il déduit une équation à deux variables vérifiée par une fonction génératrice, équation qu'il résout en développant un nouvel outil appelé *méthode quadratique*. On obtient par exemple de cette façon que le nombre de quadrangulations enracinées à  $n$  faces est :

$$\frac{2 \cdot 3^n}{(n+1)(n+2)} \binom{2n}{n}.$$

D'autres méthodes d'énumération ont depuis été élaborées. Des liens avec les intégrales de matrice ont été notés par des physiciens théoriciens dans les années 70 [132]. Une bijection cruciale, mettant en relation les cartes et un ensemble d'arbres plans, dits *bien étiquetés*, a été établie la décennie suivante par Robert Cori et Bernard Vauquelin dans [51].

## Cartes planaires et physique théorique

Une des principales motivations modernes à l'étude des cartes planaires est la théorie de la gravité quantique. À la fin du siècle dernier, il est apparu nécessaire aux physiciens théoriciens, spécialistes de la question, d'étendre aux dimensions supérieures les intégrales de chemin de Feynman, qui sont des intégrales où la somme porte sur des ensembles de chemins continus reliant deux points  $x$  et  $y$  fixés.

Construire des intégrales sur de vastes ensembles compliqués — chemins, surfaces, etc — passe généralement, dans un premier temps, par la discrétisation de ceux-ci. Pour l'intégrale de chemin, la méthode standard est de l'approcher d'abord par une somme indexée sur les chemins linéaires par morceaux reliant  $x$  et  $y$  en un nombre fini d'étapes, puis de faire tendre ce nombre vers l'infini, et enfin de vérifier que ce procédé permet d'obtenir à la limite une mesure sur l'ensemble des chemins continus reliant les deux points, mesure vérifiant un certain nombre de propriétés attendues *a priori*. C'est de cette manière, par exemple, que l'on construit la *mesure de Wiener*, la mesure de probabilité sur l'ensemble des fonctions continues associée au mouvement brownien. D'après le théorème de Donsker, ce dernier est en effet la limite en loi, universelle, de marches aléatoires. Ceci implique que la mesure de Wiener est elle-même limite de mesures chargeant uniquement les fonctions continues et linéaires par morceaux.

Pour les intégrales de surface, l'idée venue naturellement aux physiciens était de faire jouer aux cartes planaires le rôle dévolu précédemment aux chemins linéaires par morceaux. Remplacer donc une intégrale sur un ensemble de surfaces continues, mathématiquement mal définie, par une somme indexée sur un ensemble fini de surfaces discrètes — les cartes — les recouvrant de plus en plus finement. Ce qui fit alors émerger de nouvelles questions, en particulier la suivante : existe-t-il un équivalent bidimensionnel au théorème de Donsker ? Nous en reparlerons plus en détail dans la suite de ce chapitre. Pour une introduction complète aux raisons amenant à s'intéresser aux cartes planaires en théorie de la gravité quantique, nous invitons le lecteur curieux à consulter le livre [10].



## 1.2 De l'aléatoire chez les graphes, arbres et cartes

Un graphe, un arbre ou une carte *aléatoire* est tout simplement une variable aléatoire à valeurs dans un ensemble de graphes, arbres ou cartes. Les modèles aléatoires les plus élémentaires (à décrire) consistent tout simplement à tirer uniformément un graphe, un arbre ou une carte dans un ensemble fini de graphes/arbres/cartes : graphes généraux ou arbres plans à  $n$  sommets, triangulations ou quadrangulations à  $n$  faces/arêtes/sommets, etc.

La principale raison poussant à introduire du hasard en théorie des graphes est celle motivant habituellement, en science, la décision d'adopter un point de vue probabiliste : l'évidente stochasticté du réel, sa variabilité, sa complexité. Les réseaux par exemple évoqués en début d'introduction (réseaux sociaux, cérébraux, etc), qui connectent généralement un très grand nombre d'éléments, n'y échappent pas. Décrire en détail ces réseaux, leur formation, leur évolution, et ce de manière déterministe, s'avère impossible, car trop de paramètres entreraient en jeu.

Les probabilités constituent par ailleurs un cadre mathématique structurant largement de multiples pans de la physique théorique moderne (physique statistique, mécanique quantique, etc), et nous avons expliqué pourquoi les physiciens théoriciens s'intéressaient aujourd'hui aux cartes planaires.

Enfin, il faut bien le dire, la théorie mathématique des graphes aléatoires ouvre vers de nouveaux concepts et problèmes intéressants par eux-mêmes, parfois en relation avec d'autres domaines des mathématiques : théorie ergodique, combinatoire, etc.

Dans cette section, nous présentons trois modèles élémentaires de graphes, arbres et cartes aléatoires, tous occupant une place importante dans la littérature, et sur lesquels nous aurons l'occasion de revenir à plusieurs reprises dans la suite de ce manuscrit.

### Graphes de Erdős-Rényi

Considérons le graphe complet à  $n$  sommets. Par définition, tous ses sommets sont connectés les uns aux autres, il comporte donc  $\binom{n}{2} = \frac{n(n-1)}{2}$  arêtes. Introduisons maintenant un peu de hasard : pour chaque arête, à l'issue d'une épreuve de Bernoulli, nous décidons de la conserver avec probabilité  $p \in [0, 1]$ , ou de la supprimer avec probabilité  $1 - p$ , et ce indépendamment de ce qu'il se produit pour les autres arêtes. À l'issue de ce processus, nous obtenons un nouveau graphe  $\mathbf{G}_{n,p}$  à  $n$  sommets, *aléatoire*, appelé **graphe de Erdős-Rényi de paramètres  $n$  et  $p$** . Notez que le graphe n'est plus nécessairement connexe.

Si la paternité du modèle revient en réalité à Edgar Gilbert qui l'introduisit dans [70], on le nomme ainsi en l'honneur de l'article publié par Erdős et Rényi en 1960 [65], qui eut des répercussions très importantes et fonda quasiment la recherche mathématique en graphes aléatoires. En dépit de son indubitable simplicité, le modèle de Erdős-Rényi vérifie de multiples et fascinantes propriétés, que nous détaillerons plus loin dans cette introduction. Il est encore aujourd'hui l'objet d'innombrables travaux.

Pour terminer cette courte présentation, nous souhaitons faire observer que le processus de conservation/suppression des arêtes dont est issu  $\mathbf{G}_{n,p}$ , est aussi connu sous le nom de **percolation par arêtes**. Son principe peut être appliqué aux arêtes de n'importe quel autre graphe déterministe, voire aléatoire. Dans ce dernier cas, on génère aléatoirement un graphe de loi donnée, puis conditionnellement à la réalisation du graphe obtenu, on applique le processus de percolation aux arêtes. On obtient ainsi deux niveaux d'aléatoire imbriqués.

### Les arbres de Bienaymé-Galton-Watson

Le modèle des arbres de Bienaymé-Galton-Watson joue un rôle central dans la théorie des arbres aléatoires. Sa définition repose sur la structure généalogique intrinsèque des arbres que nous évoquons dans la première partie de cette introduction. Un arbre de Bienaymé-Galton-Watson code la généalogie d'un processus de branchement, introduit en 1845 par le statisticien français Irénée-Jules Bienaymé, puis trente ans plus tard, de manière indépendante, par les scientifiques britanniques Francis Galton et Henry Williams Watson, dans le but de modéliser le phénomène d'extinction des noms de familles nobles. Nous renvoyons le lecteur vers [84] ou vers [81] pour plus de détails sur l'histoire (récente) des processus généalogiques en mathématiques.

**Définition du modèle** Introduisons  $\mu = (\mu_k)_{k \geq 0}$  une mesure de probabilité sur l'ensemble des entiers naturels  $\mathbb{N}$ . On construit un **arbre plan de Bienaymé-Galton-Watson de loi de reproduction  $\mu$** , ici noté  $\mathbf{T}$ , de la façon suivante :

1. le sommet racine de  $\mathbf{T}$  donne naissance à  $\mathbf{X}$  enfants, avec  $\mathbf{X}$  une variable aléatoire distribuée selon la loi  $\mu$  ;
2. (**propriété de branchement**) conditionnellement à la valeur de  $\mathbf{X}$ , des arbres de Bienaymé-Galton-Watson **indépendants**, de même loi de reproduction  $\mu$ , émergent de chacun des  $\mathbf{X}$  enfants du sommet racine.

Autrement dit, par récurrence, tous les sommets de l'arbre plan aléatoire  $\mathbf{T}$  donnent naissance à un nombre aléatoire d'enfants, et ces nombres sont indépendants et identiquement distribués selon la loi  $\mu$ . En particulier, on obtient que pour tout arbre plan  $t$  de taille finie :

$$\mathbb{P}(\mathbf{T} = t) = \prod_{v \in t} \mu_{x_v}, \quad (1.2.1)$$

où  $x_v$  désigne le nombre d'enfants du sommet  $v \in t$ . Remarquons que  $x_v = \deg_t(v)$  lorsque  $v$  est la racine de l'arbre, mais que  $x_v = \deg_t(v) - 1$  lorsque  $v$  est un descendant de celle-ci. Voir une illustration de la construction récursive des arbres de Bienaymé-Galton-Watson en Figure 1.5.

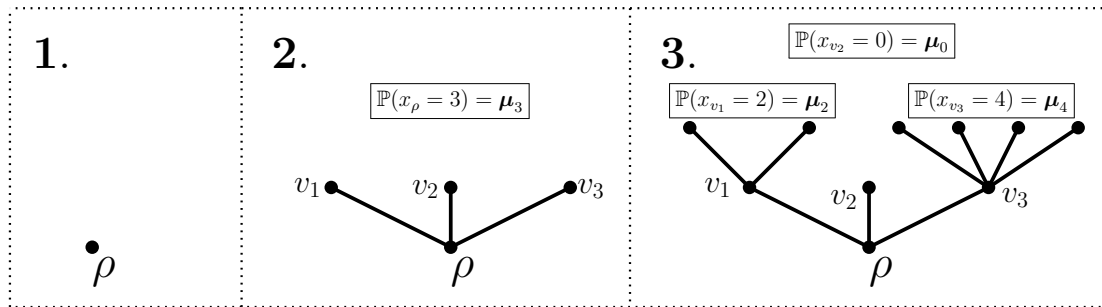


FIGURE 1.5 – Construction récursive d'un arbre de Bienaymé-Galton-Watson de loi de reproduction  $\mu$ . À gauche, le sommet racine de celui-ci, isolé. Ce sommet donne naissance à trois enfants avec probabilité  $\mu_3$  (image du milieu). Ces trois enfants donnent eux-mêmes naissance à respectivement deux, aucun et quatre enfants (image de droite). La probabilité de cet événement joint est  $\mu_2 \mu_0 \mu_4$ . On itère le procédé indéfiniment ou jusqu'à extinction de toutes les lignées.

**Extinction/non-extinction** Le modèle de Bienaymé-Galton-Watson peut également, sous certaines conditions, générer des arbres infinis. Trouver un critère nécessaire et suffisant sur la loi de reproduction  $\mu$  pour que ceci ait une chance non nulle de se produire est un exercice très classique de probabilités. En fait, nous l'avons dit, ce problème constitue même la motivation originelle derrière l'introduction de ce modèle par Bienaymé, Galton et Watson. Un arbre  $\mathbf{T}$  de taille infinie correspondra en effet, dans le processus de branchement associé, à la survie d'un patronyme, alors qu'un arbre fini signifiera son extinction.

**De l'importance de la fonction génératrice** La fonction génératrice  $\phi(z) = \sum_{k \geq 0} \mu_k z^k$  de  $\mu$  joue très souvent un rôle crucial dans l'analyse du modèle. C'est par exemple la valeur de  $m := \phi'(1)$  qui détermine si la distribution de  $\mathbf{T}$  charge ou non l'ensemble des arbres infinis :

- si  $m \leq 1$  (et si  $\mu_1 < 1$ , afin d'éviter une situation dégénérée triviale), l'arbre  $\mathbf{T}$  est fini presque sûrement : on parle alors d'arbre *sous-critique* (pour  $m < 1$ ) ou *critique* (pour  $m = 1$ ) ;
- si  $m > 1$ , la probabilité qu'il soit de taille infinie est non nulle : on parle dans ce cas d'arbre *surcritique*.

Ce n'est en réalité guère surprenant, car  $m$  correspond au nombre moyen d'enfants  $\mathbb{E}[\mathbf{X}]$  auquel donne naissance chaque individu de la lignée. Il paraît donc naturel qu'un patronyme dont les porteurs ont en moyenne moins d'un enfant soit voué à disparaître. Mais s'ils sont plus prolifiques

et que leur nombre moyen d'enfants devient strictement plus grand que 1, alors le patronyme aura une chance de survivre. Dans ce second cas, si elle survit, on peut parfois prouver que la lignée croîtra à vitesse exponentielle. Plus précisément, si on note  $\mathbf{Z}_n$  la taille de la  $n$ -ème génération d'enfants, on a sur l'événement de survie ( $|\mathbf{T}| = +\infty$ ) la convergence presque sûre de  $m^{-n} \cdot \mathbf{Z}_n$  vers une variable aléatoire strictement positive. Il est toutefois nécessaire que la loi de reproduction vérifie certaines conditions d'intégrabilité (théorème de Kesten et Stigum [89], voir aussi [102]).

La fonction génératrice permet aussi de calculer la probabilité de survie, ou de façon équivalente, la probabilité d'extinction de la lignée. On montre en effet que cette dernière est la solution minimale de l'équation de point fixe  $\phi(s) = s$  sur l'intervalle  $[0, 1]$ . La propriété de branchement est la principale explication au caractère extrêmement analytique du modèle de Bienaymé-Galton-Watson. Le premier chapitre du livre [76] propose une revue très complète des résultats connus sur le modèle de Bienaymé-Galton-Watson, vu comme processus de branchement.

**Arbres de Bienaymé-Galton-Watson et marches aléatoires** Le modèle de Bienaymé-Galton-Watson définit donc une distribution sur les arbres plans, caractérisée par l'équation 1.2.1. Or, nous avons vu en partie 1.1 que ces derniers pouvaient être codés par des processus unidimensionnels, dont le processus de contour (voir Figure 1.2). Lorsqu'on étudie le processus de contour d'un arbre de Galton-Watson, ce processus devient aléatoire. Dans le cas où la loi de reproduction  $\mu$  de l'arbre est géométrique de paramètre  $\alpha \in (0, 1)$ , autrement dit que

$$\mu_k = (1 - \alpha)\alpha^k$$

pour tout entier  $k \geq 0$ , il est facile de voir — grâce essentiellement à la propriété bien connue d'*absence de mémoire* des lois géométriques — que la loi de ce processus est en fait exactement celle d'une excursion au dessus de 0 de la marche aléatoire sur  $\mathbb{Z}$  dont les incréments suivent une loi de Bernoulli de paramètre  $\alpha$ .

La loi du processus de contour est moins clair lorsque  $\mu$  n'est plus géométrique. Il est alors plus profitable de considérer un autre processus, appelé *marche de Lukasiewicz*, défini lui aussi en parcourant, comme sur la Figure 1.2(a), l'arbre sous-jacent  $\mathbf{T}$ , mais en ne repassant pas par les sommets déjà visités. Ses variations ne sont plus seulement faites de  $+1$  ou  $-1$ , mais sont égales au nombre d'enfants du sommet visité, moins un. Quelle que soit  $\mu$ , la loi de la marche de Lukasiewicz est celle d'une excursion au dessus de 0, s'achevant en  $-1$ , d'une marche aléatoire dont la distribution des incréments est  $\mu(1 + \cdot)$ . Grâce à ce résultat général, on réussit à montrer qu'au moins à grande échelle, c'est-à-dire lorsque le nombre de sommets de l'arbre  $\mathbf{T}$  tend vers l'infini, la loi du processus de contour est approximativement celle de l'excursion d'une marche aléatoire, dont les incréments ont même variance que  $\mu$ . Ce qui a des conséquences géométriques importantes, que nous aborderons dans la section suivante. Une référence classique sur le sujet des liens entre arbres de Bienaymé-Galton-Watson et marches aléatoires est [97].

**Arbres de Bienaymé-Galton-Watson et modèles uniformes** Si le modèle d'arbres de Bienaymé-Galton-Watson s'avère si utile, et si étudié dans la littérature, c'est aussi parce qu'en choisissant intelligemment la loi de reproduction  $\mu$ , on montre qu'il s'identifie à des modèles aléatoires uniformes. Par exemple, lorsque  $\mu$  est une distribution géométrique de paramètre  $1/2$ , l'arbre *critique* de Bienaymé-Galton-Watson associé, conditionné à être de taille  $n \geq 1$ , est distribué uniformément sur l'ensemble des arbres plans à  $n$  sommets. Si nous verrons plus tard que cette identification a des conséquences théoriques importantes en termes de géométrie, nous pouvons déjà présenter, en guise d'illustration, une conséquence pratique non triviale. Grâce à la propriété de branchement, d'aucuns noteront qu'il est aisé de simuler numériquement un arbre de Bienaymé-Galton-Watson : cela revient à savoir générer des variables aléatoires indépendantes à valeurs entières, de loi  $\mu$ . Simuler un arbre aléatoire uniformément distribué sur l'ensemble des arbres plans à  $n$  sommets — pour  $n$  grand — est à l'inverse compliqué en première approche, car la taille de l'ensemble, à savoir  $\frac{1}{n+1} \cdot \binom{2n}{n}$ , croît avec  $n$  à vitesse exponentielle. Sa représentation comme arbre de Bienaymé-Galton-Watson permet de contourner cette difficulté. En effet, la probabilité que cet arbre soit exactement de taille  $n$  décroît asymptotiquement comme  $n^{-3/2}$  lorsque  $n \rightarrow +\infty$ . En appliquant alors une classique méthode de rejet, c'est-à-dire en ne conservant, à l'issue de nos simulations, que les arbres de taille  $n$ , on réduit drastiquement la vitesse de l'algorithme de génération d'un arbre plan uniforme à  $n$  sommets.

## Cartes de Boltzmann

Les cartes de Boltzmann sont nées de la volonté de définir une mesure de probabilité sur les cartes planaires, ayant la même forme produit que la distribution des marches aléatoires sur les chemins discrets ou celle des arbres de Bienaymé-Galton-Watson sur l'ensemble des arbres plans (voir l'équation (1.2.1)).

**Définition du modèle** Modèle nettement plus récent que ceux d'Erdős-Renyi et de Bienaymé-Galton-Watson, les cartes de Boltzmann ont été introduites en 2007 par Jean-François Mackert et Grégory Miermont dans [107]. Il s'agit de cartes aléatoires biparties — toutes les faces sont de degré pair — définies de la façon suivante. Considérons  $\mathbf{q} = (\mathbf{q}_k)_{k \geq 1}$  une suite de réels positifs ou nuls. Pour toute carte bipartie  $m$ , attribuons-lui un *poids de Boltzmann* :

$$w_{\mathbf{q}}(m) = \prod_{f \in \mathbf{Faces}(m)} \mathbf{q}_{\deg(f)/2},$$

où le produit est indexé sur l'ensemble des faces de la carte  $m$ , et où  $\deg(\cdot)$  désigne le degré d'une face de celle-ci. Le lecteur notera la similarité avec l'équation (1.2.1), la seule différence étant que ce ne sont pas ici les sommets, mais les faces de la carte, qui portent son "poids".

Une carte planaire aléatoire  $\mathfrak{M}$  est une **q-carte de Boltzmann** si sa distribution vérifie :

$$\mathbb{P}(\mathfrak{M} = m) = \frac{w_{\mathbf{q}}(m)}{Z_{\mathbf{q}}},$$

pour toute carte bipartie  $m$ . Le dénominateur  $Z_{\mathbf{q}}$  est une constante de normalisation, appelée aussi *fonction de partition*. Contrairement aux arbres de Bienaymé-Galton-Watson, le poids de Boltzmann d'une carte bipartie ne s'interprète pas directement comme une probabilité. En outre, rien ne garantit a priori que  $Z_{\mathbf{q}} < +\infty$ , et donc que la distribution définie ci-dessus a bien un sens. Signalons enfin que le choix de ne considérer que des cartes biparties provient du fait que leur énumération — d'après la méthode conçue par Tutte — s'avère nettement plus simple que celle des cartes générales.

**Admissibilité** On dit que la suite  $\mathbf{q}$  est *admissible* lorsque  $Z_{\mathbf{q}} < +\infty$ . Un critère nécessaire et suffisant d'admissibilité peut être déterminé, grâce encore aux méthodes d'énumération de Tutte. Ce critère est lié aux propriétés d'une certaine fonction génératrice  $f_{\mathbf{q}}$ . Lorsqu'il est vérifié, les valeurs respectives de  $f'_{\mathbf{q}}(Z_{\mathbf{q}})$  et  $f''_{\mathbf{q}}(Z_{\mathbf{q}})$  permettent de classer les cartes de Boltzmann en différentes catégories — *sous-critiques*, *critiques*, *régulières critiques*, *critiques génériques* et *critiques non génériques* — ayant chacune leurs particularités. On remarquera ici à nouveau la similarité avec le modèle de Bienaymé-Galton-Watson. En fait, il y a bien un arbre de Bienaymé-Galton-Watson sous-jacent à la carte, avec une fonction génératrice associée à la loi de reproduction proche de  $f_{\mathbf{q}}$ . L'explication repose sur l'existence d'une bijection entre cartes planaires et un ensemble d'arbres étiquetés, la bijection Bouttier-Di Francesco-Guitter [34]. Voir [107, Section 2] ou l'introduction de la thèse de Loïc Richier [122, Section 1.3] pour plus de détails.

**Cartes de Boltzmann et modèles uniformes** En choisissant  $\mathbf{q} = (\omega \cdot \mathbf{1}_{k=2})_{k \geq 0}$  pour  $\omega > 0$ , on attribue un poids de Boltzmann valant  $\omega^n$  aux quadrangulations à  $n$  faces, et un poids nul aux cartes biparties qui ne sont pas des quadrangulations. La  $\mathbf{q}$ -carte de Boltzmann est donc presque sûrement une quadrangulation. Si on la conditionne à avoir exactement  $n$  faces, sa loi sera uniforme sur l'ensemble des quadrangulations à  $n$  faces. Il faut cependant s'assurer de l'admissibilité de  $\mathbf{q}$ , qui est vérifiée lorsque  $\omega \leq \frac{1}{12}$ . Le principe peut être généralisé, en prenant  $\mathbf{q} = (\omega \cdot \mathbf{1}_{k=p})_{k \geq 0}$ , pour générer, quel que soit  $p \geq 1$ , des  $2p$ -angulations uniformes, dont on connaît aussi le seuil d'admissibilité. C'est vrai également des triangulations, qui ne sont pas des cartes biparties, mais pour lesquelles on dispose de formules d'énumération simples, obtenues à nouveau par la méthode de Tutte.

**Propriété de Markov spatiale et algorithmes d'épluchage** Contrairement à celle des arbres de Bienaymé-Galton-Watson, la définition des  $\mathbf{q}$ -cartes de Boltzmann n'est pas récursive. Elle ne fournit pas d'analogie évident à la propriété de branchement. La forme produit de leur distribution, imitant délibérément celle des marches aléatoires et des arbres de Bienaymé-Galton-Watson, laisse cependant penser qu'il en existe bel et bien un.

Apparus tout d'abord en physique théorique dans un travail de Watabiki [143], puis définis rigoureusement par Omer Angel dans [15] pour étudier des triangulations du plan, les algorithmes d'épluchage de cartes fournissent une construction dynamique des  $\mathbf{q}$ -cartes de Boltzmann, tout en mettant en évidence une *propriété de Markov spatiale* vérifiée par celles-ci. Éplucher une carte planaire consiste à la découvrir progressivement, face après face, de proche en proche. Décrivons le processus sans nous embarrasser de trop de formalisme.

Supposons observée une face d'une carte (bipartite)  $m$ . Choisissons maintenant, selon une règle quelconque, une arête  $\epsilon$  bordant cette face, puis révélons l'autre face adjacente à celle-ci dans  $m$ . Deux situations sont alors possibles :

1. la face révélée est différente de la face déjà observée. Dans ce cas, on la "colle" à la première le long de l'arête  $\epsilon$ .
2. la face révélée est la face déjà observée de  $m$  : cela signifie que l'arête  $\epsilon$  coïncide avec une autre arête bordant celle-ci. On tord alors notre face de façon à identifier les deux arêtes. L'opération coupe en deux la zone inexplorée de  $m$  (sauf si les deux arêtes pointaient un même sommet). On crée un *trou* dans la carte.

Voir une illustration en Figure 1.6.

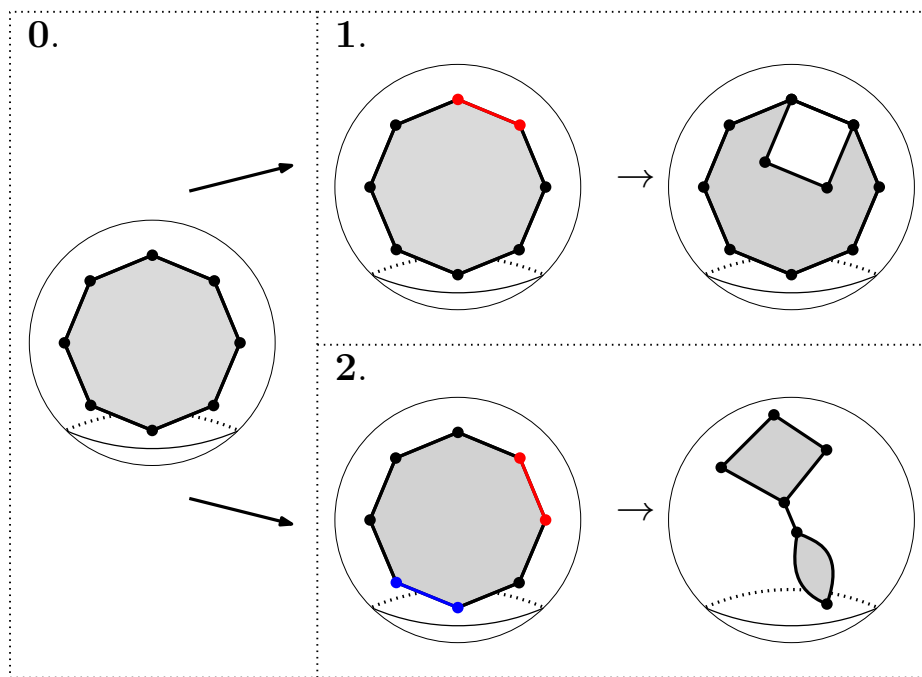


FIGURE 1.6 – Déroulés possibles d'une étape d'épluchage. La face déjà observée d'une carte est en blanc (elle est de degré 8) . La région grisée est la partie de la carte encore inexplorée par le processus, c'est-à-dire un *trou*. Nous révélons une face adjacente à une arête choisie (en rouge) sur le bord de la face observée. Si la face révélée est connue (situation n° 2), l'arête rouge coïncide forcément avec une autre arête (ici en bleu). On crée alors un second trou dans la carte courante.

On répète l'opération en choisissant à chaque étape une arête bordant un des trous de la carte, la règle de sélection de celle-ci n'important pas. À tout instant du processus, la carte courante est un assemblage connexe de faces (ou de polygones, pour reprendre la définition alternative des cartes décrite en Section 1.1), entourant un ou plusieurs trous. Ces trous peuvent être vus comme des faces de notre carte courante, mais disposant d'un statut particulier, car il ne s'agit pas de

"vraies" faces de la carte sous-jacente  $m$ . En fait, s'il y a  $p$  trous dans notre carte, dont les bords ont pour longueurs respectives  $\ell_1, \dots, \ell_p$ , on montre qu'il existe  $p$  cartes  $m_1, \dots, m_p$ , définies de manière unique, avec  $m_i$  contenant une face de degré  $\ell_i$ , telles que si on les "colle" le long des bords des trous, alors on reconstitue entièrement la carte  $m$ . Voir l'illustration en Figure 1.7.

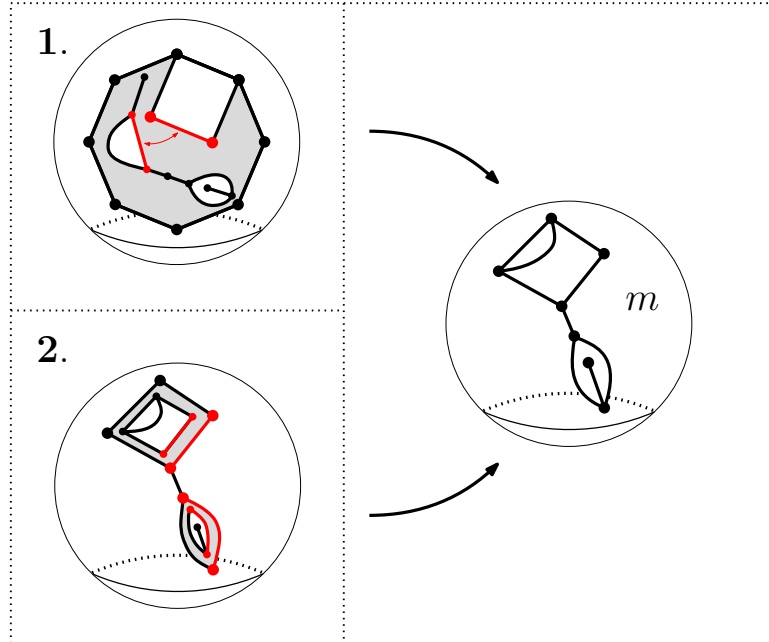


FIGURE 1.7 – Reconstitution de la carte complète (à droite) à partir des cartes avec trous issues de l'étape d'épluchage de la Figure 1.6. Les trous sont remplis en "collant" des cartes, définies de manière unique, le long de leur bord. Les arêtes en rouge indiquent comment est réalisé le "collage".

Quand  $m = \mathfrak{M}$  est une  $\mathbf{q}$ -carte de Boltzmann, ces  $p$  cartes sont elles-mêmes des  $\mathbf{q}$ -cartes de Boltzmann, indépendantes, conditionnellement aux longueurs des bords des trous. C'est l'attendue **propriété de Markov spatiale** : ce qu'il reste à découvrir de la carte  $\mathfrak{M}$  (le futur) ne dépend des régions déjà découvertes (le passé) qu'au travers des bords de celles-ci (le présent). En outre, on montre que le processus d'épluchage est une chaîne de Markov homogène, ce qui fournit donc bien un algorithme dynamique générant des  $\mathbf{q}$ -cartes de Boltzmann. Nous invitons le lecteur à consulter [53] pour une introduction aux processus d'épluchage de cartes planaires et leurs nombreuses applications.

### 1.3 De la géométrie des graphes, arbres et cartes

Dans la définition d'un graphe (voir Section 1.1), le choix de l'ensemble  $V$  pour étiqueter les sommets peut sembler arbitraire, sauf éventuellement pour certains problèmes de combinatoire. Car ce qui paraît caractériser réellement un graphe, ce sont les relations d'adjacence entre sommets. On retrouve ce type de considération en théorie des groupes. Supposons que  $(G, *)$  soit un groupe de loi de composition interne  $*$ . La nature des éléments de  $G$  — qu'ils soient des nombres, des fonctions, etc. — est indifférente. Ce qui rend "unique" ce groupe, décrit sa structure, ce sont les résultats des opérations  $g * g'$  donnés par la loi de composition interne pour toute paire d'éléments  $g, g' \in G$ . Pour traduire mathématiquement cette forme d'unicité, on introduit la notion d'isomorphisme de groupes : un groupe est isomorphe à  $G$  s'il existe une bijection  $\varphi$  entre  $G$  et ce groupe, qui préserve les résultats d'opérations par la loi de composition interne, au sens où  $\varphi(g * g') = \varphi(g) * \varphi(g')$  pour tous  $g, g' \in G$ . Un groupe est alors généralement identifié à sa classe d'isomorphismes, qui contient l'ensemble des groupes qui lui sont isomorphes.

Pour les graphes, on raisonne de manière analogue, sauf qu'ici, ce sont les relations d'adjacence induites par les arêtes que nous souhaitons préserver. On dit que deux graphes  $G_1$  et  $G_2$  sont

*isomorphes* s'il existe une bijection entre les ensembles de sommets, vérifiant la propriété suivante : l'arête  $(\varphi(x), \varphi(y))$  existe dans  $G_2$  si et seulement si l'arête  $(x, y)$  existe dans  $G_1$ . Deux graphes isomorphes partagent une même "géométrie". Nous allons préciser dans la suite ce que nous entendons exactement par "géométrie", mais pour donner dès maintenant une intuition du propos, remarquons que si  $G_1$  est un arbre, alors  $G_2$  isomorphe à  $G_1$  l'est aussi ; si  $G_1$  est connexe, alors  $G_2$  aussi, etc.

**Le graphe comme espace métrique** Dans la Section 1.1, lorsque nous avons défini la connexité d'un graphe, puis présenté la classe des arbres, nous avons parlé de *chemin d'arêtes*, dont le sens s'accorde avec l'image que l'on s'en fait : deux sommets  $x$  et  $y$  sont connectés par un chemin d'arêtes si une ligne continue peut être tracée de  $x$  à  $y$ , passant uniquement par les segments représentant les arêtes du graphe. Plus formellement, il existe une suite d'arêtes telle que : (i) la première émane du sommet  $x$  ; (ii) deux arêtes successives confluent en un même sommet ; (iii) la dernière arête atteint le sommet  $y$ .

On appelle *longueur* d'un chemin d'arêtes le nombre d'arêtes constituant ce chemin. Il s'agit donc d'un entier naturel. Si, dans un graphe, deux sommets sont connectés par un ou des chemins d'arêtes, il en existe forcément un minimisant la longueur. On nomme *géodésiques* ces chemins de longueur minimale.

Introduisons  $\mathbf{d}_{gr}$  l'application qui à toute paire de sommets d'un graphe, associe la longueur d'une géodésique reliant ces deux sommets. Pour deux sommets  $x, y$  qu'aucun chemin ne connecte — n'étant pas dans la même *composante connexe*, on pose  $\mathbf{d}_{gr}(x, y) = +\infty$ . Cette application  $\mathbf{d}_{gr}$  vérifie tous les axiomes d'une distance. Elle s'appelle *distance de graphe*. Muni de celle-ci, l'ensemble de sommets d'un graphe se transforme en espace métrique. Il devient alors naturel de s'intéresser aux notions habituelles de topologie des espaces métriques : diamètre du graphe, boules et frontières de celles-ci, etc. Remarquons pour conclure ce paragraphe que la seule donnée de  $\mathbf{d}_{gr}$  suffit à reconstituer la structure d'adjacence du graphe sous-jacent : l'arête  $(x, y)$  existe dans celui-ci si et seulement si  $\mathbf{d}_{gr}(x, y) = 1$ . Notons aussi qu'un isomorphisme de graphes réalise une *isométrie* entre deux espaces métriques. Cette interprétation confirme le caractère fondamentalement géométrique des graphes.

**Géométrie à grande échelle** Étudier dans le détail la géométrie d'un graphe particulier peut apporter de précieuses informations sur celui-ci. Dans cette thèse cependant, nous n'adoptons pas ce point de vue statique et préférons nous intéresser aux effets géométriques causés par la variation d'une quantité macroscopique descriptive des graphes, à savoir leur *volume*, ou nombre de sommets qu'ils contiennent. Plus précisément, nous souhaitons comprendre l'évolution de la géométrie d'un graphe aléatoire  $\mathbf{G}_n$  à  $n$  sommets, lorsque sa taille  $n$  tend vers l'infini par un mécanisme explicite. En déterminant par exemple une asymptotique de son diamètre, du volume des boules de rayon fixé, etc. Ces problématiques sont rassemblées sous le terme de *géométrie à grande échelle*.

**Exemples de résultats** Les caractéristiques topologiques des trois modèles fondamentaux que nous avons présentés dans la section précédente — graphes d'Erdős-Renyi, arbres de Bienaymé-Galton-Watson et cartes de Boltzmann — sont aujourd'hui bien connues. Ces modèles illustrent la variété de comportements qu'il est permis d'observer.

Prenons l'exemple du diamètre. Nous avons vu que les arbres de Bienaymé-Galton-Watson se classent en trois catégories — sous-critiques, critiques, surcritiques — selon que la moyenne de la loi de reproduction associée est strictement inférieure, égale ou strictement supérieure à 1. Un arbre surcritique a une chance non nulle d'être de taille infinie. Donc son diamètre aussi. Ce n'est pas vrai des arbres critiques, ni sous-critiques. On peut cependant tenter d'estimer l'ordre de grandeur du diamètre d'un arbre appartenant à l'une de ces deux catégories, conditionné à contenir exactement  $n$  sommets, avec  $n \rightarrow +\infty$ . Dans le cas *critique*, nous savons par exemple que le diamètre est d'ordre  $\sqrt{n}$  si la loi de reproduction est de variance finie [7, 97]. Cela inclut les arbres uniformes. Lorsque la variance est infinie mais que la queue de distribution de la loi de reproduction décroît (pour faire simple) à vitesse polynomiale, le diamètre est cette fois-ci d'ordre  $n^\delta$  pour un certain  $\delta \in (0, 1/2)$  [62, 92].

Le paysage est assez similaire dans l'univers des cartes de Boltzmann. La croissance du diamètre est généralement polynomiale, bien que les exposants ne soient pas identiques à ceux des arbres

de Bienaymé-Galton-Watson. Le diamètre des quadrangulations uniformes à  $n$  sommets est par exemple d'ordre  $n^{1/4}$  [48]. C'est vrai également pour les cartes de Boltzmann dites *génériques critiques*, sorte d'analogue aux arbres critiques avec loi de reproduction de variance finie, évoqués ci-dessus [98, 108]. Quant aux cartes de Boltzmann *non génériques critiques*, comparables cette fois aux arbres critiques avec loi de reproduction de variance infinie, on conjecture que l'ordre de leur diamètre est  $n^\delta$  pour un certain  $\delta \in (1/4, 1/2)$  [69].

Les graphes d'Erdős-Rényi présentent une plus grande diversité de comportements. L'ordre de grandeur de leur diamètre change radicalement selon qu'on fait dépendre ou non le paramètre de percolation  $p$  de la taille  $n$  du graphe. Lorsque  $p$  est constant, le diamètre du graphe converge presque sûrement vers 2! Ce qui signifie qu'à la limite, tous les sommets ont au moins un voisin en commun. Le résultat reste même valable pour une suite de paramètres  $p_n$  tendant vers 0, mais moins vite que  $(1+\epsilon)\sqrt{\frac{2\log n}{n}}$  pour un certain  $\epsilon > 0$ . Lorsqu'on accélère la vitesse de décroissance, de sorte que  $1 \leq c \leq np_n \leq C \log n$  pour  $C > 0$ , le diamètre croît vers l'infini comme  $\frac{\log n}{\log np_n}$  [50, 123]. Si  $C < 1$ , le graphe est presque sûrement non connexe à l'infini. Le diamètre est alors celui de la plus grande composante connexe, dont la taille est proportionnelle à  $n$ . On dit qu'il s'agit d'une composante *géante*. Quand  $p_n \underset{n \rightarrow +\infty}{\sim} \frac{1}{n}$ , la composante géante disparaît [65]. Dans une "fenêtre critique" mise en évidence par Aldous [8], correspondant à  $np_n = 1 + \lambda n^{-1/3}$  pour  $\lambda \in \mathbb{R}$ , les composantes connexes sont toutes de taille  $n^{2/3}$ , et leur diamètre croît comme  $n^{1/3}$  [3]. Les phénomènes décrits dans ce paragraphe ne constituent qu'une partie de ceux observés dans les graphes d'Erdős-Rényi, mais prouvent déjà leur grande variété.

**Limites locales de graphes** Lorsqu'on étudie les propriétés géométriques asymptotiques de graphes aléatoires, en faisant croître à l'infini leur volume, la question de l'existence de structures limites émerge naturellement. Introduite par Benjamini et Schramm dans [24], puis dans une version légèrement amendée par Angel et Schramm dans [20] (c'est elle que nous présentons ici), la topologie locale donne un sens à la convergence *discrète* de graphes ou cartes de taille finie vers des graphes ou cartes connectant un nombre infini (dénombrable) de sommets. Son principe peut être décrit ainsi : une suite (déterministe) de graphes  $G_n$ , *enracinés*, de volume croissant, converge pour la topologie locale si le "paysage" autour du sommet racine "se stabilise" asymptotiquement. Nous insistons sur la nécessité d'*enraciner* les graphes. Plus formellement, la topologie locale est induite par une métrique, la *distance locale*, définie de la façon suivante. Pour tous graphes finis  $G, G'$  :

$$\mathbf{d}_{loc}(G, G') = (1 + \sup\{r \geq 0 : B_r(G) = B_r(G')\})^{-1},$$

où  $B_r(G)$  désigne le sous-graphe de  $G$  contenant tous les sommets de  $G$  situés à distance au plus  $r$  de la racine (pour la distance de graphe) et les arêtes reliant ces sommets entre eux. On l'appelle *boule de rayon  $r$* . Pour résumer, deux graphes sont proches pour la distance locale s'ils coïncident (au sens des isomorphismes de graphes) sur un grand voisinage autour de la racine.

La distance locale vérifie tous les axiomes d'une distance et transforme en espace métrique l'ensemble des graphes finis. Son complété est un espace métrique polonais (séparable et complet). On vérifie que les éléments du complété qui ne sont pas des graphes finis sont effectivement des graphes contenant un nombre infini (dénombrable) de sommets, mais localement finis (tous les sommets n'ont qu'un nombre fini de voisins).

Le même résultat vaut pour les cartes planaires. Parmi les cartes planaires infinies obtenues comme limites locales, celles dites *à un bout* ont un rôle important. Une carte est *à un bout* si pour tout sous-ensemble fini de sommets, le complémentaire contient une unique composante connexe infinie. La particularité de ces cartes est de pouvoir être vues, à l'image des cartes finies, comme des graphes infinis plongés proprement dans le plan, à déformation continue près. Tout compact de  $\mathbb{R}^2$  n'intersecte alors qu'un nombre fini de sommets et d'arêtes de la carte. Une carte à un bout contient au plus une face de degré infini, c'est-à-dire bordée d'une infinité d'arêtes. Lorsqu'elle n'en contient pas, son plongement recouvre tout  $\mathbb{R}^2$  et le décompose en une infinité de domaines bornés — les faces — tous entourés d'un nombre fini d'arêtes. On parle de *carte infinie du plan*. Lorsqu'elle contient une face de degré infini, le plongement recouvre cette fois-ci  $\mathbb{R}_+ \times \mathbb{R}$  (par exemple), l'autre moitié du plan  $\mathbb{R}_- \times \mathbb{R}$  étant la face de degré infini. On parle de *carte infinie du demi-plan*. Nous invitons le lecteur à consulter [53, Chapitre 2] pour plus de détails sur la topologie locale.



**Exemples de limites locales et de leurs propriétés** Une première limite locale non triviale, couramment donnée en exemple pour introduire au sujet, concerne les composantes connexes d'un graphe d'Erdős-Renyi, considéré dans le régime de la transition de phase pour l'existence d'une composante géante, c'est-à-dire lorsque  $np_n = \lambda > 0$ . Sous cette hypothèse, on montre que la composante connexe d'un sommet particulier converge localement vers un arbre de Bienaymé-Galton-Watson, de loi de reproduction une loi de Poisson de paramètre  $\lambda$ . En effet, le nombre de connexions réalisées par ce sommet suit une loi binomiale de paramètres  $n - 1$  et  $p_n = \lambda/n$ , laquelle converge en loi, quand  $n \rightarrow +\infty$ , vers la loi de Poisson attendue, par un résultat classique d'approximation. C'est également le cas du degré des sommets voisins, qui par ailleurs se connectent entre eux avec une probabilité tendant vers zéro. Autrement dit, la probabilité de créer un cycle de longueur 3 devient négligeable asymptotiquement. Cette analyse s'étend ensuite aux échelons suivants, aux voisins des voisins, puis aux voisins de ces derniers, etc. Il n'est donc pas surprenant de retrouver un arbre à la limite, compte tenu de l'absence de cycles. Voir [78, Chapitre 2] pour une preuve détaillée du résultat.

Pour les arbres de Bienaymé-Galton-Watson, la topologie locale est l'outil adapté pour construire un équivalent *critique* aux arbres surcritiques conditionnés à survivre, seule distribution non triviale rencontrée à ce stade (voir Section 1.2), chargeant positivement les graphes infinis. Rappelons qu'à la différence d'un arbre surcritique, un arbre critique est presque sûrement de taille finie. On ne peut donc pas conditionner sur l'événement de survie du processus de branchement sous-jacent. En revanche, il est possible de conditionner l'arbre à être grand, puis faire croître son volume à l'infini. En procédant ainsi, on parvient à faire émerger une limite locale non triviale, qui est presque sûrement un arbre infini. On l'appelle *arbre de Kesten*, en référence au travail [86] de Harry Kesten, qui identifia le premier la distribution du graphe limite. Ce dernier est un arbre de Bienaymé-Galton-Watson de loi de reproduction critique  $\mu$ , où deux catégories de sommets coexistent, les sommets *normaux* et les sommets *spéciaux*. À la différence des sommets normaux, le nombre d'enfants des sommets spéciaux n'est pas distribué selon la loi de reproduction  $\mu$ , mais selon la loi biaisée par la taille  $\tilde{\mu}$ , définie par  $\tilde{\mu}_k = k\mu_k$  pour tout  $k \geq 0$ . À chaque génération de l'arbre, un unique sommet est désigné spécial. Le sommet racine de l'arbre est spécial. À la génération suivante, le sommet spécial est choisi uniformément parmi les enfants de la racine. Et ainsi de suite. L'arbre résultant de cette procédure est nécessairement infini, puisque les sommets spéciaux ont toujours au moins un enfant. Une façon simple de se figurer l'arbre est de tracer un chemin infini de sommets — les sommets spéciaux — puis de greffer sur eux des arbres de Bienaymé-Galton-Watson critiques indépendants, de loi de reproduction  $\mu$ . Le chemin infini est l'*épine dorsale* de l'arbre. La définition de l'arbre de Kesten s'étend à des lois de reproduction sous-critiques et surcritiques, mais dans ces deux cas, la limite pour la topologie locale des arbres de Bienaymé-Galton-Watson, conditionnés à être grand, est plus complexe à calculer, surtout dans le cas sous-critique. Voir [1, 2] pour plus de détails. La géométrie de l'arbre de Kesten s'avère significativement différente de l'arbre surcritique conditionné à survivre. Par exemple, le volume de la boule de rayon  $r$  croît comme  $r^2$  dans le premier — à vitesse polynomiale donc — sous la condition de variance finie de la loi de reproduction (voir [88]), alors que ce même volume croît exponentiellement dans le second, sous l'hypothèse de Kesten et Stigum [89, 102].

La définition de la topologie locale d'Angel et Schramm [20], reprise ici, est utilisée dans leur article pour démontrer la convergence quand  $n \rightarrow +\infty$  des triangulations uniformes à  $n$  sommets vers une triangulation aléatoire infinie du plan, qu'ils nomment UIPT pour *Uniform Infinite Planar Triangulation*. Ce résultat fonde la théorie probabiliste des cartes planaires infinies. À leur suite, Maxim Krikun prouve qu'une convergence similaire se produit avec les quadrangulations uniformes, la limite étant cette fois une quadrangulation infinie du plan [95]. Plus généralement, ce type de convergence a lieu avec les cartes de Boltzmann *critiques* [131]. Avec une approche différente, il est possible d'obtenir à la limite des cartes aléatoires infinies du demi-plan, voir à ce sujet [16, 53, 59]. Dans [53], sont fournies les preuves des convergences locales mentionnées ci-dessus, utilisant de façon originale des techniques d'épluchage de cartes.

Par certains aspects, la géométrie de ces cartes infinies du plan, construites comme limites locales, peut être mise en parallèle avec celle de l'arbre de Kesten. On sait ainsi [15] que le volume des boules dans l'UIPT croît avec le rayon à vitesse polynomiale, plus exactement comme  $r^4$ , à corrections logarithmiques près. Le lecteur se demandera alors sans doute s'il existe un analogue dans le domaine des cartes planaires aux arbres de Bienaymé-Galton-Watson surcritiques conditionnés

à survivre. Omer Angel et Gourab Ray dans [19], puis Nicolas Curien dans [52], ont respectivement démontré l'existence de triangulations infinies du demi-plan et du plan, appelées PSHIT pour *Planar Stochastic Hyperbolic Infinite Triangulation* dans le second cas, partageant avec ces arbres certaines propriétés géométriques. Par exemple la croissance exponentielle du volume des boules [52, 121]. La thèse de Thomas Budzinski [41] est consacrée à l'étude de la géométrie de ces cartes dites *hyperboliques*. Celles-ci ne peuvent pas être construites comme limites locales de cartes planaires. Elles sont en fait limites locales de cartes plongées sur des surfaces de genre très grand, ainsi que l'ont récemment montré Thomas Budzinski et Baptiste Louf dans [45].

## Limites d'échelle

La topologie locale donne un sens aux convergences *discrètes* de graphes et cartes, vers des objets de même nature, de taille infinie. En fin de Section 1.1, nous avons expliqué que la physique théorique fournit des raisons de s'intéresser aux limites *continues* de cartes, comme un pendant en dimensions supérieures au théorème de Donsker, qui énonce la convergence de marches aléatoires vers le mouvement brownien. Il reste à déterminer la topologie adaptée à l'émergence de telles limites. Rappelons qu'un graphe (ou une carte) muni de la distance de graphe est un espace métrique. Il est compact quand son nombre de sommets est fini. La topologie de Gromov-Hausdorff rend rigoureuse la notion de convergence pour des espaces métriques compacts.

**La topologie de Gromov-Hausdorff** Nous commençons par définir la *distance de Hausdorff*, qui compare les parties compactes d'un même espace métrique  $(E, \mathbf{d})$ . Considérons deux parties compactes  $K_1, K_2$  de  $E$ . La distance de Hausdorff entre ces deux parties est égale à :

$$\mathbf{d}_H(K_1, K_2) := \inf \{ \epsilon > 0 \mid K_1 \subseteq K_2^\epsilon \text{ et } K_2 \subseteq K_1^\epsilon \}$$

où  $K^\epsilon$  désigne, pour tout compact  $K \subseteq E$ , le  $\epsilon$ -voisinage de  $K$ , qui inclut tous les points de  $E$  à distance strictement inférieure à  $\epsilon$  de l'un des éléments de  $K$ . Par conséquent, la distance de Hausdorff entre deux parties compactes de  $E$  est d'autant plus faible que celles-ci "se superposent" dans  $E$ .

Nous voulons maintenant comparer deux compacts non nécessairement inclus dans un même espace métrique. L'idée est de les plonger dans un espace métrique tiers, puis de calculer la distance de Hausdorff des images respectives. Le plongement doit se faire de manière isométrique, afin de préserver les caractéristiques géométriques des compacts. La *distance de Gromov-Hausdorff* est définie en prenant l'infimum sur tous les plongements possibles. Plus exactement, si  $(E_1, \mathbf{d}_1)$  et  $(E_2, \mathbf{d}_2)$  sont deux espaces métriques compacts, leur distance de Gromov-Hausdorff est égale à :

$$\mathbf{d}_{GH}(E, F) = \inf \{ \mathbf{d}_H(\phi_1(E_1), \phi_2(E_2)) \},$$

où l'infimum porte sur tous les espaces métriques  $(E, \mathbf{d})$ , et les paires d'injections isométriques  $\phi_1$  et  $\phi_2$  de  $E_1$  et  $E_2$  dans  $E$ . La distance de Gromov-Hausdorff mesure ainsi le "degré d'isométrie" entre deux espaces compacts. Remarquons que cette définition implique que la distance de Gromov-Hausdorff entre deux parties compactes d'un espace métrique est plus petite que leur distance de Hausdorff. Ceci est parfaitement souhaitable. Il suffit de songer à deux cercles du plan, de même rayon, mais dont les centres sont très éloignés : ils sont distincts pour la distance de Hausdorff, mais pas pour la distance de Gromov-Hausdorff, car ils sont isométriques.

La distance de Gromov-Hausdorff n'est une véritable distance que sur l'ensemble des espaces métriques compacts quotienté par la relation d'isométrie, afin de respecter l'axiome de séparation. Elle transforme par ailleurs cet ensemble en espace polonais, et fournit donc un cadre agréable pour étudier la convergence de variables aléatoires. Nous renvoyons le lecteur à [46] pour plus de détails sur la topologie de Gromov-Hausdorff.

**Limites d'échelle** Nous avons vu précédemment des exemples de suites de graphes aléatoires, où le diamètre croissait à l'infini avec le nombre de sommets. On ne peut donc pas espérer une convergence vers un espace métrique compact, au sens de la topologie de Gromov-Hausdorff. Dans le théorème de Donsker, les marches aléatoires sont renormalisées pour converger vers le mouvement brownien, le facteur de renormalisation étant l'ordre de grandeur du déplacement maximal de

la marche, soit  $\sqrt{n}$  après  $n$  sauts. On cherche à procéder de manière similaire avec les suites de graphes de volume croissant. Au lieu d'étudier directement une suite d'espaces métriques  $(\mathbf{G}_n, \mathbf{d}_{gr})$  — où  $\mathbf{G}_n$  est, rappelons-le, un graphe aléatoire à  $n$  sommets — on va plutôt s'intéresser à une suite renormalisée  $(\mathbf{G}_n, \lambda_n^{-1} \cdot \mathbf{d}_{gr})$  — on dit aussi "remise à l'échelle" — où toutes les distances dans le graphe  $\mathbf{G}_n$  sont divisées par un facteur  $\lambda_n > 0$ , correspondant grosso modo à l'ordre de grandeur du diamètre de  $\mathbf{G}_n$ . L'idée étant que celui de l'espace métrique  $(\mathbf{G}_n, \lambda_n^{-1} \cdot \mathbf{d}_{gr})$  restera borné asymptotiquement, ce qui élimine un premier motif évident de non-convergence. On dira qu'une suite de graphes aléatoires  $(\mathbf{G}_n, \mathbf{d}_{gr})$  de volume croissant admet une *limite d'échelle* s'il existe une suite de réels positifs  $(\lambda_n)_n$  telle que :

$$(\mathbf{G}_n, \lambda_n^{-1} \cdot \mathbf{d}_{gr}) \xrightarrow[n \rightarrow +\infty]{(d)} (\mathcal{G}, \Delta),$$

où  $(\mathcal{G}, \Delta)$  est un espace métrique compact aléatoire. La convergence a lieu en distribution pour la distance de Gromov-Hausdorff.

**L'arbre continu brownien** David Aldous énonce dans [7] un résultat fondateur pour la théorie des limites d'échelle de graphes aléatoires : il identifie un espace métrique universel, vers lequel convergent tous les grands arbres de Bienaymé-Galton-Watson *critiques*, renormalisés, dont la loi de reproduction satisfait une simple condition de second moment fini. On notera la similarité avec le théorème de Donsker. Formellement, si  $\mathbf{T}_n$  est un arbre de Bienaymé-Galton-Watson critique conditionné à être de taille  $n \geq 1$ , et si la loi de reproduction  $\mu$  du modèle est de variance finie  $0 < \sigma^2 < +\infty$ , alors la suite  $(\mathbf{T}_n, \mathbf{d}_{gr})$  admet une limite d'échelle  $(\mathcal{T}, \mathbf{d})$ , universelle, au sens où sa loi ne dépend pas de  $\mu$ . Le facteur de renormalisation de la suite est  $\lambda_n = \frac{2\sqrt{n}}{\sigma}$ . On appelle *arbre continu brownien* l'espace métrique limite, ou CRT en abrégé pour *Continuum Random Tree*. Nous insistons sur le fait que le CRT est un espace métrique *aléatoire*. Il vérifie cependant deux propriétés presque sûres le rapprochant des arbres binaires, bien qu'il s'agisse d'un espace continu :

1. comme son nom l'indique, le CRT est un *arbre continu*, donc un espace géodésique — toute paire de points est relié par une unique géodésique, ou plus court chemin — et sans cycle ;
2. pour tout  $x \in \mathcal{T}$ , l'espace  $\mathcal{T} \setminus \{x\}$  a au plus trois composantes connexes.

À première vue, le lien entre  $\mathcal{T}$  et le mouvement brownien n'est pas évident. Pour le comprendre, il faut se souvenir que le processus de contour d'un arbre de Bienaymé-Galton-Watson de grande taille peut être approchée par une marche aléatoire, voire être lui-même une marche aléatoire quand la loi de reproduction  $\mu$  est géométrique (voir Section 1.2). On l'approche plus exactement par une excursion de marche aléatoire puisque le processus de contour est à valeurs positives ou nulles. Quand la variance de  $\mu$  est finie, on montre ainsi, grâce au théorème de Donsker dans sa version conditionnelle, que le processus de contour de  $\mathbf{T}_n$ , remis à l'échelle d'un facteur  $\frac{2\sqrt{n}}{\sigma}$ , converge en loi vers l'excursion brownienne normalisée  $\mathbf{e} = (\mathbf{e}_t)_{t \in [0,1]}$ . Par conséquent, de la même façon qu'un arbre plan peut être reconstitué à partir de son seul processus de contour, on peut espérer qu'une excursion brownienne code le CRT.

Dans [63, 97], Thomas Duquesne et Jean-François Le Gall ont formalisé cette idée en introduisant une classe d'arbres continus construits à partir de fonctions réelles à support compact. Soit  $g$  une fonction continue sur l'intervalle  $[0, 1]$ , à valeurs positives, vérifiant  $g(0) = g(1) = 0$ . On associe à  $g$  une pseudo-distance sur  $[0, 1]$ , notée  $\mathbf{d}_g$ , telle que pour tous  $s, t \in [0, 1]$  :

$$\mathbf{d}_g(s, t) := g(s) + g(t) - 2 \min_{u \in [s \wedge t, s \vee t]} g(u).$$

On considère maintenant la relation d'équivalence  $\sim_g$  sur  $[0, 1]$ , définie par  $s \sim_g t$  si  $\mathbf{d}_g(s, t) = 0$ . L'intervalle  $[0, 1]$  quotienté par  $\sim_g$ , muni de  $\mathbf{d}_g$ , est un espace métrique compact. Nous le notons  $\mathcal{T}_g$ . Il s'agit de surcroît d'un arbre continu, appelé *arbre réel*. Il est important de remarquer que lorsque  $g$  est le processus de contour d'un arbre plan, éventuellement reparamétré pour en faire une fonction définie sur l'intervalle  $[0, 1]$ , l'espace métrique  $\mathcal{T}_g$  correspond exactement à l'arbre en question, muni de la distance de graphe. La relation d'équivalence identifie en effet les points "se faisant face" sous le graphe de  $g$ . Voir Figure 1.8 pour une illustration. Il s'avère en outre que l'application  $g \mapsto \mathcal{T}_g$  est continue pour la topologie de Gromov-Hausdorff (et en munissant l'espace de départ de la norme uniforme). Voir par exemple [97, Lemme 2.4]. Compte tenu du résultat de convergence

relatif aux processus de contour énoncé plus haut, le CRT coïncide donc avec l'arbre réel  $\mathcal{T}_e$  codé par l'excursion brownienne normalisée, ce qui rend explicite le lien avec le mouvement brownien.

Terminons ce paragraphe en notant que depuis le résultat d'Aldous, l'arbre continu brownien est apparu comme limite universelle de nombreuses autres structures géométriques discrètes et aléatoires. De structures arborescentes, à l'instar des arbres uniformes enracinés non ordonnés (c'est-à-dire dépourvu du système d'orientations des arbres plans) [73]. Mais également de diverses structures non arborescentes : de cartes planaires [28], de "dissections" de polygones [55], de chemins browniens en milieu hyperbolique [49], etc. Le CRT continue donc d'occuper une place centrale dans la théorie des limites d'échelle de structures aléatoires.

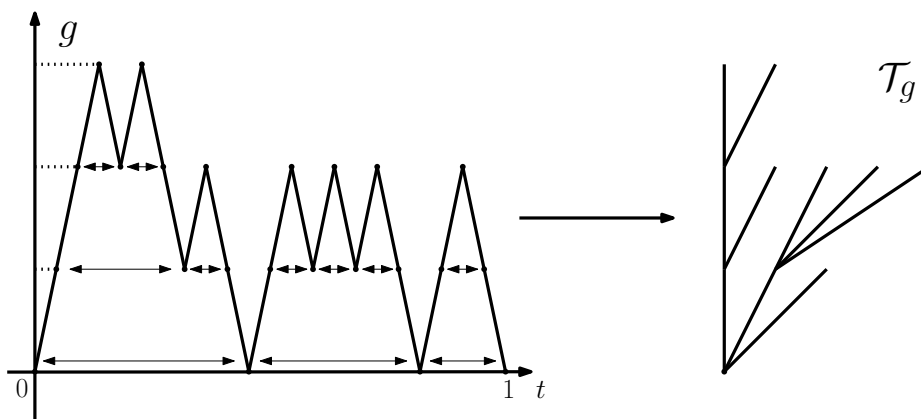


FIGURE 1.8 – La fonction  $g$ , dont le graphe est représenté à gauche de la figure, correspond au processus de contour de la Figure 1.2. La relation d'équivalence  $\sim_g$  identifie les points "se faisant face" sous le graphe de  $g$ . Le lecteur pourra vérifier que l'espace métrique  $\mathcal{T}_g$  qui en résulte, représenté à droite de la figure, correspond à l'arbre  $\mathbf{T}$  de la Figure 1.2.

**Autres exemples de limites d'échelle** Le résultat d'Aldous sur la convergence des arbres de Bienaymé-Galton-Watson vers le CRT repose de manière cruciale sur l'hypothèse de variance finie de la loi de reproduction  $\mu$ , car nous avons vu qu'il est en réalité une conséquence du théorème de Donsker. Le cas d'une variance infinie a depuis été examiné. Dans [62], Thomas Duquesne montre que si la queue de distribution décroît à vitesse polynomiale — autrement dit que  $\mu([k, +\infty)) \sim k^{-\alpha}$  quand  $k \rightarrow +\infty$ , pour un certain  $\alpha \in (1, 2)$  — alors le processus de contour d'un arbre conditionné à être de taille  $n$ , renormalisé cette fois par  $n^{-\frac{1-\alpha}{\alpha}}$ , tend vers un processus de Lévy  $\alpha$ -stable spectralement positif. Voir également [92] pour une preuve alternative de ce fait. En termes de limite d'échelle, cela signifie que les grands arbres renormalisés convergent pour la distance de Gromov-Hausdorff vers l'arbre réel codé par ce processus de Lévy. On l'appelle arbre  $\alpha$ -stable. Sa géométrie diffère notablement de celle du CRT. Nous renvoyons le lecteur vers [63] pour plus de détails sur ses propriétés.

La théorie des cartes planaires aléatoires possède elle aussi son résultat d'universalité. Dans [98], Jean-François Le Gall prouve que les grandes cartes de Boltzmann *régulières critiques* admettent une limite d'échelle commune (à un facteur multiplicatif près) lorsqu'on les renormalise par  $n^{1/4}$ , où  $n$  correspond à leur volume. Grégory Miermont démontra simultanément et indépendamment ce résultat pour les quadrangulations uniformes dans [113]. L'espace métrique apparaissant à la limite est appelé *carte brownienne*. Plus tard, c'est Cyril Marzouk qui étend dans [108] le résultat de convergence aux cartes de Boltzmann *génériques critiques*, qui sont une sorte d'analogue, dans le domaine des cartes planaires aléatoires, aux arbres de Bienaymé-Galton-Watson critiques avec loi de reproduction de variance finie. Comme avec les arbres, la preuve de la limite d'échelle repose sur un codage des cartes par des processus unidimensionnels, grâce à des bijections existant entre cartes planaires et arbres étiquetés. La carte brownienne peut elle-même être codée par un processus aléatoire, *la tête du serpent brownien*, introduit par Le Gall dans [96]. Ce dernier s'interprète comme un mouvement brownien défini sur le CRT. La géométrie de la carte brownienne se distingue nettement de celle de l'arbre continu brownien. On sait par exemple qu'elle est presque sûrement

homéomorphe à la sphère  $\mathbb{S}^2$  [99, 112]. Nous renvoyons le lecteur vers [114] pour une introduction complète à la théorie des limites d'échelle de cartes planaires aléatoires.

## 1.4 Reconnaissance de la graine dans un arbre à attachement préférentiel *affine*

L'avènement de l'informatique, puis les progrès réalisés pour accroître la puissance de calcul et les capacités de stockage des machines, a considérablement simplifié, et donc motivé, l'étude empirique des réseaux réels. Prenons l'exemple de la «Base de données cinématographiques d'Internet», ou *Internet Movie Database* en anglais, qui recense la filmographie de plusieurs millions d'acteurs et le casting détaillé d'un nombre comparable de films. Cette gigantesque masse de données est en accès libre, facilement téléchargeable, et fournit une matière précieuse pour analyser le réseau social des collaborations cinématographiques. Ce qui serait bien sûr impossible sans l'aide des ordinateurs.

### Modéliser des réseaux réels

Au début de ce chapitre d'introduction, nous avons dit que représenter un réseau réel par un graphe était naturel : on symbolise les éléments ou *nœuds* du réseau par les sommets d'un graphe, et les "relations" entre ces nœuds, qu'elles soient physiques ou immatérielles, par des arêtes. Dans l'ouvrage en deux volumes *Random graphs and complex networks* [77, 78], fondant son propos sur plusieurs exemples, Remco van der Hofstad énumère quatre caractéristiques fondamentales partagées par de nombreux réseaux réels, devant aiguiller leur modélisation par des graphes :

1. Ces réseaux connectent tous un très grand nombre d'éléments – souvent plusieurs millions, parfois des milliards, voire bien au-delà pour le réseau neuronal — et qui tend à s'accroître avec le temps. Ceci suggère de considérer des graphes à  $n$  sommets avec  $n \gg 1$  (modélisation *statique*), voire des suites croissantes de graphes avec un nombre de sommets tendant vers l'infini (modélisation *dynamique*). Cela justifie également l'adoption d'un cadre probabiliste. Les réseaux sont si grands et si complexes qu'ils est impossible de les décrire dans le détail. Mieux vaut donc essayer de cerner leur structure globale à partir de quelques règles probabilistes simples, qui produiront naturellement de la variabilité dans le modèle. Dans la suite, nous noterons  $\mathbf{G}_n$  un graphe aléatoire à  $n$  sommets.
2. En moyenne, le nœud d'un réseau n'établit qu'un nombre limité de connexions avec les autres nœuds. En fait, le nombre total de liens effectivement établis dans un réseau est généralement infime devant sa valeur maximale potentielle, à savoir  $n(n-1)/2$  (cas du graphe complet). On dit que les réseaux sont *creux*, ou *sparse* en anglais. Construit à partir de la base de données des *Mathematical Reviews*, le graphe de collaboration en mathématiques — les sommets sont les mathématiciens, deux d'entre eux sont reliés par une arête s'ils ont co-écrit un article de recherche — contient par exemple près de 400.000 sommets (ou auteurs), pour "seulement" 700.000 arêtes (ou articles écrits à plusieurs mains). Voir [77, Section 1.6.4] pour plus de détails. Une manière de formaliser cette propriété de réseau *creux* est de supposer que pour tout  $k \geq 1$ , on a :

$$P_k(\mathbf{G}_n) \xrightarrow[n \rightarrow +\infty]{(d)} \mathbf{p}_k, \quad (1.4.1)$$

où  $P_k(\mathbf{G}_n)$  désigne la proportion de sommets de degré  $k$  dans  $\mathbf{G}_n$ , et où  $(\mathbf{p}_k)_{k \geq 0}$  est une mesure de probabilité sur les entiers naturels. La convergence a lieu en distribution (ou en probabilité, puisque la limite est déterministe) lorsque la taille du graphe aléatoire sous-jacent tend vers l'infini.

3. Dans un réseau, le nombre de connexions réalisées par un nœud est souvent faible, mais cela masque en réalité de fortes hétérogénéités. Une proportion importante de nœuds — appelés *pôles*, ou *hubs* en anglais — établissent au contraire beaucoup de connexions, bien que cela reste modéré devant la taille globale du réseau. Ces nœuds impactent crucialement

la structure de ce dernier. Pour formaliser ce fait, on supposera que la mesure de probabilité limite  $p_k$ , introduite dans l'équation 1.4.1, est à queue lourde. Ainsi, on a :

$$\mathbf{p}_k \underset{k \rightarrow +\infty}{\approx} k^{-\gamma}, \quad (1.4.2)$$

pour un certain exposant  $\gamma > 1$ . La notation  $\approx$  signifiant que  $\mathbf{p}_k$  est asymptotiquement proportionnel à  $k^{-\gamma}$ . Une suite de graphes aléatoires vérifiant (1.4.2) est dite *invariante d'échelle*.

4. Enfin, une caractéristique commune à de nombreux réseaux réels est d'être *petit monde* (*small-world* en anglais) : la distance entre deux nœuds est rarement élevée, mais plutôt remarquablement faible. Une conjecture restée célèbre, issue d'un travail réalisé en 1967 par Stanley Milgram, est que les êtres humains sont dans leur vaste majorité séparés par une chaîne d'au plus six amitiés. Formellement, en termes de graphes, on exprime cette propriété ainsi :

$$\forall C > 0, \quad \lim_{n \rightarrow +\infty} \mathbb{P}(\mathbf{d}_{gr}(\mathbf{V}_1, \mathbf{V}_2) \leq C \cdot \log n) = 1, \quad (1.4.3)$$

où  $\mathbf{V}_1$  et  $\mathbf{V}_2$  sont deux sommets tirés au hasard dans le graphe aléatoire  $\mathbf{G}_n$ . Dans certaines définitions, la *distance typique*  $\mathbf{d}_{gr}(\mathbf{V}_1, \mathbf{V}_2)$  est remplacée par le diamètre du graphe.

Pour constituer une bonne approximation de réseaux réels, un modèle de graphes aléatoires devra satisfaire aux quatre propriétés recensées ci-dessus. Dans la suite, nous nous focaliserons cependant sur les points 2 et 3. Il sera peu question du caractère *petit monde* des différents modèles.

## Modèles statiques

Nous commençons par présenter trois modèles aléatoires *statiques*, c'est-à-dire cherchant à reproduire les caractéristiques d'un réseau d'une certaine taille, observé à un instant  $t$ . Pour cela, on fixe un nombre de sommets  $n \gg 1$  et on construit une règle probabiliste de répartition des arêtes au sein du graphe, de sorte que celui-ci "colle" à l'observation.

**Une première approche naïve et infructueuse : le modèle d'Erdős-Renyi** En Section 1.2, nous avons présenté les graphes d'Erdős-Renyi  $\mathbf{G}_{n,p}$  comme résultat d'une percolation par arêtes de paramètre  $p$ , appliquée au graphe complet à  $n$  sommets. Une description alternative consiste à démarrer à l'inverse d'un graphe à  $n$  sommets isolés, sans aucune arête, puis à parcourir successivement les  $n(n-1)/2$  paires de sommets, en créant une arête entre deux sommets d'une paire si le résultat d'une épreuve de Bernoulli nous indique de le faire. À l'issue de ce processus, le nombre d'arêtes émanant d'un sommet en particulier suit une loi binomiale de paramètres  $n-1$  et  $p$ . En moyenne, il crée donc  $(n-1)p$  connexions, se répartissant uniformément parmi les  $n-1$  autres sommets du graphe, car le modèle d'Erdős-Renyi est fondamentalement *homogène* : les rôles joués par les  $n$  sommets sont parfaitement interchangeables, au moins en distribution.

Pour contrôler le degré d'un sommet lorsque la taille  $n$  du graphe tend vers l'infini (car on veut que les graphes soient *creux*), la probabilité  $p$  de connexion à un autre sommet doit concomitamment décroître vers zéro. On pose ainsi  $p = p(n) = \lambda/n$ , où  $\lambda$  est un paramètre s'interprétant comme le nombre moyen de liaisons réalisées par un sommet avec ses homologues. Par un résultat d'approximation bien connu, la loi binomiale de paramètres  $n-1$  et  $\lambda/n$  converge vers une loi de Poisson de paramètre  $\lambda$ . Asymptotiquement, la probabilité qu'un sommet de  $\mathbf{G}_{n,p(n)}$  soit de degré  $k$  est donc

$$\mathbf{p}_k = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Les degrés des sommets dans un graphe d'Erdős-Renyi ne sont bien sûr pas des variables indépendantes. On ne peut donc pas utiliser la loi des grands nombres pour conclure directement que la proportion de sommets de degré  $k$  dans le graphe converge vers  $\mathbf{p}_k$  :

$$P_k(\mathbf{G}_{n,p(n)}) \xrightarrow{n \rightarrow +\infty} \mathbf{p}_k.$$

La convergence ci-dessus est toutefois vraie, en probabilité. Voir par exemple [82, Théorème 4.1]. La suite de graphes  $\mathbf{G}_{n,p(n)}$  vérifie donc (1.4.1) et est *creuse*. Mais elle n'est pas *invariante d'échelle*,

car une loi de Poisson n'est pas à queue lourde. Les graphes d'Erdős-Renyi ne sont donc malheureusement pas adaptés à la modélisation de réseaux réels.

**Une version inhomogène du modèle d'Erdős-Renyi** Le principal défaut du modèle d'Erdős-Renyi est son homogénéité. Dans la réalité, tous les nœuds d'un réseau n'ont pas la même capacité d'attraction. C'est pour cela que certains d'entre eux créent beaucoup de connexions, d'autres peu, et qu'apparaît finalement le phénomène d'invariance d'échelle. Prenons l'exemple des collaborations dans la recherche mathématique. On peut raisonnablement penser qu'un mathématicien très prolifique, reconnu dans son domaine depuis de nombreuses années, qui a dépensé temps et énergie à construire et entretenir un vaste réseau de co-auteurs, sera nettement plus susceptible qu'un jeune docteur débutant, de l'élargir encore, grâce à sa réputation acquise et l'important vivier des relations secondaires (les co-auteurs de ses co-auteurs).

Une façon de prendre en compte ce constat d'hétérogénéité est de faire varier, parmi les sommets du graphe  $\mathbf{G}_n$  devant servir de modèle au réseau, le paramètre  $\lambda$  vu au paragraphe précédent, qui correspondait au nombre moyen de relations établie par un sommet en particulier. Introduisons ainsi une suite  $(\lambda_v)_{v \in \mathbf{G}_n}$  et faisons l'hypothèse qu'une arête relie deux sommets  $v$  et  $w$  avec une probabilité

$$\frac{\lambda_v \lambda_w}{\lambda_v \lambda_w + \sum_{v' \in \mathbf{G}_n} \lambda_{v'}}, \quad (1.4.4)$$

et ce indépendamment de la présence ou non des autres arêtes. Les paramètres  $\lambda$  peuvent être, au choix, des variables aléatoires i.i.d. ou bien des quantités déterministes. Le modèle avec paramètres aléatoires fut introduit par Britton, Deijfen et Martin-Löf dans [35]. C'est van der Hofstad qui propose, quant à lui, un équivalent déterministe dans [77]. Dans les deux cas, le modèle peut s'interpréter comme un modèle d'Erdős-Renyi où la probabilité d'attachement (des arêtes) est rendue inhomogène. Une version plus générale du modèle est considérée dans [31].

Quand certaines conditions sont vérifiées par les paramètres  $\lambda_v$ , la suite de graphes aléatoires  $\mathbf{G}_n$  associée devient *creuse* et *invariante d'échelle*, contrairement aux graphes d'Erdős-Renyi. En fait, il s'agit essentiellement de s'assurer que les  $\lambda_v$  sont asymptotiquement distribués comme une loi à queue lourde, ce qui semble logique puisque nous avons dit que  $\lambda_v$  devait s'interpréter comme le degré moyen d'un sommet  $v$ . Plus exactement, si  $\mathbf{L}_n$  est une variable aléatoire uniforme sur l'ensemble  $\{\lambda_v, v \in \mathbf{G}_n\}$  (cas de paramètres *déterministe*), ou si elle a la même loi que ces paramètres (lorsqu'ils sont *aléatoires*), on doit supposer qu'elle converge en distribution vers  $\mathbf{L}$ , une variable aléatoire à queue lourde. La convergence doit également avoir lieu pour les premier et second moments. Sous ces trois hypothèses, la convergence (1.4.1) et l'asymptotique (1.4.2) sont vraies, la distribution limite  $\mathbf{p}_k$  étant par ailleurs identifiée :

$$\mathbf{p}_k = \mathbb{E} \left[ e^{-\mathbf{L}} \frac{\mathbf{L}^k}{k!} \right],$$

soit une loi de Poisson mélange de paramètre  $\mathbf{L}$ . Le lecteur trouvera une preuve de ces résultats dans [35] et [77, Chapitre 6].

**Le modèle de configuration** Avec le modèle inhomogène décrit dans le paragraphe précédent, nous disposons d'une règle probabiliste d'attachement des arêtes dans un graphe à  $n$  sommets, générant des suites de graphes creux et invariants d'échelle. C'est un résultat satisfaisant, car le modèle est en outre facilement simulable. Les paramètres  $\lambda_v$  peuvent être réglés en fonction de données issues de l'observation de réseaux réels. Le modèle présente toutefois le défaut de n'offrir aucun contrôle au modélisateur, à  $n$  fixé, sur le degré des sommets, qui sont des quantités aléatoires. Il risque notamment de générer des sommets isolés en trop grand nombre, comparé à ce que l'on constate dans de vrais réseaux.

Un modèle alternatif, palliant ce problème, consisterait à générer des graphes aléatoires copiant exactement la distribution des degrés d'un réseau réel. On pourrait procéder de la manière suivante. Admettons que l'on puisse observer un réseau réel à  $n$  nœuds : nous apprenons que les degrés respectifs de ces derniers sont  $d_1, \dots, d_n$ . On tire alors au hasard un graphe à  $n$  sommets,

par exemple de manière uniforme, parmi tous les graphes à  $n$  sommets dont les degrés sont précisément  $d_1, \dots, d_n$  (on suppose que les graphes sont étiquetés). Énoncé ainsi, cela paraît évident et facile, mais on ne sait malheureusement pas décrire dans le détail un tel ensemble, ni par conséquent générer uniformément un graphe sous de telles contraintes. Le *modèle de configuration*, décrit ci-dessous, introduit à l'origine par Bollobás pour énumérer une classe de graphes réguliers [30], remédie à ce problème.

Considérons  $n$  sommets numérotés de 1 à  $n$  et  $d_1, \dots, d_n$  une suite d'entiers positifs que l'on souhaite voir coïncider avec les degrés des  $n$  sommets dans le graphe final. Représentons le  $i$ -ème sommet avec  $d_i$  "demi-arêtes" émanant de lui et numérotons toutes les "demi-arêtes" de 1 à  $D_n := d_1 + \dots + d_n$ . Maintenant, marions-les de façon à constituer de "vraies" arête. Pour la "demi-arête" n° 1 (qui émane du sommet n° 1), on tire au hasard une des  $D_n - 1$  autres "demi-arêtes", puis on relie ces deux "demi-arêtes". Si la "demi-arête" tirée au hasard émane du sommet  $i$ , alors les sommets 1 et  $i$  seront reliés par une arête dans le graphe final. On répète ensuite la procédure jusqu'à épuisement des "demi-arêtes". L'objet obtenu à l'issue du processus n'est pas un graphe à proprement parler, car plusieurs arêtes peuvent relier la même paire de sommets, voire le même sommet (ce sont alors des *boucles*). Il s'agit en fait d'un *multigraphe*. Le degré du  $i$ -ème sommet de ce multigraphe est, comme attendu, égal à  $d_i$ .

Quel peut être l'intérêt d'un modèle générant des multigraphes, si notre objectif est d'approcher les réseaux réels? En fait, quand certaines conditions sont vérifiées par la suite des degrés, la probabilité que le modèle de configuration génère un multigraphe sans arêtes multiples ni boucles — donc un "vrai" graphe — tend vers une constante strictement positive lorsque  $n \rightarrow +\infty$ . De plus, lorsqu'on conditionne sur un tel événement, le graphe obtenu est uniformément distribué sur l'ensemble des graphes à  $n$  sommets de degrés respectifs  $d_1, \dots, d_n$ . Le modèle de configuration fournit donc un algorithme de simulation, élémentaire, de cette loi uniforme a priori difficile d'accès.

Les contraintes sur la suite de degrés, nécessaires pour que la méthode exposée ci-dessus fonctionne, sont essentiellement les mêmes que celles imposées aux paramètres  $\lambda_v$  du modèle inhomogène. Si  $\mathbf{D}_n$  est une variable aléatoire uniforme sur l'ensemble  $\{d_1, \dots, d_n\}$ , on doit faire l'hypothèse qu'elle converge en distribution vers  $\mathbf{D}$ , une variable aléatoire de variance finie. La convergence doit également avoir lieu pour les premier et second moments. Sous ces trois hypothèses, la convergence (1.4.1) et l'asymptotique (1.4.2) sont vraies, et la distribution limite  $\mathbf{p}_k$  est celle de  $\mathbf{D}$  :

$$\mathbf{p}_k = \mathbb{P}(\mathbf{D} = k).$$

Pour obtenir une suite de graphes aléatoires creux et invariants d'échelle, il suffit donc de choisir  $\mathbf{D}$  avec une distribution à queue lourde. Facile dès lors de "coller" aux données réelles! Le lecteur trouvera plus de résultats sur le modèle de configuration dans [77, Chapitre 7].

Dans cette section, nous n'avons pas parlé de la propriété *petit monde*. Les trois modèles présentés ci-dessus la vérifient tous. À ce sujet, voir par exemple [31] pour le modèle inhomogène et [138, 139] pour le modèle de configuration. Voir également le livre [78].

## Un mécanisme dynamique et explicatif : le paradigme de l'attachement préférentiel

Dans la section précédente, nous avons présenté trois modèles de graphes aléatoires, que nous avons qualifiés de *statiques*, car de taille supposée constante (bien qu'elle puisse être aussi grande que souhaité). Nous avons dit que la motivation derrière l'introduction de ces modèles était de reproduire les caractéristiques d'un réseau réel, observé à un instant donné. Tout particulièrement la distribution des degrés parmi les nœuds. Il ne s'agissait pas d'*expliquer*, par exemple, l'émergence de lois invariantes d'échelle.

Des efforts ont cependant été menés pour appréhender ce phénomène, analysant un aspect totalement négligé par les modèles statiques, qui est la *croissance* des réseaux. Savoir décrire *comment* les réseaux grandissent pourrait s'avérer clé. On connaît effectivement plusieurs exemples de grands réseaux, sur lesquels de nombreuses données sont disponibles, et dont l'invariance d'échelle, vérifiée empiriquement, se stabilise remarquablement au cours du temps. Réseau de collaborations cinématographiques, mathématiques, réseau de citations entre articles scientifiques... Le lecteur trouvera de multiples illustrations étayant ce constat dans les chapitres 1 et 8 de [77].



**L’attachement préférentiel, ou «les riches s’enrichissent»** Un principe organisateur caché pourrait donc implicitement diriger le mode de croissance des réseaux. Ce sont les physiciens Barabási et Albert qui, avec la parution de leur article pionnier [21], ont imaginé les premiers un tel mécanisme explicatif, appelé *attachement préférentiel*, se décrivant très simplement : lorsqu’à un instant donné, un nouveau nœud rejoint un réseau existant, il est plus susceptible de s’attacher aux nœuds déjà fortement connectés. On le résume parfois par le célèbre adage «*les riches s’enrichissent*». Intuitivement, un tel mécanisme paraît effectivement raisonnable pour décrire le mode de croissance de certains réseaux réels. Pensons par exemple au réseau de citations entre articles scientifiques. Un article déjà très cité, car occupant une place incontournable dans son domaine (un peu comme [21] sur le sujet de l’attachement préférentiel), aura nettement plus de chances qu’un article peu connu, d’être de nouveau cité à l’avenir. Voir [142] à ce sujet. Dans la section précédente, nous évoquions l’exemple des collaborations en recherche mathématique : un mathématicien confirmé, entouré d’un vaste réseau, patiemment construit, de co-auteurs, l’emporte sur le jeune docteur pour attirer de nouveaux collaborateurs. Ces observations ne se limitent aux seuls réseaux sociaux. Sur le Web, les pages nouvellement créées pointent davantage vers les sites les plus populaires (Google, Twitter, etc). On trouve même trace d’un tel mécanisme dans des réseaux biologiques de protéines [111] ! On notera que cette logique d’auto-renforcement favorise naturellement les éléments les plus âgés d’un réseau, puisque ceux-ci ont bénéficié de plus de temps pour forger des relations.

**Définition du modèle** Dans l’article pionnier [21], la définition du modèle de l’attachement préférentiel de Barabási et Albert est informelle, peu précise. Plus tard, les mathématiciens Bollobás, Riordan, Spencer et Tusnády ont suggéré dans [33] une formulation plus rigoureuse de celui-ci. Dans sa version la plus élémentaire, nous pouvons la décrire de la façon suivante. Partant d’un graphe constitué de deux sommets et reliés par une unique arête, on fait croître ce dernier en ajoutant un par un de nouveaux sommets. Chaque nouveau sommet est connecté par une arête à l’un des sommets déjà présents dans le graphe courant, qui est choisi de façon aléatoire. La probabilité de choisir un sommet donné est proportionnelle à une fonction croissante de son degré. Cette fonction peut être, par exemple, *linéaire* — c’est le modèle originel de Barabási et Albert — ou *affine*, comme elle le sera dans le Chapitre 2. Formellement, sous ce modèle, nous construisons une chaîne de Markov sur les arbres finis  $(\mathbf{T}_n)_{n \geq 2}$ , croissante pour la relation d’inclusion, telle que pour tout  $n \geq 2$  :

- (i) le graphe  $\mathbf{T}_n$  contient  $n$  sommets ;
- (ii) le graphe  $\mathbf{T}_{n+1}$  est obtenu en tirant au hasard un sommet  $\mathbf{u}_n \in \mathbf{T}_n$ , puis en le reliant par une arête à un nouveau sommet, créé pour l’occasion. La distribution de  $\mathbf{u}_n$  est définie de la façon suivante. Pour tout arbre  $T$  à  $n$  sommets et  $u$  sommet de  $T$ , on a :

$$\mathbb{P}(\mathbf{u}_n = u \mid \mathbf{T}_n = T) = \frac{\mathbf{f}(\deg_T(u))}{\sum_{u \in T} \mathbf{f}(\deg_T(u))}. \quad (1.4.5)$$

La fonction  $\mathbf{f} : \mathbb{N}^* \rightarrow \mathbb{R}_+$  est la *fonction d’attachement* du modèle, supposée croissante.

Le modèle d’attachement préférentiel linéaire correspond à  $\mathbf{f}(d) = d$  pour tout  $d \geq 1$ . Le modèle affine à  $\mathbf{f}(d) = d - 1 + \alpha$ , avec  $\alpha > 0$  paramètre fixé du modèle. Ce dernier fut introduit par Móri dans [115]. Une illustration du mécanisme de construction est présentée en Figure 1.9.

**Une spécificité des modèles linéaire et affine** Dans le cas d’un attachement linéaire ou affine, le dénominateur de l’équation (1.4.5) ne dépend que de  $n$ , soit la taille du graphe courant, une donnée de nature *globale*, et non de  $T$ , c’est-à-dire le détail de la structure *locale* du graphe courant. En effet, pour tout arbre  $T$  à  $n$  sommets, on a :

$$\sum_{u \in T} (\deg_T(u) - 1 + \alpha) = (1 + \alpha)n - 2,$$

car la somme  $\sum_{u \in T} \deg_T(u)$  compte deux fois le nombre total d’arêtes de  $T$  (c’est vrai pour n’importe quel graphe) et ce nombre vaut  $n - 1$ , car le graphe courant est un arbre à  $n$  sommets. Ce simple constat, spécifique aux modèles linéaire et affine, simplifie beaucoup leur étude. Nous verrons pourquoi dans la section suivante.

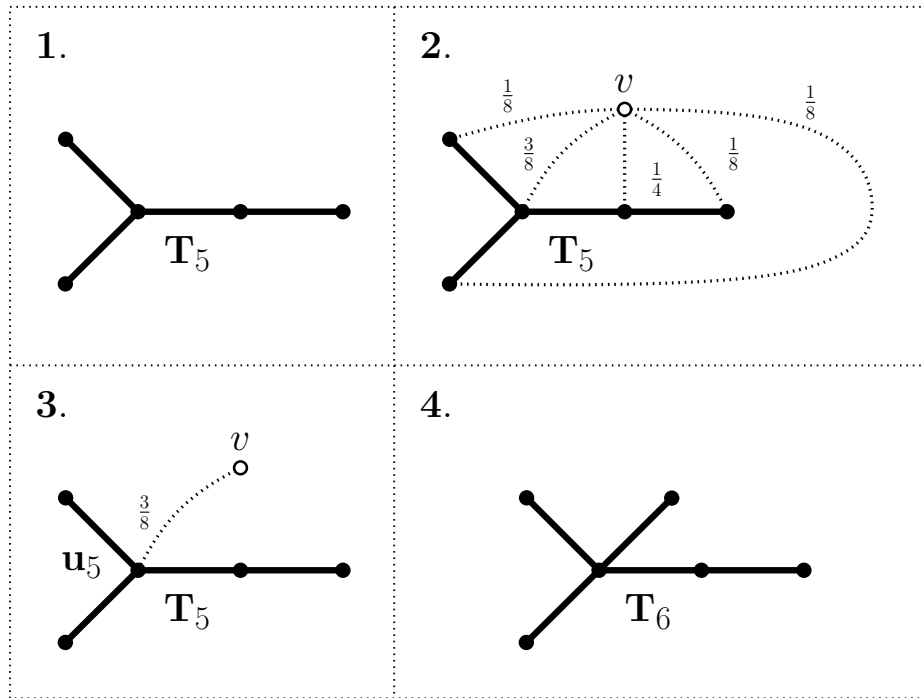


FIGURE 1.9 – Illustration de la croissance d'un arbre à attachement préférentiel linéaire à cinq sommets. Un sixième sommet, noté  $v$ , se greffe à l'arbre courant  $T_5$ . Les probabilités d'attache aux cinq sommets existant sont indiquées sur la deuxième figure, en haut à droite. Le sommet  $v$  se connecte finalement au sommet, noté  $u_5$ , de plus haut degré (voir troisième figure, en bas à gauche). Le nouvel arbre à six sommets est représenté sur la quatrième figure, en bas à droite.

### Que sait-on sur les arbres à attachement préférentiel ?

La parution de l'article [21] écrit par Barabási et Albert eut un retentissement important. Son contenu suscita le vif intérêt des physiciens des réseaux et des mathématiciens spécialistes de graphes aléatoires. Ces derniers cherchèrent à démontrer en toute rigueur les conjectures établies par les physiciens, alors appuyées par simulations numériques et raisonnements heuristiques. Compte tenu de la motivation initiale conduisant à introduire le modèle d'attachement préférentiel, c'est évidemment la question de la distribution des degrés au sein de ces graphes qui inspira les premiers travaux. Puis, la connaissance du modèle se diffusant, il devint rapidement légitime pour les mathématiciens de l'étudier comme n'importe quel autre modèle de graphes aléatoires, selon leurs propres grilles d'analyse. Examiner par exemple finement sa géométrie (limite locale, limite d'échelle, percolation, etc). Nous recensons ci-dessous quelques résultats notables de la littérature mathématique sur le modèle d'attachement préférentiel.

**Degrés des sommets** Dans [21], Barabási et Albert ont conjecturé que les arbres à attachement préférentiel *linéaires* constituaient une suite de graphes aléatoires *creux* et *invariants d'échelle*. Par un argument non rigoureux, ils ont par ailleurs estimé que l'exposant  $\gamma$  de l'asymptotique (1.4.2) devait être exactement égal à 3. Dans [33], les mathématiciens Bollobás, Riordan, Spencer et Tusnády démontrent la conjecture et identifient exactement la distribution limite  $\mathbf{p}_k$ . Ils obtiennent que pour tout  $k \geq 1$  :

$$\mathbf{p}_k = \frac{4}{k(k+1)(k+2)}.$$

Ce résultat sera ensuite généralisé au cas d'un attachement affine par Móri dans [115]. Il obtient lui aussi l'expression exacte de la distribution  $\mathbf{p}_k$ , qui dépend de la valeur du paramètre  $\alpha$  associé au modèle (voir (1.4.5)). C'est également le cas de l'exposant  $\gamma$  qui vaut  $2 + \alpha$ . Un résultat particulièrement intéressant puisqu'il montre que le modèle affine est suffisamment flexible pour "coller"

à des données réelles. On observe en effet que l'exposant de la loi limite invariante d'échelle varie selon le réseau considéré.

Dans [116], Móri analyse cette fois la croissance du degré maximal dans l'arbre  $\mathbf{T}_n$ , de nouveau dans le cas affine. Il montre que celui croît presque sûrement à vitesse polynomiale avec un exposant strictement compris entre 0 et 1 et dépendant de  $\alpha$ . Plus précisément, il obtient que :

$$n^{-\frac{1}{1+\alpha}} \cdot \mathbf{M}_n \xrightarrow[n \rightarrow +\infty]{p.s.} \xi, \quad (1.4.6)$$

où  $\mathbf{M}_n := \max_{v \in \mathbf{T}_n} \deg_{\mathbf{T}_n}(v)$  et  $\xi$  est une variable aléatoire strictement positive. La convergence 1.4.6 vaut aussi pour le degré de n'importe quel sommet fixé, seule la variable aléatoire limite changeant. La distribution (marginale puis jointe) de ces variables à densité est explicitement calculée dans [119, 120]. Le lecteur retrouvera une preuve partielle de la convergence 1.4.6 pour un sommet fixé (la stricte positivité de  $\xi$  n'est pas prouvée) dans la section 4.3 du Chapitre 2 de ce manuscrit. Elle repose essentiellement sur des arguments de martingale, dont le point de départ est la formule (1.4.5) donnant les probabilités de transition. Voyons dès maintenant, en quelques lignes, pourquoi l'ordre de grandeur en  $n^{\frac{1}{1+\alpha}}$  semble en effet être le bon.

Pour tout  $n \geq 2$ , on note  $\mathbf{d}_n$  le degré d'un sommet **fixé** de  $\mathbf{T}_n$ . D'après (1.4.5), on a la relation de récurrence conditionnelle suivante :

$$\begin{aligned} \mathbb{E}[\mathbf{d}_{n+1} \mid \mathbf{T}_n] &= \underbrace{\frac{\mathbf{d}_n - 1 + \alpha}{(1 + \alpha)n - 2} \cdot (\mathbf{d}_n + 1)}_{\text{le sommet établit une nouvelle connexion au temps } n + 1} + \underbrace{\left(1 - \frac{\mathbf{d}_n - 1 + \alpha}{(1 + \alpha)n - 2}\right) \cdot \mathbf{d}_n}_{\text{le sommet n'en établit pas}} \\ &= \left(1 + \frac{1}{(1 + \alpha)n - 2}\right) \cdot \mathbf{d}_n + \frac{1}{(1 + \alpha)n - 2}. \end{aligned}$$

Un rapide calcul de suites récurrentes permet ensuite de voir que

$$\mathbb{E}[\mathbf{d}_n] \underset{n \rightarrow +\infty}{\approx} \prod_{k \leq n} \left(1 + \frac{1}{(1 + \alpha)k - 2}\right).$$

Enfin, par des techniques de calcul asymptotique classiques, on sait que le produit ci-dessus croît comme  $n^{\frac{1}{1+\alpha}}$ . D'où l'ordre de grandeur attendu.

Dans le travail en préparation [29], Bhamidi généralise tous les résultats énoncés ci-dessus à une large classe de modèles d'attachement préférentiel, grâce à des techniques universelles issues de la théorie des processus de branchement, dont Rudas, Tóth et Valkó avaient déjà auparavant perçu la pertinence dans ce contexte [124]. Bhamidi traite par exemple le cas du modèle *sous-linéaire*, introduit par les physiciens Krapivsky, Redner et Leyvraz dans [94], correspondant à une fonction d'attachement de la forme  $\mathbf{f}(d) = d^\beta$  avec  $\beta \in (0, 1)$ . Il montre que la distribution limite  $\mathbf{p}_k$  vérifie l'asymptotique suivante :

$$\mathbf{p}_k \underset{k \rightarrow +\infty}{\approx} k^{-\beta} e^{-\frac{C}{1-\beta} k^{1-\beta}},$$

pour une certaine constante  $C = C(\beta) > 0$ , solution d'une équation explicite. La queue de distribution de  $\mathbf{p}_k$  décroît donc nettement plus vite que dans (1.4.2). Il montre aussi que le degré maximal  $\mathbf{M}_n$  croît cette fois comme  $(\log n)^{\frac{1}{1-\alpha}}$ , soit beaucoup moins vite que dans le modèle affine.

**La géométrie des arbres à attachement préférentiel** À l'instar des autres modèles statiques présentés dans la section précédente, les arbres à attachement préférentiel satisfont la propriété *petit monde*. L'ordre de grandeur de la *distance typique* (voir (1.4.3)) dans  $\mathbf{T}_n$  et du diamètre de cet arbre est exactement  $\log n$ , au sens où ces deux quantités, divisées par  $\log n$ , convergent en probabilité vers une constante explicite, dépendant seulement du paramètre  $\alpha$  associé au modèle affine. Voir [78, Chapitre 8] pour une preuve de ces résultats.

Puisque la motivation essentielle derrière l'introduction des arbres à attachement préférentiel est de modéliser des réseaux réels, il paraît légitime d'étudier leur réponse à un processus de percolation, dont la définition — voir le dernier paragraphe sur le modèle d'Erdős-Renyi en début de Section 1.2 — provient elle-même de considérations très concrètes, tester par exemple la résilience d'un

réseau subissant une panne importante. Le rôle joué par les *pôles*, ces sommets de haut degré qui font la spécificité des graphes invariants d'échelle, peut notamment susciter la curiosité. Dans [32], Bollobás et Riordan montre que lorsqu'on soumet l'arbre  $\mathbf{T}_n$  à une percolation par sites/arêtes de paramètre  $p \in [0, 1]$ , la taille du plus grand cluster croît comme  $o(n)$  lorsque  $n \rightarrow +\infty$  : l'arbre n'est donc jamais *robuste*. Dans [29], Bhamidi précise ce résultat et prouve que la taille du cluster associé à un sommet fixé de  $\mathbf{T}_n$  croît (en distribution) à vitesse polynomiale, pour un exposant strictement compris entre 0 et 1, dépendant de  $p$  et  $\alpha$ . Dans [27], Bertoin et Uribe Bravo mettent finalement en évidence une phase de percolation *surcritique*, où l'arbre devient robuste : lorsqu'on fait dépendre de  $n$  la paramètre  $p$ , de sorte que  $p = p(n) \sim 1 - \frac{c}{\log n}$  quand  $n \rightarrow +\infty$  pour une certaine constante  $c > 0$ , alors la taille du plus grand cluster de percolation dans  $\mathbf{T}_n$  est une fraction de  $n$ , et la taille des autres clusters est d'ordre  $\frac{n}{\log n}$ .

La géométrie des arbres à attachement préférentiel peut également être analysée en adoptant l'un des deux points de vue décrits dans la Section 1.3 : *limite locale* et *limite d'échelle*. Cela fonctionne bien pour la limite locale puisque Berger, Borgs, Chayes et Saberi la calcule explicitement dans [26]. En revanche, le caractère *petit monde* des arbres à attachement préférentiel fait d'eux de mauvais candidats à la limite d'échelle. Dans [54], Curien, Duquesne, Kortchemski et Manolescu montre toutefois qu'une limite d'échelle est vérifiée par l'*arbre à boucles* associé à  $\mathbf{T}_n$ , dans le cas de l'attachement *linéaire*. Le principe de construction d'un arbre à boucles est le suivant. À chaque sommet d'un arbre, on joint un cycle d'arêtes de longueur égale au degré du sommet. Ces cycles sont ensuite collés entre eux, au niveau des nœuds, en suivant la structure d'adjacence de l'arbre de départ. Cette construction implique que la distance typique dans l'arbre à boucles de  $\mathbf{T}_n$  n'est pas cette fois d'ordre  $\log n$ , mais plutôt de l'ordre de  $\mathbf{M}_n$ , le degré maximal dans l'arbre, qui croît comme  $n^{1/2}$  (voir (1.4.6)). Les auteurs de [54] montre que cet arbre à boucles, remis à l'échelle, converge pour la distance de Gromov-Hausdorff vers un espace métrique non trivial, appelé *arbre à boucles brownien*, car explicitement en lien avec l'arbre brownien d'Aldous. Dans [128], Delphin Senizergues étend ce type de résultat de limite d'échelle aux arbres à boucles associés à de nombreux modèles dynamiques de graphes aléatoires, dont les arbres à attachement préférentiel *affine*.

### Une contribution originale à une problématique nouvelle : l'influence de la *graine* dans les arbres à attachement préférentiel *affine*

Dans la définition du modèle d'attachement préférentiel que nous avons présentée plus haut, la suite d'arbres  $\mathbf{T}_n$  était initialisée en  $\circ\rightarrow$ , l'arbre à deux sommets reliés par une arête. Rien n'interdit toutefois d'amorcer le processus à partir de n'importe quel autre arbre. Ainsi, pour tout arbre  $S$  à  $k \geq 2$  sommets, nous définissons une chaîne de Markov sur les arbres finis  $(\mathbf{T}_n^S)_{n \geq k}$ , vérifiant  $\mathbf{T}_k^S = S$ , et dont les probabilités de transition sont données par (1.4.5). La condition initiale  $S$  est appelée *graine* du processus.

Une définition alternative consisterait à conditionner le processus originel sur l'événement de probabilité non nulle  $(\mathbf{T}_k = S)$ , puis à effacer de notre suite d'arbres les  $k - 2$  étapes antérieures à la réalisation de cet événement. Vu ainsi, il paraît clair que l'essentiel des résultats énoncés dans la section précédente (distribution asymptotique des degrés, vitesse de croissance du degré maximal, etc) s'étendent aux arbres à attachement préférentiel de graine quelconque.

**Une problématique nouvelle** Dans [39], Bubeck, Mossel et Rácz pose la question suivante : les arbres à attachement préférentiel *linéaire* conservent-ils en eux, à grande échelle, une trace géométrique quelconque de la *graine*? Jusqu'à présent, cet aspect du modèle était tout à fait négligé dans la littérature mathématique. Pourtant, dès 2007, des biologistes [79, 127] ont montré expérimentalement que le choix de la graine impactait fortement la topologie de plusieurs modèles de graphes aléatoires invariants d'échelle, dont l'attachement préférentiel. Pour les auteurs de [39], il convenait dans un premier temps de préciser le sens mathématique de la problématique soulevée.

Puisqu'il s'agit de comparer la géométrie à grande échelle (ou en temps long) d'arbres à attachement préférentiel issus de graines distinctes, on cherche donc une métrique capable de fournir des informations asymptotiques sur ces processus. Bubeck, Mossel et Rácz ont comme première idée de calculer leur "limite", d'abord en tant que suites croissantes de graphes, puis pour la topologie locale (voir Section 1.3). La limite "ensembliste" n'est guère intéressante, car nous savons d'après (1.4.6) que le degré de chaque sommet tend presque sûrement vers l'infini quand  $n \rightarrow +\infty$ .

La suite croissante  $\mathbf{T}_n^S$  converge donc vers un arbre infini dénombrable, où chaque sommet a lui-même un nombre infini dénombrable de voisins. Voir [90] à ce sujet. De ce point de vue, la graine n'a donc aucune influence. Bubeck, Mossel et Rácz examinent ensuite le cas de la topologie locale, la limite des arbres à attachement préférentiel initialisés en  $\circ\text{-}\circ$  ayant été identifiée dans [26]. Malheureusement, ils montrent (voir [39, Théorème 1]) que changer de graine ne modifie en rien cette limite locale. La grande majorité des sommets de  $\mathbf{T}_n^S$  se situant "loin" de la graine, celle-ci n'est pas visible à leur voisinage, et disparaît donc asymptotiquement pour la topologie locale.

Ce résultat négatif pousse à considérer une métrique nettement plus riche en informations sur  $\mathbf{T}_n^S$ . Bubeck, Mossel et Rácz propose la distance en variation totale, afin de comparer directement les distributions de  $\mathbf{T}_n^S$  et  $\mathbf{T}_n^{S'}$  pour deux graines  $S \neq S'$ . La démarche est naturelle lorsqu'on pense à reformuler la question initiale sous la forme d'un test statistique : étant donné l'observation d'un réseau  $\mathbf{G}_n$  à  $n \gg 1$  nœuds, supposé avoir crû par attachement préférentiel, existe-t-il un moyen de trancher entre les hypothèses  $\mathbf{G}_n \sim \mathbf{T}_n^S$  et  $\mathbf{G}_n \sim \mathbf{T}_n^{S'}$  sans risque trop élevé de se tromper ? En conséquence, Bubeck, Mossel et Rácz introduisent la quantité suivante :

$$\mathbf{d}(S, S') := \lim_{n \rightarrow +\infty} \mathbf{d}_{TV}(\mathbf{T}_n^S, \mathbf{T}_n^{S'}), \quad (1.4.7)$$

et cherchent à déterminer si celle-ci est non nulle lorsque les graines  $S$  et  $S'$  sont distinctes (c'est-à-dire non isomorphes, voir Section 1.3). Si c'est vrai, alors on dira qu'il y a *influence de la graine*. Notons que la limite (1.4.7) existe toujours (et est positive ou nulle). La suite  $\mathbf{d}_{TV}(\mathbf{T}_n^S, \mathbf{T}_n^{S'})$  est en effet décroissante en  $n$ , par un argument de couplage : lorsque les arbres  $\mathbf{T}_n^S$  et  $\mathbf{T}_n^{S'}$  coïncident à un instant donné, on peut coupler leur évolution future, de façon à ce qu'ils coïncident aussi à tout instant ultérieur.

**Que sait-on de l'influence de la graine ?** Le principal résultat obtenu par Bubeck, Mossel et Rácz dans [39], attestant de l'influence de la graine dans le cas d'un attachement *linéaire*, est que  $\mathbf{d}(S, S') > 0$  lorsque  $S$  et  $S'$  sont des arbres, comportant au moins 3 sommets, qui n'ont pas la même distribution des degrés, au sens où pour un certain  $d \geq 1$ , le nombre de sommets de degré  $d$  n'est pas égal dans  $S$  et dans  $S'$ . La restriction apportée à leur énoncé n'est pas surprenante, compte tenu de la forme prise par les probabilités de transition (1.4.5) quand la fonction d'attachement est linéaire (ou affine). En effet, nous avons vu que dans cette situation, le dénominateur dépend uniquement de la taille courante de l'arbre. Le modèle devient alors échangeable, à l'image d'une urne de Pólya : la loi de la distribution des degrés dans  $\mathbf{T}_n^S$  est fonction de la seule distribution initiale, celle dans  $S$ , et non des étapes suivies pour passer de  $S$  à  $\mathbf{T}_n^S$ . Il reste à voir si un "déséquilibre" initial dans la distribution des degrés de deux graines persiste à l'infini. Bubeck, Mossel et Rácz montre que oui, en se basant notamment sur le travail réalisé par Peköz, Röllin et Ross dans [119].

Le raisonnement ne tient évidemment plus si les deux graines  $S$  et  $S'$  partagent la même distribution des degrés. Dans cette situation, la loi de la distribution des degrés dans  $\mathbf{T}_n^S$  est identique à tout instant à celle dans  $\mathbf{T}_n^{S'}$ . Pour surmonter cette difficulté et montrer que l'influence de la graine vaut bien dans tous les cas, Bubeck, Mossel et Rácz suggèrent dans [39, section 3.2] de prendre en compte la structure de graphe des graines, un aspect qui est bien sûr totalement négligé lorsqu'on raisonne à partir de la seule distribution des degrés. Ils définissent et proposent d'étudier une certaine "observable", dont l'esprit est de compter pour n'importe quel arbre fini, le nombre d'injections possibles de cet arbre dans  $\mathbf{T}_n^S$  et  $\mathbf{T}_n^{S'}$ . Ils espèrent prouver l'existence d'un certain arbre pour lequel les deux "observables" associées divergeraient en variation totale. Cette idée est directement reprise et mûrie par Curien, Duquesne, Kortchemski et Manolescu dans [54], qui parviennent à établir en toute généralité, toujours pour le modèle *linéaire*, que  $\mathbf{d}(S, S') > 0$  dès que  $S \neq S'$ . Bubeck, Mossel et Rácz s'inspireront à leur tour de ce travail pour démontrer cette fois l'influence de la graine lorsque la règle d'attachement est *uniforme* [38] : quand un nouveau sommet se greffe à l'arbre courant, il choisit uniformément le sommet auquel se connecter. Ce qui correspond à  $\mathbf{f}$  constante dans (1.4.5).

**Une contribution originale : résolution du cas de l'attachement *affine*** Dans le Chapitre 2 de ce manuscrit, nous présenterons un travail mené en collaboration avec Ioan Manolescu

(voir aussi [106]), dans lequel nous prouvons l'*influence de la graine* pour le modèle d'attachement préférentiel *affine*, dont le modèle linéaire est bien sûr un cas particulier :

**Théorème 1** *Soient  $S$  et  $S'$  deux arbres finis contenant 3 sommets au moins. Pour le modèle d'attachement préférentiel affine de paramètre  $\alpha > 0$ , on a  $\mathbf{d}(S, S') > 0$  si et seulement si  $S \neq S'$  (au sens des isomorphismes de graphes).*

La démonstration de ce résultat s'inspire des arguments utilisés dans [54]. Notre méthode s'avère cependant plus robuste, c'est pourquoi nous parvenons à généraliser le résultat à toute règle d'attachement affine. Nous exposerons les points clés de cette méthode dans la section suivante. La restriction dans le Théorème 1 aux arbres à 3 sommets au moins, présentes également dans [38, 39, 54], s'explique pour une raison simple : à isomorphisme près, il n'y a qu'un seul arbre à 2 sommets et qu'un seul arbre à 3 sommets, la transition 1.4.5 est donc déterministe au temps  $n = 2$ .

**Trouver effectivement la graine** Les résultats énoncés dans [38, 39, 54, 106] peuvent s'interpréter comme suit : étant donné un grand nombre de réalisations indépendantes de  $\mathbf{T}_n^S$ , on peut déterminer, avec une bonne précision, la graine  $S$  qui en est à l'origine. En outre, le nombre de réalisations nécessaires pour atteindre un niveau de précision fixé ne dépend pas de  $n$ . En pratique, plus intéressant serait de construire un algorithme qui localiserait la graine en temps polynomial, à partir d'une unique réalisation d'un arbre de grande taille, et ce dans une région de l'arbre dont on contrôlerait l'étendue selon le niveau du risque d'erreur admis. Plusieurs résultats en ce sens ont été démontrés dans [37, 61, 83, 100] pour les modèles d'attachement uniforme, linéaire, et même sous-linéaire, demandant toutefois de fortes hypothèses sur la topologie de la graine (singleton, graphe en étoile ou en ligne). Les algorithmes décrits dans ces articles reposent le plus souvent sur la recherche d'un sous-ensemble de sommets maximisant une métrique de *centralité*, dont nous ne détaillerons pas ici la définition.

## Heuristique de la preuve de l'influence de la graine dans les arbres à attachement préférentiel *affine* (Chapitre 2, ou [106])

Dans cette section, nous résumons les principaux arguments utilisés dans le Chapitre 2 de ce manuscrit pour démontrer le Théorème 1, qui affirme l'influence de la graine dans le modèle d'attachement préférentiel affine. L'ingrédient essentiel consiste à exploiter le phénomène d'auto-renforcement intrinsèque au modèle, qui tend à maintenir voire creuser les "déséquilibres" initiaux dans la distribution des degrés des graines. Comme argumenté plus haut, deux situations se présentent, selon que les deux graines  $S$  et  $S'$  du Théorème 1 partagent ou non une même distribution des degrés. Dans un premier paragraphe, nous illustrons, via le traitement d'un cas simple, la méthode employée par Bubeck, Mossel et Rácz dans [39] pour prouver que  $\mathbf{d}(S, S') > 0$  (voir (1.4.7) pour une définition de cette quantité) lorsque  $S$  et  $S'$  n'ont pas une distribution des degrés commune. Dans les deux paragraphes suivants, nous présentons la stratégie de preuve adoptée dans le Chapitre 2 pour traiter le cas général, en la comparant à celle de Bubeck, Mossel et Rácz.

**Quand la distribution des degrés suffit à différencier les graines** Un arbre de taille  $k \geq 2$  contient exactement  $k - 1$  arêtes, pouvant se répartir de deux manières radicalement contraires parmi les sommets du graphe. Une répartition "inégalitaire" à l'extrême correspond au graphe "en étoile" : un sommet "s'accapare" les  $k - 1$  arêtes, ce qui le relie à tous les autres sommets, qui eux-mêmes ne sont connectés qu'à ce sommet "central". Ce dernier est donc de degré  $k - 1$ , quand tous les autres sommets sont de degré 1. Une répartition "égalitaire" correspond, quant à elle, au graphe "en ligne", c'est-à-dire un chemin d'arêtes de longueur  $k - 1$ . Tous les sommets sont de degré 2, sauf les deux situés en bout de chemin, de degré 1. Attention, les deux graphes sont réellement distincts seulement à partir de  $k = 4$ . Voir l'illustration en Figure 1.10 du cas  $k = 9$ .

Supposons maintenant que ces arbres soient nos deux graines, notées respectivement  $S$  et  $S'$ , que l'on fait croître à l'infini selon une règle d'attachement affine. Celle-ci favorise par définition, à chaque instant, les connexions aux sommets de plus haut degré. Tentons de comprendre l'évolution du degré maximal dans les deux suites d'arbres  $\mathbf{T}_n^S$  et  $\mathbf{T}_n^{S'}$  associées.

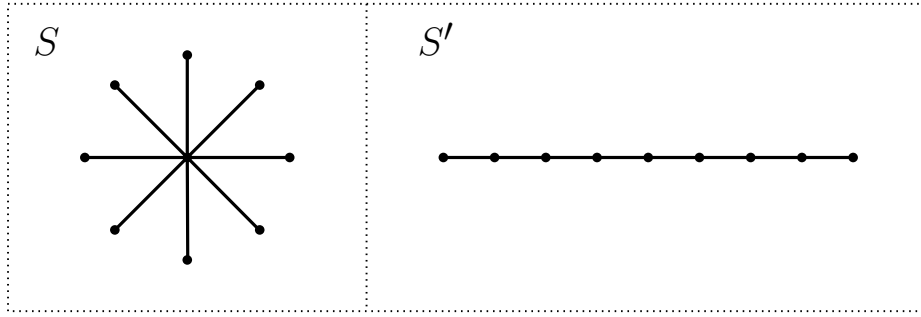


FIGURE 1.10 – Deux graines de taille 9 aux distributions des degrés radicalement opposées. À gauche, le graphe en étoile ; à droite, le graphe en ligne.

Le graphe en étoile contient un unique sommet de degré maximal, de valeur nettement plus élevée que les autres. À l'entame du processus de croissance, ce sommet captera une proportion importante des connexions avec les nouveaux arrivants. Il aura donc de fortes chances de rester le sommet de degré maximal, et la valeur de ce dernier croîtra continûment, au moins pendant les premiers instants.

En revanche, dans le graphe en ligne  $S'$ , la situation initiale est très équilibrée, les probabilités d'attachement aux différents sommets quasiment toutes égales. Les nouvelles connexions vont donc se répartir uniformément dans le graphe, ce qui risque de faire varier souvent le sommet de degré maximal, et stagner la valeur de ce dernier.

D'après (1.4.6), nous savons qu'en temps long, le degré maximal croît toujours à vitesse  $n^{\frac{1}{1+\alpha}}$ , quelle que soit la configuration de départ. Le facteur de proportionnalité  $\xi$ , aléatoire, dépend toutefois de la distribution des degrés dans la graine dont est issue la suite d'arbres croissant par attachement préférentiel. Soient  $\xi_S$  et  $\xi_{S'}$  les facteurs de proportionnalité associées aux deux graines considérées ci-dessus. D'après les arguments avancés, on s'attend à ce que

$$\mathbb{P}(\xi_S > x) > \mathbb{P}(\xi_{S'} > x),$$

au moins pour  $x > 0$  suffisamment grand. Dans [39, Corollaire 1], Bubeck, Mossel et Rácz le confirment pour le modèle linéaire  $\alpha = 1$ . Ils prouvent en effet que la queue de distribution de  $\xi_S$ , pour une graine  $S$  quelconque, vérifie l'asymptotique suivante :

$$\mathbb{P}(\xi_S > x) \underset{x \rightarrow +\infty}{\approx} x^{1-2|S|+2M(S)} \cdot e^{-x^2/4},$$

où  $|S|$  désigne la taille de la graine et  $M(S)$  la valeur du degré maximal dans celle-ci. Pour les deux graines prises en exemple, de même taille, on en déduit que

$$\frac{\mathbb{P}(\xi_{S'} > x)}{\mathbb{P}(\xi_S > x)} \underset{x \rightarrow +\infty}{\approx} x^{6-2k} \underset{x \rightarrow +\infty}{\longrightarrow} 0.$$

Ce qui donne le résultat attendu.

Comment maintenant démontrer que  $\mathbf{d}(S, S') > 0$ ? En utilisant que le degré maximal est une fonction mesurable — ou "observable" — de l'arbre courant. On a ainsi pour tout  $n \geq k$  :

$$\mathbf{d}_{VT}(\mathbf{T}_n^S, \mathbf{T}_n^{S'}) \geq \mathbf{d}_{VT}(\mathbf{M}_n(S), \mathbf{M}_n(S')),$$

où  $\mathbf{M}_n(S)$  et  $\mathbf{M}_n(S')$ , désignent respectivement le degré maximal dans  $\mathbf{T}_n^S$  et  $\mathbf{T}_n^{S'}$ . Or, pour tout  $x > 0$  :

$$\begin{aligned} \mathbf{d}_{VT}(\mathbf{M}_n(S), \mathbf{M}_n(S')) &\geq \mathbb{P}(\mathbf{M}_n(S) > x\sqrt{n}) - \mathbb{P}(\mathbf{M}_n(S') > x\sqrt{n}) \\ &\underset{n \rightarrow +\infty}{\longrightarrow} \mathbb{P}(\xi_S > x) - \mathbb{P}(\xi_{S'} > x). \end{aligned}$$

Cette dernière quantité est indépendante de  $n$  et strictement positive pour  $x$  grand. Grâce à l'inégalité précédente, on obtient donc bien que  $\mathbf{d}(S, S') = \lim_{n \rightarrow +\infty} \mathbf{d}_{VT}(\mathbf{T}_n^S, \mathbf{T}_n^{S'}) > 0$ .

La démonstration fonctionne en fait, telle quelle, pour n'importe quelle paire de graines de même taille, dont la valeur du degré maximal diffère. Bien sûr, cette situation ne se présente pas systématiquement : deux graines ne partageant pas une même distribution des degrés peuvent avoir un degré maximal égal. Selon les cas, l'observable à considérer est alors plutôt le deuxième degré le plus élevé, le troisième, le quatrième, etc. L'esprit de la preuve ne changeant pas : il s'agit toujours de comparer les queues de distribution des variables aléatoires apparaissant comme limites des observables choisies, une fois remises à l'échelle. On peut raisonnablement penser que cette stratégie de preuve fonctionnerait tout aussi bien pour  $\alpha \neq 1$ , bien que Bubeck, Mossel et Rácz ne le vérifient pas explicitement dans [39].

**Quelle observable dans le cas général ?** La méthode exposée dans le paragraphe précédent ne repose que sur la dynamique des degrés dans un arbre à attachement préférentiel. La structure du graphe sous-jacent n'est nulle part considérée. On va chercher à l'intégrer au raisonnement pour résoudre le cas de deux graines partageant une même distribution des degrés, suivant la piste suggérée par Bubeck, Mossel et Rácz dans [39, Section 3.2], puis reprise avec succès par Curien, Duquesne, Korchemski et Manolescu dans [54] pour le modèle linéaire.

Comment tenir compte de la structure de graphe pour différencier deux arbres à attachement préférentiel  $\mathbf{T}_n^S$  et  $\mathbf{T}_n^{S'}$ , issus de graines  $S$  et  $S'$  dont la distribution des degrés est identique ? On peut déjà commencer par trouver une observable permettant de différencier les deux graines. Nous savons que pour tout  $d \geq 1$ , le nombre de sommets de degré  $d$  est égal dans  $S$  et dans  $S'$ . Que se passe-t-il si on complexifie maintenant ce comptage, en le couplant avec la donnée des relations d'adjacence ? En énumérant, par exemple, dans  $S$  et  $S'$ , le nombre de paires de sommets *voisins* telles que le premier soit de degré  $d_1$  et le second de degré  $d_2$ . Sur la Figure 1.11, on voit que ce type de comptage permet effectivement de distinguer deux graines partageant pourtant une même distribution des degrés.

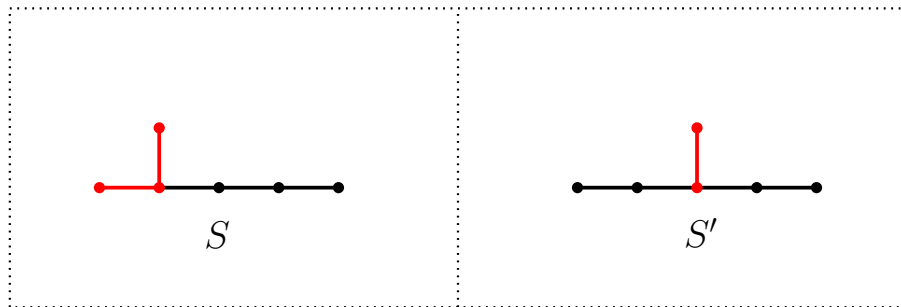


FIGURE 1.11 – Deux graines  $S$  et  $S'$  de même distribution des degrés. En revanche, il y a dans  $S$  deux paires de sommets voisins, de degrés respectifs 3 et 1, contre une seule paire dans  $S'$ . Les arêtes qui relient ces paires sont en rouge.

On généralise ce constat de la façon suivante. Pour un arbre fini  $\tau$ , une fonction à valeurs entières  $d : \tau \rightarrow \mathbb{N}$  et un autre arbre fini  $U$ , on note  $D_{\tau,d}(U)$  le nombre d'injections (de graphes)  $\varphi$  de  $\tau$  dans  $U$  vérifiant  $\deg_U \varphi(u) = d_u$  pour tout  $u \in \tau$ . On dit que l'arbre  $\tau$  est *clairvoyant* pour distinguer deux graines  $S$  et  $S'$  s'il existe  $d$  telle que  $D_{\tau,d}(S) \neq D_{\tau,d}(S')$ . Sur la Figure 1.11, l'arbre à deux sommets  $\circ-\circ$  est clairvoyant. L'existence d'un arbre clairvoyant est universelle, car si deux graines  $S$  et  $S'$  ne sont pas isomorphes, alors prendre  $\tau = S$  fonctionne toujours, avec  $d$  la distribution des degrés dans  $S$ . Il est toutefois possible de mieux faire, en choisissant un arbre de taille minimale. L'arbre  $\circ-\circ$  en Figure 1.11 est clairvoyant et minimal, car le graphe composé d'un sommet unique n'est pas clairvoyant. Il existerait sinon un  $d \geq 1$  tel que le nombre de sommets de degré  $d$  n'est pas égal dans  $S$  et  $S'$  !

Ainsi, nous cherchons un arbre clairvoyant pour distinguer deux graines  $S$  et  $S'$ . À partir de cet arbre clairvoyant  $\tau$ , supposé de taille minimale, nous construisons une observable ne prenant pas la même valeur sur  $S$  et  $S'$ , en nous inspirant de [54]. Nous introduisons à cette fin la fonction



combinatoire suivante :

$$[d]_\ell = \begin{cases} d(d-1)\dots(d-\ell+1) & \text{si } d \geq 0 \text{ et } \ell \geq 1; \\ 1 & \text{si } \ell = 0. \end{cases}$$

Puis, pour toute fonction  $\ell : \tau \rightarrow \mathbb{N}$  et tout arbre fini  $U$ , nous définissons :

$$\mathcal{F}_{\tau,\ell}(U) = \sum_{d \in \mathbb{N}^\tau} D_{\tau,d}(U) \cdot \prod_{u \in \tau} [d_u - 1]_{\ell(u)}, \quad (1.4.8)$$

où la somme est indexée sur les fonctions de  $\tau$  à valeurs entières. Il s'agit donc d'une combinaison linéaire des cardinaux  $D_{\tau,d}(\cdot)$ , où les valeurs des pondérations sont ajustées en intervenant sur la suite  $\ell$ . Si cette dernière est correctement choisie, on doit pouvoir obtenir  $\mathcal{F}_{\tau,\ell}(S) \neq \mathcal{F}_{\tau,\ell}(S')$ , car nous savons que  $D_{\tau,d}(S) \neq D_{\tau,d}(S')$  pour un certain  $d \in \mathbb{N}^\tau$ .

Une fois construite l'observable permettant de reconnaître les deux graines, la suite du raisonnement consiste à montrer qu'elle convient également pour différencier les grands arbres à attachement préférentiel qui en sont issus, autrement dit que  $\liminf_{n \rightarrow +\infty} \mathbf{d}_{VT}(\mathcal{F}_{\tau,\ell}(\mathbf{T}_n^S), \mathcal{F}_{\tau,\ell}(\mathbf{T}_n^{S'})) > 0$ .

**Preuve de l'influence générale de la graine dans le modèle affine** La définition (1.4.8) de l'observable peut être reformulée en indexant la somme sur les injections de  $\tau$  dans l'arbre  $U$  :

$$\mathcal{F}_{\tau,\ell}(U) = \sum_{\varphi: \tau \hookrightarrow U} \prod_{u \in \tau} [\deg_U \varphi(u) - 1]_{\ell(u)}. \quad (1.4.9)$$

Cette écriture est plus compatible avec l'évolution dynamique des degrés des sommets dans un arbre à attachement préférentiel. Contrairement à ce que font Bubeck, Mossel et Rácz dans [39], nous n'allons pas borner inférieurement la distance en variation totale  $\mathbf{d}_{VT}(\mathcal{F}_{\tau,\ell}(\mathbf{T}_n^S), \mathcal{F}_{\tau,\ell}(\mathbf{T}_n^{S'}))$  en étudiant la distribution asymptotique des observables remises à l'échelle. Suivant plutôt la voie tracée par Curien, Duquesne, Körtchanski et Manolescu dans [54], nous allons utiliser une inégalité générale permettant de nous restreindre à l'étude de ses premier et second moments. Pour toutes variables aléatoires réelles  $X, Y$  de carré intégrable, on sait en effet que :

$$\mathbf{d}_{VT}(X, Y) \geq \frac{(\mathbb{E}[X] - \mathbb{E}[Y])^2}{(\mathbb{E}[X] - \mathbb{E}[Y])^2 + 2 \cdot (\mathbb{E}[X^2] + \mathbb{E}[Y^2])}.$$

Ainsi, nous aurons prouvé l'influence de la graine si nous parvenons à montrer que

$$|\mathbb{E}[\mathcal{F}_{\tau,\ell}(\mathbf{T}_n^S)] - \mathbb{E}[\mathcal{F}_{\tau,\ell}(\mathbf{T}_n^{S'})]| \underset{n \rightarrow +\infty}{\approx} (\mathbb{E}[\mathcal{F}_{\tau,\ell}^2(\mathbf{T}_n^S)] + \mathbb{E}[\mathcal{F}_{\tau,\ell}^2(\mathbf{T}_n^{S'})])^{1/2}. \quad (1.4.10)$$

Dans la Section 3 du Chapitre 2, nous nous intéressons tout d'abord à l'ordre de grandeur du premier moment de  $\mathcal{F}_{\tau,\ell}(\mathbf{T}_n^S)$ , quel que soit l'arbre  $\tau$  choisi, qu'il soit clairvoyant ou non. Nous obtenons une asymptotique précise :

**Proposition 2** Soit  $\alpha > 0$  le paramètre du modèle d'attachement affine (voir (1.4.5)). Pour tout arbre fini  $\tau$  et fonction  $\ell : \tau \rightarrow \mathbb{N}$ , il existe deux entiers naturels  $w := w(\tau, \ell)$  et  $\gamma := \gamma(\tau, \ell)$ , calculables explicitement, tels que :

$$\mathbb{E}[\mathcal{F}_{\tau,\ell}(\mathbf{T}_n^S)] \underset{n \rightarrow +\infty}{\approx} n^{\max\{1, \frac{w}{1+\alpha}\}} \cdot (\log n)^\gamma.$$

On a de plus  $w \geq |\ell| := \sum_{u \in \tau} \ell(u)$ , avec égalité lorsque  $\ell(u) \geq 2$  pour tout  $u \in \tau$ . Dans ce dernier cas, on a aussi  $\gamma = 0$ .

Comment comprendre ce résultat ? D'après (1.4.9), l'observable  $\mathcal{F}_{\tau,\ell}(\mathbf{T}_n^S)$  s'écrit comme une somme sur les injections de  $\tau$  dans  $\mathbf{T}_n^S$ . Parmi ces injections, il y en a qui plongent  $\tau$  dans la graine, ou près d'elle, touchant donc des sommets déjà présents dans l'arbre aux premiers instants du processus. Or, on sait que le degré d'un sommet particulier croît en temps long comme  $n^{\frac{1}{1+\alpha}}$ .

La pondération combinatoire  $\prod_{u \in \tau} [\deg_{\mathbf{T}_n^S} \varphi(u) - 1]_{\ell(u)}$  associée à ce type d'injections croît donc comme  $n^{\frac{|\ell|}{1+\alpha}}$  pour  $|\ell| = \sum_{u \in \tau} \ell(u)$ .

Il faut également tenir compte des nouvelles injections apparaissant à mesure que l'arbre à attachement préférentiel grandit. Donnons un exemple. Soit  $\varphi$  une injection de  $\tau$  dans  $\mathbf{T}_n^S$ . Choisissons  $f$  une feuille de  $\tau$  (c'est-à-dire un sommet de degré 1) et appelons  $v$  l'unique voisin de  $f$  dans  $\tau$ . Supposons enfin que  $\ell(v) = 0$ . Dans ce cas, le degré de  $\varphi(v)$  dans  $\mathbf{T}_n^S$  ne contribue pas à la pondération combinatoire associée à l'injection  $\varphi$ , car  $[\cdot]_0 = 1$ . Toutefois, lorsqu'au cours du processus de croissance de  $\mathbf{T}_n^S$ , un nouveau sommet s'attache au sommet existant  $\varphi(v)$ , se crée aussi une nouvelle possibilité d'injection de  $\tau$  dans l'arbre : il suffit d'envoyer la feuille  $f$  voisine de  $v$  non plus sur  $\varphi(f)$  mais sur le sommet se greffant à  $\varphi(v)$ . Ces nouvelles injections apparaissent donc quand  $\varphi(v)$  réalise de nouvelles connexions. Or, leur nombre croît, on le sait, comme  $n^{\frac{1}{1+\alpha}}$ . On voit ainsi pourquoi  $w \geq 1 + |\ell| > |\ell|$  dans certaines situations.

Il existe d'autres types de contributions justifiant la présence du facteur logarithmique dans l'asymptotique de la Proposition 2. Nous ne les détaillons pas ici. Notons que les égalités  $w = |\ell|$  et  $\gamma = 0$  sont réalisées lorsque les paramètres  $\ell(u)$  sont "grands", ce qui peut signifier qu'on a accordé suffisamment de "poids" aux injections près de la graine, parmi les "vieux" sommets, pour qu'elles deviennent dominantes dans  $\mathcal{F}_{\tau, \ell}(\mathbf{T}_n^S)$ . Aussi, l'exposant  $\max\{1, \frac{w}{1+\alpha}\}$  du facteur polynomial signifie que l'observable croît toujours au moins linéairement. Ce qui est, pour résumer, une conséquence du fait que  $\mathbf{T}_n^S$  est un graphe connecté, dont le nombre d'arêtes croît lui-même linéairement, et où chaque sommet établit en temps long une infinité de connexions.

La preuve de la Proposition 2 repose sur une formule de récurrence vérifiée par  $\mathbb{E}[\mathcal{F}_{\tau, \ell}(\mathbf{T}_n^S)]$  qui, analysée finement, fournit l'asymptotique attendue. Le raisonnement pour démontrer cette formule consiste principalement à regarder si, à un instant donné  $n$ , l'attachement d'un nouveau sommet à l'arbre courant crée de nouvelles possibilités d'injections de  $\tau$  dans celui-ci. On retrouve cette formule de récurrence dans [54], bien que sa forme s'avère moins complexe car les auteurs se restreignent à l'étude du modèle linéaire. Par ailleurs, ils ne déduisent de cette formule qu'un encadrement de l'ordre de grandeur de  $\mathbb{E}[\mathcal{F}_{\tau, \ell}(\mathbf{T}_n^S)]$  et non, comme ici, le rythme exact de croissance. Enfin, les cas subtils vus ci-dessus, où par exemple  $w > |\ell|$ , n'apparaissent nulle part dans leur article.

Dans la Section 3 du Chapitre 2, nous calculons aussi une asymptotique précise du second moment de l'observable, dans un cas particulier utile à la démonstration du Théorème 1 :

**Proposition 3** *Quand  $|\ell| > 1 + \alpha$  et  $\ell(u) \geq 2$  pour tout  $u \in \tau$ , on a :*

$$\mathbb{E}[\mathcal{F}_{\tau, \ell}^2(\mathbf{T}_n^S)] \underset{n \rightarrow +\infty}{\approx} n^{\frac{2|\ell|}{1+\alpha}}.$$

La preuve de ce résultat repose sur la précédente proposition et sur des techniques de calcul identiques à celles utilisées pour la démontrer.

Dans la Section 4 du Chapitre 2, nous établissons une ultime asymptotique, dans laquelle réside une bonne partie de l'originalité de la méthode adoptée ici pour prouver l'influence de la graine, en comparaison des autres travaux sur le sujet [39, 54] :

**Proposition 4** *Soient deux graines  $S$  et  $S'$  contenant au moins 3 sommets, et  $\tau$  un arbre clairvoyant distinguant les deux graines, de taille minimale. Il existe  $\ell : \tau \rightarrow \mathbb{N}$ , vérifiant à la fois  $|\ell| > 1 + \alpha$  et  $\ell(u) \geq 2$  pour tout  $u \in \tau$ , telle que :*

$$|\mathbb{E}[\mathcal{F}_{\tau, \ell}(\mathbf{T}_n^S)] - \mathbb{E}[\mathcal{F}_{\tau, \ell}(\mathbf{T}_n^{S'})]| \underset{n \rightarrow +\infty}{\approx} n^{\frac{|\ell|}{1+\alpha}}.$$

Ce résultat dit en substance que pour  $\tau$  et  $\ell$  correctement choisis, une différence existe bien entre les distributions de  $\mathcal{F}_{\tau, \ell}(\mathbf{T}_n^S)$  et  $\mathcal{F}_{\tau, \ell}(\mathbf{T}_n^{S'})$ , et que cette différence résulte uniquement du déséquilibre initial relatif aux nombres d'injections de  $\tau$  dans les graines. La démonstration de cette proposition utilise une technique de couplage, présentée dans la Section 2 du Chapitre 2, qui isole la graine du mécanisme de croissance de l'arbre à attachement préférentiel.

En combinant la Proposition 4 avec la Proposition 3, on déduit immédiatement (1.4.10) et donc l'influence de la graine. Notre preuve est plus directe que celle de [54], grâce au critère simple trouvé sur le paramètre  $\ell$ , qui permet d'annuler le facteur logarithmique dans l'asymptotique de  $\mathbb{E}[\mathcal{F}_{\tau,\ell}(\mathbf{T}_n^S)]$  (voir Proposition 2). Dans [54], les auteurs éliminent la perturbation générée par ce facteur en introduisant une combinaison linéaire d'observables  $\mathcal{F}_{\tau,\ell}(\cdot)$ , indexée, pour résumer, sur l'arbre clairvoyant de taille minimale et sur des sous-arbres de ce dernier. La combinaison linéaire est conçue de façon à obtenir une martingale de carré intégrale, dont l'espérance (constante) est fonction de la seule graine. Bien qu'ingénieuse (et technique), leur méthode ne s'étend pas facilement au cas plus général du modèle affine, à cause de la transition de phase que connaît l'exposant  $\max\{1, \frac{w}{1+\alpha}\}$  du facteur polynomial dans l'asymptotique de la Proposition 2, un comportement imperceptible quand on se restreint à  $\alpha = 1$ .

## 1.5 Percolation orientée sur triangulations causales surcritiques

Comme expliqué en fin de Section 1.1, l'étude de la géométrie des cartes planaires est aujourd'hui profondément motivée par les réflexions menées en physique théorique à la fin du siècle dernier, à propos de la construction d'une mesure non triviale sur des ensembles de surfaces, sorte d'équivalent bidimensionnelle à la mesure de Wiener sur les chemins. Afin d'imiter la construction historique de cette dernière, les physiciens recoururent aux cartes planaires pour discrétiser les surfaces continues, ce qui mit inévitablement sur la table la question de l'existence ou non d'une forme de limite d'échelle aux grandes cartes planaires uniformes.

En fin de Section 1.3, nous avons cité plusieurs énoncés mathématiques récents [98, 108, 113], confirmant tous l'existence d'une limite aux grandes cartes planaires, commune à de nombreux modèles, la *carte brownienne*. Les physiciens — qui avaient déjà conjecturé, admis, voire calculé des quantités les renseignant sur cette limite d'échelle [14] — ont toutefois rapidement observé son inadéquation avec la théorie de la gravité quantique. Le constat restant valable dans les dimensions supérieures. Voir à ce sujet la thèse de Timothy Budd [40, Introduction et Chapitre 4]. Une raison identifiée était le traitement parfaitement symétrique des deux dimensions du modèle, alors que temps et espace jouent pourtant des rôles distincts en physique. Modifier le modèle pour y intégrer une forme de causalité parut nécessaire.

### Les triangulations causales

Jan Ambjørn et Renate Loll définissent dans [13] un nouvel ensemble de triangulations infinies du plan, appelées *triangulations causales*. Celles-ci ont la particularité d'être dotées d'une structure en couches successives, centrées autour d'un sommet racine. Chaque couche est un cycle de sommets, et le  $n$ -ème cycle contient exactement les sommets situés à distance  $n$  de la racine. Voir Figure 1.12 pour une illustration. Le modèle doit être interprété comme un processus décrivant l'évolution temporelle d'un univers de dimension 1 ayant la topologie de  $\mathbb{S}^1$  (les cycles). On parle ainsi plus volontiers d'un modèle de dimension  $1+1$ , afin de distinguer les composantes temporelle et spatiale. La représentation en Figure 1.12 d'une triangulation causale sous forme conique fait apparaître plus clairement la première. Pour des raisons physiques évidentes, il serait intéressant d'étendre la définition en dimensions supérieures. En dimension 3 (ou  $1+2$ ), les couches consisteraient par exemple en des triangulations finies de  $\mathbb{S}^2$ , avec des tétraèdres collées entre elles, de sorte que la carte entière recouvrirait tout l'espace  $\mathbb{R}^3$ . Malheureusement, l'analyse mathématique se complexifie rapidement dès que l'on quitte le cadre planaire. De nombreuses simulations numériques ont toutefois été réalisées en dimensions  $1+2$  et  $1+3$  [11, 12, 40]. Les résultats laissent penser que les triangulations causales constituent effectivement un modèle de gravité quantique pertinent.

En dimension  $1+1$ , la structure en couches de la carte masque une géométrie fondamentalement arborescente. Les triangulations causales sont en effet en bijection avec les arbres plans. À partir d'un arbre plan infini, on peut construire une triangulation causale de la façon suivante. Pour tout  $r \geq 1$ , on ajoute des arêtes "horizontales" entre les sommets situés à distance  $r$  de la racine, de façon à constituer un cycle. Cela inclut une arête entre le sommet le plus à droite (de la  $r$ -ème

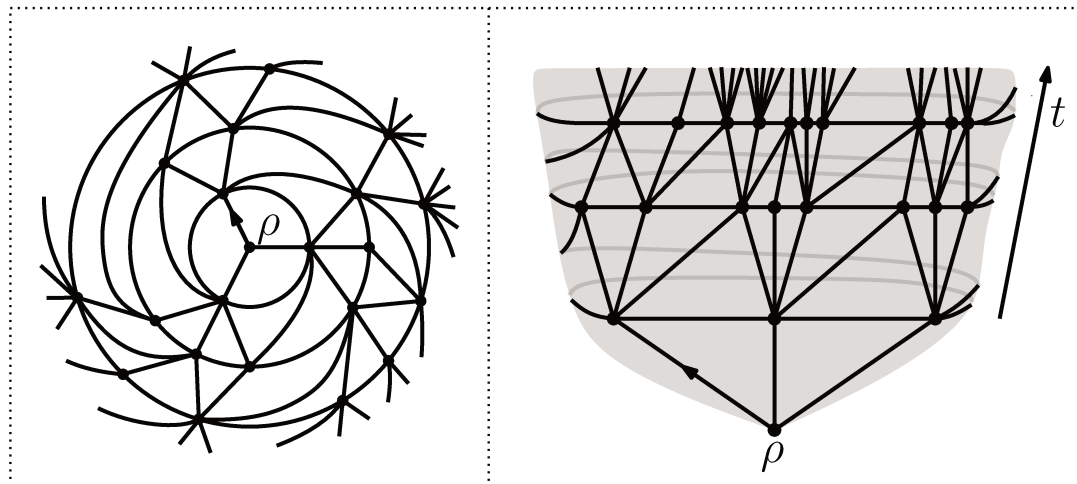


FIGURE 1.12 – Deux représentations d’une même triangulation causale. En version planaire à gauche, sous forme conique à droite. Dans cette dernière, la structure causale (ou temporelle) du modèle est orientée du bas vers le haut.

génération de sommets) et le sommet le plus à gauche. On obtient à l’issue de cette première étape une carte planaire, appelée *carte causale*, qui n’est pas encore une triangulation. On se fixe ensuite une règle (qui n’est pas la seule possible) de "triangulation" des faces : pour chaque face de la carte causale, nous relierons le sommet situé en haut à droite de la face aux autres sommets bordant celle-ci, excepté ceux qui lui sont déjà adjacents. À l’issue de cette seconde étape, nous obtenons bien une triangulation infinie du plan, dotée comme attendu d’une structure en cycles concentriques. Une illustration des différentes étapes de la construction décrite ci-dessus est présentée en Figure 1.13. Le lecteur pourra aisément vérifier que la procédure est parfaitement réversible, de sorte que triangulations causales et arbres plans sont effectivement en bijection : à toute triangulation causale, on associe un unique arbre plan en supprimant les arêtes des cycles de la carte, mais aussi, pour tout sommet, les arêtes qui relient ce dernier aux sommets plus proches de la racine, sauf une, la plus à droite.

### Que sait-on des triangulations causales aléatoires ?

Compte tenu de la relation bijective entre triangulations causales et arbres plans, construire un modèle de triangulations causales aléatoires ou d’arbres aléatoires revient strictement au même. Ce qui ne veut bien sûr pas dire que triangulations causales et arbres partagent les mêmes propriétés géométriques. On peut en effet raisonnablement penser qu’à cet égard, l’ajout des cycles et des arêtes triangulant les faces n’est pas neutre.

**Triangulations causales critiques** L’étude purement mathématique des triangulations causales est encore récente, la littérature sur le sujet embryonnaire. Bergfinnur Durhuus, Thordur Jonsson et John Weather [64], puis Nicolas Curien, Tom Hutchcroft et Asaf Nachmias à leur suite [56], ont tenté d’esquisser dans un premier temps la géométrie des triangulations causales dites *critiques*, celles dont l’arbre sous-jacent est un arbre de Kesten. Rappelons (voir Section 1.3) que les arbres de Kesten sont les arbres infinis obtenus comme limite locale de grands arbres de Bienaymé-Galton-Watson critiques. Certaines propriétés des triangulations causales critiques peuvent directement être déduites de ce que l’on sait des arbres de Kesten. On remarque par exemple que les arêtes ajoutées pour passer d’un arbre à une triangulation causale ne créent aucun raccourci pour atteindre un sommet (quelconque) depuis le sommet racine. Les boules de rayon  $r$  coïncident donc dans l’arbre et la carte. Lorsque la triangulation causale est issue d’un arbre de Kesten, lui-même limite d’arbres de Bienaymé-Galton-Watson de loi de reproduction de variance finie, le nombre de sommets à distance exactement  $r$  de la racine croît ainsi linéairement avec  $r \rightarrow +\infty$ , et le volume des boules à vitesse quadratique [88].

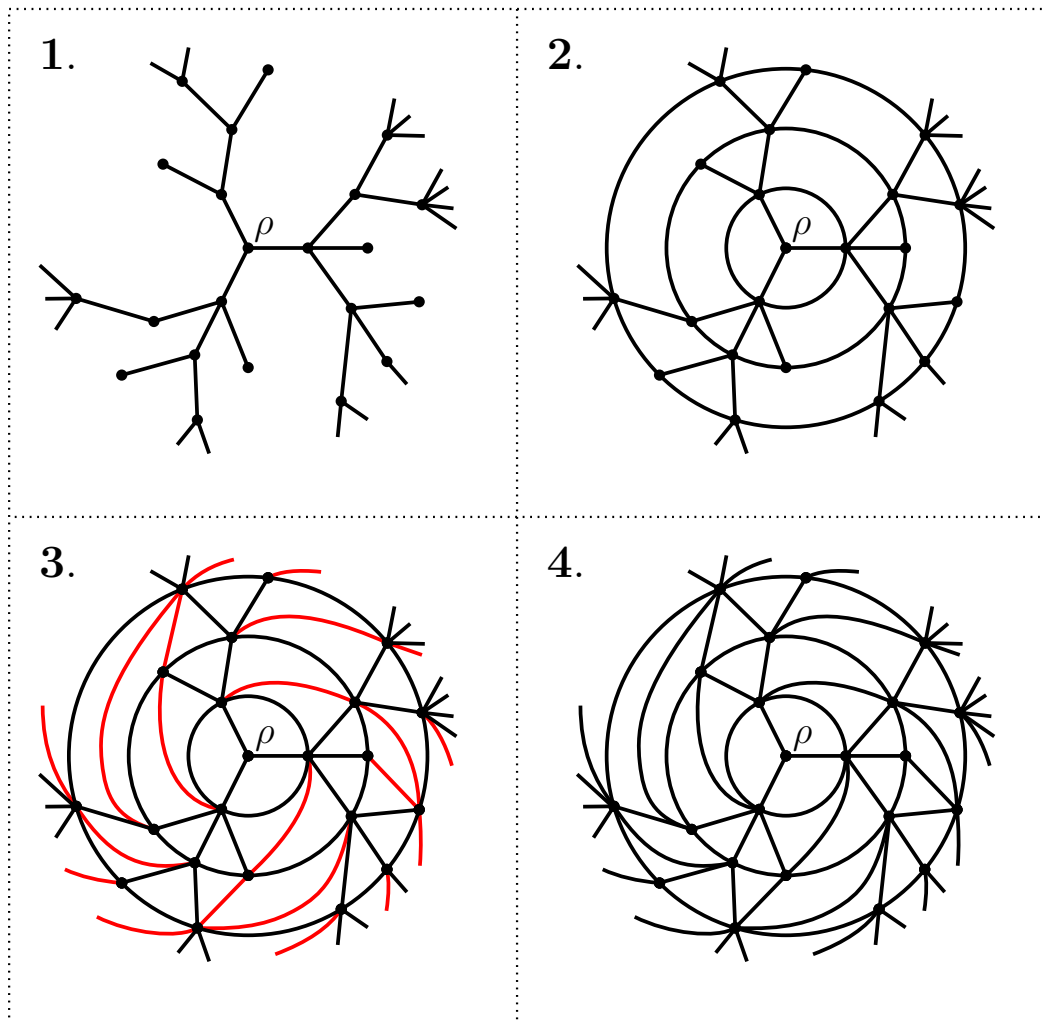


FIGURE 1.13 – Illustration de la construction en deux étapes d'une triangulation causale (représentée sur la figure n° 4) à partir d'un arbre plan (représenté sur la figure n° 1) : ajout des cycles (figure n° 2) puis triangulation des faces (via l'ajout des arêtes rouges de la figure n° 3).

La transformation de l'arbre en carte n'agit certes pas sur les distances "verticales", mais est susceptible d'affecter les distances "horizontales". Dans un arbre de Kesten, la distance maximale entre deux sommets situés à même hauteur  $r$  est asymptotiquement linéaire en  $r$ , car la géodésique reliant le sommet le plus à droite au sommet le plus à gauche descend typiquement près de la racine, avant de remonter dans l'arbre. Dans [56], Curien, Hutchcroft et Nachmias prouvent que cette distance est notablement réduite dans les triangulations causales critiques, au sens où son ordre de grandeur est maintenant  $o(r)$  quand  $r \rightarrow +\infty$ , à condition toutefois que la variance de la loi de reproduction sous-jacente à l'arbre de Kesten soit finie. L'ampleur de la baisse reste mesurée, car Curien, Hutchcroft et Nachmias montrent dans le même temps que la distance croît au moins comme  $r^{1-o(r)}$ . Quand la loi de reproduction est à queue lourde (et donc de variance infinie), ce phénomène n'est plus observé, la distance croît à nouveau linéairement (voir [56, Théorème 4]).

Une autre manière d'appréhender la géométrie d'un graphe est d'étudier le comportement de la marche aléatoire simple sur celui-ci : un "promeneur" part du sommet racine, puis à chaque instant, se déplace vers un autre sommet, choisi uniformément parmi tous les voisins de celui où il se trouve au moment présent. Lorsque la variance de la loi de reproduction sous-jacente à un arbre de Kesten est finie, la marche sur cet arbre est presque sûrement récurrente : le marcheur revient une infinité de fois au sommet racine. Durhuus, Jonsson et Weather montrent dans [64] que c'est également vrai quand le marcheur se déplace sur la triangulation causale associée. D'après Curien,

Hutchcroft et Nachmias (voir [56, Théorème 2]), le marcheur a toutefois tendance à s'éloigner du sommet racine plus rapidement. Après  $n$  déplacements, la distance qui le sépare du sommet racine est typiquement d'ordre  $\sqrt{n}$  lorsque  $n \rightarrow +\infty$ . On dit que la marche est *diffusive*. Sur l'arbre de Kesten seul, cette distance est plutôt d'ordre  $n^{1/3}$  [86]. De ce point de vue, carte et arbre diffèrent à nouveau significativement. Le comportement diffusif de la marche aléatoire simple sur la triangulation causale rapproche cette dernière des réseaux réguliers (et déterministes) du plan, à l'instar du réseau carré (voir Figure 1.1(b)).

**Triangulations causales surcritiques** Après le cas critique, il paraît naturel de s'intéresser aux triangulations causales construites à partir d'un arbre de Bienaymé-Galton-Watson surcritique, conditionné à survivre. Thomas Budzinski introduit dans [43] ce nouveau modèle de cartes aléatoires planaires, appelées *triangulations causales surcritiques*, afin d'étudier certains aspects de leur géométrie. L'article en question traite plus précisément des cartes causales surcritiques, où l'étape de triangulation des faces est omise lors du passage de l'arbre à la carte (l'étape n° 3 de la Figure 1.13). L'essentiel des résultats énoncés s'étend toutefois sans difficulté aux triangulations causales. Le travail de Budzinski s'inscrit par ailleurs dans une démarche plus globale d'analyse de différents modèles de cartes aléatoires *hyperboliques* [42, 44], dont les conclusions sont rassemblées dans le manuscrit de sa thèse [41].

En Section 1.3, nous avons présenté le modèle PSHIT comme un analogue hyperbolique de l'UIPT. Les triangulations causales surcritiques peuvent, quant à elles, être vues comme une version hyperbolique des triangulations causales critiques. Outre l'évidente croissance exponentielle du volume des boules, Budzinski montre dans [43] que les cartes causales surcritiques vérifient certaines propriétés d'hyperbolicité plus fines, relatives à la forme des géodésiques. Nous ne les détaillons pas ici. Budzinski réussit ensuite à étendre ces propriétés à d'autres modèles de cartes hyperboliques (les PSHIT notamment) car il observe que les cartes causales constituent une sorte de modèle-jouet de cartes contenant presque sûrement en leur sein un arbre de Bienaymé-Galton-Watson surcritique infini.

Imitant le travail réalisé dans le cas critique [64, 56], Budzinski étudie également le comportement de la marche aléatoire simple sur les cartes causales surcritiques, pour le comparer à celui de la marche sur l'arbre seul. Russel Lyons, Robin Pemantle et Yuval Peres ont prouvé [103] que cette dernière est presque sûrement transiente — au sens où le marcheur ne visite qu'un nombre fini de fois chaque sommet — et que le marcheur s'éloigne à vitesse positive du sommet racine — ce qui signifie que la distance entre la racine et celui-ci croît linéairement avec le nombre de déplacements effectués. Budzinski montre dans [43] que la marche se comporte de manière identique dans la carte causale. La démonstration proposée ne fonctionne toutefois que lorsque la loi de reproduction de l'arbre de Bienaymé-Galton-Watson sous-jacent ne génère aucune feuille. Autrement dit, tous les sommets de l'arbre doivent donner naissance à au moins un enfant.

## Percolation orientée sur triangulations causales surcritiques

Dans le chapitre 3 de ce manuscrit (voir aussi [105]), nous élargissons le périmètre de l'étude des triangulations causales, aux propriétés d'un processus de *percolation par arêtes*, auquel nous soumettons une partie d'entre elles. Il s'agit d'une autre manière courante d'appréhender la géométrie des graphes, qu'ils soient déterministes ou aléatoires. En début de Section 1.2, dans le paragraphe présentant le modèle d'Erdős-Renyi, nous avons énoncé la définition du processus de percolation par arêtes le plus élémentaire, à savoir le processus de *percolation de Bernoulli*, qui consiste à perturber la structure d'adjacence d'un graphe en supprimant des arêtes au hasard. Pour être plus précis, aux arêtes  $e$  d'un graphe  $G$  — toujours supposé infini (dénombrable), connexe et localement fini — nous associons des variables aléatoires de Bernoulli  $\mathbf{X}_e$ , indépendantes et identiquement distribuées, de paramètre commun  $p \in [0, 1]$ . Ce nombre  $p$  est l'unique paramètre du modèle de percolation. À partir de cette famille de variables aléatoires, nous construisons un nouveau graphe de même ensemble de sommets que  $G$ , mais dont l'ensemble d'arêtes n'inclut que celles vérifiant  $\mathbf{X}_e = 1$ . Nous notons  $G_p$  ce graphe, qui est bien sûr aléatoire par nature, et  $\mathbb{P}_p^G$  sa distribution. À l'issue du processus, chaque arête de  $G$  est présente dans  $G_p$  avec probabilité  $p$  — on dit alors qu'elle est *ouverte* — ou absente avec probabilité  $1 - p$  — on dit qu'elle est *fermée* — et ce indépendamment des autres arêtes. La percolation de Bernoulli peut s'appliquer à un graphe

étant lui-même aléatoire : on génère un graphe selon une certaine distribution  $\mathbb{P}$ , puis conditionnellement à la réalisation de ce dernier, on exécute le processus de percolation par arêtes décrit ci-dessus. Le graphe obtenu à l'issue de cette procédure suit alors la distribution

$$\mathbb{P}_p := \int \mathbb{P}(dG) \int d\mathbb{P}_p^G, \tag{1.5.1}$$

qui moyenne sur les graphes aléatoires de distribution  $\mathbb{P}$ , puis sur le processus de percolation qui leur est appliqué. Elle est appelée distribution *recuite* — ou *annealed* en anglais. Tandis que la distribution  $\mathbb{P}_p^G$  — où le graphe soumis au processus de percolation est fixé — est appelée distribution *trempee* — *quenched* en anglais. Les deux distributions coïncident évidemment quand le milieu est déterministe. Dans le chapitre 3, nous nous intéressons au cas où  $\mathbb{P}$  est la distribution d'une triangulation causale surcritique, construite à partir d'un arbre de Bienaymé-Galton-Watson dont la loi de reproduction est géométrique. Avant d'entrer dans le détail des résultats obtenus, nous allons dresser un rapide historique du modèle de percolation de Bernoulli, exposer les motivations qui ont présidé à son introduction et présenter les questions classiques qui se posent lorsqu'on procède à l'analyse du processus sur un graphe.

**Un modèle de physique statistique** Jusqu'à présent, dans ce chapitre d'introduction, nous avons fait deux fois allusion au modèle de percolation de Bernoulli. Dans la Section 1.2 pour remarquer que les graphes d'Erdős-Rényi étaient le résultat d'une percolation de Bernoulli appliquée au graphe complet. Dans la Section 1.4 pour citer de récents travaux [32, 29, 27] étudiant le processus sur les arbres à attachement préférentiel, notamment guidés par l'idée de simuler une panne d'ampleur, soudaine et aléatoire, subie par un réseau réel, et estimer ainsi sa capacité de résilience. À chaque fois, le contexte était donc celui de graphes finis, bien qu'on fasse ensuite fréquemment tendre leur volume à l'infini, pour déterminer par exemple s'il existe asymptotiquement, dans le graphe percolé, une composante connexe géante, c'est-à-dire de taille proportionnelle au nombre total de sommets du graphe. Voir à ce sujet la Section 1.3, où nous citons des travaux répondant à cette question pour les graphes d'Erdős-Rényi.

Toutefois, la formulation originelle du modèle de percolation de Bernoulli, et les motivations profondes qui ont amené à son introduction, sont différentes. Historiquement, le modèle fut conçu en 1957 par Simon Broadbent et John Hammersley [36], pour modéliser l'écoulement d'un fluide dans un milieu poreux. La dimension des pores étant généralement infimes devant la taille globale du milieu, Broadbent et Hammersley firent le choix de représenter ce dernier par un graphe infini, typiquement un réseau régulier sur  $\mathbb{Z}^d$ . Les arêtes restées ouvertes à l'issue du processus de percolation symbolisaient les points de passage possibles pour le fluide dans le milieu. Le paramètre  $p$  réglait la plus ou moins grande porosité de celui-ci.

Le problème que souhaitait initialement théoriser et résoudre Broadbent et Hammersley, était le suivant : supposons qu'une pierre poreuse soit plongée dans un seau d'eau, quelle est la probabilité pour que le centre de la pierre devienne humide ? Si on reformule cette question dans les termes du modèle décrit ci-dessus, y répondre revient en fait à calculer la probabilité qu'un chemin infini d'arêtes *ouvertes* — au travers desquelles va s'écouler l'eau environnant la pierre — atteigne l'origine du réseau. Broadbent et Hammersley ont montré que lorsque ce dernier est le réseau régulier hypercubique de sommets  $\mathbb{Z}^d$ , cette probabilité connaît une *transition de phase* pour  $d \geq 2$  : sous un certain seuil  $p_c(d)$  dit *critique*, strictement compris entre 0 et 1, la probabilité est nulle ; au dessus de ce seuil, elle est positive.

Ce résultat original et étonnant va indubitablement susciter l'engouement des physiciens statisticiens et des mathématiciens. Les premiers verront dans ce modèle de percolation une piste pertinente à explorer pour améliorer la compréhension des phénomènes de transition de phase observés dans la nature. Une relation forte et fructueuse sera ainsi découverte par Fortuin et Kasteleyn [68] entre le modèle d'Ising — un modèle de physique statistique introduit dans les années 1920, censé expliquer la température de Curie, cette température au-delà de laquelle un matériau ferromagnétique perd son aimantation — et le modèle de percolation de Bernoulli. Les mathématiciens seront, quant à eux, séduits par la définition élémentaire d'un modèle inédit, où l'absence totale de corrélations statistiques autorise de nombreux calculs. Ils entreverront dans ce modèle les ferments de possibles nouvelles théories mathématiques. Dans les décennies suivant la

parution de l'article fondateur de Broadbent et Hammersley, ils ne cesseront ainsi de le généraliser. En le sortant tout d'abord du cadre restreint des réseaux réguliers euclidiens. Ils étudieront la percolation de Bernoulli sur les arbres, les graphes de Cayley, les arbres et cartes aléatoires, etc. Les différentes hypothèses du modèle seront aussi progressivement relâchées, en considérant des probabilités de percolation inhomogènes, en effectuant une percolation par sites (ou sommets) plutôt que par arêtes, en insérant des relations de dépendances entre elles, etc. Le livre de Geoffrey Grimmett [72] constitue une excellente introduction aux mathématiques de la percolation dans son contexte d'origine, à savoir les réseaux hypercubiques de  $\mathbb{Z}^d$ , avec une insistance particulière sur le cas du plan.

**Une transition de phase** Imitant Broadbent et Hammersley, la première question à laquelle on a envie de répondre lorsqu'on soumet un graphe  $G$  à un processus de percolation de Bernoulli est l'existence (ou non) de composantes connexes — appelées aussi *clusters* — de taille infinie dans le graphe percolé  $G_p$ . Plus précisément, on cherche à déterminer, selon la valeur de  $p$ , si la probabilité de trouver un cluster infini dans  $G_p$  est positive ou non. Par un argument de couplage, il n'est pas difficile de voir que cette probabilité est une fonction croissante de  $p$ . Elle est nulle si  $p = 0$  (il n'y a aucune arête dans  $G_p$ ) et vaut 1 quand  $p = 1$  (on a alors  $G_p = G$ ). En fait, d'après la loi du zéro-un de Kolmogorov, les deux seules valeurs possibles pour cette probabilité sont soit 0, soit 1, car l'événement «il existe un cluster infini dans  $G_p$ » ne dépend jamais de l'état (ouvert ou fermé) d'un nombre fini d'arêtes de  $G$ . On définit le *seuil de percolation critique* de  $G$ , noté  $p_c(G)$ , par la relation suivante :

$$p_c(G) := \inf\{p \in [0, 1] : \mathbb{P}_p^G(\exists \text{ un cluster } \infty \text{ dans } G_p) = 1\}. \quad (1.5.2)$$

On dit qu'une transition de phase non triviale se produit si  $0 < p_c(G) < 1$ . Nous avons dit que c'était vrai dans les réseaux hypercubiques de  $\mathbb{Z}^d$  pour  $d \geq 2$ . Notons qu'en dimension 1, on a bien sûr  $p_c = 1$ .

Montrer l'existence d'une transition de phase non triviale est un objectif généralement accessible. Trouver une formule explicite pour le seuil de percolation critique l'est souvent moins. Plus de vingt années séparent par exemple l'article fondateur de Broadbent et Hammersley [36] et celui de Harry Kesten [85], dans lequel est calculée la valeur exacte — à savoir  $1/2$  — du seuil  $p_c$  correspondant au réseau carré sur  $\mathbb{Z}^2$ . La planarité et l'auto-dualité de ce graphe sont des arguments essentiels à la démonstration. Ces mêmes arguments sont utiles pour déterminer la valeur du seuil de criticité de quelques autres réseaux réguliers du plan. On sait par exemple que celui du réseau triangulaire vaut  $2 \sin(\pi/18)$  [144]. En dimensions strictement supérieures à 2, on ne dispose encore aujourd'hui d'aucune formule de ce type.

Trouver  $p_c$  s'avère nettement moins difficile lorsqu'on travaille avec des arbres. Principalement en raison de l'absence de cycles, qui permet d'estimer plus facilement le nombre de chemins d'arêtes partant d'un sommet précis, de longueur donnée. Le seuil  $p_c$  de l'arbre infini  $k$ -régulier — tous les sommets ont  $k \geq 2$  voisins — est  $\frac{1}{k}$ . Plus généralement, pour tout arbre infini  $T$ , on a  $p_c(T) = 1/\text{br}(T)$ , où  $\text{br}(T)$  est le *nombre de branchement* de  $T$ , un nombre réel quantifiant en un certain sens le nombre moyen d'enfants qu'ont les sommets de l'arbre. Ce résultat très général est dû à Russel Lyons [101].

Paradoxalement, calculer le seuil de criticité est parfois plus simple lorsque le graphe soumis au processus de percolation est lui-même aléatoire. Attention toutefois, dans ce contexte, le seuil  $p_c$  défini par (1.5.2) devient une variable aléatoire. En toute rigueur, on considère ici deux types de seuils déterministes. Un seuil  $p_c^r$  dit *recuit*, égal à l'infimum des  $p \in [0, 1]$  tel que  $\mathbb{P}(p_c(\mathbf{G}) \leq p) > 0$ , où  $\mathbf{G}$  est le graphe aléatoire et  $\mathbb{P}$  sa distribution. Autrement dit, au dessus de ce seuil, une fraction positive des graphes chargés par la distribution  $\mathbb{P}$  abrite presque sûrement un cluster infini. Le second seuil  $p_c^t$  dit *trempe*, est lui égal à l'infimum des  $p \in [0, 1]$  tel que  $\mathbb{P}(p_c(\mathbf{G}) \leq p) = 1$ . Au dessus de  $p_c^t$ , le graphe  $\mathbf{G}$  contient donc  $\mathbb{P}$ -presque sûrement un cluster infini. Cette distinction entre seuils recuit et trempé est heureusement superflue pour la grande majorité des modèles étudiés.

À l'instar des milieux déterministes, la percolation de Bernoulli sur arbres aléatoires est généralement traitée sans trop de peine. Particulièrement dans le cas des arbres de Bienaymé-Galton-Watson. En effet, le cluster de la racine d'un tel arbre  $\mathbf{T}$ , *surcritique*, dont la loi de reproduction a pour moyenne  $m > 1$ , est lui-même un arbre de Bienaymé-Galton-Watson. La moyenne de



sa loi de reproduction est égale, quant à elle, à  $pm$ . Le cluster a donc une probabilité positive d'être infini dès que  $p > 1/m$ . On déduit de cette simple observation que  $p_c(\mathbf{T})$  vaut presque sûrement  $p_c^r = p_c^t = 1/m$ . Le lecteur trouvera une démonstration rigoureuse du résultat dans [101].

La théorie de la percolation sur les cartes aléatoires infinies du plan naît avec la définition du modèle de l'UIPT par Omer Angel et Oded Schramm [20]. Rappelons que l'UIPT est une triangulation infinie du plan, aléatoire, qui est limite locale des triangulations uniformes du plan (voir Section 1.3). Dans un article paru la même année [15], Angel calcule le seuil de percolation critique de l'UIPT. Il montre que celui-ci vaut presque sûrement  $1/2$ . Il s'agit toutefois de percolation par sites (sur les sommets), non de percolation par arêtes. Plus tard, Laurent Ménard et Pierre Nolin détermineront, quant à eux, le seuil de criticité de l'UIPQ — l'analogue quadrangulaire de l'UIPT — pour la percolation par arêtes (entre autres résultats, voir [109]). Presque sûrement, ce seuil est égal à  $1/3$ . Dans [53, Théorème 12.6], Nicolas Curien obtient une expression du seuil de percolation pour toutes les cartes de Boltzmann critiques infinies du plan, un seuil à nouveau constant presque sûrement. Dans les trois travaux cités ci-dessus, les démonstrations reposent systématiquement sur l'utilisation astucieuse d'un algorithme d'épluchage de cartes (voir Section 1.2 pour une présentation succincte de ces algorithmes).

L'éventail de résultats est plus complet pour les modèles de cartes aléatoires infinies du demi-plan. La raison principale tient à la forme simplifiée prise par la propriété de Markov spatiale que vérifient ces cartes. Cette propriété réduit le problème de la détermination du seuil de percolation au calcul du *drift* d'une certaine marche aléatoire. La carte contient un cluster infini si et seulement si le drift de cette marche est strictement positif. Dans [17], Angel et Curien déduisent grâce à cette méthode le seuil de percolation par arêtes de l'UIHPT et de l'UIHPQ — les équivalents dans le demi-plan de l'UIPT et de l'UIPQ — à savoir respectivement  $1/4$  et  $1/3$ . Dans [53, Théorème 11.8], Curien généralise le résultat en démontrant une formule donnant le seuil de percolation de n'importe quelle carte de Boltzmann critique infinie du demi-plan.

La description de la transition de phase ne saurait être complète sans préciser le comportement en  $p = p_c$ . A priori, nous ne savons rien de l'existence (ou non) de clusters infinis au point critique. Résoudre ce problème revient en fait à établir la continuité ou discontinuité de la transition de phase. Qu'est-ce que cela signifie? Notons  $\Theta(p)$  la probabilité qu'un sommet quelconque du graphe soumis au processus de percolation de paramètre  $p$  soit inclus dans un cluster infini. Nous donnerons plus tard la définition rigoureuse de cette probabilité, notamment en milieu aléatoire. Intuitivement, on l'interprète aussi comme la "proportion" de sommets du graphe, recouverte par les clusters infinis. Quand  $p < p_c$ , on a bien sûr  $\Theta(p) = 0$ . C'est moins évident à première vue, mais il est vrai aussi que  $\Theta(p) > 0$  pour  $p > p_c$  : il suffit en effet de modifier l'état d'un nombre fini d'arêtes du graphe pour connecter le sommet considéré à un cluster infini, présent presque sûrement en régime surcritique. S'il y a percolation au point critique — c'est-à-dire qu'un cluster infini existe dans le graphe percolé — on aura par le même argument  $\Theta(p_c) > 0$ . La fonction  $p \mapsto \Theta(p)$  sera par conséquent discontinue en ce point. À l'inverse, s'il n'y a pas percolation, on obtient la continuité de la fonction sur tout l'intervalle  $[0, 1]$ , car on montre généralement qu'elle l'est sur  $(p_c, 1]$ . Le cas  $\Theta(p_c) = 0$  est particulièrement prisé en théorie de la percolation car on observe alors de remarquables phénomènes au seuil de criticité. Des limites d'échelle par exemple, ainsi que l'existence de constantes "universelles" — appelées *exposants critiques* — caractérisant le comportement du modèle dans un voisinage de  $p_c$ . Nous y reviendrons dans la suite. Tous les modèles cités dans ce paragraphe connaissent des transitions de phase *continues* : réseau carré du plan, arbres, UIPT, etc. Mais pour le réseau cubique de  $\mathbb{Z}^3$ , cela reste encore aujourd'hui une conjecture. Un graphe dont la structure est pourtant élémentaire! En fait, la continuité de la transition de phase du réseau hypercubique de  $\mathbb{Z}^d$  n'est prouvée que pour  $d = 2$  et  $d \geq 11$  [66].

**Transition de phase de triangulations causales surcritiques** Comme annoncé dans le préambule de cette section, nous étudions dans le chapitre 3 de ce manuscrit un processus de percolation de Bernoulli appliqué à des triangulations causales surcritiques. Plus exactement, nous considérons celles construites à partir d'un arbre de Bienaymé-Galton-Watson  $\mathbf{T}$ , surcritique, **conditionné à survivre**, dont la loi de reproduction  $\boldsymbol{\mu}$  est géométrique. Cela signifie qu'il existe  $\alpha \in (1/2, 1)$  tel que :

$$\forall k \geq 0, \quad \mu_k = (1 - \alpha) \cdot \alpha^k.$$

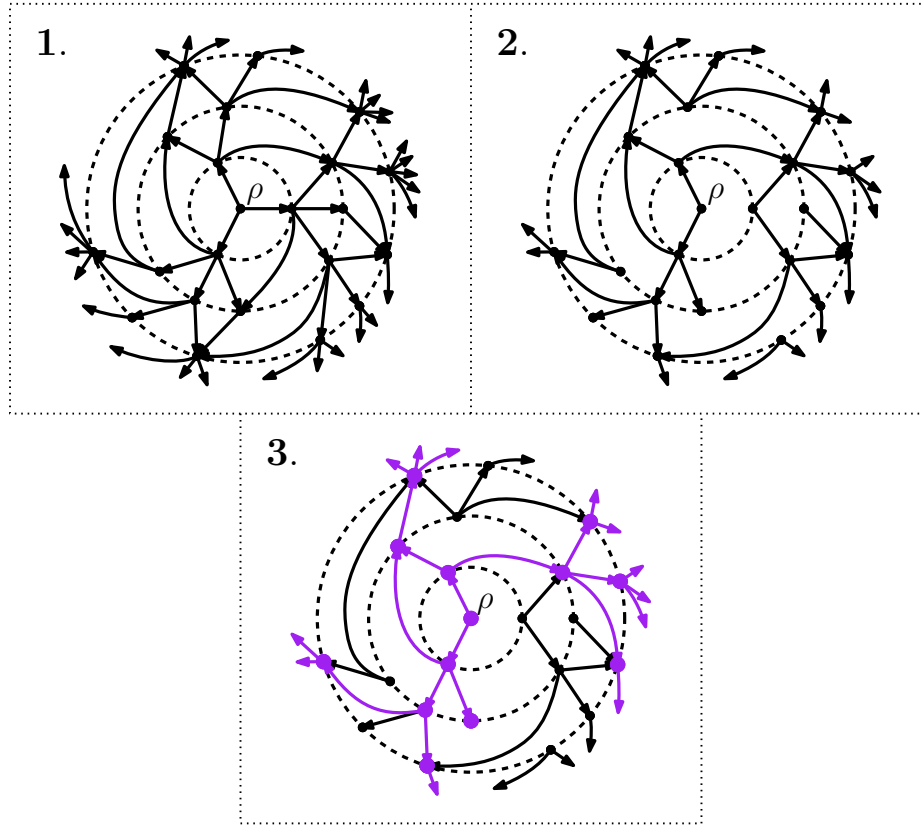


FIGURE 1.14 – Sur la première figure, en haut à gauche, est représentée une triangulation causale — identique à celle de la Figure 1.12 — dont on a "fermé" les cycles (en traits discontinus) et orienté les arêtes restantes, dans la direction opposée au sommet racine  $\rho$ . Sur la figure en haut à droite, un processus de percolation orientée de Bernoulli a été appliqué à cette carte. Enfin, sur la figure du bas, on a coloré en violet les sommets et arêtes faisant partie du cluster de percolation de  $\rho$ . Notons que des sommets noirs feraient partie de ce cluster si les arêtes n'étaient pas orientées.

La moyenne  $m$  d'une telle distribution est égale à  $\frac{\alpha}{1-\alpha}$ . On notera qu'on a bien  $m > 1$ . Le procédé permettant de passer d'un arbre à une triangulation causale a été décrit en début de section et en Figure 1.13. Nous notons  $\mathfrak{T}$  la carte obtenue à l'issue de la transformation de  $\mathbf{T}$ . Nous allons soumettre  $\mathfrak{T}$  à une percolation de Bernoulli, dans une version toutefois affaiblie de deux façons.

Tout d'abord, nous allons ignorer, tout au long de ce travail, les arêtes "horizontales" de la carte. Concrètement, **les arêtes appartenant aux cycles de la carte  $\mathfrak{T}$  ne sont pas sujettes au processus de percolation et sont d'office considérées fermées**. Tous les résultats énoncés dans cette section ne sont vrais que sous cette hypothèse. Sans elle, nous sommes à ce jour incapable d'atteindre le même niveau d'analyse du processus de percolation. Naturellement, l'objectif doit être de réussir, à terme, à relâcher cette hypothèse forte, car ce sont ces cycles qui font la particularité des triangulations causales.

Le second changement que nous apportons au modèle de percolation est de supposer **orientées** les arêtes "verticales" de la carte, celles connectant les sommets appartenant à deux cycles successifs distincts. L'orientation donnée à ces arêtes suit la direction opposée au sommet racine. Le procédé est illustré en Figure 1.14. Ce choix d'orienter les arêtes modifie les notions de chemin d'arêtes et de cluster. Désormais, un sommet  $x$  est relié à un sommet  $y$  par un chemin d'arêtes si toutes les arêtes traversées pour rejoindre  $y$  depuis  $x$  le sont dans leur sens d'orientation. Dans une triangulation causale, un chemin d'arêtes éloigne forcément du sommet racine. Si on note  $\rho$  ce dernier, on a par exemple forcément  $\mathbf{d}_{gr}(\rho, x) < \mathbf{d}_{gr}(\rho, y)$  pour les deux sommets  $x$  et  $y$  considérés ci-dessus.

Après avoir orienté les arêtes "verticales" de  $\mathfrak{T}$ , nous appliquons le processus de percolation habituel : chaque arête est déclarée ouverte avec probabilité  $p \in [0, 1]$ , fermée avec probabilité  $1 - p$ ,

et ce indépendamment des autres arêtes. Ce modèle de percolation agissant sur des arêtes orientées est connu sous le nom de *percolation orientée de Bernoulli*. Voir une illustration en Figure 1.14. Comme dans le cas non orienté, le cluster de percolation d'un sommet  $x \in \mathfrak{T}$  désigne l'ensemble des sommets atteignables depuis  $x$ , via un chemin d'arêtes orientées, ouvertes dans la carte percolée. De la même manière, on dit qu'il y a percolation s'il existe un cluster infini dans la carte. Sur la Figure 1.14, nous montrons à quoi peut ressembler le cluster de percolation orientée du sommet racine de  $\mathfrak{T}$ . Il est nécessairement moins étendu que le cluster défini de façon classique, c'est-à-dire sans orientation des arêtes.

Comme avec l'hypothèse de "fermeture" des cycles, (quasiment) tous les résultats ici énoncés ne sont vrais que dans le strict cadre d'une percolation orientée de Bernoulli. Le choix de cette dernière se justifie cependant moins par l'incapacité à analyser le processus de percolation ordinaire, que par un choix averti, fait en cohérence avec les raisons qui ont poussé à introduire ce modèle de cartes. En effet, celles-ci ayant été explicitement conçues pour avoir une structure causale, nous orientons les arêtes de  $\mathfrak{T}$  dans le sens d'écoulement du temps (voir Figure 1.12). Nous n'avons pas cherché à étendre nos résultats au cas d'une percolation de Bernoulli non orientée.

Nous énonçons maintenant notre premier résultat, qui fournit une expression sous forme de série infinie du seuil de percolation critique de  $\mathfrak{T}$ . La valeur de ce seuil ne dépend que de la moyenne  $m$  de la loi de reproduction de l'arbre de Bienaymé-Galton-Watson sous-jacent. Nous précisons également le comportement au point critique :

**Théorème 5** *Le seuil de percolation critique  $p_c(\mathfrak{T})$  est presque sûrement égal à :*

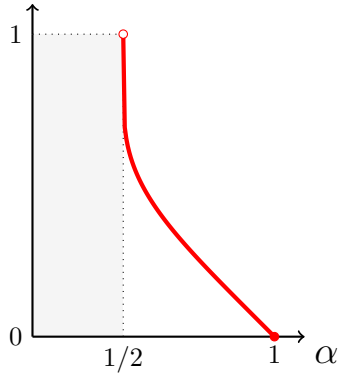
$$p_c = p_c(m) := \frac{\eta}{1 + \eta} \quad \text{avec} \quad \eta := \frac{1}{m + 1} \sum_{n \geq 0} \frac{m - 1}{m^{n+1} - 1}.$$

*De plus, il n'y a pas percolation au point critique presque sûrement.*

Nous avons tracé en Figure 3.2 (une approximation) du graphe de  $p_c$  comme fonction de  $\alpha$ . Sans surprise, sa valeur est décroissante : plus  $\alpha$  est grand, plus il y a d'arêtes dans la carte, et donc plus vite on crée un cluster infini en augmentant le paramètre de percolation  $p$ . Remarquons que  $p_c \rightarrow 1$  quand  $\alpha \rightarrow (1/2)^+$ , c'est-à-dire quand  $m \rightarrow 1^+$ .

Il est tentant d'examiner le cas  $\alpha = 1/2$ . La triangulation causale  $\mathfrak{T}$  est alors *critique*, construite à partir d'un arbre de Kesten, de loi de reproduction géométrique de paramètre  $1/2$ . Commençons déjà par observer que le théorème ci-dessus reste vrai sous cette forme lorsque la triangulation causale est issue d'un arbre de Kesten de loi de reproduction géométrique de paramètre  $\alpha > 1/2$ . Et non — comme ici — d'un arbre de Bienaymé-Galton-Watson, de même loi de reproduction, et conditionné à survivre. Les distributions de ces deux arbres ne coïncident pas, ni a fortiori celles des triangulations causales associées. En revanche, nous savons que la distribution de l'arbre de Kesten est absolument continue par rapport à celle de l'arbre de Bienaymé-Galton-Watson conditionné, car la loi géométrique est de variance finie [102]. C'est pourquoi le théorème vaut aussi lorsque la triangulation causale est issue d'un arbre de Kesten. Or, lorsqu'on fait tendre le paramètre  $\alpha$  de sa loi de reproduction vers  $1/2$ , cet arbre converge pour la topologie locale vers l'arbre de Kesten critique. La convergence se produit également pour les triangulations causales associées, car la transformation d'un arbre en triangulation causale est continue. En conclusion, le seuil de percolation critique dans le cas  $m = 1$  ne peut qu'être trivial et valoir 1 car  $\lim_{\alpha \rightarrow (1/2)^+} p_c(\alpha) = 1$ .

**Le nombre de clusters infinis dans le régime surcritique** Nous avons procédé à l'analyse de base d'un processus de percolation, en délimitant précisément la *phase surcritique* où émergent les clusters infinis. Cette analyse peut être approfondie de différentes façons. L'une d'elles consiste à déterminer si un ou plusieurs clusters infinis disjoints coexistent dans la carte lorsque  $p > p_c$ . En règle générale, lorsque le graphe soumis au processus de percolation est "suffisamment homogène", on constate une dichotomie : soit le graphe percolé admet un unique cluster infini, soit une infinité, et ce presque sûrement. Pour un graphe déterministe, être "suffisamment homogène" est synonyme de transitivité (ou quasi-transitivité [104, Chapitre 7]) : tous les sommets (ou presque) sont interchangeables dans la structure d'adjacence du graphe. Les réseaux hypercubiques de  $\mathbb{Z}^d$  ou


 FIGURE 1.15 – Tracé de  $\alpha \mapsto p_c(\alpha)$  sur l'intervalle  $(1/2, 1)$ .

les arbres  $k$ -réguliers sont des exemples typiques de graphes transitifs. Pour les graphes aléatoires, c'est la notion de stationnarité [9, 22], conjuguée avec une propriété de type Markov spatiale, qui remplace (et généralise) celle de transitivité.

Combien y a-t-il de clusters infinis dans la phase surcritique ? Un seul ou une infinité ? Deux facteurs semblent entrer en compétition. À mesure que le paramètre de percolation  $p$  augmente, le nombre d'arêtes ouvertes aussi, ce qui tend à créer de nouveaux clusters infinis. Toutefois, lorsque  $p$  est au voisinage de 1, la densité d'arêtes ouvertes risque d'être trop élevée et les clusters infinis fusionnent entre eux. Le cas extrême étant évidemment  $p = 1$ , où l'unique cluster infini est le graphe tout entier. Dans les faits, lorsque le modèle est raisonnable — c'est-à-dire génère des graphes "suffisamment homogènes" au sens donné ci-dessus — on rencontre successivement deux types de phases (voir par exemple [104, Théorème 7.20] et [9, Théorème 6.7]). Dans un premier temps, immédiatement après avoir franchi le seuil de percolation critique  $p_c$ , on observe un régime dans lequel le graphe percolé abrite presque sûrement une infinité de clusters infinis. Cette phase n'existe toutefois pas toujours et dépend fortement de la géométrie du graphe sous-jacent. Dans un deuxième temps, alors que la valeur de  $p$  continue d'augmenter, on finit par franchir un second seuil  $p_u \geq p_c$ , à partir duquel les clusters infinis ne font plus qu'un. Cette phase d'unicité existe toujours, mais peut éventuellement être réduite à  $p = 1$ .

Intuitivement, pour qu'un graphe puisse abriter une infinité de clusters infinis, il doit être "volumineux", contenir une forte "densité" de sommets et d'arêtes, capables de s'intercaler entre les clusters. Cela suggère d'examiner la croissance du volume des boules. Il s'agit pour résumer de comparer, quand  $r \rightarrow +\infty$ , le volume de la boule de rayon  $r$  et celui de sa frontière, c'est-à-dire l'ensemble de sommets à distance exactement  $r$  du sommet racine. Si le second est asymptotiquement une fraction du premier, Benjamini et Schramm ont conjecturé [23] qu'on a en effet  $p_c < p_u$ , soit une phase où coexistent une infinité de clusters infinis. Leur critère est en réalité un peu plus compliqué — sa formulation ne repose pas que sur le volume des boules et de leur frontière — mais nous omettons ces détails. Dans la littérature, les graphes déterministes qui satisfont le type de relation isopérimétrique évoquée ci-dessus sont dits *non moyennables* [104, Chapitre 7]. Il existe une notion analogue pour les graphes aléatoires, l'*expansion enracinée*<sup>1</sup> [141]. On s'attend à ce que les graphes aléatoires satisfaisant une propriété d'expansion enracinée vérifient aussi  $p_c < p_u$ .

Les arbres sont un exemple typique de graphes où on peut observer une telle phase. Par exemple les arbres  $k$ -réguliers et les arbres de Bienaymé-Galton-Watson surcritiques conditionnés à survivre. Pourquoi ? Car lorsque le processus de percolation déconnecte un sommet de son parent — ce qui concerne presque sûrement une infinité de sommets de l'arbre — il ferme l'unique point de passage possible vers le sous-arbre émanant de ce sommet. Or, celui-ci a une probabilité non nulle de percoler quand  $p > p_c$ , car il est (quasiment) distribué comme l'arbre tout entier, lequel percoler presque sûrement. On peut remarquer que dans ces deux modèles d'arbres, le volume des boules et de leur frontière croissent conjointement à vitesse exponentielle, proportionnellement l'un à l'autre. Ce qui est cohérent avec la conjecture de Benjamini et Schramm. Dans [121], Gourab Ray prouve que certaines triangulations aléatoires infinies du demi-plan — introduites par Angel et

1. *anchored expansion* en anglais, la traduction n'est pas canonique.

Ray dans [19] afin de généraliser l’UIPT — vérifient également  $p_c < p_u$ , pour une percolation par sites (voir [53, Théorème 12.9] pour la percolation par arêtes). Ces cartes se singularisent de l’UIPT par leur géométrie de nature *hyperbolique*, comparable à celle des arbres ci-dessus : boules dont le volume croît exponentiellement vite, expansion enracinée. L’hyperbolicité est en soi une notion assez large, admettant plusieurs définitions. On sait que certaines d’entre elles garantissent l’existence d’une phase où coexistent une infinité de clusters infinis. Dans [80], Tom Hutchcroft démontre la conjecture de Benjamini et Schramm, sous l’hypothèse supplémentaire de Gromov-hyperbolicité du graphe. Dans [18], Angel, Hutchcroft, Nachmias et Ray prouvent que  $p_c < p_u$  est équivalent, pour certaines cartes aléatoires planaires, à vérifier différentes propriétés d’hyperbolicité. Ce qui met parfaitement en évidence la relation forte existant entre la géométrie d’un graphe et sa réponse au processus de percolation. Tous les autres modèles cités dans cette section — réseaux sur  $\mathbb{Z}^d$ , UIPT, etc. — sont connus pour n’abriter presque sûrement qu’un seul cluster infini dans la phase surcritique, et vérifient donc  $p_c = p_u$  [5, 47, 15, 17, 53]. Rien de surprenant puisque dans ces graphes, contrairement aux modèles hyperbolique vus ci-dessus, le volume de la frontière d’une boule de rayon  $r$  est asymptotiquement négligeable (quand  $r \rightarrow +\infty$ ) devant celui de la boule entière (voir [15, 58] pour l’UIPT).

Les triangulations causales de notre modèle étant une extension d’arbres de Bienaymé-Galton-Watson surcritiques conditionnés à survivre, on s’attend légitimement à avoir  $p_c < p_u$ . Mais puisque le processus de percolation considéré agit ici sur des arêtes orientées, la notion de clusters infinis *disjoints* nécessite d’être précisée. En effet, deux clusters infinis issus de deux sommets distincts peuvent être disjoints, tout en étant inclus dans un troisième cluster plus grand, issu d’un sommet situé plus près de la racine. On ne veut pas les compter comme deux clusters infinis distincts. En fait, nous allons provisoirement revenir à un cadre *non orienté*. **Pour cet énoncé seulement**, les clusters de percolation sont définis en considérant non orientées les arêtes de la carte :

**Théorème 6** *Dans le régime surcritique  $p > p_c$  (excepté en  $p = 1$ ), il y a presque sûrement une infinité de clusters infinis disjoints dans  $\mathfrak{T}$ .*

On obtient ainsi  $p_c < p_u$ , avec  $p_u = 1$ . La phase d’unicité du cluster infini est donc triviale, comme elle l’est pour l’arbre de Bienaymé-Galton-Watson sous-jacent. Les triangulations hyperboliques du demi-plan étudiées dans [121], dont nous avons parlé plus haut, vérifient en revanche  $p_u < 1$ . Prendre en compte l’effet de la présence des cycles dans  $\mathfrak{T}$  pourrait abaisser de la même façon la valeur de  $p_u$ . C’est ce que suggère par exemple la Proposition 1 de [23], qui traite du cas de l’arbre binaire infini, augmenté d’arêtes horizontales.

**Exposants critiques** Dans les paragraphes précédents, nous avons établi des propriétés globales du processus de percolation : l’existence (ou non) de clusters infinis à l’échelle du graphe tout entier, leur nombre dans la phase surcritique. Pour affiner notre compréhension du processus — l’étendre notamment aux phases sous-critique et critique — il est naturel de s’intéresser à la distribution d’un cluster typique, par exemple via l’étude de son volume.

Dans la phase surcritique  $p > p_c$ , une quantité évidemment importante est la probabilité pour ce cluster d’être infini. Nous avons précédemment noté  $\Theta(p)$  cette probabilité, lorsque le cluster est celui d’un sommet quelconque choisi dans le graphe, lequel est soumis à un processus de percolation de paramètre  $p$ . Nous écrirons  $\mathcal{C}$  pour désigner le cluster en question. En milieu déterministe, cette définition de  $\Theta(p)$  ne souffre d’aucune ambiguïté. Le choix exact du sommet n’ayant pas d’incidence sur la suite de notre propos, nous n’en faisons pas mention dans la notation. En milieu aléatoire, il est en revanche nécessaire de préciser un peu plus cette définition. La sélection du sommet dont est issu le cluster doit tout d’abord se faire de manière déterministe. La plupart des modèles étudiés produisant des graphes enracinés, on se contente généralement de se placer au sommet racine. Une autre source d’incertitude dans la définition de  $\Theta(p)$  concerne la mesure de probabilité utilisée pour son calcul. Il peut s’agir soit de la distribution *recuite*, soit de la distribution *trempee* (voir (1.5.1)). Avec la seconde, la probabilité  $\Theta(p)$  devient une variable aléatoire. Le plus souvent, les énoncés portent toutefois sur la distribution recuite. Ce sera le cas dans la suite de ce paragraphe.

Sous le seuil critique  $p_c$ , la fonction  $\Theta$  est identiquement nulle. Elle ne fournit donc plus d’information pertinente sur le volume du cluster. Il est alors logique de reporter son attention sur le

comportement asymptotique de la queue de distribution. Autrement dit, d'examiner  $\mathbb{P}_p(|\mathcal{C}| \geq n)$  quand  $n \rightarrow +\infty$ , afin de connaître la taille typique des grands clusters dans le graphe percolé, ces derniers ne pouvant plus être infinis. Comme pour  $\Theta(p)$ , la mesure de probabilité  $\mathbb{P}_p$  désigne ici la distribution recuite, qui coïncide bien sûr avec la distribution trempée en milieu déterministe.

Il y a d'autres quantités qu'il peut s'avérer pertinent d'étudier. Nous nous limiterons cependant à celles introduites ci-dessus. C'est leur comportement au point critique (ou à proximité) qui nous intéresse tout particulièrement. En physique, un paradigme veut en effet que les variables macroscopiques, décrivant un système physique subissant une transition de phase, soient toutes gouvernées par des lois d'échelle près du seuil critique. Dans notre cas, cela se traduit concrètement par l'existence (présumée) de deux constantes  $\beta, \delta > 0$  — appelées *exposants critiques* — telles que :

$$\Theta(p) \underset{p \rightarrow (p_c)^+}{\approx} (p - p_c)^\beta \quad \text{et} \quad \mathbb{P}_{p_c}(|\mathcal{C}| \geq n) \underset{n \rightarrow +\infty}{\approx} n^{-1/\delta}. \quad (1.5.3)$$

La notation  $\approx$  signifiant ici que les fonctions à gauche et à droite du symbole sont équivalentes après passage au logarithme, au voisinage du point limite.

Montrer l'existence et calculer la valeur de ces exposants est un problème central en mathématiques de la percolation, bien que notoirement difficile à résoudre, en particulier pour les réseaux réguliers de petite dimension. La physique théorique fournit beaucoup de conjectures, mais rares sont les démonstrations rigoureuses. Un résultat significatif fut obtenu par Stanislas Smirnov et Wendelin Werner dans [130]. Ils prouvèrent (entre autres) que  $\beta = \frac{5}{36}$  pour une percolation par sites sur le réseau régulier triangulaire du plan. Leur raisonnement reposait sur un précédent travail de Harry Kesten [87], qui établit d'importantes relations — dites *d'échelle* — entre exposants critiques. Ces relations impliquaient que l'étude au point critique du modèle de percolation suffisait à caractériser son comportement au voisinage du seuil, et donc à déduire la valeur de certaines constantes, à l'instar de  $\beta$ . L'enjeu résidant principalement dans le calcul de la probabilité de voir se former (au point critique) de longs chemins d'arêtes ouvertes. Smirnov et Werner exploitèrent à cette fin les progrès récemment réalisés dans [129], où était montrée l'*invariance conforme* de la limite d'échelle du processus de percolation critique (obtenue quand la maille du réseau tend vers zéro), ce qui avait des implications pour l'asymptotique recherchée.

La relation KPZ [91], bien que non prouvée, prédit un lien entre les exposants critiques des milieux déterministes et ceux des milieux aléatoires. Connaître la valeur des seconds permettrait d'accéder aux premiers. Or, à l'image du seuil  $p_c$  marquant la transition de phase, calculer les exposants critiques (recuits) est généralement plus simple lorsque le processus de percolation agit sur une carte aléatoire qui satisfait une propriété de Markov spatiale. Omer Angel et Nicolas Curien ont été les premiers à s'intéresser au sujet des exposants critiques en milieu aléatoire. Ils ont dans [17] examiné le cas d'une percolation par sites sur l'UIHPT. Grâce à un algorithme d'épluchage, ils ont obtenu pour ce modèle  $\beta = 1$ , ainsi que plusieurs asymptotiques sur la taille au point critique de l'*enveloppe* d'un cluster, qui est l'union du cluster et des composantes connexes finies de son complémentaire dans la carte. La queue de distribution de son volume décroît ainsi à vitesse  $n^{-1/4}$ , celle de son périmètre comme  $n^{-1/3}$ . Dans l'UIPT, Matthias Gorny, Édouard Maurel-Segala et Arvind Singh ont calculé [71] que les valeurs des exposants intervenant dans ces deux lois d'échelle étaient multipliées par 2 comparées au modèle demi-plan, soit 8 pour le volume et 6 pour le périmètre. Comme dans [17], leur démonstration utilise un algorithme d'épluchage de cartes. Toujours dans le cadre de l'UIPT, Nicolas Curien et Igor Kortchemski ont étudié la longueur typique (au point critique) de l'interface séparant un cluster de sommets ouverts d'un cluster de sommets fermés. Ils ont prouvé que sa queue de distribution vérifiait elle aussi une loi d'échelle [57]. Leur approche diffère toutefois sensiblement de celles adoptées dans [17] et [71], car elle se fonde plutôt sur des arguments combinatoires. Le lecteur pourra remarquer (avec raison) qu'aucun des exposants cités ne correspond vraiment à la constante  $\delta$  de (1.5.3). Sa valeur reste en réalité inconnue à ce jour, bien que Gorny, Maurel-Segala et Singh soient parvenus à conjecturer  $\delta = 7$  (voir [71, Remarque 18]).

Un aspect fondamental de la théorie des exposants critiques est la supposée *universalité* de ces derniers. Nous savons que certains graphes ne percolent pas au même seuil, tout en se ressemblant géométriquement parlant. C'est le cas des réseaux carré et triangulaire du plan [85, 144], ou de l'UIHPT et de l'UIHPQ [17]. Le seuil de percolation  $p_c$  est donc fonction de propriétés *locales* du

milieu considéré. Il dépend aussi du type de percolation appliquée, qu'elle soit par sites ou par arêtes. À l'inverse, il y a de bonnes raisons de penser — à la fois théoriques et empiriques — que les exposants critiques ne sont pas aussi sensibles à ces variations. Dans le contexte originel des réseaux réguliers de dimension finie, on formule la conjecture d'universalité ainsi : tous les processus de percolation appliqués aux réseaux réguliers de dimension fixée  $d \geq 1$  connaissent des transitions de phase de même forme, c'est-à-dire qu'elles sont décrites par des exposants critiques identiques. On dit que ces processus forment une *classe d'universalité*. L'égalité  $\beta = \frac{5}{36}$  démontrée par Smirnov et Werner [130] pour une percolation par sites sur le réseau régulier triangulaire du plan est par exemple présumée vraie pour tout autre réseau planaire — réseau carré, hexagonal, etc. — et quel que soit le processus de percolation appliqué. On conjecture aussi l'existence de classes d'universalité parmi les graphes aléatoires. Des arguments l'étayant sont avancés par Angel et Curien dans [17] pour les modèles demi-plan.

Le principal résultat d'universalité aujourd'hui rigoureusement démontré concerne les modèles de "haute dimension". Ce terme désigne ceux dont l'ordre de grandeur du volume des boules de rayon  $r$  est  $r^d$  pour  $d$  grand. À partir d'une certaine dimension  $d_c$ , il est attendu que les valeurs des exposants critiques se stabilisent et coïncident avec celles observées sur un arbre régulier. Ce dernier peut être vu comme un réseau de dimension infinie, car sa croissance est exponentielle. Dans le régime  $d \geq d_c$  — dit de *champ moyen* — on aura par exemple toujours  $\beta = 1$  et  $\delta = 2$  (voir le calcul détaillé dans [72, Chapitre 10]). Pour les réseaux réguliers hypercubiques, Takashi Hara et Gordon Slade ont prouvé [74, 75] que le basculement vers ce régime de champ moyen se produisait au plus tard en  $d = 19$ . Robert Fitzner et Remco van der Hofstad ont abaissé ensuite ce seuil à  $d = 11$  [66]. Il est en fait généralement admis que la transition a lieu en  $d = 6$ . Pour les graphes dont la géométrie est de nature *hyperbolique* — ou qui sont simplement non moyennables — le volume des boules croît plus vite que n'importe quel polynôme, comme dans les arbres réguliers. Il paraît donc logique que les exposants critiques y prennent leur valeur de champ moyen. Ceci a été montré pour certains milieux déterministes [125, 80]. En milieu aléatoire, nous ne connaissons guère d'analogie à ces résultats, excepté pour les arbres de Bienaymé-Galton-Watson surcritiques. Lorsque leur loi de reproduction est de variance finie, il est en effet établi que  $\beta = 1$  et  $\delta = 2$ . La régularité de la fonction  $\Theta$  dans la phase surcritique et son comportement au voisinage de  $p_c$  sont analysés dans le récent papier de Marcus Michelen, Robin Pemantle et Josh Rosenberg [110] (voir en particulier la proposition 2.2). Le fait que la queue de distribution des clusters critiques soit d'ordre  $n^{-1/2}$  est une conséquence du codage des arbres par des excursions de marches aléatoires (voir par exemple [92, Lemme 1]), un codage que nous avons précédemment évoqué à plusieurs reprises. Dans le Chapitre 3 de ce manuscrit, nous calculons la valeur des exposants  $\beta$  et  $\delta$  associés au processus de percolation orienté appliqué à notre modèle de triangulations causales. Celles-ci suggèrent une transition de phase de type champ moyen. Nous obtenons plus exactement le résultat suivant :

**Théorème 7** *Soit  $\mathcal{C}$  le cluster de percolation du sommet racine de  $\mathfrak{T}$ . Il existe une constante strictement positive  $c > 0$  telle que pour tout  $p > p_c$  :*

$$c^{-1} \cdot (p - p_c) \leq \Theta(p) := \mathbb{P}_p(|\mathcal{C}| = +\infty) \leq c \cdot (p - p_c)$$

*et une autre constante  $C > 0$  telle que pour tout  $n \geq 1$  :*

$$C^{-1} \cdot n^{-1/2} \leq \mathbb{P}_{p_c}(|\mathcal{C}| \geq n) \leq C \cdot \log(n) \cdot n^{-1/2}.$$

Le premier encadrement implique que  $\beta = 1$ . Il fournit également une asymptotique de  $\Theta(p)$  plus précise que (1.5.3). On déduit du second encadrement que  $\delta = 2$ . Il est probable que l'on puisse éliminer le facteur logarithmique de la borne supérieure, mais la méthode de preuve employée ici ne nous a pas permis d'y parvenir. Enfin, rappelons que  $\mathbb{P}_p$  désigne la distribution *recuite* (voir (1.5.1)), qui moyenne sur le milieu (les triangulations causales) puis sur le processus de percolation qui lui est appliqué. Nous ignorons si des exposants similaires existent pour la distribution *trempee* et si ceux-ci coïncident avec les exposants recuits. Un problème qui reste aussi ouvert pour d'autres modèles de cartes aléatoires, voir par exemple la question concluant [17, Section 4.3]. Michelen, Pemantle et Rosenberg ont réussi dans [110] à montrer l'existence pour les arbres de Bienaymé-

Galton-Watson surcritiques, d'un exposant trempé  $\beta$  valant presque sûrement 1, à condition que la loi de reproduction de l'arbre admette un moment fini d'ordre suffisamment grand.

**Géométrie des grands clusters critiques** Après l'étude du volume d'un cluster de percolation, on va chercher à décrire plus finement sa géométrie, particulièrement au point critique. La deuxième asymptotique de (1.5.3) — qui introduit l'exposant  $\delta$  — indique le caractère fractal des grands clusters critiques, ce qui suggère de s'intéresser à leur *limite d'échelle*. Comme avec les exposants critiques, on s'attend à observer une forme d'universalité, une insensibilité de l'objet limite aux petites variations apportées au modèle de percolation sous-jacent.

On doit à Stanislas Smirnov la première avancée majeure sur le sujet. Dans [129], il démontre que la limite d'échelle d'un cluster critique issu d'une percolation par sites sur le réseau régulier triangulaire du plan est invariante conforme — autrement dit, sa loi est invariante par transformation conforme. En outre, l'interface séparant un cluster de sites ouverts d'un cluster de sites fermés converge vers une courbe aléatoire du plan, une courbe SLE — pour *Schramm-Loewner Evolution* — dont la définition a été posée par Oded Schramm dans [126]. Il est conjecturé que ceci reste vrai pour n'importe quel modèle de percolation sur un réseau de dimension 2.

Dans [57], Nicolas Curien et Igor Kortchemski établissent une sorte d'analogie en milieu aléatoire au résultat de Smirnov. Ils identifient plus exactement les différentes limites d'échelle possibles (à facteur multiplicatif près) d'une interface de percolation (par sites) sur l'UIPT. En effet, selon la valeur du paramètre  $p$ , l'aspect général de cette interface change radicalement. En régime sous-critique, sa limite est l'arbre continu brownien (voir Section 1.3 pour une définition). En régime surcritique, il s'agit cette fois du cercle unité  $\mathbb{S}^1$ . Enfin, au point critique, c'est un *arbre à boucles* qui émerge, un espace métrique continu que l'on peut grossièrement décrire comme une infinité de cercles collés le long d'un arbre *stable* (qui est, rappelons-le, la limite d'échelle d'arbres de Bienaymé-Galton-Watson de loi de reproduction de variance infinie, voir plus de détails en fin de Section 1.3).

À notre connaissance, la littérature sur la géométrie des clusters de percolation en milieu hyperbolique est à ce jour vierge de tout résultat. En première approche, analyser le cas des arbres de Bienaymé-Galton-Watson surcritiques semble être une bonne façon de se forger une intuition de ce qu'il se produit dans ces milieux. On sait en effet beaucoup de choses au sujet de ces arbres. Quand  $p = p_c$ , le cluster du sommet racine est distribué comme un arbre de Bienaymé-Galton-Watson *critique*. Sa limite d'échelle est donc l'arbre continu brownien [7, 97] (ou un arbre stable quand la variance de la loi de reproduction est infinie [62]). Comme il est conjecturé que les graphes hyperboliques font tous partie d'une même classe d'universalité — avec un comportement de champ moyen au voisinage du point critique — cette limite d'échelle devrait leur être commune. Voir [53, Question ouverte 12.12]. Notons qu'il a par ailleurs été montré que l'arbre continu brownien émerge comme limite d'échelle de processus aléatoires définis sur des espaces hyperboliques. Par exemple dans l'article [49] de Xinxin Chen et Grégory Miermont. Ce dernier propose en outre une revue des résultats connus sur le sujet (voir [49, Section 1.3]).

Dans le Chapitre 3 de ce manuscrit, nous montrons qu'au point critique de notre modèle de percolation orientée sur triangulations causales, le cluster  $\mathcal{C}$  du sommet racine conditionné à être grand, converge effectivement vers un arbre continu brownien, après renormalisation :

**Théorème 8** *En  $p = p_c$ , la convergence en distribution suivante a lieu pour la topologie de Gromov-Hausdorff :*

$$(n^{-1/2} \cdot \mathcal{C} \mid |\mathcal{C}| \geq n) \xrightarrow[n \rightarrow +\infty]{d_{GH}} a \cdot \mathcal{T}_{\geq 1},$$

*où  $a > 0$  est une constante ne dépendant que de la moyenne  $m$  de la loi de reproduction géométrique choisie pour définir  $\mathfrak{T}$ . L'espace métrique  $\mathcal{T}_{\geq 1}$  est l'arbre réel construit à partir d'une excursion brownienne de longueur supérieure à 1.*

Nous renvoyons le lecteur vers la Section 1.3 de cette introduction et [63, 97] pour une définition des arbres réels et de l'excursion brownienne de longueur quelconque. On appelle *arbre continu brownien de masse supérieure à 1* l'espace métrique  $\mathcal{T}_{\geq 1}$ . Comme le Théorème 7, l'énoncé ci-dessus ne vaut que pour la distribution *recuite*, non pour la distribution trempée.



## Heuristique des démonstrations

Nous résumons ici les principaux arguments utilisés dans le Chapitre 3 de ce manuscrit pour démontrer les Théorèmes 5, 6, 7 et 8, qui décrivent la transition de phase subie par le processus de percolation orientée introduit dans la section précédente. La stratégie adoptée pour prouver ces théorèmes se divise en deux étapes. Nous commençons par étudier le processus de percolation orientée sur un modèle-jouet générant des triangulations infinies du demi-plan, censées approcher la structure des triangulations causales surcritiques. Dans un second temps, nous transférons du modèle-jouet au modèle initial les résultats obtenus pour le premier.

**Percolation orientée dans un modèle demi-plan** Nous introduisons un ensemble de triangulations infinies du demi-plan, appelées *triangulations causales du demi-plan*, définies sur l'ensemble de sommets  $\mathbb{N} \times \mathbb{Z}$ . Leur aspect général est représenté en Figure 1.16. Ces cartes sont dotées d'une structure en couches, obtenue en quelque sorte en étirant à l'infini les cycles présents dans une triangulation causale. Chaque "couche", intercalée entre deux lignes de sommets de même ordonnée, est constituée d'une succession infinie de triangles, qui sont soit orientés vers le haut — les triangles colorés en bleu sur la Figure 1.16 — soit vers le bas — ceux colorés en rouge.

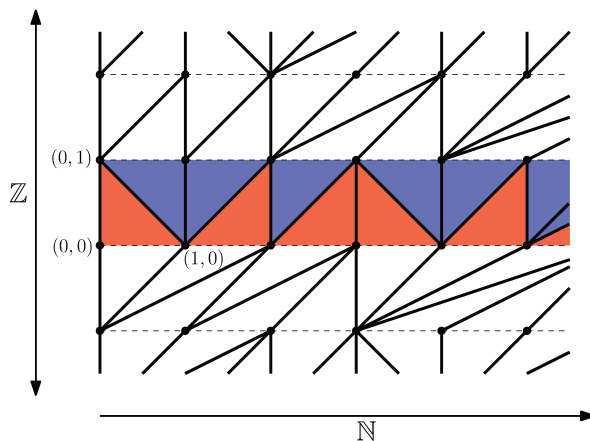


FIGURE 1.16 – Un exemple de triangulations causales du demi-plan.

Nous construisons maintenant une mesure de probabilité sur cet ensemble de triangulations. Nous supposons que chaque triangle de la carte a une probabilité  $1/2 < \alpha < 1$  d'être orienté vers le haut, respectivement une probabilité  $1 - \alpha$  d'être orienté vers le bas, et ce indépendamment de l'orientation des autres triangles. Nous notons  $\tilde{\mathfrak{T}}$  la carte tirée selon ce modèle aléatoire. Nous l'intitulons *triangulation causale surcritique du demi-plan*, par analogie avec le modèle plan défini dans la section précédente. Comment justifier cette analogie ? Le nombre de triangles bleus dont la pointe inférieure touche un sommet donné a pour distribution une loi géométrique de paramètre  $\alpha$ . Il coïncide avec le nombre d'arêtes émanant de ce sommet et dirigées vers le haut, moins une. Sur la Figure 1.16, il y a par exemple deux triangles bleus et trois arêtes, entre les lignes d'ordonnées 0 et 1, qui touchent le sommet de coordonnées  $(1, 0)$ . Considérons le sous-graphe de  $\tilde{\mathfrak{T}}$  constitué :

- d'un sommet en particulier (la *racine* du graphe) ;
- de toutes les arêtes, excepté la plus à droite (par exemple), qui partent de ce sommet et traversent la couche le surplombant immédiatement ;
- de tous les sommets pointés par ces arêtes, situés de l'autre côté de la couche ;
- enfin, de toutes les arêtes et sommets atteints en répétant ce procédé avec les nouveaux sommets, puis en l'itérant, possiblement à l'infini.

Ce graphe est un arbre, distribué comme un arbre de Bienaymé-Galton-Watson *surcritique*, de loi de reproduction géométrique de paramètre  $\alpha$ . En outre, tous les arbres obtenus de cette façon, enracinés sur une même ligne de sommets, sont i.i.d. La portion de la carte  $\tilde{\mathfrak{T}}$  située au-dessus de cette ligne peut être construite à partir de cette suite infinie d'arbres en utilisant les deux opérations qui permettent de passer d'un arbre à une triangulation causale : ajout d'arêtes "horizontales" entre

sommets de même ordonnée, puis triangulation des faces en reliant par une arête le coin inférieur gauche au coin supérieur droit. Voir Figure 1.17 pour une illustration de notre propos.

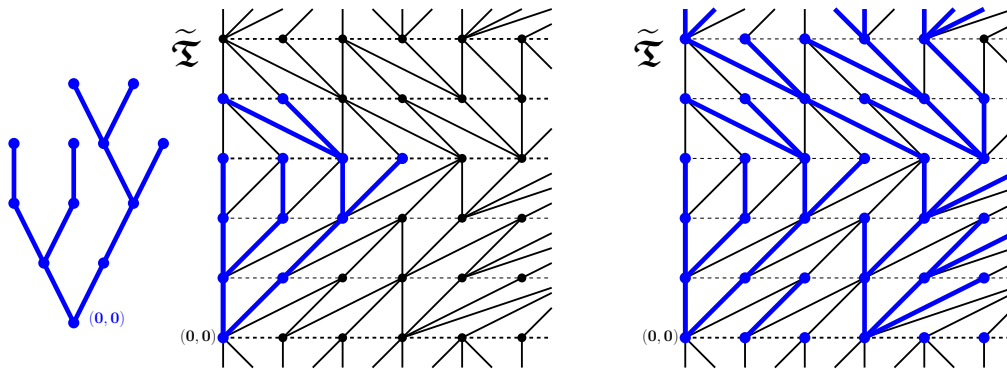


FIGURE 1.17 – L'arbre enraciné en  $(0,0)$ , coloré en bleu sur la figure centrale et isolé sur la figure de gauche, est construit selon le procédé décrit dans le paragraphe ci-dessus. Quand la carte sous-jacente  $\tilde{\mathfrak{T}}$  est une triangulation causale surcritique du demi-plan, il est distribué comme un arbre de Bienaymé-Galton-Watson *surcritique* de loi de reproduction géométrique de paramètre  $\alpha$ . Sur la figure de droite, nous observons que tous les arbres enracinés sur la ligne de sommets d'ordonnée nulle sont disjoints, donc indépendants, et recouvrent complètement la portion de la carte située au-dessus de cette ligne. Celle-ci peut être reconstruite à partir de ces arbres en connectant les sommets de même ordonnée (arêtes représentées par des lignes discontinues noires), puis en triangulant les faces (arêtes noires). Soit les deux opérations qui transforment un arbre en triangulation causale.

Nous souhaitons étudier sur  $\tilde{\mathfrak{T}}$  un processus de percolation orientée comparable à celui que nous avons défini dans la section précédente sur notre modèle de triangulations causales surcritiques. Ainsi, nous supposons **fermées** les arêtes "horizontales" reliant les sommets de même ordonnée, **orientées vers le haut** et soumises à une percolation de Bernoulli de paramètre  $p$  les arêtes restantes (celles connectant deux lignes successives de sommets). Nous montrons dans le Chapitre 3 de ce manuscrit le résultat suivant :

**Théorème 9** *Le seuil de percolation critique  $p_c(\tilde{\mathfrak{T}})$  est presque sûrement égal à  $p_c$ , quantité définie dans le Théorème 5. De plus, il n'y a pas percolation au point critique. Enfin, si nous notons  $\tilde{\mathcal{C}}$  le cluster de percolation du sommet  $(0,0)$ , alors ces quatre assertions sont vraies :*

(i) *pour tout  $p < p_c$ , il existe  $c = c(p) \in (0, 1)$  tel que  $\mathbb{P}_p(|\tilde{\mathcal{C}}| \geq n) = \mathcal{O}_{n \rightarrow +\infty}(c^n)$  ;*

(ii) *il existe  $c_1 > 0$  tel que  $\mathbb{P}_{p_c}(|\tilde{\mathcal{C}}| \geq n) \underset{n \rightarrow +\infty}{\sim} c_1 \cdot n^{-1/2}$  ;*

(iii) *il existe  $c_2 > 0$  tel que  $\mathbb{P}_p(|\tilde{\mathcal{C}}| = +\infty) \underset{p \rightarrow (p_c)^+}{\sim} c_2 \cdot (p - p_c)$  ;*

(iv) *en  $p = p_c$ , la convergence en distribution suivante a lieu pour la topologie de Gromov-Hausdorff :*

$$(n^{-1/2} \cdot \tilde{\mathcal{C}} \mid |\tilde{\mathcal{C}}| \geq n) \xrightarrow[n \rightarrow +\infty]{\mathbf{d}_{GH}} a \cdot \mathcal{T}_{\geq 1},$$

*pour une constante  $a > 0$  égale à celle apparaissant dans le Théorème 8 (à voir également pour la définition de  $\mathcal{T}_{\geq 1}$ ).*

La première partie du théorème décrit une transition de phase identique à celle que connaissent les triangulations causales surcritiques (voir Théorème 5). L'item (iv), qui traite de la géométrie à grande échelle des clusters critiques, est l'exacte réplique du Théorème 8. Des items (ii) et (iii), nous déduisons la valeur des exposants critiques recuils  $\beta$  et  $\delta$  du modèle demi-plan, à savoir respectivement 1 et 2. C'est-à-dire les valeurs de champ moyen, que l'on retrouvait déjà pour le modèle plan (voir Théorème 7). Les asymptotiques énoncées dans ces deux items signifient

que les suites  $n^{1/2} \cdot \mathbb{P}_{p_c}(|\tilde{\mathcal{C}}| \geq n)$  et  $(p - p_c)^{-1} \cdot \mathbb{P}_p(|\tilde{\mathcal{C}}| = +\infty)$  convergent vers des constantes strictement positives, au voisinage du point limite indiqué. Un résultat plus fort que le Théorème 7 qui, dans le contexte des triangulations causales surcritiques, impliquait seulement, pour les deux mêmes suites, l'existence de limites inférieure et supérieure strictement positives. Nous croyons toutefois qu'il est possible de prouver, comme ici, l'égalité de ces deux limites. Enfin, l'item (i) du Théorème 9 est une propriété classique en théorie de la percolation, connue sous le nom de *transition de phase abrupte*. Elle veut dire que la queue de distribution du volume des clusters décroît exponentiellement vite dans la phase sous-critique. Plus précisément, quand  $p < p_c$ , nous obtenons que la suite  $c^{-n} \cdot \mathbb{P}_p(|\tilde{\mathcal{C}}| \geq n)$  est bornée supérieurement pour un certain  $c \in (0, 1)$ . Contrairement aux autres items, nous n'avons pas réussi à établir une propriété analogue pour les triangulations causales surcritiques. Il est cependant raisonnable de penser que la transition de phase y soit aussi *abrupte*. En effet, elle l'est par exemple pour les arbres de Bienaymé-Galton-Watson surcritiques.

**Exploration d'un cluster de percolation** La démonstration du Théorème 9 repose sur un algorithme d'exploration du cluster  $\tilde{\mathcal{C}}$ , dont le principe s'inspire de plusieurs travaux traitant également de percolation sur des modèles de cartes aléatoires [15, 16, 17]. L'idée est la suivante : un promeneur, partant du sommet origine  $(0, 0)$ , découvre peu à peu de nouvelles portions du cluster, en cherchant systématiquement à aller le plus à gauche possible. L'objectif étant de trouver un chemin infini d'arêtes, indiquant qu'il y a percolation dans la carte. Concrètement, quand le promeneur est en un sommet donné, on lui révèle l'état de l'arête (orientée vers le haut) la plus à gauche qui émane de ce sommet. Si elle est ouverte, il la traverse pour rejoindre le sommet situé de l'autre côté. Si elle est fermée, on recommence avec les arêtes suivantes jusqu'à en trouver une ouverte. Si toutes les arêtes sont fermées, il n'a pas d'autre choix que de revenir sur ses pas et tenter de bifurquer plus à droite. Voir Figure 1.18 pour une illustration des différents cas de figure. Quand aucune voie alternative ne s'ouvre au promeneur, l'exploration s'arrête. On peut alors conclure que le sommet  $(0, 0)$  ne percole pas. La Figure 1.19 montre l'exploration complète d'un cluster de taille finie, suite et fin des premières étapes présentées sur la Figure 1.18.

Si nous notons  $\mathbf{V}_n$  le  $n$ -ième sommet visité par le promeneur et  $h(\mathbf{V}_n)$  son ordonnée, on peut remarquer que l'exploration se poursuit tant que  $h(\mathbf{V}_n)$  est positive ou nulle. Le cluster  $\tilde{\mathcal{C}}$  est donc infini si et seulement si  $h(\mathbf{V}_n) \geq 0$  pour tout  $n \geq 0$ . La suite de notre démonstration du Théorème 9 repose alors sur le fait que  $h(\mathbf{V}_n)$  est une marche aléatoire et que nous pouvons calculer la distribution exacte de ses incréments. Pour le démontrer, nous interprétons l'exploration décrite ci-dessus comme un processus d'épluchage de  $\tilde{\mathfrak{T}}$  (voir Section 1.2). La partie de la carte que nous considérons observée à un instant donné est située à gauche du chemin violet (voir Figure 1.18). Nous prouvons que ce processus d'épluchage satisfait une propriété de Markov spatiale : la région inexplorée à droite du chemin violet est indépendante de la partie déjà révélée, et sa loi est celle d'une triangulation causale surcritique du demi-plan. L'hypothèse que les arbres de Bienaymé-Galton-Watson sous-jacent à la carte (voir Figure 1.17) ont une loi de reproduction géométrique — donc *sans mémoire* — joue ici un rôle fondamental.

La probabilité de percolation du sommet  $(0, 0)$  est donc égale à la probabilité pour la marche aléatoire  $(h(\mathbf{V}_n))_n$  de rester éternellement positive ou nulle. Cette probabilité est strictement positive si le drift de la marche l'est aussi, nulle sinon. Nous calculons une expression de ce drift, linéaire en  $p$ , et déduisons la valeur du seuil de percolation critique  $p_c$ , solution d'une équation linéaire à une inconnue. Ce qui démontre la première partie du Théorème 9. La probabilité pour cette marche aléatoire de toujours rester au dessus de 0 est approximativement proportionnelle à la valeur du drift quand celui-ci est positif (voir [4, Théorème 3]). Nous déduisons ainsi l'item (iii). Pour les items (i) et (ii), nous prouvons que le volume du cluster  $\tilde{\mathcal{C}}$  croît (en gros) linéairement avec la durée totale de l'exploration, laquelle est distribuée comme la longueur d'une excursion positive de la marche aléatoire. Il est bien connu que la queue de distribution de cette dernière décroît exponentiellement vite si le drift est négatif — d'où l'item (i) — et à vitesse polynomiale quand la marche est centrée [140] — ce qui donne l'item (ii). Enfin, la démonstration de l'item (iv) repose sur l'observation que le sous-graphe des sommets visités par le promeneur — en vert sur la Figure 1.18 — est un arbre, caractérisant quasiment à lui seul la géométrie à grande échelle de  $\tilde{\mathcal{C}}$ . Le conclusion vient alors du fait que  $(h(\mathbf{V}_n))_n$  est proche du processus de contour de cet arbre, puis d'une application du théorème de Donsker, comme vu en fin de Section 1.3.

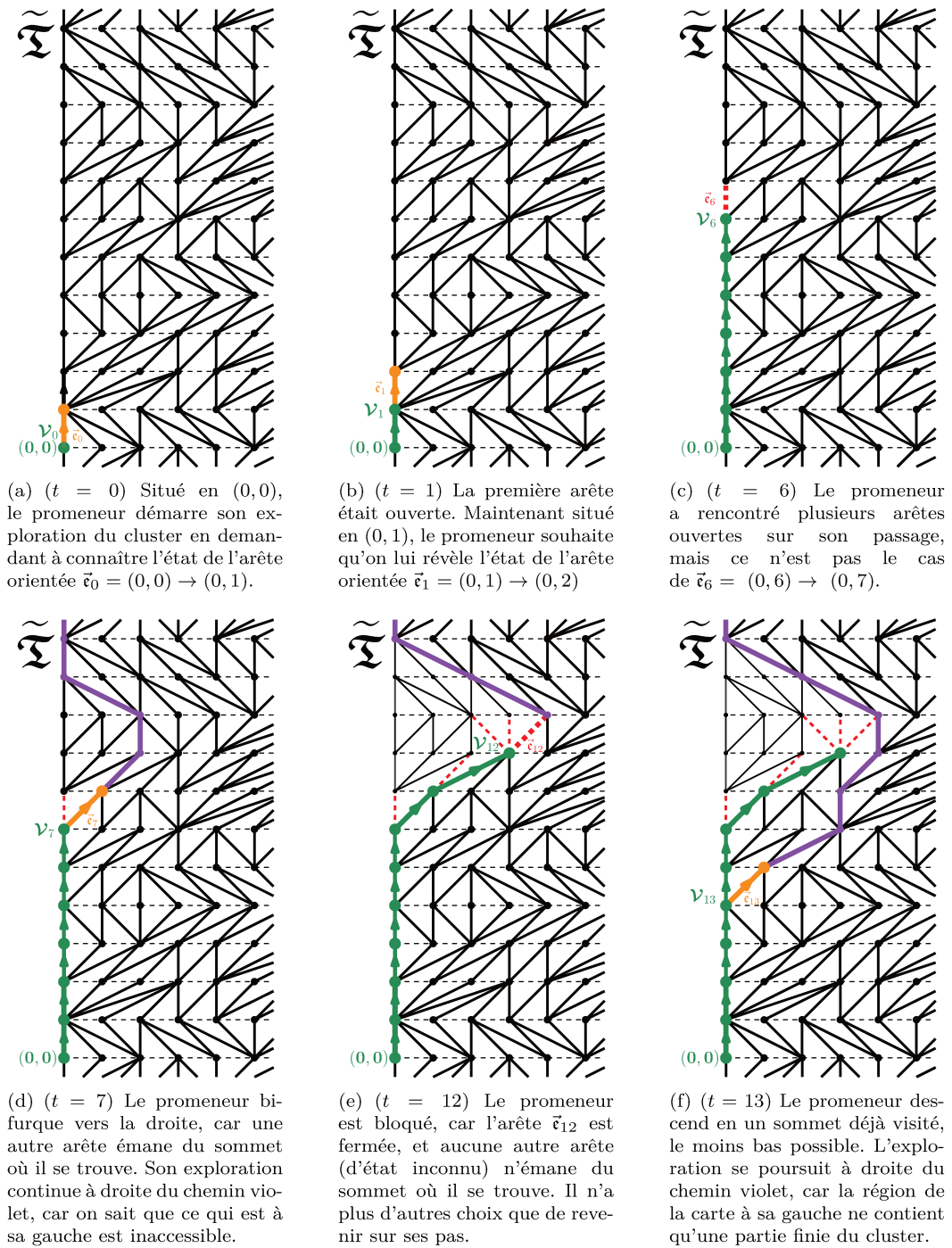


FIGURE 1.18 – Étapes de l'exploration du cluster  $\tilde{\mathcal{C}}$ . Par construction, la région de la carte  $\tilde{\mathcal{T}}$  à gauche du chemin violet ne contient à tout instant qu'une partie finie du cluster.

**Retour au modèle plan** Étudier le processus de percolation orientée sur les triangulations causales surcritiques nécessite de s'intéresser à son comportement loin du sommet racine. Nous allons expliquer pourquoi le même processus appliqué au modèle demi-plan en constitue une approximation. Dans l'arbre de Bienaymé-Galton-Watson surcritique  $\mathbf{T}$  sous-jacent à une triangulation causale surcritique — arbre qui est, rappelons-le, conditionné à être infini — nous savons qu'il y a environ  $m^r$  sommets à distance  $r \gg 1$  de la racine. Les sous-arbres émanant de ces sommets sont eux mêmes distribués comme des arbres de Bienaymé-Galton-Watson surcritiques, de loi de repro-

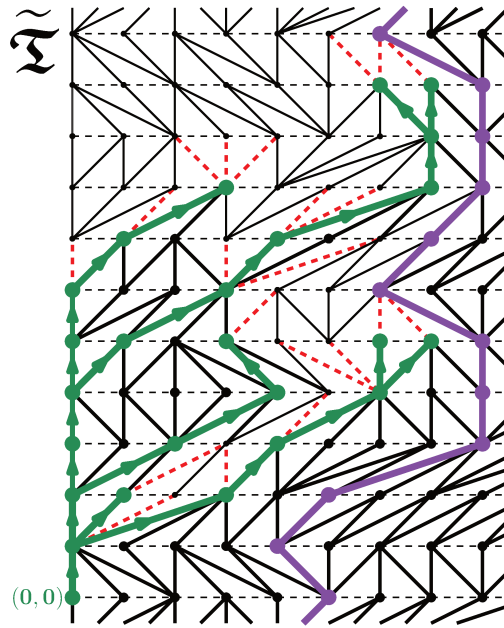
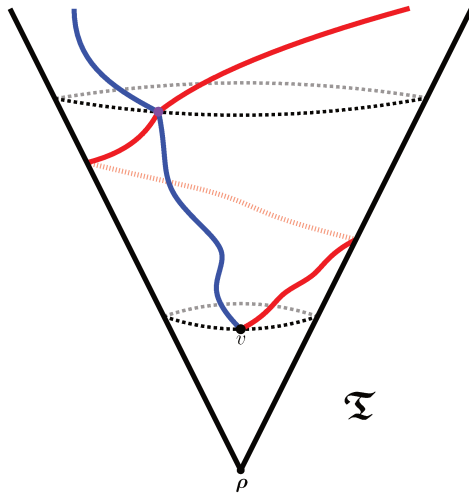
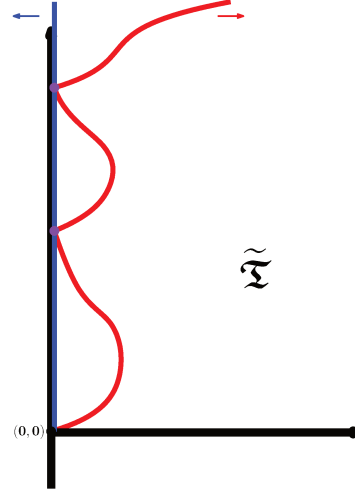


FIGURE 1.19 – Exploration complète d’un cluster de taille finie. Les arêtes en traits discontinus rouges ont été trouvées fermées. Tout sommet d’ordonnée nulle qui percolerait est nécessairement situé sur le chemin violet ou à sa droite.

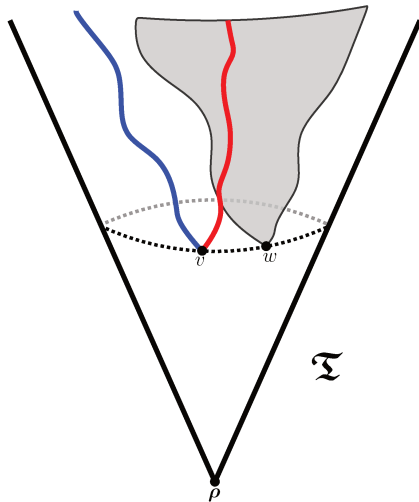
duction géométrique de paramètre  $\alpha$ , mais ne sont pas indépendants, en raison du conditionnement sur  $\mathbf{T}$ . Ce conditionnement est toutefois peu contraignant, car parmi un nombre exponentiellement grand d’arbres surcritiques indépendants, la probabilité qu’au moins l’un d’entre eux soit infini est très proche de 1. D’après l’observation faite dans le paragraphe précédent la Figure 1.17 (et illustrée sur celle-ci), le modèle demi-plan représente la situation naturellement limite : une infinité de sommet dont émanent des arbres surcritiques indépendants (le conditionnement devient alors inutile). Ce que nous montrons dans le Chapitre 3 de ce manuscrit est que le paysage vu au-dessus de lui par un sommet très éloigné de la racine de la triangulation causale, est fondamentalement identique à celui qu’il verrait dans le modèle demi-plan. En raison de la présence certaine, dans  $\mathbf{T}$ , de sous-arbres particulièrement volumineux, infini, que nous appelons en anglais *wide trees*. Si à une certaine distance  $r \gg 1$  de la racine, nous identifions un sous-arbre de ce type, celui-ci étant par définition infini, il va libérer les autres sous-arbres issus de sommets situés à cette même distance, du conditionnement sur  $\mathbf{T}$  qui nous gênait. En outre, par construction, aucun cluster émanant de ces sommets ne pourra traverser de part en part le sous-arbre *wide*, et tout se passera donc pour eux comme dans le modèle demi-plan. La Figure 1.20 illustre les points importants de ce raisonnement.



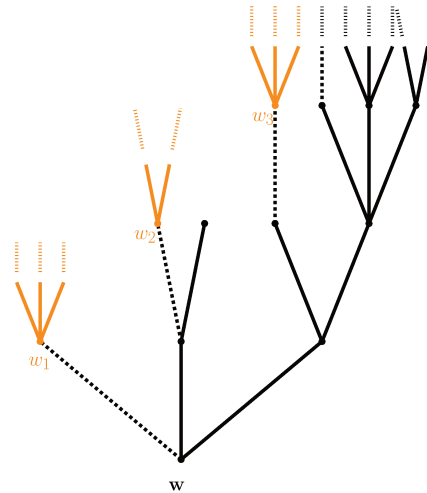
(a) Dans une triangulation causale, le chemin le plus à droite (en rouge) partant d'un sommet  $v$  peut rencontrer par la gauche le chemin le plus à gauche (en bleu).



(b) Cette situation est impossible dans le modèle demi-plan.



(c) Un sous-arbre très volumineux, dit *wide*, va en fait émerger d'un autre sommet  $w$ , et bloquer le passage du chemin rouge. Tout se passe donc pour  $v$  comme dans le modèle demi-plan. Les chemins bleu et rouge ne peuvent se rejoindre.



(d) Un arbre est *wide* si sa composante en noir sur la figure ci-dessus est infinie. Les sommets en orange forment une barrière infranchissable pour le chemin rouge émanant du sommet  $v$ .

FIGURE 1.20 – Le transfert aux triangulations causales surcritiques des résultats obtenus pour le modèle demi-plan se fonde sur la présence de sous-arbres très volumineux, qu'il est impossible de traverser en partant d'un autre sommet.

## 1.6 A brief introduction for non-French speakers

**Graphs, trees and maps** A *graph* is defined as a pair  $G = (V, E)$ , where  $V$  is a set whose elements are called *vertices* of the graph—or sometimes *nodes*—and  $E$  is a set of paired vertices whose elements are called *edges*—or *links*. Vertices are generally visualized as points with segments linking them to each other. A segment links together two vertices  $x, y \in V$  if and only if the pair  $(x, y)$  belongs to the set of edges of the graph. On Figure 1.1 are depicted two examples of graphs. The figure on the left shows the *complete graph* on six vertices : all the vertices are connected to each other. The figure on the right represents the square-grid graph on  $\mathbb{Z}^2$  : two vertices are there joined by an edge if their Euclidean distance is equal to one. There are plentiful motivations to study graphs. They are for instance used to model real networks like social networks—edges then modelling "friendship" between individuals—or the brain network. See [117] for more details. Graphs may also play a role in the field of discrete optimization. As an example, to solve maximum flow problems [93], which have concrete applications. Finally, unexpected and fruitful relations have been found between graph theory and other research topics in mathematics [60].

Some graphs exhibit properties which make them particularly interesting to study. This includes *trees*, which are *connected* graphs with no cycles—every vertex of the graph can be reached from any other vertex via a unique path of edges. Once a particular vertex of the tree is distinguished from the other ones—it is called the *root vertex*—the rest of the graph turns out to structure itself in a genealogical manner around it. See for instance the graph depicted on Figure 1.2(a). Trees hold a special place among graphs due to several reasons. They are "extremal". Indeed, any connected graph on  $n$  vertices has at least  $n - 1$  edges. Those minimizing the number of edges are exactly the trees. Furthermore, trees show striking combinatorial properties. There exists for instance an elegant bijection between plane trees—which are roughly rooted trees with an ordering on vertices—and some integer valued processes. See Figure 1.2(b) where is plotted the *contour process* of a plane tree. Another widely studied class of graphs is *planar maps*. A planar map is a connected graph on a finite number of vertices, properly embedded in the sphere  $\mathbb{S}^2$ , and viewed up to homeomorphisms that preserving orientation. On Figure 1.3 are depicted two distinct maps, sharing the same graph structure, though. An alternative representation of maps consists of seeing them as a gluing of a finite number of polygons—the *faces* of the map— along their edges, so that the resulting manifold is a topological sphere. The map on Figure 1.4 is for instance obtained by gluing one quadrangle and two pentagons. Like for trees, nice combinatorial formulas are known for maps, mainly because of the rigidity that brings the fact to consider a graph together with an embedding. On this topic, see the famous serie of papers written in the 1960s by William Thomas Tutte [133, 134, 135, 136, 137].

**Large scale geometry of random graphs** This dissertation is devoted to a study of two models of random graphs. A *random graph* is just a random variable taking value in a set of graphs. For instance, a tree drawn uniformly at random among the plane trees on  $n$  vertices, or a triangulation—a map whose all faces are triangles—among those having exactly  $n$  faces. A famous and more elaborated model, used to describe the growth of some idealized population, is the Bienaymé–Galton–Watson model (see [76, Chapter 1]). It produces random plane trees in a recursive way. Fix a distribution  $\mu = (\mu_k)_{k \geq 0}$  on positive integers, called the *offspring distribution* of the model. Start with a tree initially made of a single vertex, its root. We draw a random variable  $\mathbf{X}$  from distribution  $\mu$ , create  $\mathbf{X}$  new vertices in the tree, and link them to the root. Then, we repeat the procedure for each of these vertices, independently of the others, and so on. See Figure 1.5 for an illustration of the first two steps of the recursive construction of Bienaymé–Galton–Watson trees. When the mean  $m$  of the offspring distribution is strictly greater than 1, the tree has a positive chance to be infinite—that is to contain an infinite number of vertices. It is a *supercritical* Bienaymé–Galton–Watson tree. When  $m \leq 1$ , the tree is almost surely finite, except in the degenerate case  $\mu_1 = 1$ . It is then a *critical* (for  $m = 1$ ) or a *subcritical* Bienaymé–Galton–Watson tree.

Graphs are geometric objects by nature. For any two vertices of a graph, their graph distance is defined as the minimal number of edges that is necessary to cross to go from one to the other. Endowed with the graph distance, the vertex set then becomes a metric space. The notion of *large scale geometry* of a graph refers to features of the latter, viewed as metric space, when its number

of vertices tends to infinity. As an example, the diameter of a critical Bienaymé–Galton–Watson conditioned to have  $n$  vertices is known to be roughly  $n^{1/2}$ , when the offspring distribution has a finite variance [7, 97].

By letting grow to infinity the number of vertices of a graph, we are naturally inclined to ask the question whether some notion of limit exists. The metric point of view has led to fruitful developments in this direction. For instance, by using the Gromov–Hausdorff topology on (isometry classes of) compact metric spaces, a random tree counterpart was proved by David Aldous [7] to the well-known Donsker’s theorem for random walks. Under the hypothesis of an offspring distribution having a finite variance, a critical Bienaymé–Galton–Watson, conditioned to have  $n$  vertices and whose distances within it are all divided by  $n^{1/2}$  (namely the asymptotic order of the diameter), converges in distribution as  $n \rightarrow +\infty$  to a universal compact metric space called *Continuum Random Tree*, with respect to the Gromov–Hausdorff topology. This is called a *scaling limit* result.

**Influence of the seed in affine preferential attachment trees** In Chapter 2 of this dissertation, which is adapted from a common work with Ioan Manolescu [106], we investigate some aspect of the large scale geometry of *affine preferential attachment trees*. Introduced by the physicists Barabási et Albert in [21], the model produces sequences of trees that grow randomly according to a "rich get richer" rule, which also seems intuitively to govern the formation of many real networks. Starting from a finite tree  $S$  with  $k$  vertices, called the *seed tree*, we inductively construct a sequence  $(\mathbf{T}_n^S)_{n \geq k}$  of random trees, where  $\mathbf{T}_{n+1}^S$  is obtained from  $\mathbf{T}_n^S$  by adding a vertex and connect it to a vertex of  $\mathbf{T}_n^S$  chosen at random with a probability proportional to an affine function of its *degree*—that is the number of links already established by the vertex. The form of the affine function—also called *attachment function*—is more precisely  $f(d) = d - 1 + \alpha$  for some fixed parameter  $\alpha > 0$ . Since  $f$  is increasing in  $d$ , vertices with higher current degree are more likely to receive new connections. That is why we speak of a "rich get richer" rule. A step of the mechanism is depicted on Figure 1.9, in the case of a linear attachment (which corresponds to  $\alpha = 1$ ). Displayed numbers are the attachment probabilities to existing vertices.

First mathematical publications on preferential attachment trees focused on questions related to degrees of vertices. It has been shown [33, 115] that the degree distribution of the tree  $\mathbf{T}_n^S$ —which is the sequence made of the proportions of vertices having a given degree—asymptotically follows a power law. It is called a *scale-free property*, which turns out to be shared by a wide range of real networks. The growth rate of the degree of any given vertex was also investigated. In [116], Móri proved that its order was  $n^{\frac{1}{1+\alpha}}$ , as well as that of the current maximal degree in the tree. We refer the reader to [77, Chapter 8] for an extensive overview on these topics.

In [39], Bubeck, Mossel and Rácz raised a new kind of question, that we can formulate as follows : does the *seed tree*  $S$  have any impact on the large scale geometry of  $\mathbf{T}_n^S$ ? The answer is positive if we compare, for any two distinct seed trees  $S, S'$  having at least three vertices, the distribution of  $\mathbf{T}_n^S$  to that of  $\mathbf{T}_n^{S'}$ , with respect to the total variation distance. More precisely, the total variation distance  $\mathbf{d}_{TV}(\mathbf{T}_n^S, \mathbf{T}_n^{S'})$  is lower bounded by a positive constant, uniformly in  $n$ . This result may be understood as follows : given a large (but uniform in  $n$ ) number of samples of  $\mathbf{T}_n^S$ , one may recover  $S$  with high precision. So the seed tree has indeed an impact on the geometry of  $\mathbf{T}_n^S$ , even at large scale.

Bubeck, Mossel and Rácz actually answered the question in the particular case where the two seed trees have a distinct degree distribution. Otherwise said, for a certain  $d \geq 1$ , the number of vertices having degree  $d$  is not the same in  $S$  and  $S'$ . Given the "rich get richer" phenomenon, it is not surprising that an initial imbalance in degree distributions may persist asymptotically. In [54], Curien, Duquesne, Kortchemski and Manolescu solve the general case, but only for a linear attachment. Their proof uses an observable of  $\mathbf{T}_n^S$ —that is a measurable function of the tree—which takes into account the current graph structure, by roughly counting the number of embeddings of some given finite tree inside  $\mathbf{T}_n^S$ , in addition to the degrees of vertices.

Our main contribution, presented in Chapter 2, is the demonstration of the influence of the seed, with respect to the total variation distance, for any *affine* attachment function, not only the *linear* one. Our method is inspired of [54], but is arguably simpler, more natural and turns out to be more robust.



**Percolation on supercritical causal triangulations** In Chapter 3 of this dissertation, which is based on the work [105], we investigate properties of an oriented percolation process performed on some random causal triangulations.

*Causal triangulations* were introduced in [13] by theoretical physicists Jan Ambjørn et Renate Loll as a discrete model of Lorentzian quantum gravity in which space and time play asymmetric roles. Their definition is as follows. Take an infinite plane tree. Add "horizontal" connections between vertices located at the same distance from the root vertex in such a way that it forms cycles. At this point, the map is not a triangulation yet, and is called *causal map*. Then, we triangulate each face by linking vertices on its bottom side—except the rightmost one—to the top right vertex. See Figure 1.13 for an illustration of the construction steps of a causal triangulation from an infinite plane tree.

The mathematical study of random causal triangulation is quite recent and was launched by the two papers [64, 56]. Both examine the geometric properties of causal triangulations derived from a critical Kesten's tree, which is roughly a critical Bienaymé–Galton–Watson tree that one would have conditioned to be infinite [2]. In [56], the authors show for instance that the presence of cycles in the causal triangulation noticeably reduces "horizontal" distances between vertices being at same level, compared to what happens in the underlying tree alone, provided that the offspring distribution of the latter has a finite variance though. In [43], Thomas Budzinski introduces *supercritical causal triangulations*, which are causal triangulations obtained from a supercritical Bienaymé–Galton–Watson tree conditioned to be infinite. Among other things, he proves that the random maps satisfy several metric hyperbolicity properties, more subtle than the classic exponential growth.

In Chapter 3, we deal with supercritical causal triangulations whose the underlying Bienaymé–Galton–Watson tree has a geometric offspring distribution. We subject the map to an oriented percolation process, which is an usual method to grasp its geometry. More precisely, we remove the "horizontal" edges belonging to cycles and direct the other "vertical" edges away from the root vertex. See the top left drawing on Figure 1.14. Then, we apply a Bernoulli percolation process on the remaining edges : each one is either kept with probability  $p \in [0, 1]$  or removed with probability  $1 - p$ , independantly of the others. See now the top right drawing on Figure 1.14. Our main contribution in this chapter is to get a quite complete picture of how the oriented percolation process behaves on the map.

First, we show the existence of a non trivial percolation threshold  $p_c \in (0, 1)$  such that there is no infinite percolation cluster in the map for  $p \leq p_c$ , but infinitely many when  $p_c < p < 1$ . This kind of phase transition is generally met on hyperbolic graphs [80, 121]. Moreover, we compute an exact formula for  $p_c$ , which only depends on the mean of the geometric offspring distribution of the underlying Bienaymé–Galton–Watson tree.

In a second result, we compute several critical exponents of the percolation process. We prove that the percolation probability of the root vertex is of order  $p - p_c$  when  $p \searrow p_c$  and that the tail distribution of the number of vertices which belong to the cluster of the root vertex roughly decreases at polynomial rate  $n^{-1/2}$ .

Finally, we demonstrate that at criticality  $p = p_c$ , large clusters converge in distribution after rescaling to the Continuum Random Tree, with respect to the Gromov–Hausdorff topology.

The main ingredient of the proof is to study the same oriented percolation process on a toy model which approximates and simplifies the structure of causal triangulations. This toy model is a half-plane model, for which we know tools (like peeling process) to analyse the percolation on it [15, 16, 17]. The final step of the demonstration is then to transfer back to the supercritical causal triangulations the results that we have just obtained on the toy model.

## Chapter 2

# Influence of the seed in affine preferential attachment trees

This chapter is adapted from [106], which is a joint work with Ioan Manolescu, published in *Bernoulli*.

We study randomly growing trees governed by the affine preferential attachment rule. Starting with a seed tree  $S$ , vertices are attached one by one, each linked by an edge to a random vertex of the current tree, chosen with a probability proportional to an affine function of its degree. This yields a one-parameter family of preferential attachment trees  $(T_n^S)_{n \geq |S|}$ , of which the linear model is a particular case. Depending on the choice of the parameter, the power-laws governing the degrees in  $T_n^S$  have different exponents.

We study the problem of the asymptotic influence of the seed  $S$  on the law of  $T_n^S$ . We show that, for any two distinct seeds  $S$  and  $S'$ , the laws of  $T_n^S$  and  $T_n^{S'}$  remain at uniformly positive total-variation distance as  $n$  increases.

This is a continuation of [54], which in turn was inspired by a conjecture of [39]. The technique developed here is more robust than previous ones and is likely to help in the study of more general attachment mechanisms.

### Sommaire

---

<b>2.1</b>	<b>Introduction</b>	<b>62</b>
<b>2.2</b>	<b>Planar affine preferential attachment model</b>	<b>65</b>
2.2.1	Definition via corners	65
2.2.2	Decomposition and coupling using planted plane trees	66
<b>2.3</b>	<b>First and second moment of a class of observables</b>	<b>68</b>
2.3.1	A recurrence formula	68
2.3.2	The first moment of $\mathcal{F}_\tau(T_n)$	72
2.3.3	The second moment of $\mathcal{F}_\tau(T_n)$	75
<b>2.4</b>	<b>Observables and their difference around the seed tree</b>	<b>81</b>
2.4.1	Differences only appear around the seed	81
2.4.2	Blind trees: differences appear only in the seed	82
2.4.3	Evaluating the difference in the seed	84
2.4.4	Proof of Theorem 2.4	87
2.4.5	Proof of Theorem 2.1	89
<b>2.5</b>	<b>Open problems and future research</b>	<b>89</b>

---

## 2.1 Introduction

*Linear preferential attachment trees* - or *Barábasi-Albert trees* - are trees that grow randomly according to the following simple mechanism. Start with a finite tree  $S$  with  $k$  vertices - we call it the *seed tree* and  $k$  its size - and construct inductively a sequence  $(T_n^S)_{n \geq k}$  of random trees, where  $T_{n+1}^S$  is obtained from  $T_n^S$  by adding a new vertex and connecting it to a vertex of  $T_n^S$  chosen at random with a probability proportional to its degree. Thus the tree  $T_{n+1}^S$  has one more vertex than  $T_n^S$  - namely  $n + 1$  - and one more edge - namely  $n$ .

The reader might note that this model formalises the adage "the rich get richer" since vertices with high degrees are more likely to receive new connections. This property suits a large class of real networks (see [21]) mainly because of the emergence of power laws for the sequence of degrees of the tree as its size grows, see [33]. Such laws are observed in a wide range of contexts, like social networks (webgraph, citation graph, etc [142], [117]) or even in biological networks such as interaction protein networks (see [111]). The degree distribution in linear preferential attachment trees has been deeply investigated; for an extensive overview of these traditional topics in the preferential attachment model's analysis, the reader is directed to [77, Ch. 8].

In the literature, there are numerous variations in the definition of preferential attachment model. Here we will consider one where the new vertex is attached to an old vertex chosen with probability which is an affine function of its degree, rather than a linear one. Fix a parameter  $\alpha > 0$  and  $S$  a seed tree of size  $k \geq 2$ . The  $\alpha$ -PA trees grown from the seed  $S$  are denoted by  $(T_n^S)_{n \geq k}$  and are defined recursively as follows. Given  $(T_n^S)_{k \leq n \leq N}$  for some  $N \geq k$  with  $T_N^S = T$  for some fixed tree  $T$ , select randomly a vertex  $u_N$  in  $T$  with probability:

$$\mathbb{P}(u_N = u \mid T_k^S, \dots, T_N^S \text{ with } T_N^S = T) = \frac{\deg_T(u) - 1 + \alpha}{(1 + \alpha)N - 2} \quad \forall u \in T, \quad (2.1.1)$$

and link it to a new vertex by an edge. The resulting tree is  $T_{N+1}^S$ .

Note that  $(1 + \alpha)N - 2 = \sum_{u \in T} \deg_T(u) - 1 + \alpha$  for all trees  $T$  of size  $N$ , so (2.1.1) defines indeed a probability measure. This model, which was first introduced in [115], is a generalisation of the Barábasi-Albert model, in that the latter is obtained when choosing  $\alpha = 1$ . The authors of [25] remarked that transitions probabilities (2.1.1) provide a Pólya urn representation of  $\alpha$ -PA trees; particularly convenient properties follow, such as estimates on the degree growth and a form of exchangeability for the sequence  $(u_N)_{N \geq k}$ . When the dependence on  $S$  is not important, we will drop  $S$  from the notation.

**Henceforth  $\alpha > 0$  is fixed.**

The same questions as for the linear model may be asked of the affine one. For instance, one may study the growth of degrees of given vertices in  $T_n$ . It may be proved (see [77, Ch. 8.3 & 8.4] or Remark 2.6 of this paper) that the degree of any vertex of the seed increases at polynomial speed, with an exponent that depends on  $\alpha$  and varies between 1 and 0 as  $\alpha$  ranges from 0 to infinity. It may appear surprising that an apparently insignificant difference in the attachment mechanism leads to different scaling exponents in the power laws governing the degree sequence, and hence to preferential attachment models of different universality classes. More substantial variations of the model will be discussed in Section 2.5.

The problem of interest in this paper is that of the recognition of the seed. Precisely, we will study whether the seed tree has any influence on the law of the tree obtained after a large number of iterations of the growth procedure. This question was asked (and partly solved) for the linear model in [39]. The complete answer for the linear model was obtained in [54], and for the uniform preferential attachment model in [38]. In both cases, the seed is shown to influence the asymptotic law of the model. Our aim is to generalise the result to the  $\alpha$ -PA, of which the uniform model may be perceived as a limit.

**Theorem 2.1** *Let  $S$  and  $S'$  be two finite trees of size  $k_1, k_2 \geq 3$ . Then for any  $\alpha > 0$ , the following limit:*

$$d(S, S') = \lim_{n \rightarrow +\infty} d_{TV}(T_n^S, T_n^{S'}) \quad (2.1.2)$$

*exists and is non zero when  $S$  is not graph-isomorphic to  $S'$ .*

As a consequence of Theorem 2.1, the function  $d$  is a metric on finite trees with at least 3 vertices. It measures the statistical chance of distinguishing between two different seed trees given observations of the  $\alpha$ -PA trees grown from them.

**Remark 2.1** *The existence of the limit in Theorem 2.1 stems from the fact that the sequence  $d_{TV}(T_n^S, T_n^{S'})$  is decreasing in  $n$ , which may be proved by a coupling argument.*

In proving Theorem 2.1, we will look at the total variation between  $T_n^S$  and  $T_n^{S'}$  through some integer-valued observable obtained from them. The same is done in all previous studies [39, 54, 38].

In [39], where the model is linear (that is  $\alpha = 1$ ) and  $S$  and  $S'$  are assumed to have different degree sequences, the authors use as observable the maximum degree. Indeed, they show that the maximal degrees of  $T_n^S$  and  $T_n^{S'}$  have different tail distributions, which then easily leads to  $d(S, S') > 0$ . The key to this argument is that the degrees of vertices in the Barabási-Albert model evolve as a Pólya urn. When the degree distributions of  $S$  and  $S'$  are identical, but their geometry is different, the observable of [39] is unable to distinguish between  $T_n^S$  and  $T_n^{S'}$ . To overcome this difficulty, a more complex observable was introduced in [54]: the number of embeddings of a fixed finite tree  $\tau$  inside  $T_n^S$ , weighted by some function of the degrees in  $T_n^S$  of the embedding of  $\tau$ . The same type of observable was used in [38] and will be used below.

For  $n \geq 0$  and  $d \geq 1$ , set  $[n]_d = n(n-1)\dots(n-d+1)$ . Also set  $[n]_0 = 1$  for any  $n \geq 0$ . A *decorated tree* is a couple  $\tau = (\tau, \ell)$  where  $\tau$  is a tree and  $\ell$  is function from the vertices of  $\tau$  to the set of non-negative integers. For a decorated tree  $\tau$  and a (bigger) tree  $T$ , set

$$\mathcal{F}_\tau(T) := \sum_{\phi: \tau \hookrightarrow T} \prod_{u \in \tau} [\deg_T(\phi(u)) - 1]_{\ell(u)}, \quad (2.1.3)$$

where the sum is over all graph embeddings of  $\tau$  in  $T$  and the product is over all vertices of  $\tau$ .

For two distinct seeds  $S$  and  $S'$ , we aim to show that, for a well chosen decorated tree  $\tau$ , the difference of the expectations of  $\mathcal{F}_\tau(T_n^S)$  and  $\mathcal{F}_\tau(T_n^{S'})$  is of the same order as each of them and as their standard deviation. This allows to control the total variation between  $T_n^S$  and  $T_n^{S'}$  using the following bound.

**Lemma 2.1** *For any two real-valued, square-integrable random variables  $X$  and  $Y$ ,*

$$d_{TV}(X, Y) \geq \frac{(\mathbb{E}[X] - \mathbb{E}[Y])^2}{(\mathbb{E}[X] - \mathbb{E}[Y])^2 + 2 \cdot (\mathbb{E}[X^2] + \mathbb{E}[Y^2])}.$$

**Proof of Lemma 2.1** Let  $(X', Y')$  be a coupling of the random variables  $X$  and  $Y$ . By using Paley-Zigmond's inequality, then Jensen's one, we get:

$$\mathbb{P}(X' \neq Y') \geq \frac{(\mathbb{E}[X'] - \mathbb{E}[Y'])^2}{\mathbb{E}[(X' - Y')^2]}.$$

Furthermore, a simple decomposition gives:

$$\begin{aligned} \mathbb{E}[(X' - Y')^2] &= \mathbb{E}[(X' - \mathbb{E}[X'] + \mathbb{E}[X'] - \mathbb{E}[Y'] + \mathbb{E}[Y'] - Y')^2] \\ &\leq 2 \cdot (\text{Var}(X') + \text{Var}(Y')) + (\mathbb{E}[X'] - \mathbb{E}[Y'])^2 \\ &\leq 2 \cdot (\mathbb{E}[(X')^2] + \mathbb{E}[(Y')^2]) + (\mathbb{E}[X'] - \mathbb{E}[Y'])^2. \end{aligned}$$

It finally remains to note that  $X'$ , resp.  $Y'$ , has the same moments as  $X$ , resp.  $Y$ , and that the total variation between  $X$  and  $Y$  is obtained by taking the infimum of  $\mathbb{P}(X' \neq Y')$  over all couplings  $(X', Y')$  of  $X$  and  $Y$ .

It comes to no surprise that the evolution of the moments of  $\mathcal{F}_\tau(T_n^S)$  is of great importance for the proof. Let us fix  $\tau$  and  $S$  and discuss briefly the first moment of  $\mathcal{F}_\tau(T_n^S)$  as  $n \rightarrow \infty$ . There are two competing factors contributing to  $\mathcal{F}_\tau(T_n^S)$ . First, due to the "rich gets richer" phenomenon, the "oldest" vertices of  $T_n^S$  have degree evolving as  $n^{\frac{1}{1+\alpha}}$ . Thus, the contribution to  $\mathbb{E}[\mathcal{F}_\tau(T_n^S)]$  of embeddings included in the seed (or close to it) is of the order  $n^{\frac{|\ell|}{1+\alpha}}$  for  $|\ell| = \sum_{v \in \tau} \ell(v)$ .

Second, one should take into account the embeddings using recently acquired vertices. There are many such vertices, but they have small degrees. Depending on the form of  $\tau$ , such embeddings contribute to  $\mathbb{E}[\mathcal{F}_\tau(T_n^S)]$  by a quantity that may exceed  $n^{\frac{|\ell|}{1+\alpha}}$  by some logarithmic or even polynomial factor. Thus we find

$$\mathbb{E}[\mathcal{F}_\tau(T_n)] \approx \log^{\gamma(\tau)}(n) \cdot n^{\frac{\lambda(\tau)}{1+\alpha}}, \quad (2.1.4)$$

for some  $\gamma(\tau) \geq 0$  and  $\lambda(\tau) \geq |\ell|$ . Both values  $\gamma(\tau)$  and  $\lambda(\tau)$  are computed explicitly in Section 2.3.

Due to the coupling between  $\alpha$ -PA started from different seeds (see Section 2.2), only the first type of embeddings is sensitive to the seed. Thus, for two distinct seeds  $S, S'$ , we may expect that

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})] \approx n^{\frac{|\ell|}{1+\alpha}}, \quad (2.1.5)$$

provided that a difference exists. Let us, for the sake of this explanation, assume that the standard deviations of  $\mathcal{F}_\tau(T_n^S)$  and  $\mathcal{F}_\tau(T_n^{S'})$  are of the same order as their expectations. Then, in order to successfully apply Lemma 2.1, we should have  $\gamma = 0$  and  $\lambda(\tau) = |\ell|$ .

Such asymptotics (slightly weaker actually) were already observed in [54] for  $\alpha = 1$ ; they were proved using an amenable recursive relation for  $\mathbb{E}[\mathcal{F}_\tau(T_n)]$ . Rather than showing that  $\tau$  may be chosen to satisfy (2.1.5) and so that  $\gamma(\tau) = 0$  and  $\lambda(\tau) = |\ell|$ , the authors of [54] constructed a linear combination over trees  $\tau$  of observables  $\mathcal{F}_\tau(T_n)$  for which the logarithmic factors cancel out. The resulting observable, properly rescaled, turns out to be a martingale that is bounded in  $L^2$  and whose expectation depends on the seed tree.

We will employ a different, arguably simpler strategy: we will prove that for any two distinct seeds  $S \neq S'$ , the decorated tree  $\tau$  may be chosen so as to observe a difference as in (2.1.5) and such that  $\gamma(\tau) = 0$  and  $\lambda(\tau) = |\ell|$ . As such, our strategy is more likely to apply to other attachment mechanisms, as will be discussed in Section 2.5. Additional differences with [54], are the recurrence formula used to prove (2.1.4), which is more complex in the affine case, and the fact that the exponent  $\lambda(\tau)$  is not always equal to  $|\ell|$ , as opposed to the linear case, where  $\lambda(\tau) = |\ell|$  always.

Let us finally mention that the exact definition (2.1.3) of the observable is somewhat arbitrary. Indeed, it is also possible to use slight variations instead, such as  $\sum_{\phi: \tau \hookrightarrow T} \prod_{u \in \tau} (\deg_T \phi(u))^{\ell(u)}$ . We chose (2.1.3) as it is inspired by the equivalent construction in [54] and has an interpretation in the planar version of the model (see Section 2.2).

**Organisation of the paper** In Section 2.2, we introduce a planar version of the affine preferential attachment model. This construction is not formally necessary to prove our main result, but we believe it is interesting in its own right and helps clarify the subsequent proofs. In particular, the planar version will give rise to a natural coupling between two  $\alpha$ -PA trees starting from two different seeds of same size.

In Section 2.3 we study the asymptotics of the first and second moments of  $\mathcal{F}_\tau(T_n^S)$  as  $n \rightarrow \infty$ . The only role of this section within the proof of Theorem 2.1 is to show a bound on the second moment of  $\mathcal{F}_\tau(T_n^S)$  for certain trees  $\tau$ . These estimates use a recurrence relation for  $\mathcal{F}_\tau(T_n^S)$ , which is similar to that of [54], but more complicated due to the affine probabilities.

Section 2.4 contains lower bounds on the first-moment difference of  $\mathcal{F}_\tau(T_n^S)$  for two distinct seeds  $S$ . Precisely, we prove (2.1.5) for well-chosen decorated trees  $\tau$ . This is then used to prove Theorem 2.1.

Finally, certain variations of the model and potential extensions of our result are discussed in Section 2.5.

**Notation** In the rest of the paper, we will use the following notations:

- for two sequences  $f, g : \mathbb{N} \rightarrow (0, +\infty)$ , write  $f(n) \approx g(n)$  if there exists some constant  $C > 0$  such that  $\frac{1}{C} \leq \frac{f(n)}{g(n)} \leq C$ ;
- also  $f(n) = \mathcal{O}(g(n))$  if  $\limsup_{n \rightarrow +\infty} \left| \frac{f(n)}{g(n)} \right| < +\infty$ ;
- and  $f(n) \ll g(n)$  if  $\lim_{n \rightarrow +\infty} \left| \frac{f(n)}{g(n)} \right| = 0$ .

- The size of a tree  $T$  is written  $|T|$  and stands for the number of vertices of  $T$ .
- For a graph  $S$ , write  $V_S$  and  $E_S$  for its sets of vertices and edges, respectively. For  $v \in V_S$ , write  $\deg_S(v)$  or  $\deg(v)$  for the degree of  $v$  in  $S$ .

**Acknowledgements** The authors thank Nicolas Curien for helpful discussions and corrections on an earlier version of this paper.

The first author is supported by a PhD Fellowship from the MESR, France and was supported in the early stages of this project by the NCCR SwissMAP. The second author is a member of the NCCR SwissMAP.

## 2.2 Planar affine preferential attachment model

In this section, we first define the planar version of affine preferential attachment trees, then we present a useful coupling between any two  $\alpha$ -PA trees derived from different seed trees of same size. The coupling is based on the decomposition of  $T_n^S$  into *planted plane trees*, which isolates the roles of the seed and of the growth mechanism, respectively.

We reiterate that this planar version is not formally necessary for any of the results; it is used merely to illustrate the arguments.

### 2.2.1 Definition via corners

We first need to introduce the notion of *colouring of corners* in a plane tree. A plane tree is a tree embedded in the plane up to continuous deformation, or equivalently a tree whose vertices are equipped with a cyclic order of their neighbours. A *corner* of a plane tree is an angular sector of the plane contained between two consecutive half-edges around a vertex. In particular, the number of corners surrounding a vertex is equal to its degree.

**Definition 2.1 (Colouring of corners)** *A colouring of the corners of a plane tree  $\mathcal{T}$  is a choice of one distinguished corner for each vertex of the tree. The distinguished corners are called “red”; all other corners are called “blue”.*

*When indexing  $\{c_{v,i} : v \in V_{\mathcal{T}}, 1 \leq i \leq \deg(v_i)\}$  the corners of  $\mathcal{T}$ , we will always suppose that the red vertices are  $(c_{v,1})_{v \in V_{\mathcal{T}}}$ .*

It will be useful to keep in mind that any plane tree with  $n$  vertices and coloured corners has  $n$  red corners and  $n - 2$  blue ones. We are now ready to introduce the planar version of the affine preferential attachment model.

**Definition 2.2** *Fix a parameter  $\alpha > 0$  and a plane tree  $S$  - namely the seed tree - of size  $|S| = k \geq 2$  endowed with a colouring of its corners. The planar  $\alpha$ -PA with seed  $S$  is a growing sequence of random plane trees  $(T_n^S)_{n \geq k}$  (or  $(T_n)_{n \geq k}$  when the role of  $S$  is clear) with coloured corners, evolving according to the following inductive principle:*

- (i)  $T_k = S$ ;
- (ii) *Suppose that our sequence of plane trees is built until step  $n \geq k$ . Then, independently from the past of the process, select at random a corner  $c_n$  of  $T_n$  with*

$$\mathbb{P}(c_n = c_{u,1}) = \frac{\alpha}{n(1+\alpha) - 2} \quad \text{and} \quad \mathbb{P}(c_n = c_{u,i}) = \frac{1}{n(1+\alpha) - 2} \quad \forall 2 \leq i \leq \deg_{T_n}(u),$$

*for any vertex  $u \in V_{T_n}$ . That is, red corners are chosen with probability proportional to  $\alpha$  and blue corners with probability proportional to 1.*

- (iii) *Write  $u_n$  for the vertex of  $T_n$  of which  $c_n$  is a corner. To obtain  $T_{n+1}$  add a new vertex  $v_{n+1}$  to  $T_n$  and we connect it by an edge to  $u_n$  through the corner  $c_n$ .*
- (iv) *All corners of  $T_n$  except  $c_n$  subsist in  $T_{n+1}$  and conserve their colour. The unique corner  $c'_n$  of the new vertex  $v_{n+1}$  is coloured red, as required by Definition 2.1. The randomly selected corner  $c_n$  is divided in two angular areas by the edge connecting  $u_n$  to  $v_{n+1}$ . If  $c_n$  was blue in  $T_n$ , then both these corners of  $T_{n+1}$  are blue. If  $c_n$  was red in  $T_n$ , we arbitrarily colour the corner to the right of the edge  $(u_n, v_{n+1})$  in red and the other in blue.*

The probabilities appearing in step (ii) indeed sum to 1 since:

$$\underbrace{n}_{\text{number of red corners}} \cdot \frac{\alpha}{n \cdot (1 + \alpha) - 2} + \underbrace{(n - 2)}_{\text{number of blue corners}} \cdot \frac{1}{n \cdot (1 + \alpha) - 2} = 1$$

Observe that, the attachment principle in the planar  $\alpha$ -PA is such that the probability for  $v_{n+1}$  to be attached to a vertex  $u$  of  $T_n$  is  $\frac{\deg_{T_n}(u) - 1 + \alpha}{n \cdot (1 + \alpha) - 2}$ . Thus, we obtain :

**Lemma 2.2** *If  $(T_n)_{n \geq k}$  is a planar  $\alpha$ -PA started from some plane seed tree  $S$ , then the trees  $(T_n)_{n \geq k}$  (stripped of their planar embedding) have the law of a  $\alpha$ -PA.*

**Remark 2.2** *Changing the rules of the assignment of labels and colours to the new corners of a planar  $\alpha$ -PA affects the law of the plane tree, but only via its embedding.*

*In the special case  $\alpha = 1$ , the weights of red and blue corners are equal, and the planar  $\alpha$ -PA is simply a uniform embedding of the abstract  $\alpha$ -PA. This is not the case when  $\alpha \neq 1$ . For more on this see [54, Conj. 2].*

As mentioned in the introduction, the observables  $\mathcal{F}_\tau$  have a particular interpretation in terms of the planar model. Let  $\tau = (\tau, \ell)$  be a decorated tree and  $T$  be a plane tree with coloured corners. Then, for any embedding  $\phi : \tau \hookrightarrow T$ , the factor  $\prod_{u \in \tau} [\deg_T(\phi(u)) - 1]_{\ell(u)}$  is the number of ordered choices of  $\ell(u)$  different blue corners around each vertex  $\phi(u)$  for  $u \in \tau$ . As in [54], one may imagine that each vertex  $u \in \tau$  is endowed with  $\ell(u)$  distinct arrows. Call a decorated embedding an embedding of  $\phi$  of  $\tau$  in  $T$  together with, for each  $u \in \tau$  and each arrow of  $u$ , a blue corner of  $\phi(u)$  to which that arrow points, in such a way that distinct arrows point to distinct corners. Then  $\mathcal{F}_\tau(T)$  is the number of decorated embeddings of  $\tau$  in  $T$ .

## 2.2.2 Decomposition and coupling using planted plane trees

Apart from its intrinsic interest, the decomposition described below will be used in Section 2.4. We begin with the definition of a planted plane tree:

**Definition 2.3** *A planted plane tree  $T_{\circ}$  is a plane tree  $T$  with a distinguished vertex called the root and an additional half-edge emerging from the root. This half edge divides the corner of the root delimited by the two edges linking the root to its leftmost and its rightmost children. When  $|T| = 1$ , the planted plane tree  $T_{\circ}$  is merely a single vertex with an half-edge attached to it.*

It should be noted that there is one more corner in any planted version  $T_{\circ}$  of a plane tree  $T$ . A colouring of corners for a planted plane tree  $T_{\circ}$  is defined as for plane trees (see Definition 2.1) with the exception that the root is allowed to have no red corners or one red corner, which will always be the corner to the right of the half-edge. In the former situation, we say that  $T_{\circ}$  is a **blue planted plane tree**, in the latter we say it is a **red planted plane tree**.

Since the recursive procedure used to define the planar  $\alpha$ -PA tree is simply based on the colouring of corners, we straightforwardly adapt it to define a planted planar version of the same model. Note by  $T_n^{\circ b}$ , resp.  $T_n^{\circ r}$ , a blue planted plane tree, resp. a red planted plane tree, obtained through the preferential attachment algorithm with initial condition  $\circ$ , which is the tree composed of a single vertex with a half-edge attached to it and a blue corner, resp. a red corner, surrounding it.

Let  $S$  be a plane seed tree of size  $k \geq 2$  with coloured corners indexed  $\{c_{v,i} : v \in V_S, 1 \leq i \leq \deg_S(v)\}$ . Fix  $n \geq k$  and  $T_n$  be a realisation of the planar  $\alpha$ -PA tree at step  $n$  starting from  $S$ . For  $v \in V_S$  and  $1 \leq i \leq \deg_S(v)$ , say that a vertex  $u \in V_{T_n} \setminus V_S$  is a descendent of the corner  $c_{v,i}$  if the unique path linking  $u$  to  $S$  arrives at  $S$  through  $c_{v,i}$ . Additionally, say that  $v$  is also a descendent of  $c_{v,i}$ . The tree  $T_n^{v,i}$  stemming from  $c_{v,i}$  is the planted plane tree containing all vertices of  $T_n$  that are descendent of  $c_{v,i}$  and all edges of  $T_n$  between such vertices; it is rooted at  $v$ . The half-edge attached to  $v$  is such that it does not split the corner  $c_{v,i}$ . Call these trees the planted plane subtrees of  $T_n^S$ . See Figure 2.1 for an illustration.

**Proposition 2.1** *Let  $S$  be a plane seed tree of size  $k \geq 2$  with coloured corners indexed  $\{c_{v,i} : v \in V_S, 1 \leq i \leq \deg_S(v)\}$ . For  $n \geq k$ , let  $T_n^{v,i}$  be the planted plane tree stemming from  $c_{v,i}$  and let  $k_n^{v,i}$  denote its size. Write  $x_n^{v,1} = (1 + \alpha)k_n^{v,1} - 1$  and  $x_n^{v,i} = (1 + \alpha)k_n^{v,i} - \alpha$  for all  $v \in V_S$  and  $i \geq 2$ . Then*

- *the vector  $(x_n^{v,i})_{v \in V_S, 1 \leq i \leq \deg_S(v)}$  has the distribution of a Pólya urn with  $2k - 2$  colours and diagonal replacement matrix  $(1 + \alpha)I_{2k-2}$ , starting from  $(\alpha \mathbb{1}_{\{i=1\}} + \mathbb{1}_{\{i \neq 1\}})_{v \in V_S, 1 \leq i \leq \deg_S(v)}$ ;*
- *conditionally on  $(x_n^{v,i})_{v \in V_S, 1 \leq i \leq \deg_S(v)}$ , the trees  $(T_n^{v,i})_{v \in V_S, 1 \leq i \leq \deg_S(v)}$  are independent, with  $T_n^{v,i}$  having law  $T_{k_n^{v,i}}^{-or}$  if  $i = 1$  and  $T_{k_n^{v,i}}^{-ob}$  if  $i \neq 1$ .*

Proposition 2.1 may be restated as follows. Given a plane seed  $S$  with  $|S| = k$ , we would like to construct  $T_n^S$  for some  $n \geq k$ . This may be done in the following steps.

1. Generate a vector  $(x_n^{v,i})_{v \in V_S, 1 \leq i \leq \deg_S(v)} \in \mathbb{R}^{2k-2}$  with the law of the Pólya urn of Proposition 2.1. Define sizes  $k_n^{v,1} := \frac{x_{\ell,p}^{(b)} + \alpha}{1 + \alpha}$  and  $k_n^{v,i} := \frac{x_i^{(r)} + 1}{1 + \alpha}$  for  $v \in V_S$  and  $i \geq 2$ .
2. Randomly draw independent realisations  $\tau_n^{v,1}$  of  $T_{k_n^{v,1}}^{-or}$  and  $\tau_n^{v,i}$  of  $T_{k_n^{v,i}}^{-ob}$  for each  $v \in V_S$  and  $i \geq 2$ .
3. Graft each tree  $\tau_n^{v,i}$  in the corner  $c_{v,i}$  of  $S$ .

The resulting tree has the law of a planar  $\alpha$ -PA started from  $S$ .

This decomposition allows to couple the evolution of two planar  $\alpha$ -PA trees emerging from distinct seeds of same size. Indeed, in spite of the notation, the first two steps do not depend on  $S$ , only on  $k = |S|$ : in the first step, the vector  $(x_n^{v,i})$  contains  $2k - 2$  entries and the Pólya urn generating it starts with  $k$  entries  $\alpha$  and  $k - 2$  entries 1; in the second step,  $k$  trees are of type  $T^{-or}$  and  $k - 2$  of type  $T^{-ob}$ . If  $S, S'$  are two seeds of same size  $k$ , then  $T_n^S$  and  $T_n^{S'}$  are coupled as follows. Simulate a common vector for the sizes of the planted plane subtrees of  $T_n^S$  and  $T_n^{S'}$ , then simulate a common set of planted plane trees. Finally graft these trees to  $S$  to obtain  $T_n^S$  and the same trees to  $S'$  to obtain  $T_n^{S'}$ .

**Proof of Proposition 2.1** Fix  $S$ . Let us first prove that the vector  $(x_n^{v,i})_{v \in V_S, 1 \leq i \leq \deg_S(v)}$  has the distribution of the Pólya urn described in the statement. We do this by induction on  $n$ . When  $n = k$ , all planted plane subtrees of  $T_k^S = S$  have size 1. The formula relating the sizes of the subtrees to  $x_n^{v,i}$  then yields  $x_n^{v,1} = \alpha$  and  $x_n^{v,i} = 1$  for  $i \geq 2$ .

Suppose now that the statement is proved up to step  $n$  and let  $T_n^S$  be a realisation of the planar  $\alpha$ -PA tree of size  $n$ . Then, each red planted plane subtree  $T_n^{v,1}$  has  $k_n^{v,1}$  red corners and  $k_n^{v,1} - 1$  blue ones. Each blue planted plane subtree  $T_n^{v,i}$  with  $i \geq 2$  has  $k_n^{v,i} - 1$  red corners and  $k_n^{v,i}$  blue corners. It follows that, for each  $v$

$$\begin{aligned} \mathbb{P}(v_{n+1} \text{ descendent of } c_{v,1}) &= \frac{(\alpha + 1)k_n^{v,1} - 1}{2n - 2} = \frac{x_n^{v,1}}{2n - 2} \quad \text{and} \\ \mathbb{P}(v_{n+1} \text{ descendent of } c_{v,i}) &= \frac{(\alpha + 1)k_n^{v,i} - \alpha}{2n - 2} = \frac{x_n^{v,i}}{2n - 2} \quad \text{for all } i \geq 2. \end{aligned}$$

Finally, if  $v_{n+1}$  is attached to some descendant of a corner  $c_{v,i}$ , then  $x_{n+1}^{v,i} = 1 + \alpha + x_n^{v,i}$ . This shows that the vector  $(x_{n+1}^{v,i})$  also has the claimed distribution.

Let us now condition on  $(x_n^{v,i})_{v,i}$ , or equivalently on  $(k_n^{v,i})_{v,i}$  for some  $n$ . Then, for each  $v \in V_S$  and  $i \leq \deg_S(v)$ , the  $k_n^{v,i}$  vertices of  $T_n^{v,i}$ , independently of the times they join  $T_k, \dots, T_n$ , attach themselves to the planted plane subtree of the corner  $c_{v,i}$  with a (conditional) distribution that assigns to any red corner a weight proportional to  $\alpha$  and to any blue corner a weight proportional to 1. It follows readily that  $T_n^{v,i}$  has the law of  $T_{k_n^{v,i}}^{-or}$  or  $T_{k_n^{v,i}}^{-ob}$ , depending on whether  $c_{v,i}$  is red or blue, respectively. It is also immediate, that the resulting trees in different corners are independent<sup>1</sup>.

---

1. This independence only holds *conditionally on the size vector*  $(k_n^{v,i})_{v,i}$ . Moreover, we do not claim that the evolution between step  $k$  and  $n$  is independent in different corners, only the resulting subtrees are.



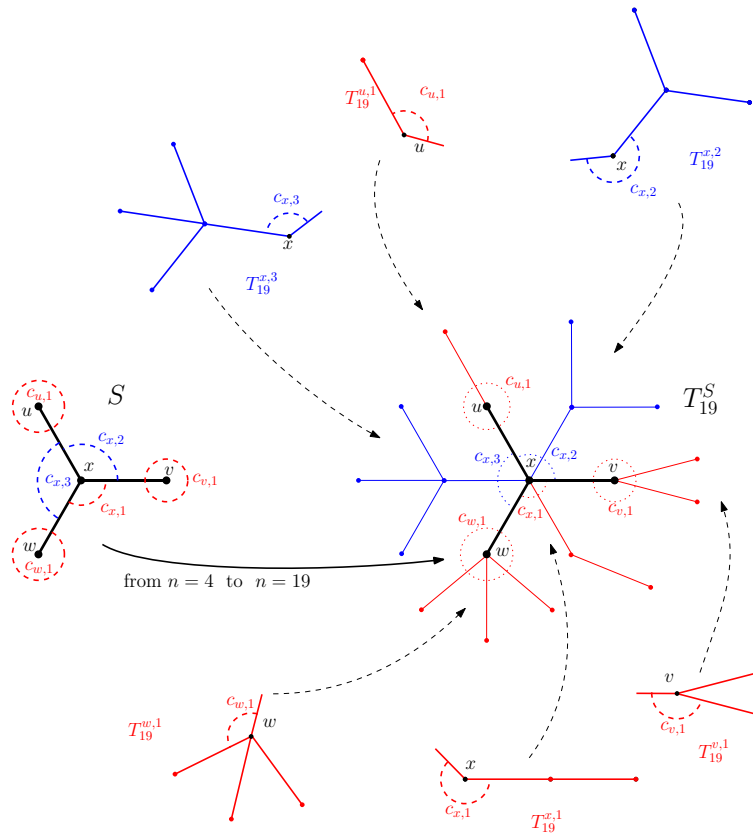


Figure 2.1: An example of decomposition into red and blue planted plane tree.

## 2.3 First and second moment of a class of observables

The ultimate goal of this section is to obtain precise estimates on the second moment of  $\mathcal{F}_\tau(T_n)$  for particular decorated trees  $\tau$ . In doing so we will also prove a general result on the first moment of  $\mathcal{F}_\tau(T_n)$  for any  $\tau$ . The latter is longer to state, and is deferred to later in the section. Below is the minimal result required in the proof of Theorem 2.1.

**Theorem 2.2** *Let  $\tau$  be a finite decorated tree with  $\ell(u) \geq 2$  for any  $u \in \tau$  and such that  $|\ell| = \sum_{u \in \tau} \ell(u) > 1 + \alpha$ . For any seed  $S$ , we have the following asymptotic as  $n \rightarrow \infty$ :*

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S)^2] \approx n^{\frac{2|\ell|}{1+\alpha}}.$$

To prove the above, we will proceed in several steps, each occupying a subsection below. First we prove a recurrence relation on the first moment of our observables (see Section 2.3.1), which is then used in Section 2.3.2 to obtain an accurate estimate of  $\mathbb{E}[\mathcal{F}_\tau(T_n^S)]$  as  $n$  goes to infinity (see Theorem 2.3). Finally, in Section 2.3.3, we use the previous results to prove Theorem 2.2.

### 2.3.1 A recurrence formula

Recall from the introduction that the asymptotic of  $\mathbb{E}[\mathcal{F}_\tau(T_n^S)]$  as  $n \rightarrow \infty$  is polynomial with an exponent depending on  $\tau$ . This exponent will be determined by the weight of  $\tau$  defined below.

**Definition 2.4 (Weight of decorated trees)** *Let  $\tau$  be a decorated tree. If  $\tau$  has size 1 or 2 and all decorations of its vertices are 0, then the weight of  $\tau$  is 1. In all other cases, the weight of  $\tau$  is*

$$w(\tau) = \sum_{v \in V_\tau} \ell(v) + \mathbb{1}_{\{\deg_\tau(v)=1 \text{ and } \ell(v)=0\}}.$$

*Call a vertex  $v \in \tau$  with  $\ell(v) = 0$  and  $\deg_\tau(v) = 1$  a loose leaf of  $\tau$ . Then the weight of  $\tau$  is  $|\ell|$  plus the number of loose leaves.*

Definition 2.4 induces a partial order on the set of decorated trees.

**Definition 2.5 (Partial order on decorated trees)** For any two decorated trees  $\tau = (\tau, \ell)$ ,  $\tau' = (\tau', \ell')$  we write  $\tau' \prec \tau$  if and only if:

- $w(\tau') < w(\tau)$  and  $|\tau'| \leq |\tau|$ ;
- or if  $w(\tau') = w(\tau)$  and  $|\tau'| < |\tau|$ ;
- or if  $w(\tau') = w(\tau)$ ,  $|\tau'| = |\tau|$  and  $\sum_{v \in V_{\tau'}} \ell'(v) = |\ell'| < |\ell| = \sum_{v \in V_{\tau}} \ell(v)$ .

Clearly,  $\prec$  is a strict partial order on the set of decorated trees. We denote by  $\preceq$  the associated partial order.

There are three trees  $\tau$  of weight 1: those with a single vertex and decoration 0 or 1 (they are denoted by  $\textcircled{0}$  and  $\textcircled{1}$ , respectively) and that with two vertices and decorations 0 for both of them (denoted by  $\textcircled{0} - \textcircled{0}$ ).

**Proposition 2.2** There exists a family of nonnegative real numbers  $\{c(\sigma, \tau) \mid \sigma \prec \tau\}$  such that, for every seed  $S$ , every decorated tree  $\tau$  with  $w(\tau) > 1$  and every  $n \geq |S|$ :

$$\mathbb{E}[\mathcal{F}_{\tau}(T_{n+1}^S) | \mathcal{E}_n] = \left(1 + \frac{w(\tau)}{(1+\alpha)n-2}\right) \cdot \mathcal{F}_{\tau}(T_n^S) + \frac{1}{(1+\alpha)n-2} \sum_{\sigma \prec \tau} c(\sigma, \tau) \cdot \mathcal{F}_{\sigma}(T_n^S), \quad (2.3.1)$$

where  $\mathcal{E}_n$  is the  $\sigma$ -algebra generated by  $T_k^S, \dots, T_n^S$ . In addition, when:

- $\tau = \textcircled{0}$ ,  $\mathcal{F}_{\tau}(T_n) = n$ ;
- $\tau = \textcircled{1}$ ,  $\mathcal{F}_{\tau}(T_n) = n - 2$ ;
- $\tau = \textcircled{0} - \textcircled{0}$ ,  $\mathcal{F}_{\tau}(T_n) = 2n - 2$ .

It is worth mentioning that the constants  $\{c(\sigma, \tau) \mid \sigma \prec \tau\}$  do not depend on  $n$  or  $S$ , but do depend on the parameter  $\alpha$  of the model.

The proof below is somewhat algebraic and does not use the planar  $\alpha$ -PA. A more visual proof that uses the notion of decorated embedding (as described at the end of Section 2.2.1) may be given. It is similar to that of [54, Lem. 6] with some additional difficulties due to the colouring of corners.

**Proof** Fix the seed tree  $S$  and drop it from the notation. First let us prove the three particular cases.

- For  $\tau = \textcircled{0}$ , then  $\mathcal{F}_{\tau}(T_n)$  is simply the number of vertices of  $T_n$ , hence is equal to  $n$ .
- For  $\tau = \textcircled{1}$ , then  $\mathcal{F}_{\tau}(T) = \sum_{u \in T} \deg_T(u) - 1 = 2|E_T| - |V_T| = |V_T| - 2$ . Hence  $\mathcal{F}_{\tau}(T_n) = n - 2$ .
- For  $\tau = \textcircled{0} - \textcircled{0}$ , then  $\mathcal{F}_{\tau}(T) = 2|E_T|$  since the  $\tau$  may be embedded over any edge of  $T$  in one of two directions. Hence  $\mathcal{F}_{\tau}(T_n) = 2n - 2$ .

Let us now prove the recurrence formula. Fix  $\tau = (\tau, \ell)$  a decorated tree with  $w(\tau) > 1$ . For a tree  $T$  and an embedding  $\phi : \tau \hookrightarrow T$ , let  $\pi(\tau, \phi, T) = \prod_{u \in \tau} [\deg_T(\phi(u)) - 1]_{\ell(u)}$ , so that

$$\mathcal{F}_{\tau}(T) = \sum_{\phi: \tau \hookrightarrow T} \prod_{u \in \tau} [\deg_T(\phi(u)) - 1]_{\ell(u)} = \sum_{\phi: \tau \hookrightarrow T} \pi(\tau, \phi, T). \quad (2.3.2)$$

Recall that, in passing from  $T_n$  to  $T_{n+1}$ , a new vertex  $v_n$  is attached to a randomly chosen vertex  $u_n$  of  $T_n$ . Our purpose is to compute the sum above over embeddings  $\phi$  of  $\tau$  in  $T_{n+1}$ . We may restrict the sum only to embeddings with  $\pi(\tau, \phi, T_{n+1}) > 0$ . We separate such embeddings into three categories:

1. those who do not include  $u_n$  or  $v_n$  in their image;
2. those who include  $u_n$  but not  $v_n$  in their image;
3. those who include both  $u_n$  and  $v_n$  in their image.

The embedding cannot contain  $v_n$  without  $u_n$  in its image. Indeed, we have  $\deg_{T_{n+1}}(v_n) = 1$ , hence, if  $\phi$  is an embedding that maps a vertex  $v \in \tau$  to  $v_n$ , then for  $\pi(\tau, \phi, T_{n+1})$  to be non-zero, it is necessary that  $\ell(v) = 0$ . By choice of  $\tau$ , if such a vertex exists, it necessarily has a neighbour

$u$ , which then is mapped by  $\phi$  onto  $u_n$ . Moreover, the vertex  $v$  needs to have a single neighbour in  $\tau$ , hence needs to be a loose leaf of  $\tau$ .

Write  $\mathcal{F}_\tau^{(i)}(T_{n+1})$  with  $i = 1, 2, 3$  for the contribution to (2.3.2) of embeddings from each of the categories above. For the first two categories, the embeddings considered are in one to one correspondence with embedding of  $\tau$  in  $T_n$  (although their weights are different whether considered in  $T_n$  or  $T_{n+1}$ ). Thus

$$\begin{aligned}\mathcal{F}_\tau^{(1)}(T_{n+1}) &= \sum_{\phi: \tau \hookrightarrow T_n} \pi(\tau, \phi, T_n) \cdot \prod_{u \in \tau} \mathbb{1}_{\phi(u) \neq u_n} \quad \text{and} \\ \mathcal{F}_\tau^{(2)}(T_{n+1}) &= \sum_{\phi: \tau \hookrightarrow T_n} \sum_{u \in \tau} \mathbb{1}_{\phi(u) = u_n} \left( \prod_{v \in \tau \setminus \{u\}} [\deg_{T_n} \phi(v) - 1]_{\ell(v)} \right) \cdot [\deg_{T_n} \phi(u)]_{\ell(u)}.\end{aligned}$$

Now a basic algebraic manipulation shows that  $[k]_\ell = [k-1]_\ell + \ell[k-1]_{\ell-1}$  for all  $k \geq 1$  and  $\ell \geq 0$ <sup>2</sup>. Moreover  $\prod_{u \in \tau} \mathbb{1}_{\phi(u) \neq u_n} = 1 - \sum_{u \in \tau} \mathbb{1}_{\phi(u) = u_n}$ . Hence

$$\begin{aligned}\mathcal{F}_\tau^{(1)}(T_{n+1}) + \mathcal{F}_\tau^{(2)}(T_{n+1}) \\ = \sum_{\phi: \tau \hookrightarrow T_n} \left[ \pi(\tau, \phi, T_n) + \sum_{u \in \tau} \mathbb{1}_{\phi(u) = u_n} \ell(u) \left( \prod_{v \in \tau \setminus \{u\}} [\deg_{T_n} \phi(v) - 1]_{\ell(v)} \right) \cdot [\deg_{T_n} \phi(u) - 1]_{\ell(u)-1} \right].\end{aligned}$$

Notice that  $\sum_{\phi: \tau \hookrightarrow T_n} \pi(\tau, \phi, T_n) = \mathcal{F}_\tau(T_n)$ . Moreover, recall that, for any fixed  $u \in \tau$  and embedding  $\phi: \tau \hookrightarrow T_n$ , we have  $\mathbb{P}(\phi(u) = u_n | \mathcal{E}_n) = \frac{\deg_{T_n} \phi(u) - 1 + \alpha}{(1+\alpha)^{n-2}}$ . Thus, when taking the expectation in the above we find

$$\begin{aligned}\mathbb{E}[\mathcal{F}_\tau^{(1)}(T_{n+1}) + \mathcal{F}_\tau^{(2)}(T_{n+1}) | \mathcal{E}_n] - \mathcal{F}_\tau(T_n) \\ = \sum_{u \in \tau} \ell(u) \cdot \sum_{\phi: \tau \hookrightarrow T_n} \left( \prod_{v \in \tau \setminus \{u\}} [\deg_{T_n} \phi(v) - 1]_{\ell(v)} \right) \cdot [\deg_{T_n} \phi(u) - 1]_{\ell(u)-1} \cdot \frac{\deg_{T_n} \phi(u) - 1 + \alpha}{(1+\alpha)^{n-2}}.\end{aligned}$$

The sum over  $u \in \tau$  in the right-hand side above may be limited to vertices  $u$  with  $\ell(u) > 0$ .

For  $u \in \tau$  with  $\ell(u) > 0$ , let  $\tau^{(u^-)}$  be the decorated tree  $(\tau, \ell^{(u^-)})$  with decorations identical to those of  $\tau$  except at the vertex  $u$  for which  $\ell^{(u^-)} = \ell(u) - 1$ . Then  $\tau^{(u^-)} \prec \tau$ . Write  $\deg_{T_n} \phi(u) - 1 + \alpha = [\deg_{T_n} \phi(u) - \ell(u)] + \ell(u) + \alpha - 1$  to find that the summand in the right hand side is

$$\frac{1}{(1+\alpha)^{n-2}} \left[ \pi(\tau, \phi, T_n) + (\ell(u) + \alpha - 1) \cdot \pi(\tau^{(u^-)}, \phi, T_n) \right].$$

Thus we find

$$\begin{aligned}\mathbb{E}[\mathcal{F}_\tau^{(1)}(T_{n+1}) + \mathcal{F}_\tau^{(2)}(T_{n+1}) | \mathcal{E}_n] &= \mathcal{F}_\tau(T_n) + \sum_{u \in \tau} \frac{\ell(u)}{(1+\alpha)^{n-2}} \cdot [\mathcal{F}_\tau(T_n) + (\ell(u) + \alpha - 1) \cdot \mathcal{F}_{\tau^{(u^-)}}(T_n)] \\ &= \left[ 1 + \frac{|\ell|}{(1+\alpha)^{n-2}} \right] \mathcal{F}_\tau(T_n) + \sum_{u \in \tau} \frac{\ell(u)(\ell(u) + \alpha - 1)}{(1+\alpha)^{n-2}} \cdot \mathcal{F}_{\tau^{(u^-)}}(T_n). \quad (2.3.3)\end{aligned}$$

Finally we turn to  $\mathcal{F}_\tau^{(3)}(T_{n+1})$ . Let  $\phi: \tau \hookrightarrow T_{n+1}$  be an embedding contributing to  $\mathcal{F}_\tau^{(3)}(T_{n+1})$ , let  $v$  be the loose leaf mapped to  $v_{n+1}$  and  $u$  its only neighbour in  $\tau$ . Define the following three modifications of  $\tau$ :

- $\tau \setminus v$  is the decorated tree obtained from  $\tau$  by removing the leaf  $v$  and conserving the same decorations for all other vertices;
- $(\tau \setminus v)^-$  is the decorated tree obtained from  $\tau$  by removing the leaf  $v$ , decreasing the decoration of  $u$  by one, and conserving the same decorations for all other vertices<sup>3</sup>;
- $(\tau \setminus v)^+$  is the decorated tree obtained from  $\tau$  by removing the leaf  $v$ , increasing the decoration of  $u$  by one, and conserving the same decorations for all other vertices.

---

2. when  $\ell = 0$ , the second term in the RHS is not defined. However, its prefactor cancels it out, and we allow this abuse of notation.

3. This is only defined when  $\ell(u) > 0$ ; it will implicitly only appear in such cases in the upcoming formulas

It is immediate to check that all trees above are smaller than  $\tau$  for the order  $\prec$ .

Write  $\tau \setminus \mathbf{v}$  for the tree (stripped of decoration) of all of the above. To  $\phi$  associate its restriction  $\tilde{\phi} : \tau \setminus \mathbf{v} \rightarrow T_n$  to  $\tau \setminus \mathbf{v}$ . Then, by the same type of computation as above

$$\begin{aligned} \pi(\tau, \phi, T_{n+1}) &= \left( \prod_{\mathbf{w} \in \tau \setminus \mathbf{v}} [\deg_{T_{n+1}} \phi(\mathbf{w}) - 1]_{\ell(\mathbf{w})} \right) \\ &= \left( \prod_{\mathbf{v} \in \tau \setminus \{\mathbf{u}, \mathbf{v}\}} [\deg_{T_n} \phi(\mathbf{v}) - 1]_{\ell(\mathbf{v})} \right) \cdot \left( [\deg_{T_n} \phi(\mathbf{u}) - 1]_{\ell(\mathbf{u})} + \ell(\mathbf{u}) [\deg_{T_n} \phi(\mathbf{u}) - 1]_{\ell(\mathbf{u})-1} \right). \end{aligned} \quad (2.3.4)$$

In the first line, since  $\ell(\mathbf{v}) = 0$ , we removed the term coming from  $v$  from the product. The quantity above will be weighted by  $\mathbb{P}(\phi(\mathbf{u}) = u_n \mid \mathcal{E}_n) = \frac{\deg_{T_n} \phi(\mathbf{u}) - 1 + \alpha}{(1+\alpha)^{n-2}}$ . In preparation, observe that

$$\begin{aligned} (d-1+\alpha) \cdot [d-1]_{\ell(\mathbf{u})} &= [d-1]_{\ell(\mathbf{u})+1} + (\ell(\mathbf{u}) + \alpha) [d-1]_{\ell(\mathbf{u})} \quad \text{and} \\ (d-1+\alpha) \cdot [d-1]_{\ell(\mathbf{u})-1} &= [d-1]_{\ell(\mathbf{u})} + (\ell(\mathbf{u}) - 1 + \alpha) [d-1]_{\ell(\mathbf{u})-1}. \end{aligned}$$

Applying the above with  $d = \deg_{T_n}(\mathbf{u})$  to (2.3.4), multiplied by  $\mathbb{P}(\phi(\mathbf{u}) = u_n \mid \mathcal{E}_n)$ , we find

$$\begin{aligned} &\mathbb{P}(\phi(\mathbf{u}) = u_n \mid \mathcal{E}_n) \cdot \pi(\tau, \phi, T_{n+1}) \\ &= \frac{1}{(1+\alpha)^{n-2}} \cdot \left[ \pi((\tau \setminus \mathbf{v})^+, \tilde{\phi}, T_n) + (2\ell(\mathbf{u}) + \alpha) \cdot \pi(\tau \setminus \mathbf{v}, \tilde{\phi}, T_n) + \ell(\mathbf{u})(\ell(\mathbf{u}) + \alpha - 1) \cdot \pi((\tau \setminus \mathbf{v})^-, \tilde{\phi}, T_n) \right]. \end{aligned}$$

Summing over all embeddings and all values of  $\mathbf{u}, \mathbf{v}$  we find

$$\begin{aligned} &\mathbb{E} \left[ \mathcal{F}_{\tau}^{(3)}(T_{n+1}) \mid \mathcal{E}_n \right] \\ &= \frac{1}{(1+\alpha)^{n-2}} \sum_{\mathbf{v} \text{ loose leaf}} \mathcal{F}_{(\tau \setminus \mathbf{v})^+}(T_n) + (2\ell(\mathbf{u}) + \alpha) \cdot \mathcal{F}_{\tau \setminus \mathbf{v}}(T_n) + \ell(\mathbf{u})(\ell(\mathbf{u}) + \alpha - 1) \cdot \mathcal{F}_{(\tau \setminus \mathbf{v})^-}(T_n), \end{aligned} \quad (2.3.5)$$

where the sum in the right-hand side is over all loose leaves  $\mathbf{v}$  of  $\tau$  and  $\mathbf{u}$  denotes their unique neighbour.

By summing (2.3.3) and (2.3.5), we may obtain a recurrence formula similar to (2.3.1), but with one flaw. Indeed, in such a formula the trees  $\sigma$  would potentially be of the type  $(\tau \setminus \mathbf{v})^+$ , hence have same weight as  $\tau$ . We reduce the contribution of such trees to ones of lower weight via the following lemma.

**Lemma 2.3** *Let  $\tau = (\tau, \ell)$  be a decorated tree with  $w(\tau) \geq 2$ , let  $\mathbf{v}$  be a loose leave of  $\tau$  and  $\mathbf{u}$  its only neighbour in  $\tau$ . Then, for any tree  $T$ :*

$$\mathcal{F}_{(\tau \setminus \mathbf{v})^+}(T) = \mathcal{F}_{\tau}(T) + (\deg_{\tau}(\mathbf{u}) - \ell(\mathbf{u}) - 2) \cdot \mathcal{F}_{\tau \setminus \mathbf{v}}(T). \quad (2.3.6)$$

**Proof of Lemma 2.3** We proceed in two steps. Let us first express  $\mathcal{F}_{\tau}(T)$  using  $\mathcal{F}_{\tau \setminus \mathbf{v}}(T)$ . To any  $\phi : \tau \hookrightarrow T$ , associate its restriction  $\tilde{\phi} : \tau \setminus \mathbf{v} \rightarrow T$  to  $\tau \setminus \mathbf{v}$ . Conversely, any tree embedding  $\tilde{\psi} : \tau \setminus \mathbf{v} \hookrightarrow T$  may be extended to some embedding  $\psi : \tau \hookrightarrow T$  in as many different ways as there are neighbouring vertices of  $\tilde{\psi}(\mathbf{u})$  not reached by  $\psi$ , that is to say  $\deg_T(\tilde{\psi}(\mathbf{u})) - \deg_{\tau \setminus \mathbf{v}}(\mathbf{u})$  ways. Therefore:

$$\mathcal{F}_{\tau}(T) = \sum_{\tilde{\psi} : \tau \setminus \mathbf{v} \rightarrow T} (\deg_T(\tilde{\psi}(\mathbf{u})) - \deg_{\tau \setminus \mathbf{v}}(\mathbf{u})) \cdot \pi(\tau \setminus \mathbf{v}, \tilde{\psi}, T). \quad (2.3.7)$$

In the above equation, we further use that  $\pi(\tau, \psi, T) = \pi(\tau \setminus \mathbf{v}, \tilde{\psi}, T)$  given that  $\mathbf{v}$  is a loose leaf.

Next we express  $\mathcal{F}_{(\tau \setminus \mathbf{v})^+}(T)$  in terms of  $\mathcal{F}_{\tau \setminus \mathbf{v}}(T)$ . Recall that  $(\tau \setminus \mathbf{v})^+$  is obtained from  $\tau \setminus \mathbf{v}$  by increasing the decoration of the vertex  $\mathbf{u}$  by one. Thus:

$$\mathcal{F}_{(\tau \setminus \mathbf{v})^+}(T) = \sum_{\tilde{\psi} : \tau \setminus \mathbf{v} \rightarrow T} (\deg_T(\tilde{\psi}(\mathbf{u})) - 1 - \ell(\mathbf{u})) \cdot \pi(\tau \setminus \mathbf{v}, \tilde{\psi}, T). \quad (2.3.8)$$

Writing  $\deg_T(\tilde{\psi}(\mathbf{u})) - 1 - \ell(\mathbf{u}) = [\deg_T(\tilde{\psi}(\mathbf{u})) - \deg_{\tau \setminus \mathbf{v}}(\mathbf{u})] + [\deg_{\tau \setminus \mathbf{v}}(\mathbf{u}) - 1 - \ell(\mathbf{u})]$  and observing that  $\deg_{\tau \setminus \mathbf{v}}(\mathbf{u}) = \deg_{\tau}(\mathbf{u}) - 1$ , we deduce Lemma 2.3 by subtracting (2.3.7) from (2.3.8).

By Lemma 2.3, equation (2.3.5) may be re-expressed as

$$\begin{aligned} \mathbb{E}[\mathcal{F}_\tau^{(3)}(T_{n+1}) \mid \mathcal{E}_n] &= \frac{1}{(1+\alpha)n-2} \sum_{\mathbf{v} \text{ loose leaf}} \mathcal{F}_\tau(T_n) + (\deg_\tau(\mathbf{u}) + \ell(\mathbf{u}) + \alpha - 2) \cdot \mathcal{F}_{\tau \setminus \mathbf{v}}(T_n) \\ &\quad + \ell(\mathbf{u})(\ell(\mathbf{u}) + \alpha - 1) \cdot \mathcal{F}_{(\tau \setminus \mathbf{v})^-}(T_n). \end{aligned} \quad (2.3.9)$$

Since  $w(\tau) > 1$ , the vertex  $\mathbf{u}$  may never be a loose leaf of  $\tau$ , hence  $\deg_\tau(\mathbf{u}) \geq 2$  or  $\deg_\tau(\mathbf{u}) = 1$  but  $\ell(\mathbf{u}) \geq 1$ . In both cases, the multiplicative factor  $\deg_\tau(\mathbf{u}) + \ell(\mathbf{u}) + \alpha - 2$  is non-negative. Equations (2.3.3) and (2.3.9) together yield (2.3.1).

**Remark 2.3** *Following the proof we find that the constants  $c(\sigma, \tau)$  appearing in (2.3.1) are non-zero only if  $\sigma$  may be obtained from  $\tau$  by*

- (i) *decreasing the value of one decoration by 1 or*
- (ii) *removing a loose leaf and conserving all other decorations or*
- (iii) *removing a loose leaf and modifying the decoration of its unique neighbour by  $-1$ .*

*It is direct that all the trees  $\sigma$  are smaller than  $\tau$  for  $\prec$ . Indeed, in most cases we have  $w(\sigma) < w(\tau)$ . However, there are two cases where  $w(\sigma) = w(\tau)$ : when the decoration of a leaf is 1 in  $\tau$  and decreases to 0 in  $\sigma$  (by the procedure (i)) and when a loose leaf is removed from  $\tau$  (as in (ii)), with its unique ancestor having decoration 0, hence becoming a loose leaf of  $\sigma$ . In both these cases,  $\sigma \prec \tau$  due to the second condition of Definition 2.5.*

### 2.3.2 The first moment of $\mathcal{F}_\tau(T_n)$

We are ready to state the full estimate of the first moment of  $\mathcal{F}_\tau(T_n^S)$ .

**Theorem 2.3** *For any  $\alpha > 0$ , any seed tree  $S$  of size  $k \geq 2$  and any decorated tree  $\tau$ , we have:*

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S)] \approx n^{\max\{1, \frac{w(\tau)}{1+\alpha}\}} \cdot (\log n)^{\gamma(\tau)},$$

where  $\gamma(\tau)$  is a nonnegative exponent equal to zero when  $w(\tau) < 1 + \alpha$  and otherwise recursively defined by:

$$\gamma(\tau) = \sup_{\sigma \prec \tau, c(\sigma, \tau) > 0, w(\sigma) = w(\tau)} (\gamma(\sigma) + 1) \quad \text{if } w(\tau) > 1 + \alpha,$$

or by (critical case):

$$\gamma(\tau) = \gamma_c(\tau) := \max \left\{ 1, \sup_{\sigma \prec \tau, c(\sigma, \tau) > 0, w(\sigma) = w(\tau)} (\gamma_c(\sigma) + 1) \right\} \quad \text{if } w(\tau) = 1 + \alpha.$$

with the convention  $\sup \emptyset = 0$ .

**Remark 2.4** *Let  $\tau$  be a decorated tree with  $\ell(u) \geq 2$  for all its leaves. In particular, it contains no loose leaves and  $w(\tau) = |\ell| = \sum_{u \in \tau} \ell(u)$ . Then Remark 2.3 indicates that  $c(\sigma, \tau) > 0$  only for decorated trees  $\sigma$  obtained by lowering the decoration of some vertex  $u \in \tau$  by 1. Moreover, no such tree  $\sigma$  has any loose leaf either. Assuming in addition that  $w(\tau) = |\ell| > 1 + \alpha$ , we have  $w(\sigma) < w(\tau)$  and  $\gamma(\tau) = 0$ .*

To prove Theorem 2.3, we proceed by induction on the set of decorated trees (for the partial order  $\preceq$ ) and use the recurrence formula (2.3.1).

**Proof of Theorem 2.3** We first note that for the three decorated trees  $\tau$  with  $w(\tau) = 1$  (thus  $w(\tau) < 1 + \alpha$  for any  $\alpha > 0$ ) the first moment estimate is satisfied as indicated by the explicit formula of Proposition 2.2.

Consider now a decorated tree  $\tau$  with  $w(\tau) \geq 2$  and suppose by induction that Theorem 2.3 is valid for all decorated trees  $\sigma \prec \tau$ . Let us define for any  $n \geq k$  the quantity  $\omega_{n+1}^{(\tau)}$  equal to:

$$\omega_{n+1}^{(\tau)} := \prod_{\ell=k}^n \left( 1 + \frac{w(\tau)}{(1+\alpha)\ell - 2} \right)^{-1} \approx n^{-\frac{w(\tau)}{1+\alpha}}, \quad (2.3.10)$$

where the latter equivalent is obtained by a standard computation. Then, by multiplying (2.3.1) by this factor and taking the expectation, we get:

$$\mathbb{E}[\omega_{n+1}^{(\tau)} \cdot \mathcal{F}_\tau(T_{n+1})] = \mathbb{E}[\omega_n^{(\tau)} \cdot \mathcal{F}_\tau(T_n)] + \sum_{\sigma \prec \tau} c(\sigma, \tau) \cdot \frac{\omega_{n+1}^{(\tau)}}{(1+\alpha)n-2} \cdot \mathbb{E}[\mathcal{F}_\sigma(T_n)].$$

By iterating the above over  $n$ , we find

$$\mathbb{E}[\omega_{n+1}^{(\tau)} \cdot \mathcal{F}_\tau(T_{n+1})] = \mathcal{F}_\tau(S) + \sum_{\sigma \prec \tau} c(\sigma, \tau) \sum_{\ell=k}^n \frac{\omega_{\ell+1}^{(\tau)}}{(1+\alpha)\ell-2} \cdot \mathbb{E}[\mathcal{F}_\sigma(T_\ell)]. \quad (2.3.11)$$

Thus, the asymptotic behaviour of  $\mathbb{E}[\mathcal{F}_\tau(T_{n+1})]$  can be derived from that of  $\mathbb{E}[\mathcal{F}_\sigma(T_\ell)]$  for  $\sigma \prec \tau$  and  $\ell \leq n$ . Define the following variables

$$\mathcal{S}_n(\sigma, \tau) := \sum_{\ell=k}^n \frac{\omega_{\ell+1}^{(\tau)}}{(1+\alpha)\ell-2} \cdot \mathbb{E}[\mathcal{F}_\sigma(T_\ell)].$$

The growth rate in  $n$  of  $\mathcal{S}_n(\sigma, \tau)$  depends on  $w(\sigma)$  through  $\mathbb{E}[\mathcal{F}_\sigma(T_\ell)]$ . We distinguish three cases according to the value of  $w(\tau)$ .

**1st case:**  $w(\tau) < 1 + \alpha$

In this situation, we necessarily have  $w(\sigma) < 1 + \alpha$  for any  $\sigma \prec \tau$ . Thus, by the induction hypothesis:

$$\mathbb{E}[\mathcal{F}_\sigma(T_\ell)] \approx \ell.$$

The above together with (2.3.10) imply that

$$\frac{\omega_{\ell+1}^{(\tau)}}{(1+\alpha)\ell-2} \cdot \mathbb{E}[\mathcal{F}_\sigma(T_\ell)] \approx \ell^{-\frac{w(\tau)}{1+\alpha}}.$$

Since  $w(\tau) < 1 + \alpha$ , the sum over  $\ell$  of the above - which constitutes  $\mathcal{S}_n(\sigma, \tau)$  - diverges at rate:

$$\mathcal{S}_n(\sigma, \tau) = \sum_{\ell=k}^n \frac{1}{(1+\alpha)\ell-2} \cdot \omega_{\ell+1}^{(\tau)} \cdot \mathbb{E}[\mathcal{F}_\sigma(T_\ell)] \approx n^{1-\frac{w(\tau)}{1+\alpha}}.$$

We then sum over every  $\sigma \prec \tau$  to get an asymptotic estimate for the quantity on the right of (2.3.11):

$$\mathbb{E}[\omega_{n+1}^{(\tau)} \cdot \mathcal{F}_\tau(T_{n+1})] = \mathcal{F}_\tau(S) + \sum_{\sigma \prec \tau} c(\sigma, \tau) \sum_{\ell=k}^n \frac{\omega_{\ell+1}^{(\tau)}}{(1+\alpha)\ell-2} \cdot \mathbb{E}[\mathcal{F}_\sigma(T_\ell)] \approx n^{1-\frac{w(\tau)}{1+\alpha}}, \quad (2.3.12)$$

since there exists at least one  $\sigma$  with  $c(\sigma, \tau) > 0$ . Finally, dividing (2.3.12) by  $\omega_{n+1}^{(\tau)}$  and using (2.3.10), we obtain the expected estimate:

$$\mathbb{E}[\mathcal{F}_\tau(T_{n+1})] \approx n.$$

**2nd case:**  $w(\tau) = 1 + \alpha$

Now, when  $\sigma \prec \tau$ , we can either have  $w(\sigma) < 1 + \alpha$  or  $w(\sigma) = 1 + \alpha$ . In the former situation, we have

$$\frac{1}{(1+\alpha)n-2} \cdot \omega_{n+1}^{(\tau)} \cdot \mathbb{E}[\mathcal{F}_\sigma(T_n)] \approx n^{-1},$$

hence,

$$\mathcal{S}_n(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \sum_{\ell=k}^n \frac{1}{(1+\alpha)\ell-2} \cdot \omega_{\ell+1}^{(\boldsymbol{\tau})} \cdot \mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_{\ell})] \approx \log(n).$$

In the latter situation, by the induction hypothesis,

$$\mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_{n+1})] \approx n \cdot (\log n)^{\gamma_c(\boldsymbol{\sigma})}.$$

Using (2.3.10), we find,

$$\frac{1}{(1+\alpha)n-2} \cdot \omega_{n+1}^{(\boldsymbol{\tau})} \cdot \mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_n)] \approx n^{-\frac{w(\boldsymbol{\tau})}{1+\alpha}} \cdot (\log n)^{\gamma_c(\boldsymbol{\sigma})} = n^{-1} \cdot (\log n)^{\gamma_c(\boldsymbol{\sigma})}.$$

As a consequence,  $\mathcal{S}_n(\boldsymbol{\sigma}, \boldsymbol{\tau})$  diverges at rate:

$$\mathcal{S}_n(\boldsymbol{\sigma}, \boldsymbol{\tau}) \approx \sum_{\ell=k}^n \ell^{-1} \cdot (\log \ell)^{\gamma_c(\boldsymbol{\sigma})} \approx (\log n)^{\gamma_c(\boldsymbol{\sigma})+1}.$$

Thus, only the terms  $\boldsymbol{\sigma}$  with maximal weight and  $c(\boldsymbol{\sigma}, \boldsymbol{\tau}) > 0$  contribute to (2.3.11) significantly:

$$\mathbb{E}[\omega_{n+1}^{(\boldsymbol{\tau})} \cdot \mathcal{F}_{\boldsymbol{\tau}}(T_{n+1})] \approx \sum_{\substack{\boldsymbol{\sigma} \prec \boldsymbol{\tau} \\ w(\boldsymbol{\sigma}) < 1+\alpha}} c(\boldsymbol{\sigma}, \boldsymbol{\tau}) \cdot \log n + \sum_{\substack{\boldsymbol{\sigma} \prec \boldsymbol{\tau} \\ w(\boldsymbol{\sigma}) = 1+\alpha}} c(\boldsymbol{\sigma}, \boldsymbol{\tau}) \cdot (\log n)^{\gamma_c(\boldsymbol{\sigma})+1} \approx (\log n)^{\gamma_c(\boldsymbol{\tau})}.$$

where  $\gamma_c(\boldsymbol{\tau})$  is defined as in Theorem 2.3. Dividing the last equation by  $\omega_{n+1}^{(\boldsymbol{\tau})}$  leads to the result.

**3rd case:**  $w(\boldsymbol{\tau}) > 1 + \alpha$

In this case, the trees  $\boldsymbol{\sigma} \prec \boldsymbol{\tau}$  can satisfy either  $w(\boldsymbol{\sigma}) < 1 + \alpha$ ,  $w(\boldsymbol{\sigma}) = 1 + \alpha$ ,  $1 + \alpha < w(\boldsymbol{\sigma}) < w(\boldsymbol{\tau})$ , or  $1 + \alpha < w(\boldsymbol{\sigma}) = w(\boldsymbol{\tau})$ . In the two first situations, by the induction hypothesis, there exists  $\delta > 0$  such that:

$$\frac{1}{(1+\alpha)\ell-2} \cdot \omega_{\ell+1}^{(\boldsymbol{\tau})} \cdot \mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_{\ell})] \ll \ell^{-\frac{w(\boldsymbol{\tau})}{1+\alpha}} \cdot (\log \ell)^{\delta}.$$

Since  $w(\boldsymbol{\tau}) > 1 + \alpha$ , the sum over  $\ell$  of the above converges and  $\mathcal{S}_n(\boldsymbol{\sigma}, \boldsymbol{\tau}) \approx 1$ .

When  $1 + \alpha < w(\boldsymbol{\sigma}) < w(\boldsymbol{\tau})$ , the induction hypothesis implies:

$$\frac{1}{(1+\alpha)\ell-2} \cdot \omega_{\ell+1}^{(\boldsymbol{\tau})} \cdot \mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_{\ell})] \approx \ell^{\frac{w(\boldsymbol{\sigma})-w(\boldsymbol{\tau})}{1+\alpha}-1} \cdot (\log \ell)^{\gamma(\boldsymbol{\sigma})}.$$

Since  $w(\boldsymbol{\sigma}) < w(\boldsymbol{\tau})$ , the sum over  $\ell$  of the above converges again, and  $\mathcal{S}_n(\boldsymbol{\sigma}, \boldsymbol{\tau}) \approx 1$ .

Finally, when  $w(\boldsymbol{\sigma}) = w(\boldsymbol{\tau})$ , the induction hypothesis gives us:

$$\frac{1}{(1+\alpha)\ell-2} \cdot \omega_{\ell+1}^{(\boldsymbol{\tau})} \cdot \mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_{\ell})] \approx \ell^{-1} \cdot (\log \ell)^{\gamma(\boldsymbol{\sigma})}.$$

Consequently, by a direct computation, the divergence rate of  $\mathcal{S}_n(\boldsymbol{\sigma}, \boldsymbol{\tau})$  may be shown to be:

$$\mathcal{S}_n(\boldsymbol{\sigma}, \boldsymbol{\tau}) \approx \sum_{\ell=k}^n \ell^{-1} \cdot (\log \ell)^{\gamma(\boldsymbol{\sigma})} \approx (\log n)^{\gamma(\boldsymbol{\sigma})+1}.$$

In conclusion, by considering the asymptotic of  $\mathcal{S}_n(\boldsymbol{\sigma}, \boldsymbol{\tau})$  for all  $\boldsymbol{\sigma} \prec \boldsymbol{\tau}$  with  $c(\boldsymbol{\sigma}, \boldsymbol{\tau}) > 0$  according to the above, we obtain

$$\mathbb{E}[\omega_{n+1}^{(\boldsymbol{\tau})} \cdot \mathcal{F}_{\boldsymbol{\tau}}(T_{n+1})] \approx \sum_{\substack{\boldsymbol{\sigma} \prec \boldsymbol{\tau} \\ w(\boldsymbol{\sigma}) < w(\boldsymbol{\tau})}} c(\boldsymbol{\sigma}, \boldsymbol{\tau}) + \sum_{\substack{\boldsymbol{\sigma} \prec \boldsymbol{\tau} \\ w(\boldsymbol{\sigma}) = w(\boldsymbol{\tau})}} c(\boldsymbol{\sigma}, \boldsymbol{\tau}) \cdot (\log n)^{\gamma(\boldsymbol{\sigma})+1} \approx (\log n)^{\gamma(\boldsymbol{\tau})},$$

where the last equivalent is due to how  $\gamma(\boldsymbol{\tau})$  is defined in Theorem 2.3. Divide by  $\omega_{n+1}^{(\boldsymbol{\tau})}$  to obtain the expected result.

### 2.3.3 The second moment of $\mathcal{F}_\tau(T_n)$

We are now ready to prove the second moment estimate on  $\mathcal{F}_\tau(T_n)$  of Theorem 2.2. We will build on the analogous result on the first moment obtained in the previous section as well as on its proof. First, remark that the square of the observables may be written as:

$$\mathcal{F}_\tau(T_n)^2 = \sum_{\phi_1, \phi_2: \tau \hookrightarrow T_n} \pi(\tau, \phi_1, T_n) \cdot \pi(\tau, \phi_2, T_n), \quad (2.3.13)$$

where the sum is this time over all pairs  $\phi_1, \phi_2$  of graph embeddings of  $\tau$  in  $T_n$ . We will decompose the sum appearing in (2.3.13) in two parts, according to whether the embeddings  $\phi_1(\tau)$  and  $\phi_2(\tau)$  overlap or not. The same is done in [54]. Call  $\mathcal{F}_{\tau, \tau}(T_n)$  the first resulting quantity and  $\mathcal{F}_{\tau+\tau}(T_n)$  the second one:

$$\mathcal{F}_{\tau, \tau}(T_n) = \sum_{\substack{\phi_1, \phi_2: \tau \rightarrow T_n \\ \phi_1(\tau) \cap \phi_2(\tau) = \emptyset}} \pi(\tau, \phi_1, T_n) \pi(\tau, \phi_2, T_n), \quad \mathcal{F}_{\tau+\tau}(T_n) = \sum_{\substack{\phi_1, \phi_2: \tau \rightarrow T_n \\ \phi_1(\tau) \cap \phi_2(\tau) \neq \emptyset}} \pi(\tau, \phi_1, T_n) \pi(\tau, \phi_2, T_n).$$

The first moments of the two quantities above are bounded separately. First we estimate  $\mathbb{E}[\mathcal{F}_{\tau+\tau}(T_n)]$ , which turns out to be the the easiest of the two. Indeed, it may be expressed as a sum of first moments of observables for some decorated trees derived from  $\tau$ . These are computed using Theorem 2.3. To deal with  $\mathbb{E}[\mathcal{F}_{\tau, \tau}(T_n)]$  we will prove a recurrence inequality on such quantities, similar to (2.3.1) and using the same techniques.

#### 2.3.3.1 An estimate on $\mathbb{E}[\mathcal{F}_{\tau+\tau}(T_n)]$

The goal of this section is to show the following:

**Proposition 2.3** *Let  $\tau$  be a decorated tree with  $\ell(u) \geq 2$  for any  $u \in \tau$  and such that  $|\ell| = \sum_{u \in \tau} \ell(u) > 1 + \alpha$ . Then, for any seed  $S$ ,*

$$\mathbb{E}[\mathcal{F}_{\tau+\tau}(T_n^S)] = \mathcal{O}(n^{\frac{2|\ell|}{1+\alpha}}). \quad (2.3.14)$$

The proposition is based on the following lemma, which we prove below.

**Lemma 2.4** *Let  $\tau = (\tau, \ell)$  be a decorated tree. There exists a finite set  $\mathcal{U}(\tau)$  of decorated trees  $\sigma$  with  $w(\sigma) \leq 2w(\tau)$  and positive constants  $C(\tau, \sigma)$  for  $\sigma \in \mathcal{U}(\tau)$  (see the proof for an explicit description) such that, for any tree  $T$ ,*

$$\mathcal{F}_{\tau+\tau}(T) = \sum_{\sigma \in \mathcal{U}(\tau)} C(\tau, \sigma) \cdot \mathcal{F}_\sigma(T). \quad (2.3.15)$$

In the planar setting the above is very intuitive; we sketch a proof below. If  $T$  is plane, then  $\mathcal{F}_{\tau+\tau}(T)$  is the number of decorated embeddings  $(\phi_1, \phi_2)$  of two copies  $\tau_1$  and  $\tau_2$  of  $\tau$  in  $T$ , which overlap.

Let us first forget about the decorations and focus on graph embeddings. The union of the images of  $\tau_1$  and  $\tau_2$  via such embeddings is a tree  $\sigma$ ; one may see  $\sigma$  as a merger of  $\tau_1$  and  $\tau_2$ . Thus, the pairs of overlapping embeddings of  $\tau_1$  and  $\tau_2$  in  $T$  are in bijection with the embeddings of  $\sigma$  in  $T$  where  $\sigma$  ranges over all possible mergers of  $\tau_1$  and  $\tau_2$ .

Now consider overlapping decorated embeddings of  $\tau_1$  and  $\tau_2$  in  $T$ . Then, each corner of  $T$  may have no arrow pointing to it, an arrow of  $\tau_1$ , an arrow of  $\tau_2$  or one arrow of  $\tau_1$  and one of  $\tau_2$  pointing to it. In the first three cases, the arrows pointing to the corner will be considered as arrows of  $\sigma$ ; in the last case, the arrow of  $\tau_1$  and that of  $\tau_2$  merge into a single arrow of  $\sigma$ . Thus, any such pair of decorated embeddings corresponds to a decorated embedding of some decorated tree  $\sigma = (\sigma, m)$  obtained as a merger of  $\tau_1$  and  $\tau_2$ . In particular  $|m| \leq |\ell_1| + |\ell_2| = 2|\ell|$ . The constants  $C(\tau, \sigma)$  are combinatorial factors that account for the different ways to merge arrows of  $\tau_1$  and  $\tau_2$ .

The actual proof given below avoids the use of the planar model and is more algebraic.



**Proof of Lemma 2.4** Fix  $\tau$  and  $T$  as in the lemma. Consider two embeddings  $\Phi_1, \Phi_2$  of  $\tau$  in  $T$  whose images intersect. Since  $T$  is a tree,  $\Phi_1(\tau) \cup \Phi_2(\tau)$  is itself a subtree of  $T$ . Moreover,  $\Phi_1(\tau) \cap \Phi_2(\tau)$  is the image via  $\Phi_1$  and  $\Phi_2$ , respectively, of two isomorphic subtrees  $\sigma_1$  and  $\sigma_2$  of  $\tau$ .

For two isomorphic subtrees  $\sigma_1$  and  $\sigma_2$  of  $\tau$ , define the  $(\sigma_1, \sigma_2)$ -merger of two copies of  $\tau$  as the tree obtained by “gluing” two copies of  $\tau$  along  $\sigma_1$  and  $\sigma_2$  respectively. Write  $\mathcal{M}(\sigma_1, \sigma_2)$  for this tree. To identify the two copies of  $\tau$  merged to obtain  $\mathcal{M}(\sigma_1, \sigma_2)$ , call them  $\tau_1 = (\tau_1, \ell_1)$  and  $\tau_2 = (\tau_2, \ell_2)$ .

Each vertex of  $\mathcal{M}(\sigma_1, \sigma_2)$  is clearly identified to either one vertex in  $\tau_1 \setminus \sigma_1$ , a vertex in  $\tau_2 \setminus \sigma_2$  or to a vertex in  $\sigma_1$  and simultaneously to one in  $\sigma_2$ . For a vertex  $u \in \mathcal{M}(\sigma_1, \sigma_2)$  write  $\ell_1(u)$  for its decoration in  $\tau_1$ , if it is identified to a vertex of  $\tau_1$ , otherwise set  $\ell_1(u) = 0$ . Define  $\ell_2(u)$  for  $u \in \mathcal{M}(\sigma_1, \sigma_2)$  in the same way.

Then, the pairs of embeddings  $\Phi_1, \Phi_2$  of  $\tau$  in  $T$  with  $\Phi_1(\tau) \cap \Phi_2(\tau) = \Phi_1(\sigma_1) = \Phi_2(\sigma_2)$  are in bijection with the embeddings of  $\mathcal{M}(\sigma_1, \sigma_2)$  in  $T$ . It follows that

$$\mathcal{F}_{\tau+\tau}(T) = \sum_{(\sigma_1, \sigma_2)} \sum_{\phi: \mathcal{M}(\sigma_1, \sigma_2) \hookrightarrow T} \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} [\deg_T(\phi(u)) - 1]_{\ell_1(u)} \cdot [\deg_T(\phi(u)) - 1]_{\ell_2(u)}, \quad (2.3.16)$$

where the first sum is over all pairs of isomorphic subtrees  $(\sigma_1, \sigma_2)$  of  $\tau$ . To reduce the above to a formula of the type (2.1.3), we use the following combinatorial identity.

**Fact 2.1** Fix non-negative integers  $n, \ell_1, \ell_2$ . Then

$$[n]_{\ell_1} \cdot [n]_{\ell_2} = \sum_{j=0}^{\min\{\ell_1, \ell_2\}} [n]_{\ell_1 + \ell_2 - j} \cdot \frac{\ell_1! \cdot \ell_2!}{(\ell_1 - j)! (\ell_2 - j)! j!}.$$

**Proof** First observe that  $[n]_{\ell} = \binom{n}{\ell} \cdot \ell!$ . Now let us inspect the quantity  $\binom{n}{\ell_1} \binom{n}{\ell_2}$  which is the number of pairs of subsets  $A_1, A_2$  of  $\{1, \dots, n\}$  with  $\ell_1$  and  $\ell_2$  elements, respectively. These may be counted as follows. First establish the set  $A_1 \cup A_2$  which can have  $\ell_1 + \ell_2 - j$  elements with  $j \in \{0, \dots, \min\{\ell_1, \ell_2\}\}$ . Once  $A_1 \cup A_2$  is chosen, split it into  $A_1 \setminus A_2, A_2 \setminus A_1$  and  $A_1 \cap A_2$ . This forms a partition of  $A_1 \cup A_2$  into three sets of cardinality  $\ell_1 - j, \ell_2 - j$  and  $j$ , respectively. Thus

$$\binom{n}{\ell_1} \cdot \binom{n}{\ell_2} = \sum_{j=0}^{\min\{\ell_1, \ell_2\}} \binom{n}{\ell_1 + \ell_2 - j} \frac{(\ell_1 + \ell_2 - j)!}{(\ell_1 - j)! (\ell_2 - j)! j!}.$$

Multiply by  $\ell_1! \cdot \ell_2!$  to find the desired result.

Let us return to the proof of Lemma 2.4. A valid decoration  $m$  for a tree in  $\mathcal{M}(\sigma_1, \sigma_2)$  is one such that, for all  $u \in \mathcal{M}(\sigma_1, \sigma_2)$ ,

$$\max\{\ell_1(u), \ell_2(u)\} \leq m(u) \leq \ell_1(u) + \ell_2(u). \quad (2.3.17)$$

Observe that, if  $u$  is not in the images of  $\sigma_1$  and  $\sigma_2$ , then  $m(u)$  is entirely determined by the above. For vertices that are in the overlap of  $\tau_1$  and  $\tau_2$ ,  $m(u)$  may take one of several values.

Applying Fact 2.1 to the summand in (2.3.16), we find

$$\begin{aligned} & \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} [\deg_T(\phi(u)) - 1]_{\ell_1(u)} \cdot [\deg_T(\phi(u)) - 1]_{\ell_2(u)} \\ &= \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} \sum_{j=0}^{\min\{\ell_1(u), \ell_2(u)\}} [\deg_T(\phi(u)) - 1]_{\ell_1(u) + \ell_2(u) - j} \cdot \frac{\ell_1(u)! \cdot \ell_2(u)!}{(\ell_1(u) - j)! (\ell_2(u) - j)! j!} \\ &= \sum_m \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} [\deg_T(\phi(u)) - 1]_{m(u)} \cdot \frac{\ell_1(u)! \cdot \ell_2(u)!}{(m(u) - \ell_2(u))! (m(u) - \ell_1(u))! (\ell_1(u) + \ell_2(u) - m(u))!} \\ &= \sum_m C(\tau, (\mathcal{M}(\sigma_1, \sigma_2), m)) \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} [\deg_T(\phi(u)) - 1]_{m(u)}, \end{aligned}$$

where the sum in the last two terms is over all valid decorations  $m$  of  $\mathcal{M}(\sigma_1, \sigma_2)$  and

$$C(\tau, (\mathcal{M}(\sigma_1, \sigma_2), m)) := \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} \frac{\ell_1(u)! \cdot \ell_2(u)!}{(m(u) - \ell_2(u))! (m(u) - \ell_1(u))! (\ell_1(u) + \ell_2(u) - m(u))!}.$$

Write  $\mathcal{M}$  for the tree  $\mathcal{M}(\sigma_1, \sigma_2)$  with decoration  $m$ . Then

$$\sum_{\phi: \mathcal{M}(\sigma_1, \sigma_2) \hookrightarrow T} \prod_{u \in \mathcal{M}(\sigma_1, \sigma_2)} [\deg_T(\phi(u)) - 1]_{m(u)} = \mathcal{F}_{\mathcal{M}}(T).$$

Inserting this into (2.3.16), we find

$$\mathcal{F}_{\tau+\tau}(T) = \sum_{\mathcal{M}} C(\tau, \mathcal{M}) \mathcal{F}_{\mathcal{M}}(T),$$

where the sum is over all trees of the form  $\mathcal{M}(\sigma_1, \sigma_2)$  with valid decorations  $m$ . These form the set  $\mathcal{U}(\tau)$ ; it is immediate that they have weight at most  $2w(\tau)$ .

We are finally ready to prove Proposition 2.3

**Proof of Proposition 2.3** Fix a decorated tree as in the statement. By Theorem 2.3, for any tree  $\sigma \in \mathcal{U}(\tau)$  of weight strictly smaller than  $2w(\tau)$ ,

$$\mathbb{E}[\mathcal{F}_{\sigma}(T_n^S)] \ll n^{\frac{2w(\tau)}{1+\alpha}}.$$

Moreover, if  $\sigma = (\sigma, m) \in \mathcal{U}(\tau)$  is such that  $w(\sigma) = 2w(\tau)$  then its decorations satisfy  $m(u) = \ell_1(u) + \ell_2(u) \geq 2$  for all  $u \in \sigma$  (see (2.3.17) for how  $m$  is defined). As explained in Remark 2.4, for any such tree

$$\mathbb{E}[\mathcal{F}_{\sigma}(T_n^S)] \approx n^{\frac{2w(\tau)}{1+\alpha}}.$$

Using (2.3.15) and observing that  $\tau$  has no loose leaf, we conclude that

$$\mathbb{E}[\mathcal{F}_{\tau+\tau}(T_n^S)] = \mathcal{O}(n^{\frac{2w(\tau)}{1+\alpha}}) = \mathcal{O}(n^{\frac{2|\ell|}{1+\alpha}}).$$

### 2.3.3.2 An estimate on $\mathbb{E}[\mathcal{F}_{\tau, \tau}(T_n)]$

We start this section by defining a wider class of observables that will be involved in a recurrence relation which will eventually allow us to estimate  $\mathbb{E}[\mathcal{F}_{\tau, \tau}(T_n)]$ . Let  $\tau, \sigma$  be two decorated trees and  $T$  a (bigger) tree. Then we denote by  $\mathcal{F}_{\tau, \sigma}(T)$  the following integer-valued observable:

$$\mathcal{F}_{\tau, \sigma}(T) := \sum_{\substack{\phi_1, \phi_2 \\ \phi_1(\tau) \cap \phi_2(\sigma) = \emptyset}} \pi(\tau, \phi_1, T) \cdot \pi(\sigma, \phi_2, T), \quad (2.3.18)$$

where the sum is over all graph embeddings  $\phi_1$ , resp.  $\phi_2$ , of  $\tau$ , resp.  $\sigma$ , in  $T$  with no overlap in their image. The quantity of interest to us is that with  $\sigma = \tau$  and  $T = T_n$ . This section is concerned with proving the following bound.

**Proposition 2.4** *Let  $T_n$  be an  $\alpha$ -PA tree. Then it holds that:*

$$\mathbb{E}[\mathcal{F}_{\tau, \sigma}(T_n)] = \mathcal{O}(\mathbb{E}[\mathcal{F}_{\tau}(T_n)] \cdot \mathbb{E}[\mathcal{F}_{\sigma}(T_n)]). \quad (2.3.19)$$

Proposition 2.4 implies directly the bound necessary for the proof of Theorem 2.2.

**Corollary 2.1** *Let  $\tau$  be a decorated tree with  $\ell(u) \geq 2$  for any  $u \in \tau$  and such that  $|\ell| > 1 + \alpha$ . Then, for any seed  $S$ ,*

$$\mathbb{E}[\mathcal{F}_{\tau, \tau}(T_n^S)] = \mathcal{O}(n^{\frac{2|\ell|}{1+\alpha}}). \quad (2.3.20)$$

**Proof of Corollary 2.1** Apply Proposition 2.4 with  $\sigma = \tau$  and observe that, due to Remark 2.4 and the conditions on  $\tau$ , we have  $\mathbb{E}[\mathcal{F}_\tau(T_n)] \approx n^{\frac{|\ell|}{1+\alpha}}$ .

Proposition 2.4 is obtained through a recurrence relation, similarly to how Theorem 2.3 follows from Proposition 2.2. Since we only need an upper bound, we state only a (simpler) recursive inequality.

**Proposition 2.5** *There exists a family of nonnegative real numbers  $\{c^*(\theta', \theta) : \theta' \prec \theta\}$  such that for any two decorated trees  $\tau, \sigma \neq \odot$ :*

$$\begin{aligned} \mathbb{E}[\mathcal{F}_{\tau, \sigma}(T_{n+1}) | \mathcal{F}_n] &\leq \left(1 + \frac{w(\tau) + w(\sigma)}{(1+\alpha) \cdot n - 2}\right) \cdot \mathcal{F}_{\tau, \sigma}(T_n) \\ &+ \frac{1}{(1+\alpha) \cdot n - 2} \left[ \sum_{\mathbf{t} \prec \tau} c^*(\mathbf{t}, \tau) \cdot \mathcal{F}_{\mathbf{t}, \sigma}(T_n) + \sum_{\mathbf{s} \prec \sigma} c^*(\mathbf{s}, \sigma) \cdot \mathcal{F}_{\tau, \mathbf{s}}(T_n) \right]. \end{aligned} \quad (2.3.21)$$

In the rest of the section, we show how Proposition 2.5, implies Proposition 2.4, then prove Proposition 2.5. Both proofs follow similar arguments to those in Sections 2.3.2 and 2.3.1, respectively.

**Proof of Proposition 2.4** We are going to proceed by induction on pairs  $(\tau, \sigma)$  of decorated trees, for the partial order induced by  $\prec$  on such pairs (precisely  $(\mathbf{t}, \mathbf{s}) \prec (\tau, \sigma)$  if either  $\mathbf{t} \preceq \tau$  and  $\mathbf{s} \prec \sigma$  or  $\mathbf{t} \prec \tau$  and  $\mathbf{s} \preceq \sigma$ ).

**Base case** We show (2.3.19) when  $\tau$  is any decorated tree and  $\sigma = \odot$ . If  $\phi_1$  is a graph embedding of  $\tau$  in  $T_n$ , its image consists of  $|\tau|$  vertices of  $T_n$ . Hence, the number of ways to embed  $\sigma$  in  $T_n$  without overlapping with  $\phi_1(\tau)$  is  $n - |\tau|$ . Thus

$$\mathbb{E}[\mathcal{F}_{\tau, \odot}(T_n)] = \mathbb{E}[\mathcal{F}_\tau(T_n)] \cdot (n - |\tau|) = \mathcal{O}(\mathbb{E}[\mathcal{F}_\tau(T_n)] \cdot \mathbb{E}[\mathcal{F}_\odot(T_n)]),$$

as required.

**Induction step** Let  $\tau, \sigma$  be two decorated trees, both different from  $\odot$ . Assume that (2.3.19) holds for all pairs  $(\mathbf{t}, \mathbf{s})$  with either  $\mathbf{t} \preceq \tau$  and  $\mathbf{s} \prec \sigma$  or  $\mathbf{t} \prec \tau$  and  $\mathbf{s} \preceq \sigma$ . In the following, we set:

$$\omega_n^{(\tau, \sigma)} := \prod_{\ell=k}^n \left(1 + \frac{w(\tau) + w(\sigma)}{(1+\alpha) \cdot \ell - 2}\right)^{-1}$$

and

$$\mathcal{S}_n(\mathbf{t}, \tau; \sigma) := \sum_{\ell=k}^n \frac{1}{(1+\alpha)\ell - 2} \cdot \omega_{\ell+1}^{(\tau, \sigma)} \cdot \mathbb{E}[\mathcal{F}_{\mathbf{t}, \sigma}(T_\ell)]$$

for any decorated tree  $\mathbf{t} \prec \tau$ . Iterating (2.3.21), we obtain:

$$\mathbb{E}[\omega_{n+1}^{(\tau, \sigma)} \cdot \mathcal{F}_{\tau, \sigma}(T_{n+1})] \leq \mathcal{F}_{\tau, \sigma}(S) + \sum_{\mathbf{t} \prec \tau} c^*(\mathbf{t}, \tau) \cdot \mathcal{S}_n(\mathbf{t}, \tau; \sigma) + \sum_{\mathbf{s} \prec \sigma} c^*(\mathbf{s}, \sigma) \cdot \mathcal{S}_n(\mathbf{s}, \sigma; \tau).$$

Thus, to prove (2.3.19), it suffices to show that  $\mathcal{S}_n(\mathbf{t}, \tau; \sigma) = \mathcal{O}(\omega_n^{(\tau, \sigma)} \cdot \mathbb{E}[\mathcal{F}_\tau(T_n)] \cdot \mathbb{E}[\mathcal{F}_\sigma(T_n)])$  for all  $\mathbf{t} \prec \tau$  (by symmetry, the same will also hold for  $\mathcal{S}_n(\mathbf{s}, \sigma; \tau)$  with  $\mathbf{s} \prec \sigma$ ). Recall from Theorem 2.3 the asymptotic

$$\omega_n^{(\tau, \sigma)} \cdot \mathbb{E}[\mathcal{F}_\tau(T_n)] \cdot \mathbb{E}[\mathcal{F}_\sigma(T_n)] \approx n^{\max\{0, 1 - \frac{w(\tau)}{1+\alpha}\} + \max\{0, 1 - \frac{w(\sigma)}{1+\alpha}\}} \cdot (\log n)^{\gamma(\tau) + \gamma(\sigma)}. \quad (2.3.22)$$

Fix  $\mathbf{t} \prec \tau$ . According to the induction hypothesis and the above, the terms of  $\mathcal{S}_n(\mathbf{t}, \tau; \sigma)$  are bounded as

$$\frac{\omega_{n+1}^{(\tau, \sigma)}}{(1+\alpha)n - 2} \cdot \mathbb{E}[\mathcal{F}_{\mathbf{t}, \sigma}(T_n)] = \mathcal{O}\left(n^{\max\left\{\frac{w(\mathbf{t}) - w(\tau)}{1+\alpha}, 1 - \frac{w(\tau)}{1+\alpha}\right\} + \max\left\{0, 1 - \frac{w(\sigma)}{1+\alpha}\right\} - 1} (\log n)^{\gamma(\mathbf{t}) + \gamma(\sigma)}\right). \quad (2.3.23)$$

The sum of the above has different asymptotics depending on the value of the exponent of  $n$ :

$$\boxed{\text{1st case : } \max \left\{ \frac{w(\mathbf{t})-w(\boldsymbol{\tau})}{1+\alpha}, 1 - \frac{w(\boldsymbol{\tau})}{1+\alpha} \right\} + \max \left\{ 0, 1 - \frac{w(\boldsymbol{\sigma})}{1+\alpha} \right\} < 0}$$

Then the sum of (2.3.23) converges, thus  $\mathcal{S}_n(\mathbf{t}, \boldsymbol{\tau}; \boldsymbol{\sigma}) = \mathcal{O}(1) = \mathcal{O}(\omega_n^{(\boldsymbol{\tau}, \boldsymbol{\sigma})} \cdot \mathbb{E}[\mathcal{F}_{\boldsymbol{\tau}}(T_n)] \cdot \mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_n)])$ .

$$\boxed{\text{2nd case : } \max \left\{ \frac{w(\mathbf{t})-w(\boldsymbol{\tau})}{1+\alpha}, 1 - \frac{w(\boldsymbol{\tau})}{1+\alpha} \right\} + \max \left\{ 0, 1 - \frac{w(\boldsymbol{\sigma})}{1+\alpha} \right\} > 0}$$

Then the sum of (2.3.23) diverges, and a standard estimate provides the precise rate of growth:

$$\mathcal{S}_n(\mathbf{t}, \boldsymbol{\tau}; \boldsymbol{\sigma}) = \mathcal{O}\left(n^{\max \left\{ \frac{w(\mathbf{t})-w(\boldsymbol{\tau})}{1+\alpha}, 1 - \frac{w(\boldsymbol{\tau})}{1+\alpha} \right\} + \max \left\{ 0, 1 - \frac{w(\boldsymbol{\sigma})}{1+\alpha} \right\}} (\log n)^{\gamma(\mathbf{t}) + \gamma(\boldsymbol{\sigma})}\right).$$

Compare the above to (2.3.22) to find

$$\frac{\mathcal{S}_n(\mathbf{t}, \boldsymbol{\tau}; \boldsymbol{\sigma})}{\omega_n^{(\boldsymbol{\tau}, \boldsymbol{\sigma})} \mathbb{E}[\mathcal{F}_{\boldsymbol{\tau}}(T_n)] \mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_n)]} = \mathcal{O}\left(n^{\max \left\{ \frac{w(\mathbf{t})-w(\boldsymbol{\tau})}{1+\alpha}, 1 - \frac{w(\boldsymbol{\tau})}{1+\alpha} \right\} - \max \left\{ 0, 1 - \frac{w(\boldsymbol{\tau})}{1+\alpha} \right\}} \cdot (\log n)^{\gamma(\mathbf{t}) - \gamma(\boldsymbol{\tau})}\right).$$

Now, recall that

- $w(\mathbf{t}) \leq w(\boldsymbol{\tau})$  always;
- if  $w(\mathbf{t}) \leq w(\boldsymbol{\tau}) < 1 + \alpha$ , then  $\gamma(\mathbf{t}) = \gamma(\boldsymbol{\tau}) = 0$ ;
- if  $w(\mathbf{t}) = w(\boldsymbol{\tau}) \geq 1 + \alpha$ , then  $\gamma(\mathbf{t}) < \gamma(\boldsymbol{\tau})$ .

A separate analysis of the three different situations above shows that  $\mathcal{S}_n(\mathbf{t}, \boldsymbol{\tau}; \boldsymbol{\sigma}) = \mathcal{O}(\omega_n^{(\boldsymbol{\tau}, \boldsymbol{\sigma})} \cdot \mathbb{E}[\mathcal{F}_{\boldsymbol{\tau}}(T_n)] \cdot \mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_n)])$ .

$$\boxed{\text{3rd case : } \max \left\{ \frac{w(\mathbf{t})-w(\boldsymbol{\tau})}{1+\alpha}, 1 - \frac{w(\boldsymbol{\tau})}{1+\alpha} \right\} + \max \left\{ 0, 1 - \frac{w(\boldsymbol{\sigma})}{1+\alpha} \right\} = 0}$$

Then the sum of (2.3.23) diverges, and a standard estimate provides the precise rate of growth:

$$\mathcal{S}_n(\mathbf{t}, \boldsymbol{\tau}; \boldsymbol{\sigma}) = \mathcal{O}\left((\log n)^{\gamma(\mathbf{t}) + \gamma(\boldsymbol{\sigma}) + 1}\right).$$

By (2.3.22) and our assumption, we find

$$\frac{\mathcal{S}_n(\mathbf{t}, \boldsymbol{\tau}; \boldsymbol{\sigma})}{\omega_n^{(\boldsymbol{\tau}, \boldsymbol{\sigma})} \mathbb{E}[\mathcal{F}_{\boldsymbol{\tau}}(T_n)] \mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_n)]} = \mathcal{O}\left(n^{\max \left\{ \frac{w(\mathbf{t})-w(\boldsymbol{\tau})}{1+\alpha}, 1 - \frac{w(\boldsymbol{\tau})}{1+\alpha} \right\} - \max \left\{ 0, 1 - \frac{w(\boldsymbol{\tau})}{1+\alpha} \right\}} (\log n)^{\gamma(\mathbf{t}) - \gamma(\boldsymbol{\tau}) + 1}\right). \quad (2.3.24)$$

The power of  $n$  in the right-hand side above is negative or null. When it is negative, equation (2.3.24) is bounded, as required. It can only be 0 in two cases: when  $w(\boldsymbol{\tau}) < 1 + \alpha$  or when  $w(\mathbf{t}) = w(\boldsymbol{\tau}) \geq 1 + \alpha$ . The former is incoherent with the assumption of this 3rd case; when the latter occurs,  $\gamma(\mathbf{t}) \leq \gamma(\boldsymbol{\tau}) - 1$ , hence (2.3.24) is bounded. In conclusion, (2.3.24) is always bounded, which is to say that  $\mathcal{S}_n(\mathbf{t}, \boldsymbol{\tau}; \boldsymbol{\sigma}) = \mathcal{O}(\omega_n^{(\boldsymbol{\tau}, \boldsymbol{\sigma})} \cdot \mathbb{E}[\mathcal{F}_{\boldsymbol{\tau}}(T_n)] \cdot \mathbb{E}[\mathcal{F}_{\boldsymbol{\sigma}}(T_n)])$ .

It remains to prove Proposition 2.5:

**Proof of Proposition 2.5** The strategy followed here is the same as for Proposition 2.2. We start by fixing  $\boldsymbol{\tau} = (\tau, \ell_{\tau})$  and  $\boldsymbol{\sigma} = (\sigma, \ell_{\sigma})$  two decorated trees, both different from  $\odot$ . Recall our notation: in passing from  $T_n$  to  $T_{n+1}$  a new vertex denoted by  $v_n$  is attached to a randomly chosen vertex  $u_n$  of  $T_n$ . Our purpose is to compute the sum (2.3.18) over pairs of non-overlapping graph embeddings  $\phi_1, \phi_2$  of  $\tau$  and  $\sigma$  into  $T_{n+1}$ . We restrict the sum to embeddings with  $\pi(\boldsymbol{\tau}, \phi_1, T_{n+1}) > 0$  and  $\pi(\boldsymbol{\sigma}, \phi_2, T_{n+1}) > 0$ .

As it has already been noted in the proof of Proposition 2.2, such embeddings fall into three distinct categories: those which do not include  $u_n$  nor  $v_n$  in their image, those which include  $u_n$  but not  $v_n$  in their image and those which include both  $u_n$  and  $v_n$  in their image. Embeddings that include  $v_n$  but not  $u_n$  in their image are not authorised as they only have positive weight when the decorated tree is  $\odot$ .

A crucial remark is that two graph embeddings  $\phi_1 : \tau \rightarrow T_{n+1}$ ,  $\phi_2 : \sigma \rightarrow T_{n+1}$  with no overlap cannot be at once in one of the two last categories of embeddings. We thus enumerate three kind of situations:

1. the images of  $\phi_1$  and  $\phi_2$  are both in  $T_{n+1} \setminus \{u_n, v_n\}$ ;
2. the image of  $\phi_1$  contains  $u_n$  but not  $v_n$ , that of  $\phi_2$  does not include any of them; or the same situation but reversing the roles of  $\phi_1$  and  $\phi_2$ ;
3. the image of  $\phi_1$  contains  $u_n$  and  $v_n$ , that of  $\phi_2$  does not include any of them; or the same situation but reversing the roles of  $\phi_1$  and  $\phi_2$ .

Write  $\mathcal{F}_{\tau, \sigma}^{(i)}(T_{n+1})$  with  $i = 1, 2, 3$  for the contribution to (2.3.18) of pairs of embeddings being in the corresponding situation above. As in the proof of Proposition 2.2, the pairs of embeddings in the two first situations are in one-to-one correspondance with non-overlapping embeddings  $\phi_1 : \tau \rightarrow T_n$ ,  $\phi_2 : \sigma \rightarrow T_n$ , although their weights  $\pi(\tau, \phi_1, T_{n+1})$  and  $\pi(\sigma, \phi_2, T_{n+1})$  may differ from that in  $T_n$ . The same algebraic manipulations used for  $\mathcal{F}_{\tau}^{(1)}(T_{n+1})$  and  $\mathcal{F}_{\tau}^{(2)}(T_{n+1})$  in the proof of Proposition 2.2 may also be applied here to find

$$\begin{aligned} & \mathbb{E}[\mathcal{F}_{\tau, \sigma}^{(1)}(T_{n+1}) + \mathcal{F}_{\tau, \sigma}^{(2)}(T_{n+1}) \mid \mathcal{E}_n] \\ &= \left[ 1 + \frac{|\ell_{\tau}| + |\ell_{\sigma}|}{(1+\alpha)n-2} \right] \mathcal{F}_{\tau, \sigma}(T_n) + \sum_{u \in \tau} \frac{\ell_{\tau}(u)(\ell_{\tau}(u) + \alpha - 1)}{(1+\alpha)n-2} \cdot \mathcal{F}_{\tau^{u-}, \sigma}(T_n) + \sum_{v \in \sigma} \frac{\ell_{\sigma}(v)(\ell_{\sigma}(v) + \alpha - 1)}{(1+\alpha)n-2} \cdot \mathcal{F}_{\tau, \sigma^{v-}}(T_n) \end{aligned} \quad (2.3.25)$$

Moreover, the quantity  $\mathcal{F}_{\tau, \sigma}^{(3)}(T_{n+1})$  may be treated as  $\mathcal{F}_{\tau}^{(3)}(T_{n+1})$  in (2.3.5), and we find

$$\begin{aligned} & \mathbb{E}[\mathcal{F}_{\tau, \sigma}^{(3)}(T_{n+1}) \mid \mathcal{E}_n] \\ &= \frac{1}{(1+\alpha)n-2} \cdot \left( \sum_{\mathbf{v} \text{ l.l. of } \tau} \mathcal{F}_{(\tau \setminus \mathbf{v})^+, \sigma}(T_n) + (2\ell_{\tau}(\mathbf{u}) + \alpha) \cdot \mathcal{F}_{\tau \setminus \mathbf{v}, \sigma}(T_n) + \ell_{\tau}(\mathbf{u})(\ell_{\tau}(\mathbf{u}) + \alpha - 1) \cdot \mathcal{F}_{(\tau \setminus \mathbf{v})^-, \sigma}(T_n) \right. \\ & \left. + \sum_{\mathbf{v} \text{ l.l. of } \sigma} \mathcal{F}_{\tau, (\sigma \setminus \mathbf{v})^+}(T_n) + (2\ell_{\sigma}(\mathbf{u}) + \alpha) \cdot \mathcal{F}_{\tau, \sigma \setminus \mathbf{v}}(T_n) + \ell_{\sigma}(\mathbf{u})(\ell_{\sigma}(\mathbf{u}) + \alpha - 1) \cdot \mathcal{F}_{\tau, (\sigma \setminus \mathbf{v})^-}(T_n) \right) \end{aligned} \quad (2.3.26)$$

where the two sums are over the set of loose leaves of  $\tau$  and  $\sigma$  respectively. The above equation is a derivative of (2.3.5). At the same stage, the last ingredient for the proof of Proposition 2.2 was a combinatorial lemma - Lemma 2.3 - claiming that  $\mathcal{F}_{(\tau \setminus \mathbf{v})^+}(T_n)$  is actually a linear combination of  $\mathcal{F}_{\tau}(T_n)$  and  $\mathcal{F}_{\tau \setminus \mathbf{v}}(T_n)$  when  $\mathbf{v}$  is a loose leaf of  $\tau$ . Here that is almost what we do for  $\mathcal{F}_{(\tau \setminus \mathbf{v})^+, \sigma}(T_n)$  and of course for  $\mathcal{F}_{\tau, (\sigma \setminus \mathbf{v})^+}(T_n)$  by symmetry. The equality (2.3.6) is just turned into an inequality:

**Lemma 2.5** *Let  $\tau = (\tau, \ell_{\tau})$  and  $\sigma = (\sigma, \ell_{\sigma})$  be two decorated trees,  $\mathbf{v}$  a loose leaf of  $\tau$  and  $\mathbf{u}$  its only neighbour in  $\tau$ . Then, for any tree  $T$ :*

$$\mathcal{F}_{\tau, \sigma}(T) + C(\tau, w) \cdot \mathcal{F}_{\tau \setminus \mathbf{v}, \sigma}(T) \leq \mathcal{F}_{(\tau \setminus \mathbf{v})^+, \sigma}(T) \leq \mathcal{F}_{\tau, \sigma}(T) + C^*(\tau, w) \cdot \mathcal{F}_{\tau \setminus \mathbf{v}, \sigma}(T),$$

with  $C(\tau, w)$  being the constant defined in Lemma 2.3 and  $C^*(\tau, w) := C(\tau, w) + 1$ .

**Proof** As for Lemma 2.3, we proceed in two steps. First, we express  $\mathcal{F}_{\tau, \sigma}(T)$  in terms of pairs of non-overlapping embeddings of  $\tau \setminus \mathbf{v}$  and  $\sigma$ , respectively, in  $T$ .

To any non-overlapping pair of embeddings  $\phi_1 : \tau \rightarrow T$ ,  $\phi_2 : \sigma \rightarrow T$  associate the restriction  $\tilde{\phi}_1 : \tau \setminus \mathbf{v} \rightarrow T$  of  $\phi_1$  to the set of vertices of  $\tau \setminus \mathbf{v}$  together with the same embedding  $\phi_2$  of  $\sigma$ . Conversely, if a pair of non-overlapping embeddings  $\tilde{\psi}_1 : \tau \setminus \mathbf{v} \rightarrow T$  and  $\phi_2 : \sigma \rightarrow T$  is given, we may extend  $\tilde{\psi}_1$  to some  $\psi_1 : \tau \rightarrow T$  which preserves the non-overlapping property with  $\phi_2$ . The number of such extensions depends on how many neighbours of  $\tilde{\psi}_1(u)$  are not contained in the images of  $\tilde{\psi}_1$  and  $\phi_2$ . The former occupies  $\deg_{\tau \setminus \mathbf{v}}(u)$  neighbours of  $\tilde{\psi}_1(u)$ ; the latter may occupy 0 or 1 neighbour due to the non-overlapping requirement. From this observation, and given that  $\pi(\tau, \phi_1, T) = \pi(\tau \setminus \mathbf{v}, \tilde{\phi}_1, T)$  ( $\mathbf{v}$  being a loose leaf), we deduce that:

$$\mathcal{F}_{\tau, \sigma}(T) \leq \sum_{\substack{\tilde{\phi}_1, \phi_2 \\ \tilde{\phi}_1(\tau \setminus \mathbf{v}) \cap \phi_2(\sigma) = \emptyset}} \left( \deg_T(\tilde{\phi}_1(u)) - \deg_{\tau \setminus \mathbf{v}}(u) \right) \cdot \pi(\tau \setminus \mathbf{v}, \tilde{\phi}_1, T) \cdot \pi(\sigma, \phi_2, T) \quad (2.3.27)$$

and

$$\mathcal{F}_{\tau, \sigma}(T) \geq \sum_{\substack{\tilde{\phi}_1, \phi_2 \\ \tilde{\phi}_1(\tau \setminus \mathbf{v}) \cap \phi_2(\sigma) = \emptyset}} \left( \deg_T(\tilde{\phi}_1(u)) - \deg_{\tau \setminus \mathbf{v}}(u) - 1 \right) \cdot \pi(\tau \setminus \mathbf{v}, \tilde{\phi}_1, T) \cdot \pi(\sigma, \phi_2, T). \quad (2.3.28)$$

The two above equations correspond to (2.3.7) in the proof of Lemma 2.3. Finally, similarly to (2.3.8), we find

$$\mathcal{F}_{(\tau \setminus \nu)^+, \sigma}(T) = \sum_{\substack{\tilde{\phi}_1, \phi_2 \\ \tilde{\phi}_1(\tau \setminus \nu) \cap \phi_2(\sigma) = \emptyset}} \left( \deg_T(\tilde{\phi}_1(u)) - 1 - \ell_\tau(u) \right) \cdot \pi(\tau \setminus \nu, \tilde{\phi}_1, T) \cdot \pi(\sigma, \phi_2, T). \quad (2.3.29)$$

Subtracting (2.3.27) and (2.3.28) from (2.3.29), we obtain the desired lower and upper bounds, respectively.

To finish the proof of Proposition 2.5, insert the upper bound of Lemma 2.5 in (2.3.26) to find

$$\begin{aligned} \mathbb{E} \left[ \mathcal{F}_{\tau, \sigma}^{(3)}(T_{n+1}) \mid \mathcal{E}_n \right] &\leq \frac{|\{\nu \text{ loose leaf of } \tau\}| + |\{\nu \text{ loose leaf of } \sigma\}|}{(1+\alpha)n-2} \cdot \mathcal{F}_{\tau, \sigma}(T_n) \\ &+ \frac{1}{(1+\alpha)n-2} \sum_{\nu \text{ l.l. of } \tau} (\deg_\tau(u) + \ell_\tau(u) + \alpha - 1) \cdot \mathcal{F}_{\tau \setminus \nu, \sigma}(T_n) + \ell_\tau(u)(\ell_\tau(u) + \alpha - 1) \cdot \mathcal{F}_{(\tau \setminus \nu)^-, \sigma}(T_n) \\ &+ \frac{1}{(1+\alpha)n-2} \sum_{\nu \text{ l.l. of } \sigma} (\deg_\sigma(u) + \ell_\sigma(u) + \alpha - 1) \cdot \mathcal{F}_{\tau, \sigma \setminus \nu}(T_n) + \ell_\sigma(u)(\ell_\sigma(u) + \alpha - 1) \cdot \mathcal{F}_{\tau, (\sigma \setminus \nu)^-}(T_n) \end{aligned} \quad (2.3.30)$$

where the two sums are again over the set of loose leaves of  $\tau$  and  $\sigma$  respectively. Equations (2.3.25) and (2.3.30) eventually lead to the expected inequality (2.3.21).

### 2.3.3.3 Proof of Theorem 2.2

Theorem 2.2 follows directly from the estimates of the two previous sections and from the decomposition

$$\mathcal{F}_\tau(T_n)^2 = \mathcal{F}_{\tau, \tau}(T_n) + \mathcal{F}_{\tau+\tau}(T_n). \quad (2.3.31)$$

**Proof of Theorem 2.2** Take the expectation of (2.3.31) and insert the bounds (2.3.14) and (2.3.20).

## 2.4 Observables and their difference around the seed tree

The goal of this section is to produce decorated trees  $\tau$  that can distinguish between two different seeds  $S$  and  $S'$ . Moreover, we wish the asymptotic of  $\mathcal{F}_\tau(T_n^S)$  for such trees (both for the first and the square root of the second moment) be of the type  $n^{\frac{w(\tau)}{1+\alpha}}$ , with a null logarithmic correction. The main result is the following.

**Theorem 2.4** *For any two seed trees  $S \neq S'$  of size  $k, k' \geq 3$ , there exists a decorated tree  $\tau$  with  $w(\tau) > 1 + \alpha$  and  $\ell(u) \geq 2$  for all  $u \in \tau$  such that*

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})] \approx n^{\frac{w(\tau)}{1+\alpha}}.$$

The next four sections are concerned with the proof of Theorem 2.4. To start, we will consider seeds  $S, S'$  of same size. Finally, in Section 2.4.5, we prove our main result, Theorem 2.1.

### 2.4.1 Differences only appear around the seed

We are here interested in the variation of the first moment of (2.1.3), depending on the value taken by the seed tree  $S$ . We state that :

**Proposition 2.6** *For any two seed trees  $S, S'$  of common size  $k \geq 3$ , any decorated tree  $\tau$  and any  $n \geq k$  :*

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})] = \mathbb{E}[\mathcal{F}_\tau(T_n^S, \{\cdot\} \cap S \neq \emptyset)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'}, \{\cdot\} \cap S' \neq \emptyset)],$$

where:

$$\mathcal{F}_\tau(T_n^S, \{\cdot\} \cap S \neq \emptyset) := \sum_{\substack{\phi: \tau \rightarrow T_n^S \\ \phi(\tau) \cap S \neq \emptyset}} \prod_{u \in \tau} [\deg_{T_n^S}(\phi(u)) - 1]_{\ell(u)},$$

and the same for  $S'$ . In other words,  $\mathcal{F}_\tau(T_n^S, \{\cdot\} \cap S \neq \emptyset)$  is the contribution to  $\mathcal{F}_\tau$  of embeddings that “intersect the seed”.

The reader may be surprised that, in the statement above, we assume the location of the seed inside  $T_n$  known, all while trying to prove that the seed may be determined. It should be clear that, while Proposition 2.6 and other steps of the proof of Theorem 2.4 use the knowledge of the seed, their ultimate result (that is Theorem 2.4) does not.

The proof is based on the coupling of Section 2.2.2, hence we use the planar  $\alpha$ -PA formalism.

**Proof** Fix seed trees  $S, S'$  of equal size  $k \geq 3$ . We start by decomposing the sum over embeddings  $\phi$  defining (2.1.3) according to whether  $\phi(\tau) \cap S$  is empty or not :

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S)] = \mathbb{E}[\mathcal{F}_\tau(T_n^S, \{\cdot\} \cap S \neq \emptyset)] + \mathbb{E}[\mathcal{F}_\tau(T_n^S, \{\cdot\} \cap S = \emptyset)].$$

If an embedding  $\phi$  of  $\tau$  in  $T_n^S$  does not intersect  $S$ , then  $\phi(\tau)$  is necessarily a proper subset of one of the planted plane subtrees  $T_n^{v,i}$ , and does not intersect the root of said tree. Write  $\mathcal{F}_\tau(T_n^{v,i})$  for the contribution to  $\mathcal{F}_\tau(T_n^S)$  of all such embeddings. Then

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S, \{\cdot\} \cap S = \emptyset)] = \sum_{v \in V_S} \sum_{i=1}^{\deg_S(v)} \mathbb{E}[\mathcal{F}_\tau(T_n^{v,i})].$$

The right-hand side of the above does not actually depend on  $S$ , only on its size. Indeed, the coupling described at the end of Section 2.2.2 indicates that the same planted plane subtrees may be used to construct  $T_n^S$  and  $T_n^{S'}$ . While it is not necessary for this argument, one may notice that  $T_n^S$  and  $T_n^{S'}$  may even be coupled so that

$$\mathcal{F}_\tau(T_n^S, \{\cdot\} \cap S = \emptyset) = \mathcal{F}_\tau(T_n^{S'}, \{\cdot\} \cap S' = \emptyset) \quad \text{a.s.}$$

The result follows readily.

## 2.4.2 Blind trees: differences appear only in the seed

Next we aim to improve Proposition 2.6 by showing that  $\mathbb{E}[\mathcal{F}_\tau(T_n^S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})]$  only depends on embeddings totally contained in the seed, not just intersecting it. This will not be true for all trees  $\tau$ , only for special ones. Some definitions are required.

**Definition 2.6 (Perfect embedding)** Let  $\tau = (\tau, d)$  be a decorated tree and  $T$  a (larger) tree. An embedding  $\phi : \tau \hookrightarrow T$  is called **perfect** if

$$d(u) = \deg_T(\phi(u)) \quad \forall u \in \tau.$$

Write  $D_{\tau,d}(T)$  for the number of perfect embeddings from  $(\tau, d)$  to  $T$ . Note that any embedding of  $\tau$  in  $T$  is perfect for exactly one decoration, hence contributes exactly to one  $D_{\tau,d}(T)$ .

**Definition 2.7 (Blind tree)** Let  $\tau, T_1$  and  $T_2$  be finite trees with  $|T_1| = |T_2|$ . We say that  $\tau$  is a  **$(T_1, T_2)$ -blind tree** if for any decoration  $d : \tau \rightarrow \mathbb{N}^*$  :

$$D_{\tau,d}(T_1) = D_{\tau,d}(T_2).$$

Intuitively,  $\tau$  is  $(T_1, T_2)$ -blind if it can not distinguish between  $T_1$  and  $T_2$  using observables involving the number of embeddings and the environment around these embeddings (as are our observables  $\mathcal{F}_\tau$ ). For instance, the tree formed of a single vertex is  $(T_1, T_2)$ -blind if and only if  $T_1$  and  $T_2$  have the same degree sequence.

Let  $(\tau, d)$  be a decorated tree,  $\sigma \subseteq \tau$  a non empty subtree of  $\tau$ , and  $S$  be a seed tree. Write  $D_{\sigma,\tau,d}(T_n^S)$  for the number of perfect embeddings  $\Phi$  of  $(\tau, d)$  in  $T_n^S$  with  $\Phi(\tau) \cap S = \Phi(\sigma)$ . We obviously have :

$$\mathcal{F}_\tau(T_n^S, \{\cdot\} \cap S \neq \emptyset) = \sum_{\sigma \subseteq \tau} \sum_{d : \tau \rightarrow \mathbb{N}^*} D_{\sigma,\tau,d}(T_n^S) \cdot \prod_{u \in \tau} [d(u) - 1]_{\ell(u)}. \quad (2.4.1)$$

The next proposition constitutes the essential step for the upgrade of Proposition 2.6.

**Proposition 2.7** *Let  $S$  and  $S'$  be two seed trees of common size  $k \geq 3$ . For trees  $\sigma \subseteq \tau$  with  $\sigma$  which is  $(S, S')$ -blind,*

$$\forall d : \tau \rightarrow \mathbb{N}^* \text{ and } n \geq k, \quad \mathbb{E}[D_{\sigma, \tau, d}(T_n^S)] = \mathbb{E}[D_{\sigma, \tau, d}(T_n^{S'})].$$

**Proof** The idea of the proof is to decompose the embeddings contributing to  $D_{\sigma, \tau, d}(T_n^S)$  according to degrees of the vertices belonging to its image *in the seed tree*  $S$ . For illustration, we will start with the simpler case when  $\sigma = \tau$ , then move on to the general case.

**Particular case  $\sigma = \tau$**

For any seed  $S$  and  $n \geq |S|$ ,

$$D_{\tau, \tau, d}(T_n^S) = \sum_{\phi: \tau \hookrightarrow S} \prod_{u \in \tau} \mathbb{1}_{\deg_{T_n^S}(\phi(u))=d(u)}.$$

Due to the coupling explained in Section 2.2.2, the probability of the event  $\bigcap_{u \in \tau} \{\deg_{T_n^S}(\phi(u)) = d(u)\}$  only depends on the degrees of the vertices  $(\phi(u))_{u \in \tau}$  in  $S$ . Write  $f_d[\deg_S \phi(u) : u \in \tau]$  for this probability. Then

$$\mathbb{E}[D_{\tau, \tau, d}(T_n^S)] = \sum_{\phi: \tau \hookrightarrow S} f_d[\deg_S \phi(u) : u \in \tau] = \sum_{\ell: \tau \rightarrow \mathbb{N}^*} D_{\tau, \ell}(S) \cdot f_d[\ell(u) : u \in \tau]. \quad (2.4.2)$$

Since  $\tau$  is assumed  $(S, S')$ -blind, the quantity above remains unaltered when  $S$  is replaced by  $S'$ .

**General case  $\sigma \subset \tau$**

For a seed  $S$  and  $v \in S$ , write  $(T_n^S \setminus S)_v$  for the subtree of  $T_n^S$  formed of all vertices that may be connected to  $v$  without using any edge of  $S$ . In terms of vertices  $\{(T_n^S \setminus S)_v : v \in S\}$  is a partition of the vertices of  $T_n^S$ ; in terms of edges, it is a partition of the edges of  $T_n^S$  not contained in  $S$ . In the language of Section 2.2.2,  $(T_n^S \setminus S)_v$  is the tree obtained by gluing  $T_n^{v,1}, \dots, T_n^{v, \deg_S(v)}$  together at the root. Use the same notation for  $\tau$  and  $\sigma$ : for  $v \in \sigma$  write  $(\tau \setminus \sigma)_v$  for the subtree of  $\tau$  formed of all vertices that may be connected to  $v$  in  $\tau$  without using any edge of  $\sigma$ .

For a decoration  $d$  of  $\tau$ , and vertices  $v \in \sigma$  and  $v \in S$ , write  $\mathcal{D}(d, v, v, T_n^S)$  for the number of perfect embeddings  $\phi$  of  $(\tau \setminus \sigma)_v, d$  in  $(T_n^S \setminus S)_v$  with  $\phi(v) = v$ :

$$\mathcal{D}(d, v, v, T_n^S) = \sum_{\phi: (\tau \setminus \sigma)_v \rightarrow (T_n^S \setminus S)_v} \mathbb{1}_{\phi(v)=v} \prod_{u \in (\tau \setminus \sigma)_v} \mathbb{1}_{\{\deg_{T_n^S}(\phi(u))=d(u)\}}.$$

It may not be explicit in the above, but the last product does only depend on  $(T_n^S \setminus S)_v$  for  $u \neq v$ , since  $\deg_{T_n^S}(\phi(u)) = \deg_{(T_n^S \setminus S)_v}(\phi(u))$ . For  $v \in S$ , it should be noted that  $\deg_{T_n^S}(v) = \deg_{(T_n^S \setminus S)_v}(v) + \deg_S(v)$ .

Now express  $D_{\sigma, \tau, d}(T_n^S)$  as follows:

$$D_{\sigma, \tau, d}(T_n^S) = \sum_{\phi: \sigma \hookrightarrow S} \prod_{v \in \sigma} \mathcal{D}(d, v, \phi(v), T_n^S). \quad (2.4.3)$$

Indeed, any embedding contributing  $D_{\sigma, \tau, d}(T_n^S)$  comes from an embedding of  $\sigma$  in  $S$  and embeddings of the subtrees  $(\tau \setminus \sigma)_v$  in the corresponding subtrees of  $T_n^S \setminus S$ . The conditions that the embedding is perfect may be verified separately in each subtree.

Now, due to the coupling of Section 2.2.2 and to the expression of  $(T_n^S \setminus S)_v$  in terms of  $(T_n^{v,i})_{i \geq 1}$ , the expectation of the product above only depends on the degrees of the vertices  $\phi(v)$  in  $S$ :

$$\mathbb{E}\left[\prod_{v \in \sigma} \mathcal{D}(d, v, \phi(v), T_n^S)\right] = f_{\sigma, \tau, d}[\deg_S(\phi(v)) : v \in \sigma],$$

for some explicit function  $f_{\sigma, \tau, d}$ . Injecting this in (2.4.3), we find

$$\mathbb{E}[D_{\sigma, \tau, d}(T_n^S)] = \sum_{\phi: \sigma \hookrightarrow S} f_{\sigma, \tau, d}[\deg_S(\phi(v)) : v \in \sigma] = \sum_{\ell: \sigma \rightarrow \mathbb{N}^*} D_{\sigma, \ell}(S) \cdot f_{\sigma, \tau, d}[\ell(v) : v \in \sigma].$$

Since  $\sigma$  is assumed  $(S, S')$ -blind, the quantity above is equal when  $S$  is replaced by  $S'$ .



We are now ready to give a finer version of Proposition 2.6. Consider two distinct seeds  $S, S'$  of equal size  $k \geq 4$  (non two distinct seeds of smaller size exist). Then there exists at least one tree which is not  $(S, S')$ -blind, for instance  $S$  or  $S'$  have this property. It follows that there exists at least one minimal tree which is not  $(S, S')$ -blind, that is a tree  $\tau$  which is not  $(S, S')$ -blind but for which any proper subtree  $\sigma \subsetneq \tau$  is  $(S, S')$ -blind.

**Corollary 2.2** *Let  $\tau$  be a minimal tree that is not  $(S, S')$ -blind. Then for any decoration  $\ell$  of  $\tau$*

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})] = \mathbb{E}[\mathcal{F}_\tau(T_n^S, \{\cdot\} \subseteq S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'}, \{\cdot\} \subseteq S')],$$

where

$$\mathcal{F}_\tau(T_n^S, \{\cdot\} \subseteq S) := \sum_{\phi: \tau \hookrightarrow S} \prod_{u \in \tau} [\deg_{T_n^S}(\phi(u)) - 1]_{\ell(u)},$$

and the same for  $S'$ .

**Proof** We apply (2.4.1) together with Proposition 2.7 and use the minimality of  $\tau$  to obtain

$$\begin{aligned} & \mathbb{E}[\mathcal{F}_\tau(T_n^S, \{\cdot\} \cap S \neq \emptyset)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'}, \{\cdot\} \cap S' \neq \emptyset)] \\ &= \sum_{d: \tau \rightarrow \mathbb{N}^*} \left( \prod_{u \in \tau} [d(u) - 1]_{\ell(u)} \right) \cdot (\mathbb{E}[D_{\tau, \tau, d}(T_n^S)] - \mathbb{E}[D_{\tau, \tau, d}(T_n^{S'})]). \end{aligned}$$

Furthermore, it is clear that:

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S, \{\cdot\} \subseteq S)] = \sum_{d: \tau \rightarrow \mathbb{N}^*} \left( \prod_{u \in \tau} [d(u) - 1]_{\ell(u)} \right) \cdot \mathbb{E}[D_{\tau, \tau, d}(T_n^S)].$$

Of course, the above equality also holds when  $S$  is replaced by  $S'$ . Hence the result.

### 2.4.3 Evaluating the difference in the seed

In light of the above, the quantities of interest for the proof of Theorem 2.4 are of the type  $\mathbb{E}[\mathcal{F}_\tau(T_n^S, \{\cdot\} \subseteq S)]$ . We give below a more convenient expression for them based on (2.4.2).

**Lemma 2.6** *Let  $S$  be a seed tree of size  $k$  and  $\tau = (\tau, \ell)$  a decorated tree. Then, for any  $n \geq k$*

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S, \{\cdot\} \subseteq S)] = \sum_{d: \tau \rightarrow \mathbb{N}^*} \mathbf{f}(k, n; d, \ell) \cdot D_{\tau, d}(S), \quad (2.4.4)$$

where  $\mathbf{f}(k, n; d, \ell)$  are functions defined as follows (for decorations  $d$  such that  $|d| < 2k - 2$ ).

Let  $(y_n(u) : u \in \tau \cup \{\text{other}\}, n \geq k)$  be a Pólya urn with  $|\tau| + 1$  colours, replacement matrix  $M = (m_{uv})_{u, v \in \tau \cup \{\text{other}\}}$  with

$$m_{uv} = \begin{cases} 1 & \text{if } u = v \neq \text{other}, \\ 1 + \alpha & \text{if } u = v = \text{other}, \\ \alpha & \text{if } u \neq v = \text{other}, \\ 0 & \text{otherwise,} \end{cases}$$

and initial states  $y_k(u) = \alpha + d(u) - 1$  for  $u \in \tau$  and  $y_k(\text{other}) = (1 + \alpha)k - 2 - \sum_{u \in \tau} y_k(u)$ . Then

$$\mathbf{f}(k, n; d, \ell) = \mathbb{E} \left[ \prod_{u \in \tau} [y_n(u) - \alpha]_{\ell(u)} \right].$$

**Remark 2.5** *The actual definition of the variables  $(y_n(u) : u \in \tau \cup \{\text{other}\}, n \geq k)$  is not very important; a more intuitive expression will be used (see (2.4.6)). The important aspect of (2.4.4) is that  $\mathbb{E}[\mathcal{F}_\tau(T_n^S, \{\cdot\} \subseteq S)]$  is factorised between a part that depends on  $n$  but not the seed structure (namely  $\mathbf{f}(k, n; d, \ell)$ ) and one that depends on the seed structure but not on  $n$  (namely  $D_{\tau, d}(S)$ ).*

**Proof** Fix  $S$  and  $\tau = (\tau, \ell)$ . Recall that

$$\mathbb{E}[\mathcal{F}_\tau(T_n, \{\cdot\} \subseteq S)] = \sum_{\phi: \tau \hookrightarrow S} \mathbb{E} \left[ \prod_{u \in \tau} [\deg_{T_n^S}(\phi(u)) - 1]_{\ell(u)} \right], \quad (2.4.5)$$

where the sum is over all embeddings of  $\tau$  in  $S$ .

Let  $\phi$  be an embedding of  $\tau$  in  $S$ . Then the family  $\{\deg_{T_n^S}(\phi(u)) : u \in \tau\}$  has a markovian dynamics as  $n$  increases described as follows:

- for each  $u \in \tau$ , with probability  $\frac{\deg_{T_n^S}(\phi(u)) - 1 + \alpha}{(1 + \alpha)n - 2}$ , we have  $\deg_{T_{n+1}^S}(\phi(u)) = \deg_{T_n^S}(\phi(u)) + 1$  and all other entries remain the same;
- otherwise  $\deg_{T_{n+1}^S}(\phi(u)) = \deg_{T_n^S}(\phi(u))$  for all  $u \in \tau$ .

Then, if we set

$$\begin{aligned} y_n(u) &= \deg_{T_n^S}(\phi(u)) + \alpha - 1, & \text{for } u \in \tau \text{ and} \\ y_n(\text{other}) &= (1 + \alpha)n - 2 - \sum_{u \in \tau} y_n(u) = \sum_{v \in T_n^S \setminus \phi(\tau)} \deg_{T_n^S}(v) + \alpha - 1, \end{aligned}$$

we deduce readily that the family  $(y_n(u) : u \in \tau \cup \{\text{other}\}, n \geq k)$  has the dynamics of a Pólya urn with the replacement matrix  $M$  and initial conditions given in the statement. Thus

$$\mathbb{E} \left[ \prod_{u \in \tau} [\deg_{T_n^S}(\phi(u)) - 1]_{\ell(u)} \right] = \mathbb{E} \left[ \prod_{u \in \tau} [y_n(u) - \alpha]_{\ell(u)} \right] = \mathbf{f}(k, n; d, \ell). \quad (2.4.6)$$

By inserting the above in (2.4.5) and grouping the terms of the sum by the degrees in  $S$  of the embedding, we obtain (2.4.4).

Corollary 2.2 and Lemma 2.6 state that, for seeds  $S \neq S'$  of same size and  $\tau$  which is a minimal non- $(S, S')$ -blind tree, the difference of the observables  $\mathcal{F}_\tau$  for  $T_n^S$  and  $T_n^{S'}$  is a linear combination of functions  $\mathbf{f}(k, n; d, \ell)$ :

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})] = \sum_{d: \tau \rightarrow \mathbb{N}^*} \mathbf{f}(k, n; d, \ell) \cdot [D_{\tau, d}(S) - D_{\tau, d}(S')] \quad (2.4.7)$$

for any decoration  $\ell$  of  $\tau$ . Thus, for our proof of Theorem 2.4, it will be of great interest to study the asymptotics of the functions  $\mathbf{f}(k, n; d, \ell)$  as  $n \rightarrow \infty$ . The relevant result is the following.

**Proposition 2.8** *Let  $S$  be a seed tree of size  $k \geq 2$ ,  $\tau = (\tau, \ell)$  be a decorated tree and  $d: \tau \rightarrow \mathbb{N}^*$  be some decoration of  $\tau$ . Then*

$$n^{-\frac{|\ell|}{1+\alpha}} \cdot \mathbf{f}(k, n; d, \ell) \xrightarrow[n \rightarrow \infty]{} C(k, |\ell|) \cdot \prod_{u \in \tau} [d(u) + \ell(u) + \alpha - 2]_{\ell(u)}, \quad (2.4.8)$$

where  $C(k, |\ell|) > 0$  is some constant depending only on  $k$  and  $|\ell|$ .

The above is a technical result based on the study of Pólya urns; the rest of the section is dedicated to proving it. It is possible to prove (2.4.8) using only the abstract quantities  $(y_n(u) : u \in \tau \cup \{\text{other}\}, n \geq k)$ . We prefer however to use the more visual interpretation of these quantities in terms of the degrees in  $T_n^S$  of some embedding of  $\tau$  in  $S$ .

Fix a seed  $S$  of size  $k$  and  $R$  a subtree of  $S$ . Also fix some decoration  $\ell$  of  $R$ . Set

$$M_n^\ell = \prod_{u \in R} [\deg_{T_n^S}(u) + \alpha + \ell(u) - 2]_{\ell(u)}$$

and

$$\mathcal{W}_n^\ell = \begin{cases} 1 & \text{if } n = k, \\ \prod_{t=k}^{n-1} \left(1 + \frac{|\ell|}{(1 + \alpha)t - 2}\right)^{-1} & \text{if } n \geq k + 1. \end{cases}$$

When the dependance on  $\ell$  is not important, we will drop it from the notation.

**Lemma 2.7** *The sequence  $(M_n \cdot \mathcal{W}_n)_{n \geq k}$  is a martingale and is bounded in  $L^2$ .*

**Proof** We start by proving that  $(M_n \cdot \mathcal{W}_n)_{n \geq k}$  is a martingale. We introduce  $\mathcal{E}_n$  the  $\sigma$ -algebra generated by  $M_k, \dots, M_n$ . When going from  $n$  to  $n+1$  one of two things can happen: either the new vertex of  $T_{n+1}^S$  is attached to a vertex  $u \in R$  or it is attached to some other vertex of  $T_n^S$ . In the first case  $M_n$  is multiplied by  $\frac{\deg_{T_n^S}(u) + \alpha + \ell(u) - 1}{\deg_{T_n^S}(u) + \alpha - 1}$ ; this occurs with probability  $\frac{1}{(1+\alpha)n-2}(\deg_{T_n^S}(u) + \alpha - 1)$ . In the latter case the value of  $M_n$  remains unchanged. Thus

$$\frac{\mathbb{E}(M_{n+1}|\mathcal{E}_n) - M_n}{M_n} = \sum_{u \in R} \frac{\deg_{T_n^S}(u) + \alpha - 1}{(1+\alpha)n-2} \cdot \left( \frac{\deg_{T_n^S}(u) + \alpha - 1 + \ell(u)}{\deg_{T_n^S}(u) + \alpha - 1} - 1 \right) = \frac{|\ell|}{(1+\alpha)n-2}.$$

It follows that  $\mathbb{E}(M_{n+1}|\mathcal{E}_n) = \left(1 + \frac{|\ell|}{(1+\alpha)n-2}\right) \cdot M_n$ , which is to say that  $M_n \cdot \mathcal{W}_n$  is a martingale.

That the martingale  $M_n \cdot \mathcal{W}_n$  is bounded in  $L^2$  follows from:

$$\forall u \in R, [d + \alpha + \ell(u) - 2]_{\ell(u)}^2 \leq [d + \alpha + 2\ell(u) - 2]_{2\ell(u)}$$

and also:

$$(\mathcal{W}_n^\ell)^2 \leq \mathcal{W}_n^{2\ell},$$

where  $2\ell$  is the decoration such that for every  $u \in R$ ,  $(2\ell)(u) = 2 \cdot \ell(u)$ . Indeed, the two above inequalities together imply that  $(M_n^\ell \cdot \mathcal{W}_n^\ell)^2 \leq M_n^{2\ell} \cdot \mathcal{W}_n^{2\ell}$  for every  $n \geq k$ . But since  $M_n^{2\ell} \cdot \mathcal{W}_n^{2\ell}$  is a martingale according to what we have just proved, it is bounded in  $L^1$ , thus  $M_n \cdot \mathcal{W}_n$  is for its part bounded in  $L^2$ .

**Corollary 2.3** *The following convergence holds almost surely and in  $L^1$ :*

$$n^{-\frac{|\ell|}{1+\alpha}} \cdot M_n \xrightarrow[n \rightarrow \infty]{} C(k, |\ell|) \cdot \xi(R), \quad (2.4.9)$$

where  $\xi(R)$  is a random variable of expectation  $\mathbb{E}[\xi(R)] = \prod_{u \in R} [\deg_S(u) + \ell(u) + \alpha - 2]_{\ell(u)}$  and  $C(k, |\ell|) > 0$  is a universal constant depending only on  $k$  and  $|\ell|$ .

**Proof** As a martingale that is bounded in  $L^2$ ,  $(M_n \cdot \mathcal{W}_n)$  converges a.s. and in  $L^1$  when  $n \rightarrow \infty$  to some random variable  $\xi(R)$ . By the  $L^1$  convergence,

$$\mathbb{E}[\xi(R)] = \mathbb{E}[M_k \cdot \mathcal{W}_k] = \prod_{u \in R} [\deg_S(u) + \ell(u) + \alpha - 2]_{\ell(u)}.$$

Finally, as in (2.3.10), a straightforward computation proves that  $\mathcal{W}_n \cdot n^{\frac{|\ell|}{1+\alpha}}$  converges as  $n \rightarrow \infty$  to some constant depending only on  $|\ell|$  and  $k$ . This implies (2.4.9).

**Remark 2.6** *If we apply Corollary 2.3 to  $R$  being formed of a single vertex  $u$  with  $\ell(u) = 1$ , we obtain*

$$n^{-\frac{1}{1+\alpha}} \deg_{T_n^S}(u) \xrightarrow[n \rightarrow \infty]{} C(k, 1) \cdot \xi(u),$$

where  $\mathbb{E}[\xi(u)] = \deg_S(u) + \alpha - 1$  and  $C(k, 1) > 0$  is defined as above. Moreover, it may be shown that  $\xi(u) > 0$  a.s. (see [77, Lemma 8.17]), which is to say that  $\deg_{T_n^S}(u)$  grows as  $n^{\frac{1}{1+\alpha}}$ .

We are now ready to prove Proposition 2.8.

**Proof of Proposition 2.8** Fix some  $S$ ,  $\tau$ ,  $\ell$  and  $d$  as in the proposition. Let  $\phi$  be some perfect embedding of  $(\tau, d)$  in  $S$ . Write  $R = \phi(\tau)$  and keep in mind that  $\deg_S(\phi(u)) = d(u)$  for all  $u \in \tau$ . We will use the notation of Corollary 2.3 for  $R$ .

Using (2.4.6) and the fact that  $\deg_{T_n^S}(u) \rightarrow \infty$  for all  $u \in \tau$ , we have

$$\frac{\prod_{u \in \tau} [y_n(u) - \alpha]_{\ell(u)}}{M_n} = \prod_{u \in \tau} \frac{[\deg_{T_n^S}(\phi(u)) - 1]_{\ell(u)}}{[\deg_{T_n^S}(\phi(u)) + \ell(u) + \alpha - 2]_{\ell(u)}} \xrightarrow{n \rightarrow \infty} 1.$$

Moreover, the ratio is always bounded from above by 1<sup>4</sup>. Thus, by the dominated convergence theorem,

$$\lim_{n \rightarrow \infty} n^{-\frac{|\ell|}{1+\alpha}} \cdot \mathbf{f}(k, n; d, \ell) = \lim_{n \rightarrow \infty} n^{-\frac{|\ell|}{1+\alpha}} \cdot \mathbb{E}[M_n] = C(k, |\ell|) \cdot \prod_{u \in \tau} [d(u) + \ell(u) + \alpha - 2]_{\ell(u)},$$

where the last equality is due to Corollary 2.3.

#### 2.4.4 Proof of Theorem 2.4

In proving Theorem 2.4 we analyse differently the case where the two seeds  $S, S'$  have same size and that where they have distinct sizes. We start with the former.

**Proof of Theorem 2.4 for seeds of common size** Fix  $S, S'$  two distinct seeds of same size  $k \geq 4$ . Let  $\tau$  be a minimal tree which is not  $(S, S')$ -blind. Then, for any decoration  $\ell$  of  $\tau$ , equation (2.4.7) and Proposition 2.8 imply that

$$\lim_{n \rightarrow \infty} n^{-\frac{|\ell|}{1+\alpha}} \cdot (\mathbb{E}[\mathcal{F}_\tau(T_n^S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})]) = \sum_{d: \tau \rightarrow \mathbb{N}^*} \mathbf{f}(\infty; d, \ell) \cdot [D_{\tau, d}(S) - D_{\tau, d}(S')], \quad (2.4.10)$$

where  $\mathbf{f}(\infty; d, \ell) := \lim_{n \rightarrow \infty} n^{-\frac{|\ell|}{1+\alpha}} \cdot \mathbf{f}(k, n; d, \ell) = C(k, |\ell|) \cdot \prod_{u \in \tau} [d(u) + \ell(u) + \alpha - 2]_{\ell(u)}$ . Thus, our goal is to find a decoration  $\ell: \tau \rightarrow \mathbb{N}$  with  $\ell(u) \geq 2$  for all  $u \in \tau$ ,  $|\ell| > 1 + \alpha$  and such that the right-hand side of (2.4.10) is non-zero.

Fix some arbitrary order  $u_1, \dots, u_r$  for the vertices of  $\tau$  (where  $r := |\tau|$ ) and write  $\prec_\ell$  for the lexicographical order induced on the set of decorations of  $\tau$ . Consider the set  $\Delta$  of decorations  $d$  of  $\tau$  with  $D_{\tau, d}(S) \neq D_{\tau, d}(S')$ . This set is finite, since  $D_{\tau, d}(S) = D_{\tau, d}(S') = 0$  for all large enough decorations. Moreover, it is not void since  $\tau$  is not  $(S, S')$ -blind. Let  $d_{\max}$  be the maximal element of  $\Delta$  for  $\prec_\ell$ .

Consider some  $d \in \Delta$  with  $d \neq d_{\max}$  and let  $j = \min\{i : d(u_i) \neq d_{\max}(u_i)\}$ . Then

$$\frac{\mathbf{f}(\infty; d, \ell)}{\mathbf{f}(\infty; d_{\max}, \ell)} = \prod_{i \geq j} \frac{[d(u_i) + \ell(u_i) + \alpha - 2]_{\ell(u_i)}}{[d_{\max}(u_i) + \ell(u_i) + \alpha - 2]_{\ell(u_i)}} \rightarrow 0 \quad (2.4.11)$$

when all  $\ell(u_i)$  for  $i > j$  are fixed and  $\ell(u_j) \rightarrow \infty$ .

Write  $\Delta_j$  for the set of decorations  $d \in \Delta$  with  $d \neq d_{\max}$  and such that  $\min\{i : d(u_i) \neq d_{\max}(u_i)\} = j$ . Then  $\Delta \setminus \{d_{\max}\} = \bigsqcup_{j=1}^r \Delta_j$ . Now let us construct a decoration  $\ell$  with the necessary requirements by first choosing  $\ell(u_r)$ , then  $\ell(u_{r-1})$  etc.

First fix  $\ell(u_r) > 1 + \alpha$  large enough so that

$$\left| \sum_{d \in \Delta_r} \frac{\mathbf{f}(\infty; d, \ell)}{\mathbf{f}(\infty; d_{\max}, \ell)} \cdot [D_{\tau, d}(S) - D_{\tau, d}(S')] \right| \leq \frac{1}{2r}.$$

This is possible by (2.4.11). Observe that the values of  $\ell(u_i)$  for  $i < r$  are irrelevant, as they do not appear in the above. Once  $\ell(u_r)$  is fixed, fix  $\ell(u_{r-1}) \geq 2$  so that

$$\left| \sum_{d \in \Delta_{r-1}} \frac{\mathbf{f}(\infty; d, \ell)}{\mathbf{f}(\infty; d_{\max}, \ell)} \cdot [D_{\tau, d}(S) - D_{\tau, d}(S')] \right| \leq \frac{1}{2r}.$$

Again, (2.4.11) shows that the values  $\{\ell(u_i); i < r - 1\}$  do not appear in the above and that such a choice of  $\ell(u_{r-1})$  is possible regardless of the value of  $\ell(u_r)$  fixed before.

---

4. This is proved individually for each term in the product: it is clear when  $\ell(u) \geq 1$ ; when  $\ell(u) = 0$ , the ratio is 1

Continue as such until all values  $\{\ell(u) : u \in \tau\}$  are fixed. The resulting decoration  $\ell$  has  $\ell(u) \geq 2$  for all  $u \in \tau$ <sup>5</sup>, it is such that  $|\ell| \geq \ell(u_r) > 1 + \alpha$ , and satisfies

$$\left| \sum_{d \in \Delta_j} \frac{\mathbf{f}(\infty; d, \ell)}{\mathbf{f}(\infty; d_{\max}, \ell)} \cdot [D_{\tau, d}(S) - D_{\tau, d}(S')] \right| \leq \frac{1}{2r}, \quad \forall j = 1, \dots, r.$$

When summing the above, we find

$$\left| \sum_{d \in \Delta \setminus \{d_{\max}\}} \mathbf{f}(\infty; d, \ell) \cdot [D_{\tau, d}(S) - D_{\tau, d}(S')] \right| \leq \frac{1}{2} \mathbf{f}(\infty; d_{\max}, \ell).$$

Now, since  $D_{\tau, d_{\max}}(S) \neq D_{\tau, d_{\max}}(S')$ , this implies that

$$\left| \sum_{d: \tau \rightarrow \mathbb{N}^*} \mathbf{f}(\infty; d, \ell) \cdot [D_{\tau, d}(S) - D_{\tau, d}(S')] \right| \geq \frac{1}{2} \mathbf{f}(\infty; d_{\max}, \ell) > 0,$$

which was the desired condition.

**Proof of Theorem 2.4 for seeds of different sizes** Fix two seeds  $S, S'$  of sizes  $k$  and  $k'$ , respectively, with  $3 \leq k' < k$ . Due to the Markov property, the law of  $(T_n^{S'})_{n \geq k}$  is a linear combination of laws  $(T_n^R)_{n \geq k}$  with  $R$  ranging over the different values taken by  $T_k^{S'}$ . Write  $\mathcal{T}_k$  for the set of trees of size  $k$  with  $\mathbb{P}(T_k^{S'} = R) \neq 0$ .

First we claim that there exist two trees in  $\mathcal{T}_k$  with different maximal degree. Indeed, one possible way of going from  $S'$  to a tree in  $\mathcal{T}_k$  is to always attach the new vertices to leaves. In this scenario, the maximal degree of the resulting tree is the same as that of  $S'$  (since that in  $S'$  is at least 2). In conclusion

$$\mathbb{P}(\max\{\deg_{T_k^{S'}}(u) : u \in T_k^{S'}\} = \max\{\deg_{S'}(u) : u \in S'\}) > 0.$$

Another is to always attach the new vertex to the one of maximal degree. The resulting maximal degree in  $T_k^{S'}$  would then be  $k - k'$  more than that in  $S'$ :

$$\mathbb{P}(\max\{\deg_{T_k^{S'}}(u) : u \in T_k^{S'}\} = \max\{\deg_{S'}(u) : u \in S'\} + k - k') > 0.$$

Thus, there exist two trees in  $\mathcal{T}_k$  with distinct maximal degree, as claimed.

Fix  $\tau$  to be the tree formed of a single vertex. A decoration for  $\tau$  is then simply an integer number. This tree has no subtree, hence Proposition 2.7 applies to it. For any decoration  $\ell \in \mathbb{N}$  of  $\tau$ , equation (2.4.10) adapts to

$$\lim_{n \rightarrow \infty} n^{-\frac{\ell}{1+\alpha}} (\mathbb{E}[\mathcal{F}_\tau(T_n^S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})]) = \sum_{d \in \mathbb{N}^*} \mathbf{f}(\infty; d, \ell) \sum_{R \in \mathcal{T}_k} \mathbb{P}(T_k^{S'} = R) \cdot [D_d(S) - D_d(R)],$$

where  $\mathbf{f}(\infty; d, \ell) = C(k, \ell) \cdot [d + \ell + \alpha - 2]_\ell$  and  $D_d(S)$  is simply the number of vertices of degree  $d$  in  $S$ .

Write  $d_{\max}$  for the maximal number  $d \in \mathbb{N}^*$  with

$$\sum_{R \in \mathcal{T}_k} \mathbb{P}(T_k^{S'} = R) \cdot [D_d(S) - D_d(R)] \neq 0. \quad (2.4.12)$$

First observe that the set of values  $d$  satisfying (2.4.12) is finite since  $D_d(S) = D_d(R) = 0$  for all  $R \in \mathcal{T}_k$  provided  $d$  is large enough. Second, notice that there exist at least one  $d$  satisfying (2.4.12) since the maximal degree of  $R \in \mathcal{T}_k$  is not constant.

Now, by (2.4.11),

$$\begin{aligned} \lim_{\ell \rightarrow \infty} \sum_{d \in \mathbb{N}^*} \frac{\mathbf{f}(\infty; d, \ell)}{\mathbf{f}(\infty; d_{\max}, \ell)} \sum_{R \in \mathcal{T}_k} \mathbb{P}(T_k^{S'} = R) \cdot [D_d(S) - D_d(R)] \\ = \sum_{R \in \mathcal{T}_k} \mathbb{P}(T_k^{S'} = R) \cdot [D_{d_{\max}}(S) - D_{d_{\max}}(R)] \neq 0. \end{aligned}$$

In conclusion, one may fix  $\ell > 1 + \alpha$  (and implicitly  $\ell \geq 2$ ) so that

$$\lim_{n \rightarrow \infty} n^{-\frac{\ell}{1+\alpha}} (\mathbb{E}[\mathcal{F}_\tau(T_n^S)] - \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})]) \neq 0.$$

---

5.  $\ell(u_r) \geq 2$  is ensured by the fact that  $\ell(u_r) > 1 + \alpha$

### 2.4.5 Proof of Theorem 2.1

**Proof** Fix  $S, S'$  two distinct seeds of sizes at least 3. Let  $\tau$  be the decorated tree given by Theorem 2.4 for these two seeds. Due to our assumptions on  $\tau$ , Theorem 2.2 applies to it, and we have

$$\mathbb{E}[\mathcal{F}_\tau(T_n^S)^2] = \mathcal{O}\left(n^{\frac{2|\ell|}{1+\alpha}}\right) \quad \text{and} \quad \mathbb{E}[\mathcal{F}_\tau(T_n^{S'})^2] = \mathcal{O}\left(n^{\frac{2|\ell|}{1+\alpha}}\right).$$

Lemma 2.1 then yields  $\liminf_{n \rightarrow \infty} d_{TV}(\mathcal{F}_\tau(T_n^S), \mathcal{F}_\tau(T_n^{S'})) > 0$ , which in turn implies Theorem 2.1.

## 2.5 Open problems and future research

**Other attachment mechanisms** As mentioned in the introduction, our approach uses the fact that the attachment mechanism is affine to couple the evolution of trees starting from distinct seeds of same size; see the coupling of Section 2.2.2 and its use in Section 2.4.1 and 2.4.2 to deduce Corollary 2.2.

Suppose now that we consider a different attachment model, where the new vertex is attached to a vertex  $u \in T_n$  with probability proportional to  $g(\deg_{T_n}(u))$  for some function  $g : \mathbb{N} \rightarrow (0, +\infty)$ . Thus (2.1.1) becomes

$$\mathbb{P}(u_N = u \mid T_k^S, \dots, T_N^S \text{ with } T_N^S = T) = \frac{g(\deg_T(u))}{\sum_{v \in T} g(\deg_T(v))} \quad \forall u \in T.$$

Notice that, when  $g$  is not affine, the denominator above depends on the structure of  $T$ , and the attachment probabilities cease to be a local function of  $u$ . In other words, the sequence of drawn vertices loses its exchangeability. As a consequence, it is not possible anymore to construct a coupling between sequences  $(T_n^S)_{n \geq k}$  and  $(T_n^{S'})_{n \geq k}$  as in Section 2.2.2, where  $S$  and  $S'$  are two seeds of size  $k$ . Nevertheless, if  $S$  and  $S'$  have same degree sequences, then regardless of the form of  $g$ , one may couple the evolution of  $T_n^S$  and  $T_n^{S'}$  so that the sets of connected components of  $T_n^S \setminus S$  and  $T_n^{S'} \setminus S'$  are identical (here  $T_n^S \setminus S$  stands for the graph obtained from  $T_n^S$  by removing all edges of  $S$ ). The equivalent of Corollary 2.2 may then be proved in the same manner.

This hints to the possibility of seed recognition (*i.e.* Theorem 2.1) for a much larger set of models. One particular example of interest is when  $g(k) = k^\beta$  for some  $\beta \in (0, 1)$  [94]. In such models, the largest degree in  $T_n^S$  is of order  $(\log n)^{\frac{1}{1-\beta}}$ , hence much smaller than in the affine case [29, Thm. 22]. The case  $\beta = 1$  is that of the linear preferential attachment model treated in [54]; when  $\beta = 0$  we obtain the uniform attachment model of [38]. For  $\beta \in (0, 1)$  we expect the same type of result to hold, and plan to investigate this in future work.

Let us also mention that, when  $\beta > 1$  a single vertex of  $T_n^S$  has degree tending to infinity, all other degrees are a.s. bounded by a constant, as is the diameter of  $T_n^S$  [118, Thm. 1.2].

**Finding the seed** Our result may be understood as follows: given a large (but uniform in  $n$ ) number of samples of  $T_n^S$ , one may recover  $S$  with high precision. A related question is to locate  $S$  given a single instance of  $T_n^S$ . One may not hope to do this with high probability, but is it possible to do it with uniformly positive probability? Results in this direction were obtained in [37, 61, 100] for the uniform and linear preferential attachment models.

One may ask whether having a uniformly positive chance of locating the seed is equivalent to Theorem 2.1. Moreover, for specific cases such as the  $\alpha$ -PA, is there an explicit algorithm that locates the seed?

**Beyond trees** Finally, one may consider randomly growing graphs, rather than trees. Indeed, imagine a model where vertices are added one by one to a growing graph, with each new vertex being attached to each old vertex independently, with a probability depending on the size of the graph and on the degree of the old vertex. This offers great freedom for the choice of the attachment probability, but the resulting graph ceases to be a tree; it may even have multiple connected components. These aspects render the study of such models more delicate. However, we believe that for certain attachment mechanisms - affine for instance - an equivalent of Theorem 2.1 would remain valid.



## Chapter 3

# Percolation on supercritical causal triangulations

This chapter is adapted from the work in progress [105].

We study oriented percolation on random causal triangulations, those are random planar graphs obtained roughly speaking by adding horizontal connections between vertices of an infinite tree. When the underlying tree is a geometric Galton–Watson tree with mean  $m > 1$ , we prove that the oriented percolation undergoes a phase transition at  $p_c(m)$ , where

$$p_c(m) = \frac{\eta}{1 + \eta} \quad \text{with} \quad \eta = \frac{1}{m + 1} \sum_{n \geq 0} \frac{m - 1}{m^{n+1} - 1}.$$

We establish that strictly above the threshold  $p_c(m)$ , infinitely many infinite components coexist in the map. This is a typical percolation result for graphs with a hyperbolic flavour. We also demonstrate that large critical oriented percolation clusters converge after rescaling towards the Brownian continuum random tree. The proof is based on a Markovian exploration method, similar in spirit to the peeling process of random planar maps.

### Sommaire

---

<b>3.1</b>	<b>Introduction and main results . . . . .</b>	<b>92</b>
<b>3.2</b>	<b>The supercritical causal triangulation of the half-plane . . . . .</b>	<b>95</b>
3.2.1	Definition of the model . . . . .	95
3.2.2	Basic properties . . . . .	96
3.2.3	A peeling exploration and a spatial Markov property . . . . .	98
<b>3.3</b>	<b>Percolation on SCTHP . . . . .</b>	<b>99</b>
3.3.1	Exploration of the percolation cluster via a peeling process . . . . .	101
3.3.2	The volume of the percolation cluster . . . . .	105
3.3.3	Scaling limit of the critical percolation cluster . . . . .	107
3.3.4	From SCTHP to SCT: an union of critical percolation clusters . . . . .	110
<b>3.4</b>	<b>Percolation on SCT . . . . .</b>	<b>117</b>
3.4.1	SCT and wide trees . . . . .	117
3.4.2	Large critical percolation clusters in SCT . . . . .	120
3.4.3	The number of clusters in the supercritical phase . . . . .	125

---



### 3.1 Introduction and main results

In these pages we study oriented percolation on a random "hyperbolic" type of planar graph which is constructed as follows. Let  $t$  be any infinite rooted plane tree. For any  $r \geq 1$ , we add horizontal edges between consecutive vertices belonging to the boundary of the ball of radius  $r$  in  $t$ , so that it forms a cycle. Then, we triangulate each face by linking vertices on its bottom side—except the rightmost one—to the top-right vertex. See Figure 3.1 for an illustration. The planar triangulation that we get is called a *causal triangulation*.

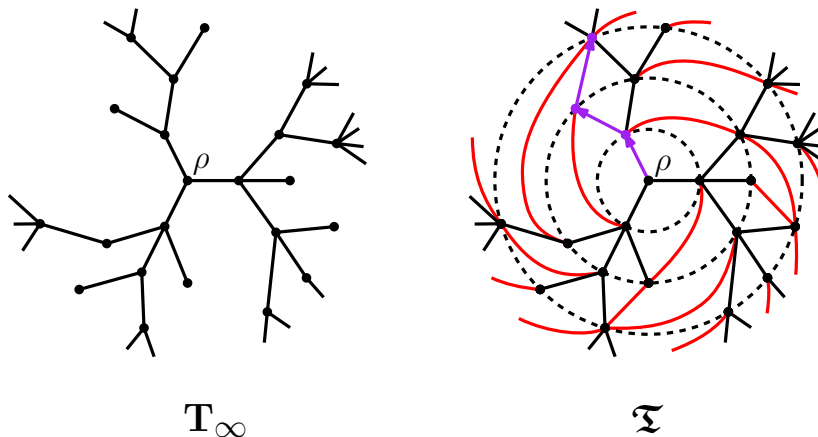


Figure 3.1: A piece of an infinite supercritical Galton–Watson tree  $\mathbf{T}_\infty$ , rooted in  $\rho$ , and its associated causal triangulation  $\mathfrak{T}$ . In dashed lines are the horizontal edges connecting at each level consecutive vertices. They do not play any role in our work. Red edges are added to triangulate the map. Edges in purple illustrate the orientation—away from the root—chosen in our percolation model.

We are interested in the case where  $t$  is a random supercritical Galton–Watson tree conditioned to survive, denoted by  $\mathbf{T}_\infty$ . Specifically, we shall focus on the case where the offspring distribution  $\boldsymbol{\mu}$  of the tree is geometric. That is for some  $\alpha \in (1/2, 1)$ , we have for every  $k \geq 0$ ,

$$\boldsymbol{\mu}(k) = \boldsymbol{\mu}_\alpha(k) := \alpha^k(1 - \alpha).$$

We recall that the mean of  $\boldsymbol{\mu}_\alpha$  is

$$m := \alpha \cdot (1 - \alpha)^{-1} > 1.$$

We will denote by  $\mathfrak{T}$  the random causal triangulation resulting from  $\mathbf{T}_\infty$ , and  $\mathbb{P}_\mathfrak{T}$  the associated probability measure. We call the model **supercritical causal triangulation**, abbreviated to SCT. In what follows, we will write  $\mathbf{GW}_\beta$  to designate the distribution of a Galton–Watson tree with offspring law  $\boldsymbol{\mu}_\beta$ —without conditioning on the survival event—and where  $\beta$  is more generally in  $(0, 1)$ .

**In the rest of the paper, we assume that  $\alpha \in (1/2, 1)$  is fixed. We will omit  $\alpha$  in our notation most of the time, although what will be stated will often depend on it.**

Causal triangulations first appeared in the physics literature [13]. Random versions have been recently studied from a mathematical point of view [56, 43]. In particular, the author of [43] focuses on the causal maps derived from a supercritical (non necessarily geometric) Galton–Watson tree conditioned to survive. Causal maps are constructed from a rooted tree in the same manner as causal triangulations, except that they are not triangulated. It is shown in the mentioned paper that they have "hyperbolic" properties (anchored Gromov hyperbolicity, positive speed of the associated simple random walk, a non-trivial Poisson boundary, etc).

**Results.** Our contribution here is to grasp the SCT model through the lens of *directed percolation*. Fix a (deterministic) causal triangulation  $\theta$ . The edges are directed away from the root,

**the horizontal connections play no role.** A Bernoulli bond percolation is applied on them: for some fixed parameter  $p \in [0, 1]$ , each (directed) edge is declared either *open* or *closed* with probabilities  $p \in [0, 1]$  and  $1 - p$  respectively, independently of the others. We denote by  $\mathbb{P}_{p,\theta}$  the distribution of the percolation process on  $\theta$ . Now, we introduce  $\mathbb{P}_p$  the overall distribution

$$\int \mathbb{P}_{\mathfrak{T}}(d\theta) \int d\mathbb{P}_{p,\theta}$$

which averages the percolation process on supercritical causal triangulations. This is an *annealed* percolation distribution, while  $\mathbb{P}_{p,\theta}$  is a *quenched* one.

In our first main result, we show the existence of a non trivial *annealed* phase transition:

**Theorem 3.1** *Let  $\mathcal{C}$  be the (directed) percolation cluster of the root in  $\mathfrak{T}$ . We define  $\Theta(p)$  the annealed probability that  $\mathcal{C}$  is infinite and set*

$$p_c = p_c(m) := \frac{\eta}{1 + \eta} \quad \text{with} \quad \eta = \frac{1}{m + 1} \sum_{n \geq 0} \frac{m - 1}{m^{n+1} - 1}. \quad (3.1.1)$$

Then:

$$\Theta(p) = \begin{cases} 0 & \text{if } p \leq p_c \\ > 0 & \text{if } p > p_c. \end{cases} \quad (3.1.2)$$

Furthermore, when  $p_c < p < 1$ , there are  $\mathbb{P}_p$ -almost surely infinitely many disjoint infinite clusters in  $\mathfrak{T}$ .

Since the environment is directed, the notion of disjoint clusters may seem ambiguous. Imagine indeed two clusters  $\mathcal{C}_u$  and  $\mathcal{C}_v$  associated to two distinct vertices  $u$  and  $v$ , which do not share any vertex in common, but are both included in a bigger cluster emanating from a third vertex nearer to the root. We do not definitely want to consider them as disjoint. So, to be more precise, our statement is that as  $p_c < p < 1$ , there are  $\mathbb{P}_p$ -almost surely infinitely many disjoint *non-directed* clusters (i.e. when edges are not directed anymore), each containing at least one infinite *directed* cluster within it.

We easily derive from Theorem 3.1.1 a *quenched* version of it. In a classic way, we define for any causal triangulation  $\theta$ :

$$p_c(\theta) := \inf \{p \in [0, 1] : \mathbb{P}_{p,\theta}(\text{there exists an infinite (directed) cluster in } \theta) = 1\}.$$

The theorem ensures that  $p_c(\theta)$  is  $\mathbb{P}_{\mathfrak{T}}$ -almost surely equal to  $p_c$ . Also, we have  $\mathbb{P}_{\mathfrak{T}}$ -almost surely infinitely many disjoint infinite clusters in  $\theta$  with probability one.

The graph of  $p_c$  as a function of  $\alpha$  is plotted in Figure 3.2. We remark that  $p_c(\alpha) \rightarrow 1$  as  $\alpha \rightarrow 1/2$ . It is worth to investigate the case  $\alpha = 1/2$ . The Galton–Watson tree  $\mathbf{T}$  is then critical and almost surely finite. If we condition its size to be increasingly large, an infinite random tree still emerges as local limit in distribution, namely the Kesten’s tree with offspring distribution  $\mu_{1/2}$  [2]. Since the latter is infinite, a causal triangulation can be defined from it. This infinite tree is furthermore the local limit as  $\alpha \rightarrow (1/2)^+$  of Kesten’s trees with offspring distribution  $\mu_\alpha$ . Naturally, the convergence also holds for the related causal triangulations. Now, we claim that the percolation threshold of the causal triangulation derived from the Kesten’s tree with offspring distribution  $\mu_\alpha$  is equal to  $p_c$  as defined in Theorem 3.1. This is due to the fact that the distribution of the tree is absolutely continuous with respect to that of a Galton–Watson tree with offspring distribution  $\mu_\alpha$ , conditioned to survive [102]. Finally, since the critical percolation threshold is continuous with respect to the local topology, we deduce that the phase transition indeed degenerates and  $p_c = 1$  when  $\alpha = 1/2$ . Although the percolation problem is trivial on the critical causal triangulation, its geometric and spectral properties are not and have been studied in [64] and [56].

The annealed phase transition described in Theorem 3.1 is of the same nature when bond percolation is performed on the underlying Galton–Watson tree alone. Only the value of the threshold  $p_c$  change into  $m^{-1}$ . It stems from the fact that in such context, the cluster of the root vertex is itself a Galton–Watson tree of mean  $p \cdot m$ . The conclusion is identical by taking for instance an infinite regular tree instead of a supercritical Galton–Watson tree. Such kind of phase

transition—including a non degenerate phase  $p_c < p < p_u$  where infinitely many infinite clusters coexist in the map—is typical of percolation on "hyperbolic" type of graphs—that is roughly those where the volume of balls of radius  $r$  grows exponentially fast as  $r \rightarrow +\infty$ . See [23] or [80] for a deterministic environment. Similar results have also been established in some random environment context. In [121] for site percolation on hyperbolic half-planar triangulations, or in [53] for bond percolation on hyperbolic bipartite half-planar maps. In the former reference, the author furthermore shows the existence of another non degenerate phase  $p_u < p$ , where there is exactly one (and only one) infinite cluster in the map. In our case, the phase is unfortunately trivial with  $p_u = 1$ , as it is also in the tree set-up mentioned above.

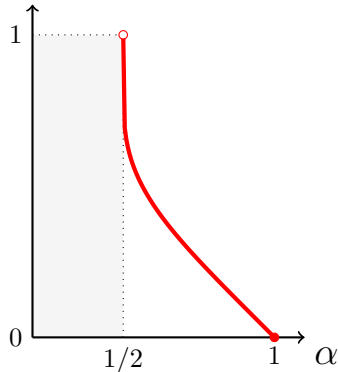


Figure 3.2: Plot of the map  $\alpha \mapsto p_c(\alpha)$  on the interval  $(1/2, 1)$ .

In a second result, we compute several annealed exponents related to percolation cluster:

**Theorem 3.2** *We have:*

- (i)  $\mathbb{P}_{p_c}(|\mathcal{C}| \geq n) = \Omega(n^{-1/2})$  and  $\mathbb{P}_{p_c}(|\mathcal{C}| \geq n) = \mathcal{O}(\log(n) \cdot n^{-1/2})$ ;
- (ii)  $C^{-1} \cdot (p - p_c) \leq \Theta(p) \leq C \cdot (p - p_c)$  for any  $p > p_c$  and some constant  $C > 0$ .

Here and later, for any two sequences  $f, g : \mathbb{N} \mapsto (0, +\infty)$ , we use the notation  $f(n) = \mathcal{O}(g(n))$ , resp.  $f(n) = \Omega(g(n))$ , if the ratio  $f/g$  is bounded below, resp. above, by some positive constant  $C > 0$ . Sometimes, we also write  $f(n) = o(g(n))$  to mean that  $f(n)/g(n) \rightarrow 0$ . We expect the second asymptotic of item (i) to hold with no logarithmic correction to the polynomial term. We insist on the fact that exponents of Theorem 3.2 are annealed versions, i.e. averaged on the random map and the percolation process. We do not know if quenched exponents exist and if their values coincide with the annealed ones. The same critical and off-critical exponents arise in the context of bond percolation performed on the underlying Galton–Watson tree, or on some deterministic hyperbolic graphs as well [80]. We do not have knowledge of any equivalent result in the literature about other model of random hyperbolic maps. See [17, 57, 71] or [53] to get a more general background, which are works dealing with non-hyperbolic random environment, where it turns out that critical exponents significantly differ from what is observed in our model. We expect the second asymptotic of item (i) to hold with no logarithmic correction to the polynomial term.

In our last main result, we establish that a large critical oriented percolation cluster admits the Brownian continuum random tree as scaling limit:

**Theorem 3.3** *At  $p = p_c$ , we have:*

$$(n^{-1/2} \cdot \mathcal{C} \mid |\mathcal{C}| \geq n) \xrightarrow[n \rightarrow +\infty]{d_{GH}} \kappa' \cdot \mathcal{T}_{\geq 1}, \quad (3.1.3)$$

where  $\kappa' = \kappa'(\alpha) > 0$  is a positive number only depending on  $\alpha$  and  $\mathcal{T}_{\geq 1}$  is the Continuum Random Tree of mass greater than 1 [6, 7]. The convergence (3.1.3) holds in distribution for the Gromov–Hausdorff distance.

Again, a comparable behaviour is witnessed in the Galton–Watson tree cleared of the additional edges of the causal triangulation. Indeed, at the critical threshold  $p_c = m^{-1}$ , the percolation cluster

of the root is distributed as a critical Galton–Watson. When the latter is conditioned to have a large size, it converges to the CRT [7, 97]. The emergence of such scaling limit in a hyperbolic context is not new. See for instance [49] where the authors prove that the CRT is the limit of long Brownian bridges in hyperbolic space. For a precise definition of the Gromov–Hausdorff topology and of the metric space  $\mathcal{T}_{\geq 1}$ , we refer the reader to Section 3.3.3.

**Techniques.** The main tool to prove Theorems 3.1, 3.2 and 3.3 is a Markovian exploration of  $\mathfrak{T}$  along the directed cluster of the origin. More precisely, we show that in a simplified "half-plane" model  $\tilde{\mathfrak{T}}$  of  $\mathfrak{T}$ , we can explore the underlying map step-by-step using an algorithm tailored to the percolation. This exploration yields a random walk with independent increments which roughly does the contour of our percolation cluster. This is reminiscent of the peeling process in the theory of random planar maps, and in particular of the fact that percolation on random half-planar maps can be studied (using the peeling process) via a random walk with independent increments, see [15, 16] for the pioneer works of Angel on the subject and [53] for a comprehensive treatment. With this tool at hands, a version of Theorems 3.1, 3.2 and 3.1.3 is easily established in our simplified model  $\tilde{\mathfrak{T}}$ . The results are then transferred back to  $\mathfrak{T}$ .

**Organisation of the paper** In Section 3.2, we introduce the toy model of random triangulations of the half-plane, and outline its main properties. We describe a peeling procedure for these maps which exhibits a spatial Markov property. In Section 3.3, we study percolation on the toy model and prove an equivalent of our three main theorems. Section 3.3.4 plays a special role, as an intermediate section where we generalize some of the previous results and pave the way to the analysis of the SCT model, which is finally carried out in Section 3.4.

**Acknowledgement** I deeply thank Nicolas Curien for his careful reading of the earlier versions of this work.

## 3.2 The supercritical causal triangulation of the half-plane

We present in a first section the definition of a random model of triangulations of the half-plane, which is supposed to approach SCT. The idea is that it will capture essential features of  $\mathfrak{T}$  at large distances from the root vertex. In the next two sections, we outline some basic properties of this model, and introduce a Markovian exploration of the map.

### 3.2.1 Definition of the model

We start by introducing a family of random triangulation of the cylinder  $\mathbb{R} \times [0, 1]$ , with vertex set  $\mathbb{N} \times \{0, 1\}$ . The maps are then *labelled*. Any face of such a map is a triangle, which has to be either *top-oriented*—if it shares an edge with the upper boundary of the cylinder—or *bottom-oriented*—if it does with the lower boundary. We can rank them according to their order of apparition from left to right. In our random model, we assume that each triangle is top-oriented with probability  $\alpha$ , or bottom-oriented with probability  $1 - \alpha$ , independently of the others. Then, the triangulation can be summed up by a sequence of i.i.d. biased Bernoulli trials of parameter  $\alpha$ , where the  $i$ -th trial indicates whether the  $i$ -th leftmost triangle is top or bottom-oriented.

Consider now an i.i.d. sequence (indexed by  $\mathbb{Z}$ ) of such random maps. We glue them so that for any  $i \in \mathbb{Z}$ , the lower boundary of the  $i$ -th cylinder coincides with the upper boundary of the  $i - 1$ -th. After a simple relabelling operation of vertices, it yields a random triangulation of the half-plane  $\mathbb{R}_+ \times \mathbb{R}$ , denoted by  $\tilde{\mathfrak{T}}$ , with vertex set  $\mathbb{N} \times \mathbb{Z}$ . We call it  $\alpha$ -**supercritical causal triangulation of the half-plane**, abbreviated to  $\alpha$ -SCTHP or just SCTHP. See Figure 3.3 for an illustration.

As in the SCT model, **horizontal edges here never play an effective role** since our aim is to study directed percolation with edges being oriented in the bottom-up direction. This is why they are drawn in dashed lines in our figures. Their presence is however helpful to understand the peeling algorithm that we will introduce later.

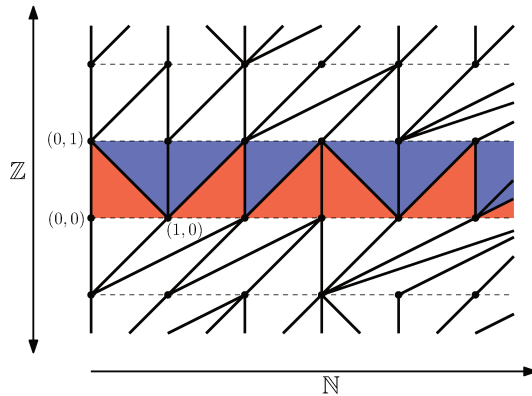


Figure 3.3: A portion of a causal triangulation of the half-plane  $\tilde{\mathfrak{T}}$ . In the middle cylinder between heights 0 and 1, top-oriented triangles are coloured in blue, bottom-oriented triangles in red.

### 3.2.2 Basic properties

Some natural notion of direction—both vertically and horizontally—arises in SCTHP. For any vertex  $v = (i, j)$  of a SCTHP, we will designate by  $h(v)$ —the *height of  $v$* —the second coordinate  $j$ . Vertices  $w = (i', j)$ —with same height than  $v$ —shall be said *on the left* of  $v$  if  $i' \leq i$ , or *on the right* if  $i' \geq i$ . Given the structure of SCTHP, the vertex  $v$  is necessarily adjacent to vertices  $w$  which are either such as

$$h(w) = h(v) - 1, \quad \text{or} \quad h(w) = h(v), \quad \text{or also} \quad h(w) = h(v) + 1.$$

Vertices of the first kind shall be called *parents of  $v$* , those of the last kind *offspring of  $v$* . We remark that if  $w$  is on the left, resp. on the right, of  $v$ , then its offspring are on the left, resp. on the right, of offspring of  $v$ . The same holds for the parents.

Of course, the parent and offspring notions can be reversed, and it allows to exhibit some nice duality property of  $\alpha$ -SCTHP. Indeed, by flipping vertically the map, we transform a top-oriented triangle into a bottom-oriented, and vice versa. The  $\alpha$ -SCTHP becomes an  $(1 - \alpha)$ -SCTHP.

It is worthwhile to note that some tree structure underlies SCTHP:

**Definition 3.1 (Ascending and descending tree)** *For any vertex  $v$  of a SCTHP, we define its associated ascending tree  $\mathfrak{t}_v^\uparrow$  as the plane tree rooted in  $v$ , containing all its offspring except the rightmost one, and the ascending trees associated to these offspring. Conversely, we also define its associated descending tree  $\mathfrak{t}_v^\downarrow$  as the plane tree rooted in  $v$ , containing all its parents except the rightmost one, and the descending trees associated to these offspring. See Figure 3.4 below for an illustration.*

At any given height  $h \in \mathbb{Z}$ , the collection of ascending and descending trees emerging from vertices with height  $h$  is enough data to recover the whole map. Indeed, these ascending trees, resp. descending trees, form a partition of the set of vertices with height greater than  $h$ , resp. less than  $h$ . Edges to add to get the associated SCTHP are deterministically deduced from the trees, in the same way that the SCT  $\mathfrak{T}$  was obtained from the tree  $\mathbf{T}$ . See Figure 3.4 for an illustration.

As with parent and offspring, the notions of ascending and descending trees switch by flipping vertically the SCTHP: the ascending tree of some vertex  $v$  becomes its descending tree in the new map, and conversely.

The ascending and descending trees have actually a well-known distribution:

**Proposition 3.1** *For any height  $h \in \mathbb{Z}$ , the ascending trees  $(\mathfrak{t}_v^\uparrow)_v$  and the descending trees  $(\mathfrak{t}_v^\downarrow)_v$  emerging from vertices with height  $h$  are all mutually independant. Moreover, the ascending trees are distributed like  $\mathbf{GW}_\alpha$ , and the descending trees like  $\mathbf{GW}_{1-\alpha}$ .*

**Proof of Proposition 3.1** As explained in Section 3.2.1, triangles in a layer of the SCTHP are i.i.d. and top-oriented with probability  $\alpha$ . This gives some trivial spatial Markov property: assume

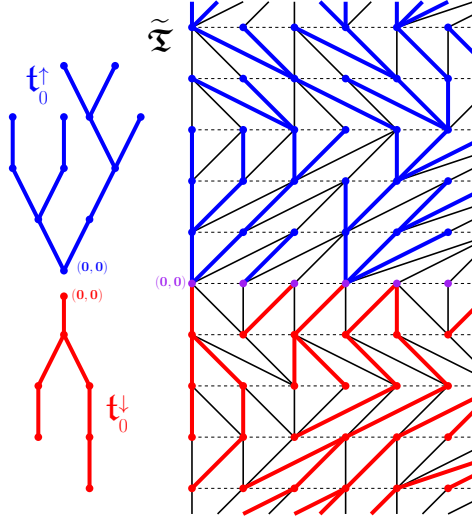


Figure 3.4: Ascending (in blue) and descending (in red) trees emerging from vertices of  $\mathbb{N} \times \{0\} \subset \tilde{\mathfrak{T}}$ . Note how large is  $\mathfrak{t}_{(3,0)}^\uparrow$  while  $\mathfrak{t}_{(2,0)}^\uparrow$  and  $\mathfrak{t}_{(1,0)}^\downarrow$  are for instance made of a single vertex. On the left-hand side of the figure are depicted the ascending and tree associated to the origin vertex. The thin continuous or dashed black lines are the edges deterministically added to recover the whole map, in case where the only provided data are the blue and red trees. The operation is the same than that to build  $\mathfrak{T}$  from  $\mathbf{T}$  in the SCT model. See Figure 3.1.

that we know the orientation of the  $K$  first triangles (ordered from left to right), the next triangles remain i.i.d. with the same distribution. In particular, the probability that the next  $L$  first triangles are top-oriented is  $\alpha^L = \mu_\alpha(\cdot \geq L)$ . Since the number of offspring of some vertex  $v$  is equal to the number of top-oriented triangles which emerge from it, the above facts imply that it is distributed as  $\mu_\alpha$  and independent of the corresponding quantity for the other vertices. Triangulations of the cylinders being independent, we deduce that the ascending trees  $(\mathfrak{t}_v^\uparrow)_v$  emerging from vertices with some height  $h \in \mathbb{Z}$  are i.i.d. with distribution  $\mathbf{GW}_\alpha$ . Due to duality, a similar conclusion holds for the descending trees, but the offspring distribution is in this case  $\mu_{1-\alpha}$ . Finally, the ascending and descending trees are mutually independent since the former are measurable with respect to cylinders above height  $h$ , while the latter are with respect to those below height  $h$ .

A straightforward consequence of Proposition 3.1 is that ascending trees in SCTHP are supercritical Galton–Watson trees—since  $\sum_{k>0} k \cdot \mu_\alpha(k) = \frac{\alpha}{1-\alpha} > 1$  when  $\alpha \in (\frac{1}{2}, 1)$ —while descending trees are subcritical. Later on, it will be useful to have some more information on the latter. Below, we prove that the distribution of their height is exponentially tailed. We also express the first moment as an infinite series.

**Proposition 3.2** Consider  $\mathfrak{t}^\downarrow \stackrel{(d)}{\sim} \mathbf{GW}_{1-\alpha}$  for  $\alpha > 1/2$ . We note  $\mathbf{h}(\cdot)$  the function returning the height of any rooted tree. Recall that  $m := \frac{\alpha}{1-\alpha} > 1$ . Then:

$$\mathbb{P}(\mathbf{h}(\mathfrak{t}^\downarrow) \geq n) = \frac{m-1}{m^{n+1}-1} \quad \text{and so} \quad \mathbb{E}[\mathbf{h}(\mathfrak{t}^\downarrow)] = \sum_{n \geq 1} \frac{m-1}{m^{n+1}-1}.$$

**Proof of Proposition 3.2** Let  $(\mathbf{Z}_n)_{n \geq 0}$  be a Galton–Watson process started from one particle and with  $\mu_{1-\alpha}$  as offspring distribution. We have:

$$\mathbb{E}[\mathbf{h}(\mathfrak{t}^\downarrow)] = \sum_{n \geq 0} \mathbb{P}(\mathbf{h}(\mathfrak{t}^\downarrow) \geq n) \stackrel{\text{Proposition 3.1}}{=} \sum_{n \geq 0} 1 - \underbrace{\mathbb{P}(\mathbf{Z}_n = 0)}_{:= u_n} \quad (3.2.1)$$

The stake is to derive an exact formula for the  $u_n$ . It is classic and easy to see that  $u_n$  satisfies  $u_0 = 0$ , and for every  $n \geq 0$ , the recurrence relation

$$u_{n+1} = \phi(u_n), \quad (3.2.2)$$

where  $\phi$  is the generative function  $\phi : [0, \frac{1}{1-\alpha}) \rightarrow \mathbb{R}_+$  of the distribution  $\mu_{1-\alpha}$ , which is equal to

$$\phi(z) := \sum_{n \geq 0} \mu_{1-\alpha}(n) \cdot z^n = \frac{\alpha}{1 - (1-\alpha) \cdot z}. \quad (3.2.3)$$

The next step is then to find the fixed points of the homography:

$$\phi(z) = z \Leftrightarrow \alpha \cdot z^2 - z + (1-\alpha) = 0 \Leftrightarrow z = 1 \text{ or } z = \frac{\alpha}{1-\alpha} = m > 1.$$

Let us introduce now the new sequence  $v_n := \frac{u_n - m}{u_n - 1}$ . We check that it is geometric of ratio  $m$  and immediately deduce that for every  $n \geq 1$ :

$$v_n = m^{n-1} \cdot v_1 = m^{n-1} \cdot \frac{\alpha^2}{(1-\alpha)^2} = m^{n+1}.$$

Which yields the conclusion.

### 3.2.3 A peeling exploration and a spatial Markov property

In this section, we describe a particular way to *peel* a SCTHP, that exhibits a nice spatial Markov property. By peeling, we mean that the map is "revealed" step by step through some process: at each time, we only know and observe a portion of the map—a *submap* of it. In the next definition—in which we introduce the fundamental pattern of our *peeling process*, it will become clear for the reader what is a submap.

**Definition 3.2** *Let  $\tilde{\mathfrak{T}}$  be a SCTHP and  $\mathbf{e}$  be an edge of its left boundary  $\sigma_0 := \{0\} \times \mathbb{Z}$ . Note that  $\mathbf{e} = (x, y)$  is necessarily such that  $x = (0, n)$  and  $y = (0, n+1)$  for some  $n \in \mathbb{Z}$ . Then:*

- (**upward revelation**) *if the triangle bordering  $\mathbf{e}$  on its right is bottom-oriented, we reveal all the vertices of the ascending tree  $\mathfrak{t}_y^\uparrow$ , and all the edges emerging from them;*
- (**downward revelation**) *otherwise, the triangle is top-oriented. Here, we reveal all the vertices of the descending tree  $\mathfrak{t}_x^\downarrow$ , and again all the edges emerging from them.*

*In both cases, our procedure is actually equivalent to reveal a certain number (maybe infinite) of triangles in  $\tilde{\mathfrak{T}}$ . Glued together with the unique face of infinite degree of  $\tilde{\mathfrak{T}}$  (that on the left of the infinite path  $\{0\} \times \mathbb{Z}$ ), we obtain a planar map that we denote by  $\text{Peel}(\sigma_0, \mathbf{e}, \tilde{\mathfrak{T}})$ . It is the submap of  $\tilde{\mathfrak{T}}$  obtained after the peeling of the edge  $\mathbf{e}$ . See Figure 3.5 for an illustration. The map which instead results from the gluing of the non-revealed faces of the SCTHP is called the complement submap of  $\text{Peel}(\sigma_0, \mathbf{e}, \tilde{\mathfrak{T}})$  and is written  $\tilde{\mathfrak{T}} \setminus \text{Peel}(\sigma_0, \mathbf{e}, \tilde{\mathfrak{T}})$ .*

We label vertices both in  $\text{Peel}(\sigma_0, \mathbf{e}, \tilde{\mathfrak{T}})$  and in the complement submap. In the former, we simply keep labels of the underlying SCTHP. In the latter, heights are unchanged but we translate in a trivial way the first coordinates so that to get a labelling in bijection with  $\mathbb{N} \times \mathbb{Z}$ . Note that the submap  $\text{Peel}(\sigma_0, \mathbf{e}, \tilde{\mathfrak{T}})$  has two faces of infinite degree, which surround triangles. We define the *right boundary* as the path bordering the right infinite face.

**Remark 3.1** *An alternative way to construct  $\text{Peel}(\sigma_0, \mathbf{e}, \tilde{\mathfrak{T}})$  is as follows. If the triangle bordering  $\mathbf{e}$  is bottom-oriented, we set  $\gamma$  as the leftmost (vertical) path crossing the edge  $(x, z)$ , where  $z$  is the right neighbour of  $y$ . Otherwise, we set  $\gamma$  as the leftmost (vertical) path crossing the edge  $(z, y)$ , where  $z$  is this time the right neighbour of  $x$ . In both cases, the submap  $\text{Peel}(\sigma_0, \mathbf{e}, \tilde{\mathfrak{T}})$  is obtained by gluing together all the faces of  $\tilde{\mathfrak{T}}$  located on the left of  $\gamma$ . See Figure 3.5 for an illustration.*

As announced earlier, the above peeling procedure highlights a spatial Markov property of SCTHP, summed up in two points:

1. the submap  $\text{Peel}(\sigma_0, \mathbf{e}, \tilde{\mathfrak{T}})$  and its complement  $\tilde{\mathfrak{T}} \setminus \text{Peel}(\sigma_0, \mathbf{e}, \tilde{\mathfrak{T}})$  are independent;
2. the complement submap  $\tilde{\mathfrak{T}} \setminus \text{Peel}(\sigma_0, \mathbf{e}, \tilde{\mathfrak{T}})$  is distributed as a SCTHP.

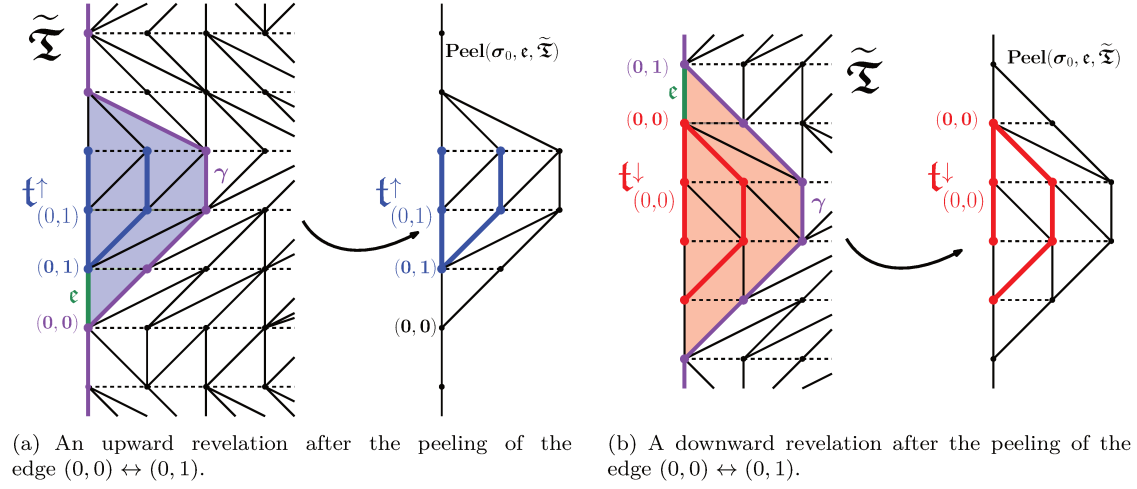


Figure 3.5: Illustration of the peeling of the edge  $\epsilon = (0,0) \leftrightarrow (0,1)$ . Filled in light blue or red are the faces glued to form  $\text{Peel}(\sigma_0, \epsilon, \tilde{\mathfrak{T}})$ . They are bordered on the right by the leftmost path  $\gamma$  defined in Remark 3.1.

Both are simple consequence of the alternative construction of SCTHP, via the collection of ascending and descending trees emerging from all the vertices with same given height. The peeling operation can actually be regarded as follows. It consists of removing from the latter collection an ascending or a descending tree associated to some vertex located on the left boundary of  $\tilde{\mathfrak{T}}$  (either  $x$  or  $y$ , see Definition 3.2). Then look at the new map that we get from the amputated collection: it is merely the complement submap  $\tilde{\mathfrak{T}} \setminus \text{Peel}(\sigma_0, \epsilon, \tilde{\mathfrak{T}})$ . Proposition 3.1 ensures that this map is distributed as a SCTHP and is independent of  $\text{Peel}(\sigma_0, \epsilon, \tilde{\mathfrak{T}})$ .

The second point of the spatial Markov property makes easy to iterate the peeling pattern of Definition 3.2. We just need to choose an edge on the right boundary of  $\text{Peel}(\sigma_0, \epsilon, \tilde{\mathfrak{T}})$ , then peel the latter in the complement submap. It reveals new faces that we glue to our current submap, along the right boundary. With a deterministic function—the *peeling algorithm*—that associates to any submap an edge on its right boundary, we can indefinitely pursue the process—or, say, the *peeling exploration*. See Figure 3.6 for an illustration. The edge to peel can also be picked randomly, with a source of randomness independent of the current submap. This is actually what we will do in the next section to study percolation on SCTHP.

### 3.3 Percolation on SCTHP

Before dealing with SCT, a preliminary step is to grasp oriented percolation in the context of SCTHP, **with horizontal edges playing no role** and the others directed from bottom to the top. We perform on the latter a Bernoulli bond percolation process of parameter  $p \in [0, 1]$ , as in SCT. We write  $\tilde{\mathbb{P}}_p$  for the *annealed* distribution that averages on the random maps and the percolation process, and  $\tilde{\mathbb{E}}_p$  for the expectation computed under  $\tilde{\mathbb{P}}_p$ . We first establish that a phase transition occur in SCTHP and is similar to that in SCT—see Theorem 3.1:

**Theorem 3.4** *Let  $\tilde{\mathcal{C}}$  be the percolation cluster of the origin in  $\tilde{\mathfrak{T}}$  and  $\tilde{\Theta}(p)$  the annealed probability that  $\tilde{\mathcal{C}}$  is infinite. We have:*

$$\tilde{\Theta}(p) = \begin{cases} 0 & \text{if } p \leq p_c \\ > 0 & \text{if } p > p_c. \end{cases} \quad (3.3.1)$$

where the quantity  $p_c$  is defined in Equation (3.1.1).

The method to prove the above theorem is based on the spatial Markov property raised in Section 3.2.3 and is broadly inspired of the approach developed in previous works like [15, 16, 17].



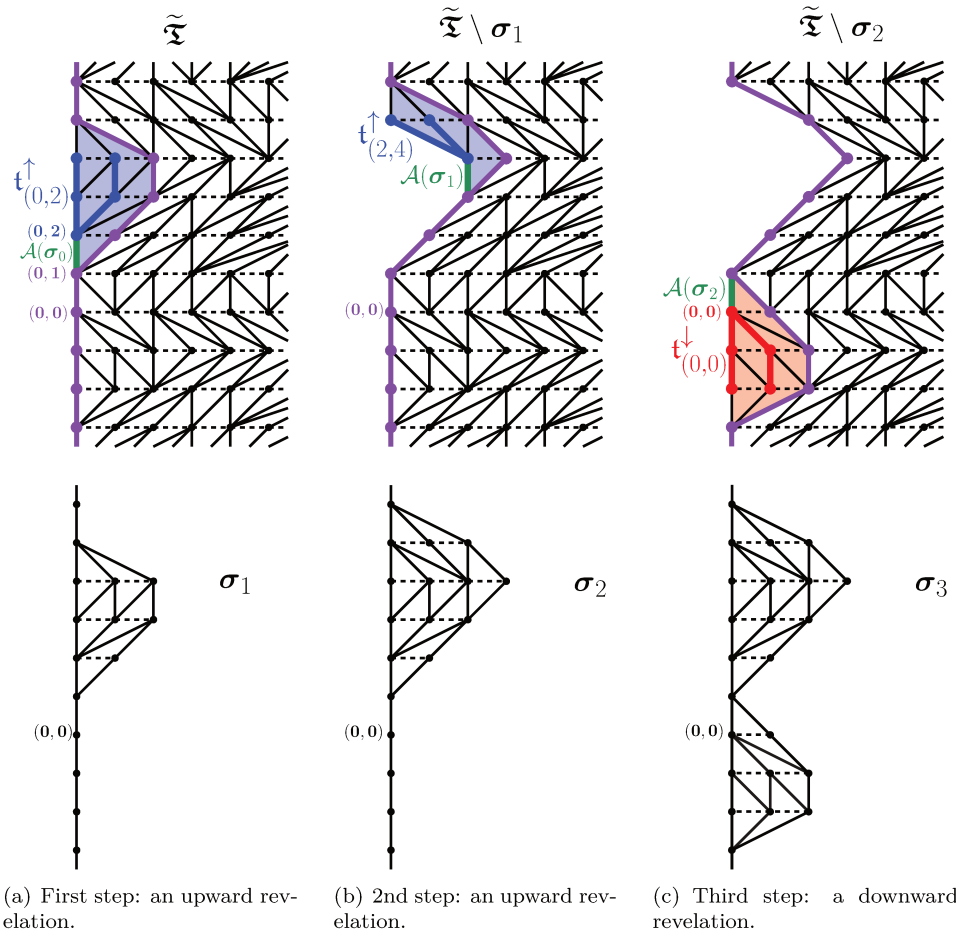


Figure 3.6: An example of the first three steps of a peeling exploration for some algorithm  $\mathcal{A}$ .

To sum up, we will conceive a smart exploration of the cluster  $\tilde{\mathcal{C}}$ —roughly by following its leftmost interface—from which we will exhibit some real random walk whose the asymptotic behaviour—does it diverge or not?—actually governs the percolation phenomenon. Successive moves of the walk will be interpreted as resulting of peeling actions. That will allow us to derive the exact distribution of the process and its drift. Details are outlined in Section 3.3.1.

We then go deeper into the analysis of the phase transition and compute several related annealed exponents which are as expected—at least for two of them—the same than in the SCT model:

**Theorem 3.5** *We have:*

- (i)  $\tilde{\mathbb{P}}_p(|\tilde{\mathcal{C}}| \geq n) = \mathcal{O}(c^n)$  for some  $c = c(p) \in (0, 1)$  when  $p < p_c$ ;
- (ii)  $\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| \geq n) \sim A \cdot n^{-1/2}$  for some  $A > 0$ ;
- (iii)  $\tilde{\Theta}(p) \sim a \cdot (p - p_c)$  for some  $a > 0$ , as  $p \rightarrow (p_c)^+$ .

The first estimate is commonly called "sharpness of the phase transition". Note that we do not prove in these pages any equivalent for SCT, although we believe that it is true in such context as well. All the three asymptotics stated in Theorem 3.5 are computed with the help of the exploration of  $\tilde{\mathcal{C}}$  above mentioned. We notably use the fact that the size of the cluster is roughly proportional to the length of the first non-negative excursion of the related random walk, which makes the estimates on the tail distribution of  $|\tilde{\mathcal{C}}|$  look more familiar. Indeed, the length of a non-negative excursion is for instance known to be exponentially-tailed when the drift of the underlying random walk is negative (a simple large deviations result), and polynomially-tailed when the drift is null [140]. The third item is the easiest to get, see the end of Section 3.3.1. In Section 3.3.2, we prove the first and second items.

Our third concern is to investigate the geometry of a large critical cluster and demonstrate that the Brownian continuum random tree emerges—like in SCT—as its scaling limit:

**Theorem 3.6** *At  $p = p_c$ , we have:*

$$(n^{-1/2} \cdot \tilde{\mathcal{C}} \mid |\tilde{\mathcal{C}}| \geq n) \xrightarrow[n \rightarrow +\infty]{d_{GH}} \kappa' \cdot \mathcal{T}_{\geq 1}, \quad (3.3.2)$$

where  $\kappa' = \kappa'(\alpha)$  is a positive number only depending on  $\alpha$  and  $\mathcal{T}_{\geq 1}$  is the CRT of mass greater than 1 [6, 7]. The convergence (3.3.2) holds in distribution for the Gromov–Hausdorff distance.

Again, the proof rests on a clever use of the cluster’s exploration. While the convergence in (3.3.2) involves metric spaces, we will rewrite it in terms of random processes instead. The result will decisively stem from the Donsker’s theorem. Details are tackled in Section 3.3.3. Definitions of the Gromov–Hausdorff topology and of the CRT are also recalled in the latter.

Finally, in Section 3.3.4, we look at how interact percolation clusters emerging from vertices on the  $x$ -axis of the SCTHP. We generalise Theorem 3.5(ii) and Theorem 3.6. This will be mostly helpful to transfer back the above result on the scaling limit to SCT.

### 3.3.1 Exploration of the percolation cluster via a peeling process

A simple way to find an infinite *directed* path—provided that it exists—in the cluster  $\tilde{\mathcal{C}}$  is by tracing its contours from left to right. We start by following the leftmost *directed* path emerging from the origin vertex, as far as possible. At some point, we may reach a dead end—a vertex, say  $x$ , with no offspring in the cluster: all the opportunities to go further have been exhausted in that direction. A clear consequence is that at our current height in the SCTHP, any infinite directed path in the cluster necessarily crosses an edge located strictly on our right. The leftmost such edge is that connecting the first right neighbour of  $x$  to the rightmost offspring of  $x$ —say  $z$  and  $y$ . We look now for a way to reach the latter edge  $(z, y)$  from  $(0, 0)$ .

For this purpose, it is natural to consider  $\gamma$  the leftmost directed path in the map passing through the edge  $(z, y)$ . The friendly situation is when  $\gamma$  meets somewhere downward a vertex  $x^*$  (take the highest one), on the segment explored by the walk from  $(0, 0)$  to  $x$ . In such case, we get a possible path from the origin to  $z$ . So we restart the walk at  $x^*$  and continue the exploration on the right of  $\gamma$ , according to the same principle (following the leftmost directed path up to a dead end, etc). We insist again on the fact that an infinite path in  $\tilde{\mathcal{C}}$  cannot be found anywhere else but on the right of  $\gamma$ . By dropping to  $x^*$ , some vertices of the cluster are set aside by the walk, but we do know that they are in finite number, enclosed between the explored segment  $[x^*, x]$  and  $\gamma$ . See Figure 3.7 for an illustration.

An unfriendly situation may also happen when  $\gamma$  do not intersect vertices visited earlier. The finding is in fact more definitive: the cluster  $\tilde{\mathcal{C}}$  do not merely contain any infinite directed path since then, there is no way to get through  $\gamma$  from the origin and connect itself to an infinite directed component.

**The link with peeling** A crucial point in our analysis is to read the above exploration of  $\tilde{\mathcal{C}}$  through the prism of peeling. All along the process, is in fact extended a *submap* of  $\tilde{\mathfrak{T}}$  containing for sure a finite portion of the cluster, while we may still find an infinite component in its complement. Indeed, when we are at some vertex  $x$ , we successively check from left to right whether the edges emerging from  $x$  are in  $\tilde{\mathcal{C}}$  or not. This operation is performed until we perhaps find one open. Each time that we find a closed edge—say  $(x, y)$ —we derive a set of vertices becoming inaccessible in a directed fashion from  $x$  and, a fortiori, from the origin. Those are exactly the vertices in the ascending tree of  $y$ . So, in case of such event, the submap in question can be enlarged like after an upward revelation. When, finally, we end up not having any possibility to go further from  $x$ , the walker dropping to the vertex  $x^*$  defined in the foregoing, we enlarge this time our current submap with the finite area bounded by the explored segment  $[x^*, x]$  on the left and  $\gamma$  on the right. It is not hard to see that such operation actually corresponds to a downward revelation. See indeed Remark 3.1.

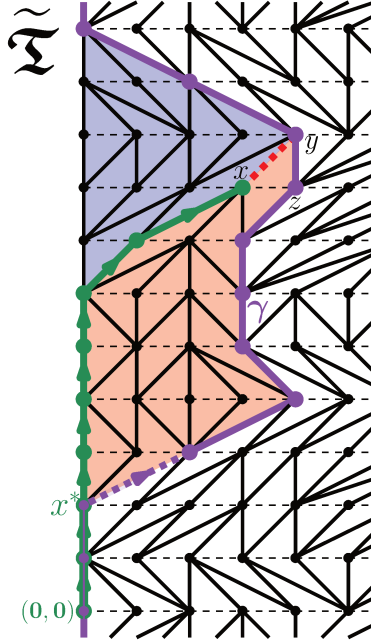


Figure 3.7: In green is the leftmost directed path in  $\tilde{\mathcal{C}}$ , going from the origin to  $x$ . Filled in light blue is a part of  $\tilde{\mathfrak{T}}$  which is inaccessible from  $(0, 0)$  in a directed fashion, given the first steps of the exploration. The area filled in light red may contain vertices of  $\tilde{\mathcal{C}}$ , but they are not visited by the process. In purple is the path  $\gamma$ , which is the leftmost crossing the edge  $(z, y)$ . The walk restarts at  $x^*$ , the highest already visited vertex which is met by  $\gamma$ , and continues on the right of the path.

**The peeling exploration of  $\Theta_p^0$**  To be more formal, from a SCTHP  $\tilde{\mathfrak{T}}$ , we define an infinite sequence  $(\mathbf{v}_n; \vec{\epsilon}_n; \Sigma_n)_{n \geq 0}$  of vertices, directed edges and submaps, obtained by applying a *randomized* peeling algorithm. Indeed, contrary to our example in Figure 3.6, the algorithm  $\mathcal{A}$  will also be a function of random external data—independent of the SCTHP—provided by the Bernoulli percolation process. We start with

$$\mathbf{v}_0 = (0, 0), \vec{\epsilon}_0 = (0, 0) \rightarrow (0, 1) \text{ and } \Sigma_0 = \{0\} \times \mathbb{Z} \text{ (the left boundary of } \tilde{\mathfrak{T}}).$$

The update of the sequence depends on the state of the edge  $\vec{\epsilon}_0$  in  $\tilde{\mathcal{C}}$ :

- if  $\vec{\epsilon}_0$  is *open* in  $\tilde{\mathcal{C}}$ , we simply update the vertex sequence as

$$\mathbf{v}_1 := \mathbf{v}_0 + \vec{\epsilon}_0,$$

i.e.  $\mathbf{v}_1$  is the next vertex on the right boundary of  $\Sigma_0$ —here  $(0, 1)$ . The edge  $\vec{\epsilon}_1$  is the edge emerging from  $\mathbf{v}_1$  on the same boundary—namely  $(0, 1) \rightarrow (0, 2)$ . The submap does not change.

- otherwise, if  $\vec{\epsilon}_0$  is *closed* in  $\tilde{\mathcal{C}}$ , we peel the (not directed) edge  $\epsilon_0$ . The new submap is then

$$\Sigma_1 = \text{Peel}(\Sigma_0, \epsilon_n, \tilde{\mathfrak{T}}).$$

In case of an *upward revelation*, the current vertex does not change but the new directed edges  $\vec{\epsilon}_1$  is now the second emerging from  $\mathbf{v}_0$  (and is on the right boundary of  $\Sigma_1$ ). It is  $(0, 0) \rightarrow (1, 1)$ . See Figure 3.8(a) for an illustration.

In case of an *downward revelation*, triangles added to the current submap are directly got from the descending tree of  $\mathbf{v}_0$  in  $\tilde{\mathfrak{T}}$  (see Definition 3.2). If the latter has height  $h \geq 0$ , we fix  $\mathbf{v}_1$  as the rightmost vertex of  $\Sigma_1$  (or  $\Sigma_0$ ) with height  $\mathbf{h}(\mathbf{v}_0) - 1 - h$ . Here, it is of course the vertex  $(0, -1 - h)$ . The edge  $\vec{\epsilon}_1$  is set as that emerging from  $\mathbf{v}_1$  to its offspring which is on the right boundary of  $\Sigma_1$ . See Figure 3.8(b) for an illustration.

The update is well-defined since the height of the descending tree  $\mathfrak{t}_{(0,0)}^\downarrow$  is finite almost surely, given it is distributed as a subcritical Galton–Watson tree according to Proposition 3.1. The algorithm is then iterated. This is possible because, by construction, the vertex  $\mathbf{V}_n$  and the edge  $\vec{\mathbf{e}}_n$  are always on the right boundary of  $\Sigma_n$ . See Figure 3.8(c) for an illustration of a complete exploration of the cluster  $\tilde{\mathcal{C}}$ .

**The height process** Derived from the peeling exploration, the height process  $\mathcal{H} := (h(\mathbf{V}_n))_{n \geq 0}$  plays a key role in our work. For instance, we show below that it condenses many information about the size of the cluster  $\tilde{\mathcal{C}}$ . Given the spatial Markov property, combined with the fact that updates during the peeling exploration all obey the same rules, it is straightforward that the height process is a random walk with i.i.d. increments. Their common distribution can be explicitly computed:

**Proposition 3.3** *Recall that  $\mathbf{h}(\cdot)$  is the function returning the height of any rooted tree. Then we have:*

$$\tilde{\mathbb{P}}_p(h(\mathbf{V}_1) = h) = \begin{cases} p & \text{for } h = 1 \\ \alpha \cdot (1 - p) & \text{for } h = 0 \\ (1 - \alpha) \cdot (1 - p) \cdot \mathbb{P}(\mathbf{h}(\mathfrak{t}^\downarrow) = -h - 1) & \text{for } h < 0 \end{cases} \quad (3.3.3)$$

where  $\mathfrak{t}^\downarrow \stackrel{d}{\sim} \mathbf{GW}_{1-\alpha}$ .

The proof of (3.3.3) is totally transparent if we carefully read the instructions guiding updates during the peeling exploration, and if we remember that the graph structure of the SCTHP is independent of the percolation process. From (3.3.3) and Proposition 3.2, we easily deduce that the distribution of increments has a finite first moment and:

$$\tilde{\mathbb{E}}_p[h(\mathbf{V}_1)] = p - (1 - p) \cdot (1 - \alpha) \cdot (1 + \mathbb{E}[\mathbf{h}(\mathfrak{t}^\downarrow)]) = p - (1 - p) \cdot (1 - \alpha) \cdot \sum_{n \geq 0} \frac{m - 1}{m^{n+1} - 1}. \quad (3.3.4)$$

Recall that  $m$  is the mean of a geometric law of parameter  $\alpha$ , so is equal to  $\alpha \cdot (1 - \alpha)^{-1} > 1$ .

**Percolation is governed by the drift of  $\mathcal{H}$ : proof of Theorem 3.4.** Let  $T$  be the first hitting time of *negative* integers  $\{-1, -2, \dots\}$  by the height process  $\mathcal{H}$ . We have:

$$|\tilde{\mathcal{C}}| = +\infty \quad \text{if and only if} \quad T = +\infty. \quad (3.3.5)$$

Indeed, vertices visited up to the time right before  $T$  are all in the cluster. At time  $T < +\infty$ , we drop to negative heights after a downward revelation. It means that the right boundary of the new submap  $\Sigma_T$ —or the path  $\gamma$  defined at the beginning of the current section—did not meet any vertex visited earlier. And as it has been argued, it implies that  $\tilde{\mathcal{C}}$  is finite. See Figure 3.8(c) for an illustration. The converse statement is trivial. With (3.3.5), the proof of Theorem 3.4 is reduced to the well-known fact that:

$$\tilde{\mathbb{P}}_p(T = +\infty) > 0 \quad \text{if and only if} \quad \tilde{\mathbb{E}}_p[h(\mathbf{V}_1)] > 0.$$

Given (3.3.5), this yields the conclusion of Theorem 3.4.

**The off-critical percolation probability** Let  $T'$  be the first hitting time of *non-positive* integers  $\{0, -1, -2, \dots\}$  by the height process  $\mathcal{H}$ . In the case  $p > p_c$ , the drift of the latter is positive and a standard result on skip-free ascending walk—which are random walks with integer valued increments and  $+1$  as the only possible positive one, like  $\mathcal{H}$ —ensures that:

$$\tilde{\mathbb{P}}_p(T' = +\infty) = \tilde{\mathbb{E}}_p[h(\mathbf{V}_1)].$$

See for instance [4, Theorem 3]. When  $T = +\infty$ , the height process may touch 0 a finite number of times  $\mathbf{n}$  before never come back again. Thanks to the strong Markov property, the distribution

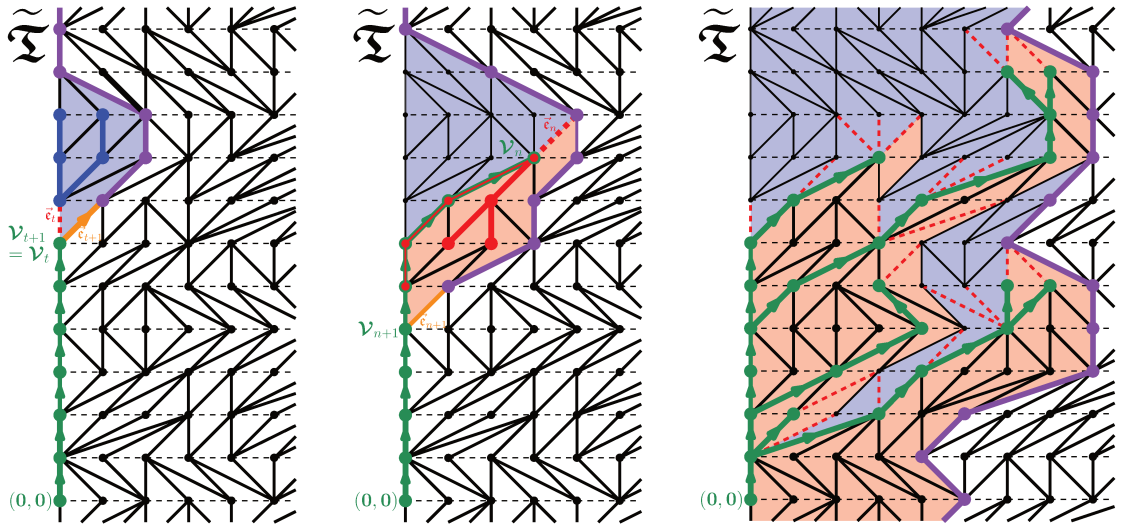
of  $\mathbf{n}$  is geometric. Its parameter is the probability for  $\mathcal{H}$  to carry out a positive excursion and finish it at 0. As a consequence:

$$\tilde{\Theta}(p) = \tilde{\mathbb{E}}_p[\mathbf{n}] \cdot \tilde{\mathbb{E}}_p[h(\mathbf{V}_1)] = (1 - \tilde{\mathbb{P}}_p(T' < +\infty; h(\mathbf{V}_{T'}) = 0))^{-1} \cdot \tilde{\mathbb{E}}_p[h(\mathbf{V}_1)]. \quad (3.3.6)$$

Since  $T' < +\infty$  almost surely at  $p = p_c$ , we clearly have

$$\tilde{\mathbb{P}}_p(T' < +\infty; h(\mathbf{V}_{T'}) = 0) \xrightarrow{p \rightarrow p_c} \tilde{\mathbb{P}}_{p_c}(h(\mathbf{V}_{T'}) = 0) < 1.$$

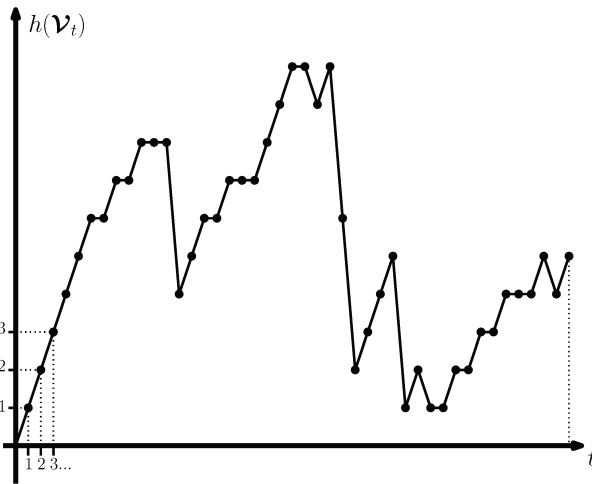
According to (3.3.4), we also have  $\tilde{\mathbb{E}}_p[h(\mathbf{V}_1)] \sim c \cdot (p - p_c)$  for some  $c > 0$ . Together with (3.3.6), the last two estimates implies Theorem 3.5(iii).



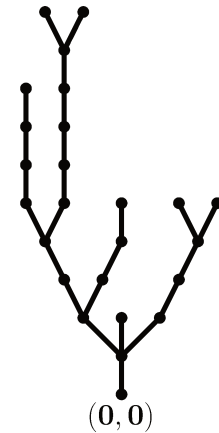
(a) Updates after an upward revelation occurring at time  $t = 6$ . The area filled in light blue is inaccessible from the origin vertex: we remove it and pursue the exploration on the complement submap.

(b) Updates after a downward revelation occurring at time  $n = 12$ . The area filled in light red may contain vertices of the cluster (among those stroke in red here), but which are not visited by the process.

(c) A complete exploration of  $\tilde{\mathcal{C}}$ . Filled in light blue/red are triangles removed after an upward/downward revelation. The cluster may contain vertices in red areas above the null height. Dashed red lines are for the edges turning out to be closed after checking.



(d) Graph of the height process along the exploration of  $\tilde{\mathcal{C}}$ .



(e) The trace of the walk is tree-shaped.

Figure 3.8: Illustration of the peeling process used to explore  $\tilde{\mathcal{C}}$ .

### 3.3.2 The volume of the percolation cluster

Recall that we have defined earlier  $T$  as the first hitting time of negative numbers by the height process  $\mathcal{H}$ . At the critical threshold  $p = p_c$ , the cluster  $\tilde{\mathcal{C}}$  is finite almost surely and is fully contained in the submap  $\Sigma_T$ . Actually, it is merely the union of vertices explored up to the time right before  $T$ —with a new vertex visited at each  $+1$  jump in the associated height process—plus some others left behind, at times of downward revelation, or negative jump for  $\mathcal{H}$ . See Figure 3.8(c). We know that these vertices are all enclosed in areas whose volume is distributed as the total progeny of a subcritical Galton–Watson tree with offspring law  $\mu_{1-\alpha}$ . See Figure 3.5(b). It is well known that the latter quantity is exponentially-tailed, by considering for instance its generative function. All these facts allow us to write the following equality, holding in distribution:

$$|\tilde{\mathcal{C}}| \stackrel{(d)}{=} 1 + \sum_{t=1}^T \theta_t. \quad (3.3.7)$$

where  $(\theta_t)_{t \geq 1}$  is a sequence of i.i.d. random variables, satisfying:

$$\theta_t = \begin{cases} 1 & \text{if } h(\mathbf{V}_t) - h(\mathbf{V}_{t-1}) = 1; \\ 0 & \text{if } h(\mathbf{V}_t) - h(\mathbf{V}_{t-1}) = 0; \\ \geq 0 & \text{otherwise.} \end{cases}$$

The random variable  $\theta_t$  is furthermore, in the third case (corresponding to a negative jump in the height process), stochastically dominated by the volume of a Galton–Watson tree with offspring law  $\mu_{1-\alpha}$ , conditioned on having height  $|h(\mathbf{V}_t) - h(\mathbf{V}_{t-1})| - 1$ . In (3.3.7), pay attention to the fact that  $\theta_T$  plays a special role since vertices in  $\tilde{\mathcal{C}}$  must have a nonnegative height. However, it does not change anything for our further analysis since the above stochastic domination holds even more for  $\theta_T$ .

Set now  $\kappa := \mathbb{E}[\theta_1] < +\infty$ . Equation 3.3.7 suggests that the cluster  $\tilde{\mathcal{C}}$  roughly contains  $\kappa \cdot T$  vertices, as  $T \rightarrow +\infty$ . This is indeed true and it turns out to be a crucial point in order to prove the two first items of Theorem 3.5. Large deviations arguments will help us to make it rigorous.

Before entering into details, we introduce a new notation which will simplify the presentation:

**Definition 3.3** We write  $x_n = oe_\delta(n)$  for some sequence  $(x_n)$  and  $\delta > 0$  if there exists constants  $c, C > 0$  such that  $x_n \leq Ce^{-cn^\delta}$  for any  $n$ .

**Remark 3.2** If  $x_n = oe_\delta(n)$  and if  $y_n := \sum_{k \geq n} x_k$ , then  $y_n = oe_\delta(n)$ .

#### 3.3.2.1 The sharpness of the phase transition

Fix  $p < p_c$ . We aim to prove the first item of Theorem 3.5. Let  $n \geq 1$  and  $n' \leq n/2\kappa$ . We have:

$$\tilde{\mathbb{P}}_p(|\tilde{\mathcal{C}}| \geq n; T = n') \stackrel{(3.3.7)}{\leq} \tilde{\mathbb{P}}_p\left(1 + \sum_{t=1}^{n'} \theta_t \geq n\right) \leq \tilde{\mathbb{P}}_p\left(1 + \sum_{t=1}^{\lfloor n/2\kappa \rfloor} \theta_t \geq n\right).$$

The last inequality comes from the fact that as  $n'$  increases, so does  $\tilde{\mathbb{P}}_p\left(1 + \sum_{t=1}^{n'} \theta_t \geq n\right)$ . By a standard large deviations argument, we know that:

$$\tilde{\mathbb{P}}_p\left(1 + \sum_{t=1}^{\lfloor n/2\kappa \rfloor} \theta_t \geq n\right) = oe_1(n).$$

Given both inequalities above, by summing over  $n' \leq n/2\kappa$ , we get:

$$\tilde{\mathbb{P}}_p(|\tilde{\mathcal{C}}| \geq n; T \leq n/2\kappa) = oe_1(n).$$

Also, it is clear that  $\tilde{\mathbb{P}}_p(T > n/2\kappa) \leq \tilde{\mathbb{P}}_p(h(\mathbf{V}_{\lfloor n/2\kappa \rfloor}) \geq 0)$ , and the latter probability is  $oe_1(n)$  because the drift of the height process is negative when  $p < p_c$  (so the event  $\{h(\mathbf{V}_{\lfloor n/2\kappa \rfloor}) \geq 0\}$  is a large deviations event). Finally, it holds that:

$$\tilde{\mathbb{P}}_p(|\tilde{\mathcal{C}}| \geq n) \leq \tilde{\mathbb{P}}_p(|\tilde{\mathcal{C}}| \geq n; T \leq n/2\kappa) + \tilde{\mathbb{P}}_p(T > n/2\kappa),$$

and the two quantities on the right-hand side of the inequality are  $oe_1(n)$ , as it has just been argued. So is then  $\tilde{\mathbb{P}}_p(|\tilde{\mathcal{C}}| \geq n)$ , which is the expected conclusion.

### 3.3.2.2 The volume of a critical cluster

We focus now on the second item of Theorem 3.5 with  $p = p_c$ . It is well known that the distribution of the hitting time  $T$  is polynomially-tailed as the underlying random walk has no drift (like here). The speed of decreasing is explicit, see for instance [140, Theorem 11] applied to an integer-valued random walks with square-integrable increments:

**Proposition 3.4** *There exists  $A > 0$  such that:*

$$\tilde{\mathbb{P}}_{p_c}(T = n) \underset{n \rightarrow +\infty}{\sim} A \cdot n^{-3/2} \quad \text{and} \quad \tilde{\mathbb{P}}_{p_c}(T \geq n) \underset{n \rightarrow +\infty}{\sim} 2A \cdot n^{-1/2}.$$

Now we recall a classic moderate deviations result on exponentially-tailed random variables:

**Lemma 3.1** *Let  $(X_t)_{t \geq 1}$  be a sequence of i.i.d. random variables with mean  $\bar{x}$  and an exponentially-tailed distribution, meaning that  $\mathbb{P}(X_1 \geq n) = oe_1(n)$ . Then, for any  $C > 0$ :*

$$\mathbb{P}\left(\left|\frac{1}{n} \sum_{t=1}^n X_t - \bar{x}\right| > C \cdot n^{-1/4}\right) = oe_{1/2}(n).$$

This is a particular case of [97, Lemma 1.12]. With the help of the latter lemma, we are going to prove that:

$$\tilde{\mathbb{P}}_{p_c}((|\tilde{\mathcal{C}}| \geq \kappa \cdot n) \triangle (T \geq n)) = \mathcal{O}(n^{-3/4}) = o(n^{-1/2}) = o(\tilde{\mathbb{P}}_{p_c}(T \geq n)). \quad (3.3.8)$$

The symbol  $\triangle$  denotes the symmetric difference of two sets. Together with Proposition 3.4, the estimate (3.3.8) is enough to get the conclusion since it implies that

$$\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| \geq \kappa \cdot n) \underset{n \rightarrow +\infty}{\sim} \tilde{\mathbb{P}}_{p_c}(T \geq n).$$

Fix  $C > 0$ . Proposition 3.4 ensures that

$$\tilde{\mathbb{P}}_{p_c}(|T - n| \leq C \cdot n^{3/4}) = \mathcal{O}(n^{-3/4}) = o(n^{-1/2}) = o(\tilde{\mathbb{P}}_{p_c}(T \geq n)). \quad (3.3.9)$$

Moreover, for any  $n \geq 1$  and  $n' \geq n + C \cdot n^{3/4}$ :

$$\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| < \kappa \cdot n; T = n') \underset{(3.3.7)}{\leq} \tilde{\mathbb{P}}_{p_c}\left(\frac{1 + \sum_{t=1}^{n'} \boldsymbol{\theta}_t}{n'} < \kappa \cdot n/n'\right) \leq \tilde{\mathbb{P}}_{p_c}\left(\frac{1 + \sum_{t=1}^{n'} \boldsymbol{\theta}_t}{n'} < \kappa \cdot n/(n + C \cdot n^{3/4})\right).$$

There exists some  $C' > 0$  such that  $n/(n + C \cdot n^{3/4}) \leq 1 - \kappa^{-1} \cdot C' \cdot n^{-1/4}$  for any  $n$ . Since  $n' \geq n$ , and given the above inequality, we get that:

$$\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| < \kappa \cdot n; T = n') \leq \tilde{\mathbb{P}}_{p_c}\left(\frac{1 + \sum_{t=1}^{n'} \boldsymbol{\theta}_t}{n'} < \kappa - C' \cdot (n')^{-1/4}\right) \stackrel{\text{Lemma 3.1}}{=} oe_{1/2}(n').$$

By summing over  $n'$ , we deduce that:

$$\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| < \kappa \cdot n; T \geq n + C \cdot n^{3/4}) = oe_{1/2}(n).$$

Then, by using (3.3.9):

$$\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| < \kappa \cdot n; T \geq n) = \mathcal{O}(n^{-3/4}). \quad (3.3.10)$$

Furthermore, for any  $n' \leq \lfloor n - C \cdot n^{3/4} \rfloor := N$ , we have:

$$\tilde{\mathbb{P}}_{p_c} \left( |\tilde{\mathcal{C}}| \geq \kappa \cdot n; T = n' \right) \stackrel{(3.3.7)}{\leq} \tilde{\mathbb{P}}_{p_c} \left( 1 + \sum_{t=1}^{n'} \theta_t \geq \kappa \cdot n \right) \leq \tilde{\mathbb{P}}_{p_c} \left( 1 + \sum_{t=1}^N \theta_t \geq \kappa \cdot n \right).$$

The last inequality comes from the fact that as  $n'$  increases, so does  $\tilde{\mathbb{P}}_p \left( 1 + \sum_{t=1}^{n'} \theta_t \geq \kappa \cdot n \right)$ . Since there exists some  $C' > 0$  such that  $n/N \geq 1 + \kappa^{-1} \cdot C' \cdot n^{-1/4}$ , we derive that:

$$\tilde{\mathbb{P}}_{p_c} \left( |\tilde{\mathcal{C}}| \geq \kappa \cdot n; T = n' \right) \leq \tilde{\mathbb{P}}_{p_c} \left( \frac{1 + \sum_{t=1}^N \theta_t}{N} \geq \kappa + C' \cdot n^{-1/4} \right) \stackrel{\text{Lemma 3.1}}{=} oe_{1/2}(n).$$

By summing over  $n'$ , we obtain that:

$$\tilde{\mathbb{P}}_{p_c} \left( |\tilde{\mathcal{C}}| \geq \kappa \cdot n; T < n - C \cdot n^{3/4} \right) = oe_{1/2}(n).$$

Then, by using again (3.3.9):

$$\tilde{\mathbb{P}}_{p_c} \left( |\tilde{\mathcal{C}}| < \kappa \cdot n; T \geq n \right) = \mathcal{O} \left( n^{-3/4} \right). \quad (3.3.11)$$

The estimates (3.3.10) and (3.3.11) are together equivalent to (3.3.8).

### 3.3.3 Scaling limit of the critical percolation cluster

We have shown that the height process condenses many information on the size of  $\tilde{\mathcal{C}}$ . It turns out that it captures the geometry as well. Consider the graph made of vertices and edges of the SCTHP, visited by the peeling exploration up to the time right before  $T$ . This graph is tree-shaped since is basically the result of the gluing of non-intersecting paths of the map. See Figure 3.8. Henceforth, we call it *tree skeleton*, denoted by  $\mathcal{S}$ . We choose  $(0, 0)$  as root vertex.

As it has been explained at the beginning of the previous section, the percolation cluster is the union of the tree skeleton and some other vertices left behind. This suggests to divide the proof of Theorem 3.2 in two parts. Adopting a method common in the literature—see for instance [62, 97]—based on the coding of rooted trees via discrete processes, we are going to demonstrate that the (properly rescaled) tree skeleton admits some close to (3.3.2) scaling limit, as the length  $T$  of the exploration tends to infinity. Then, we will extend the convergence to the whole cluster, observing that the "left-behind" vertices are wiped out at large scale, because are confined in small areas. Finally, we use results of Section 3.3.2 to transform the conditioning on  $T$  into a conditioning on  $|\tilde{\mathcal{C}}|$ .

**Scaling limit of the tree skeleton** The purpose of the paragraph is to establish the following convergence, holding in distribution at  $p = p_c$  for the Gromov–Hausdorff distance:

$$\left( n^{-1/2} \cdot \mathcal{S} \mid T \geq n \right) \xrightarrow[n \rightarrow +\infty]{d_{GH}} \sigma \cdot \mathcal{T}_{\geq 1}, \quad (3.3.12)$$

where  $\sigma = \sigma(\alpha)$  is the standard deviation of  $h(\mathcal{V}_1)$  and  $\mathcal{T}_{\geq 1}$  is the CRT of mass greater than 1.

Let us start by clarifying notions and objects. Any finite (or *discrete*) tree can be viewed as a compact metric space by endowing its set of vertices with the graph distance. The Gromov–Hausdorff distance is a distance on the set of metric spaces, inducing some notion of convergence. If  $(E, \mathbf{d})$  and  $(E', \mathbf{d}')$  are two *compact* metric spaces, the Gromov–Hausdorff distance between them is set as

$$\mathbf{d}_{GH} \left( (E, \mathbf{d}), (E', \mathbf{d}') \right) = \inf \left\{ \mathbf{d}_H(\phi(E), \phi'(E')) \right\}, \quad (3.3.13)$$

where the infimum is taken over all metric spaces  $(F, \delta)$  and isometric embeddings  $\phi : E \rightarrow F$  and  $\phi' : E' \rightarrow F$ , and where  $\mathbf{d}_H$  designates the Hausdorff distance between compact sets in  $F$ .



The limit in distribution in (3.3.12) of the tree skeleton for  $\mathbf{d}_{GH}$  is claimed to be the CRT (up to a positive factor) of mass greater than 1. The CRT is a *real tree*, that is roughly a *compact metric space* containing no cycle within it. A very convenient method to represent real trees is by using nonnegative real functions with compact support in  $\mathbb{R}_+$ . Let  $g$  be as such, with  $g(0) = 0$ . We set for every  $0 \leq s \leq t$ :

$$\mathbf{d}_g(s, t) := g(s) + g(t) - 2 \inf_{u \in [s, t]} g(u),$$

and the following equivalence relation  $\sim_g$  on  $[0, +\infty[$ :

$$s \sim_g t \iff \mathbf{d}_g(s, t) = 0.$$

The quotient space  $\mathcal{T}_g := [0, +\infty[ / \sim_g$  endowed with the distance  $\mathbf{d}_g$  is a compact real tree. In Figure 3.9, we show how is derived a real tree from a piecewise linear function. The CRT of mass greater than 1 is the real tree obtained by taking for  $g = \mathbf{e}_{\geq 1}$ , where  $\mathbf{e}_{\geq 1}$  is the Brownian excursion of duration greater than 1 (we have  $\mathbf{e}_{\geq 1}(t) = 0$  for  $t > t_{exc}$  and some  $t_{exc} \geq 1$ ). See [63, 97] for more details.

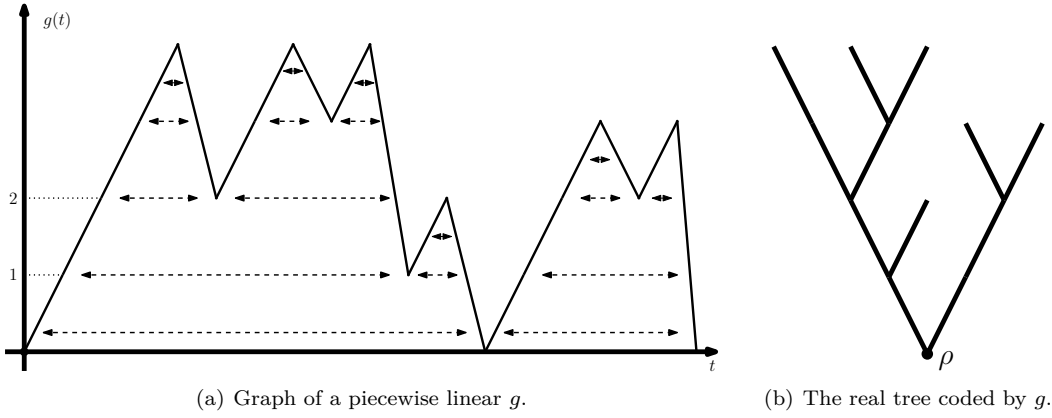


Figure 3.9: Construction of a real tree via the coding by a piecewise linear function. Imagine that we apply glue all along the bottom side of the curve, then press it on both sides.

Discrete trees—like  $\mathcal{S}$ —may also be viewed as real trees if we imagine them in the plane an union of line segments of length one (the edges), equipped with the obvious distance (the length of the shortest path between two vertices). As real tree, the tree skeleton is isomorphic to that coded by a linearly interpolated version of the height process. Set for instance for all  $t \geq 0$ :

$$\mathcal{H}(t) = \begin{cases} \left( h(\mathcal{V}_n) + (t - n) \cdot (h(\mathcal{V}_{n+1}) - h(\mathcal{V}_n)) \right)_+ & \text{for } n \leq T - 1 \text{ and } t \in [n, n + 1]; \\ 0 & \text{elsewhere.} \end{cases} \quad (3.3.14)$$

Then, up to some isometry, we have:

$$(\mathcal{S}, \mathbf{d}) = (\mathcal{T}_{\mathcal{H}}, \mathbf{d}_{\mathcal{H}}), \quad (3.3.15)$$

where  $\mathbf{d}$  stands for the shortest-path distance in  $\mathcal{S}$  viewed as real tree. The statement is clearly supported by Figure 3.8 and 3.9.

Once (3.3.15) is known, the scaling limit (3.3.12) merely results from a suited application of the Donsker's theorem. Indeed, a consequence of the latter is that the continuous-time (and properly rescaled) height process defined in (3.3.14) converges in distribution as follows:

$$\left( (n^{-1/2} \cdot \mathcal{H}(n \cdot t))_{t \geq 0} \mid T \geq n \right) \xrightarrow[n \rightarrow +\infty]{(d)} \sigma \cdot \mathbf{e}_{\geq 1}, \quad (3.3.16)$$

where  $\sigma = \sigma(\alpha)$  is the standard deviation of  $h(\mathbf{V}_1)$  and  $\mathbf{e}_{\geq 1}$  a Brownian excursion of duration greater than 1. Then it becomes clear—by using for instance [97, Lemma 2.4]—that:

$$(n^{-1/2} \cdot \mathcal{S} \mid T \geq n) = (\mathcal{T}_{n^{-1/2} \cdot \mathcal{H}} \mid T \geq n) \xrightarrow[n \rightarrow +\infty]{d_{GH}} \sigma \cdot \mathcal{T}_{\mathbf{e}_{\geq 1}},$$

where the convergence holds in distribution for the Gromov–Hausdorff distance. This is exactly the expected conclusion.

**A large cluster is roughly the tree skeleton** Now we aim to extend (3.3.12) to the whole cluster. It is enough to prove that

$$(n^{-1/2} \cdot \mathbf{d}_{GH}(\tilde{\mathcal{C}}, \mathcal{S}) \mid T \geq n) \xrightarrow[n \rightarrow +\infty]{(d)} 0, \quad (3.3.17)$$

where the convergence holds in distribution. Since  $\tilde{\mathcal{C}}$  and  $\mathcal{S}$  are two subsets of the same SCTHP, we have:

$$\mathbf{d}_{GH}(\tilde{\mathcal{C}}, \mathcal{S}) \leq \mathbf{d}_H(\tilde{\mathcal{C}}, \mathcal{S}),$$

where  $\mathbf{d}_H$  is the Hausdorff distance between compact sets of  $\Theta$ . See indeed (3.3.13). By construction, the tree skeleton is contained in the percolation cluster. Some vertices are in the latter, but not in the former. As it has already been explained, they are those left behind by the peeling exploration when occurs a downward revelation, or a drop for the height process. The distance between such vertices and the tree skeleton cannot exceed the amplitude of the jump carried out. See Figure 3.7 or Figure 3.8. As a consequence:

$$\mathbf{d}_H(\tilde{\mathcal{C}}, \mathcal{S}) \leq \max_{t \leq T-1} \{|h(\mathbf{V}_{t+1}) - h(\mathbf{V}_t)|\}. \quad (3.3.18)$$

The distribution of  $|h(\mathbf{V}_1)|$  is exponentially tailed. Then, for any  $\epsilon > 0$ , any integer  $n \leq n' \leq n^{1+\epsilon}$  and any real  $C > 0$ :

$$\begin{aligned} & \tilde{\mathbb{P}}_{p_c} \left( \max_{t \leq T-1} \{|h(\mathbf{V}_{t+1}) - h(\mathbf{V}_t)|\} > C \cdot \log n; T = n' \right) \\ & \leq \tilde{\mathbb{P}}_{p_c} \left( \max_{t \leq n'-1} \{|h(\mathbf{V}_{t+1}) - h(\mathbf{V}_t)|\} > C \cdot \log n \right) \\ & = 1 - (1 - \tilde{\mathbb{P}}_{p_c}(|h(\mathbf{V}_1)| > C \cdot \log n))^{n^{1+\epsilon}}. \end{aligned}$$

By using Propositions 3.2 and 3.3, we get the estimate

$$\tilde{\mathbb{P}}_{p_c}(|h(\mathbf{V}_1)| > C \cdot \log n) = \mathcal{O}(m^{-C \cdot \log n}) = \mathcal{O}(n^{-C \cdot \log m}).$$

Recall that  $m = \alpha \cdot (1 - \alpha)^{-1} > 1$ . We can choose  $C$  large enough so that

$$\tilde{\mathbb{P}}_{p_c}(|h(\mathbf{V}_1)| > C \cdot \log n) = o(n^{-5/2-2\epsilon}).$$

A classic asymptotic expansion then ensures that:

$$\begin{aligned} & 1 - (1 - \tilde{\mathbb{P}}_{p_c}(|h(\mathbf{V}_1)| > C \cdot \log n))^{n^{1+\epsilon}} \\ & = 1 - e^{n^{1+\epsilon} \cdot \log(1 - \tilde{\mathbb{P}}_{p_c}(|h(\mathbf{V}_1)| > C \cdot \log n))} = o(n^{-3/2-\epsilon}). \end{aligned}$$

From the foregoing, we deduce that:

$$\begin{aligned} & \tilde{\mathbb{P}}_{p_c} \left( \max_{t \leq T-1} \{|h(\mathbf{V}_{t+1}) - h(\mathbf{V}_t)|\} > C \cdot \log n; n \leq T \leq n^{1+\epsilon} \right) \\ & = \sum_{n \leq n' \leq n^{1+\epsilon}} \tilde{\mathbb{P}}_{p_c} \left( \max_{t \leq T-1} \{|h(\mathbf{V}_{t+1}) - h(\mathbf{V}_t)|\} > C \cdot \log n; T = n' \right) \\ & \leq n^{1+\epsilon} \cdot (1 - \tilde{\mathbb{P}}_{p_c}(|h(\mathbf{V}_1)| > C \cdot \log n))^{n^{1+\epsilon}} = o(n^{-1/2}) = o(\tilde{\mathbb{P}}_{p_c}(T \geq n)). \end{aligned} \quad (3.3.19)$$

Furthermore, according to Proposition 3.4:

$$\tilde{\mathbb{P}}_{p_c}(T > n^{1+\epsilon}) = \mathcal{O}(n^{-1/2-\epsilon/2}) = o(n^{-1/2}) = o(\tilde{\mathbb{P}}_{p_c}(T \geq n)). \quad (3.3.20)$$

Given (3.3.19) and (3.3.20), we eventually derive that

$$\tilde{\mathbb{P}}_{p_c} \left( \max_{t \leq T-1} \{|h(\mathcal{V}_{t+1}) - h(\mathcal{V}_t)|\} > C \cdot \log n \mid T \geq n \right) \xrightarrow{n \rightarrow +\infty} 0,$$

for  $C > 0$  large enough. With (3.3.18), this yields the limit (3.3.17).

**Scaling limit of the percolation cluster: proof of Theorem 3.6** Let  $F$  be a measurable and bounded function on the set of compact metric spaces, which is moreover continuous with respect to the Gromov–Hausdorff distance. We have:

$$\begin{aligned} \tilde{\mathbb{E}}_{p_c}[F(n^{-1/2} \cdot \tilde{\mathcal{C}}) \mid |\tilde{\mathcal{C}}| \geq n] &= \frac{1}{\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| \geq n)} \cdot \tilde{\mathbb{E}}_{p_c}[F(n^{-1/2} \cdot \tilde{\mathcal{C}}) \cdot \mathbf{1}_{|\tilde{\mathcal{C}}| \geq n}] \\ &= \frac{1}{\tilde{\mathbb{P}}_{p_c}(T \geq n/\kappa)} \cdot \tilde{\mathbb{E}}_{p_c}[F(n^{-1/2} \cdot \tilde{\mathcal{C}}) \cdot \mathbf{1}_{T \geq n/\kappa}] + o(1). \end{aligned} \quad (3.3.21)$$

In the last estimate, we use that  $F$  is bounded and (3.3.8). The limit (3.3.17) implies that:

$$\tilde{\mathbb{E}}_{p_c}[F(n^{-1/2} \cdot \tilde{\mathcal{C}}) \mid T \geq n/\kappa] \xrightarrow{n \rightarrow +\infty} \mathbb{E}[F(\kappa^{1/2} \cdot \sigma \cdot \mathcal{T}_{\geq 1})].$$

Together with (3.3.21), we get the expected conclusion for  $\kappa' = \kappa^{1/2} \cdot \sigma$ .

### 3.3.4 From SCTHP to SCT: an union of critical percolation clusters

In this section, we generalise our results on large critical percolation cluster in SCTHP, namely the second item of Theorem 3.5 and Theorem 3.6. This is a necessary intermediate step before tackling these problematics in the SCT context. To be more specific, here we study the interactions at criticality, existing between (directed) percolation clusters emanating from a growing number of vertices located on the horizontal axis  $\mathbb{N} \times \{0\}$ . Fix an infinite sequence of such vertices  $(v_k = (i_k, 0))_{k \geq 1}$ , ordered from left to right—that is  $i_k < i_{k+1}$  for every  $k \geq 1$ . We denote by  $\tilde{\mathcal{C}}_{v_k}$  the directed cluster associated to  $v_k$ . We write for any  $N \geq 1$ :

$$\tilde{\mathcal{C}}_N := \bigcup_{1 \leq k \leq N} \tilde{\mathcal{C}}_{v_k},$$

which is literally the union of the  $N$  directed percolation clusters emanating from the  $N$  leftmost vertices  $v_k$  of our sequence.

We are interested in the tail distribution of  $\tilde{\mathcal{C}}_N$  and its geometry at large scale, as  $N$  goes to infinity, when  $p = p_c$ . The growth rate of  $N$  has to remain slow compared to the volume of  $\tilde{\mathcal{C}}_N$ , though. Our main statement is the following:

**Theorem 3.7** *We fix some  $\epsilon \in (0, 1/20)$  and a sequence  $(N_n)_{n \geq 1}$  such that  $N_n = \mathcal{O}(n^\epsilon)$ . Then we have:*

$$\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}_{N_n}| \geq n) = \mathcal{O}(N_n \cdot n^{-\frac{1}{2}}) \quad (3.3.22)$$

and also:

$$(n^{-1/2} \cdot \tilde{\mathcal{C}}_{N_n} \mid |\tilde{\mathcal{C}}_{N_n}| \geq n) \xrightarrow[n \rightarrow +\infty]{d_{GH}} \kappa' \cdot \mathcal{T}_{\geq 1}, \quad (3.3.23)$$

where  $\kappa' = \kappa'(\alpha)$  is the positive number defined in Theorem 3.6 and  $\mathcal{T}_{\geq 1}$  is the CRT of mass greater than 1 [6, 7]. The convergence (3.3.23) holds in distribution for the Gromov–Hausdorff distance.

Remark that the scaling limit (3.3.23) does not depend on the specific choice of vertices  $v_k$  made in the original definition of  $\tilde{\mathcal{C}}_N$ . It turns out to be true for the asymptotic (3.3.22) as well, in the sense that we provide in the proof an universal bound to  $\limsup_{n \rightarrow +\infty} N_n^{-1} n^{1/2} \cdot \tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}_{N_n}| \geq n)$ . The convergence (3.3.23) is a bit confused, since  $\tilde{\mathcal{C}}_{N_n}$  may contain several disjoint components emanating from different vertices  $v_k$  on the  $x$ -axis, and so not to be compact. For two vertices  $v, w \in \tilde{\mathcal{C}}_{N_n}$  belonging to disjoint components, say for instance associated to  $v_k$  and  $v_{k'}$ , we set

$$\mathbf{d}(v, w) := \mathbf{d}(v, v_k) + \mathbf{d}(w, v_{k'}).$$

By doing so, we turn  $(\tilde{\mathcal{C}}_{N_n}, \mathbf{d})$  into a compact metric space.

Note that Theorem 3.5(ii) trivially implies that  $\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}_{N_n}| \geq n) = \Omega(n^{-1/2})$ , since we have

$$\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}_{N_n}| \geq n) \geq \tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}_{v_1}| \geq n) = \tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| \geq n).$$

The proof of the above theorem rests on a wider application of the peeling exploration designed in Section 3.3.1. Fix  $n \geq 1$ . We aim to discover one by one the clusters  $\mathcal{C}_{v_k}$  for  $k \leq N_n$ . Let us outline the guiding principle.

First, the vertex  $v_1$  may not be the origin vertex of the map  $\tilde{\mathfrak{T}}$ . In such case, we simply remove vertices (in the underlying SCTHP  $\tilde{\mathfrak{T}}$ ), inaccessible from  $v_1$  in a directed fashion. It merely consists of erasing the ascending trees associated to vertices on the  $x$ -axis of  $\tilde{\mathfrak{T}}$  which are on the left of  $v_1$ . The new map obtained through this operation is distributed as a SCTHP with  $v_1$  as origin vertex—see indeed Section 3.2.2. Once the pruning achieved (if needed), we perform the usual peeling exploration of  $\tilde{\mathcal{C}}_{v_1}$  described in Section 3.3.1.

Since  $p = p_c$ , the exploration of  $\tilde{\mathcal{C}}_{v_1}$  ends at some time  $T < +\infty$ . The cluster  $\tilde{\mathcal{C}}_{v_1}$  is finite, considered *explored*, as well as those of vertices  $v_s$  located in the interior of  $\Sigma_T$ . Indeed, paths emerging from such vertices cannot overcome the right boundary of the submap, so the clusters are entirely contained in  $\Sigma_T$  and we thus know everything about them. See Figure 3.10 for an illustration. We do not stop the process however, and refresh it at the leftmost vertex of our sequence, say  $v_r$ , belonging to the complement submap of  $\Sigma_T$ . We then iterate the procedure. It means that we explore now the part of  $\tilde{\mathcal{C}}_{v_r}$  contained in  $\tilde{\mathfrak{T}} \setminus \Sigma_T$ , again after a pruning of the complement submap if  $v_r$  is not its origin vertex. We carry on as long as there remains some cluster  $\tilde{\mathcal{C}}_{v_k}$ —for  $k \leq N_n$ —*unexplored*. The iteration is allowed since the complement submap is itself distributed as a SCTHP, given the spatial Markov property.

Vertices of our sequence are of two kinds. Either the peeling exploration refreshes at them—like  $v_r$  in Figure 3.10, or they are skipped—like  $v_s$ . We name the former *core vertices*. We guess indeed that clusters emanating from them contribute the most to the size and the shape of  $\tilde{\mathcal{C}}_N$ , since directed paths starting from the other vertices are enclosed in quite limited area, as it has been observed before. We will clarify what we mean below.

Obviously, the status of any vertex  $v_k$  of our sequence is not deterministic, except  $v_1$  which is a core vertex by convention. From now on, we denote by  $\mathcal{R} \in [1, N_n]$  the random number of core vertices (or refreshment times). For any  $1 \leq r \leq \mathcal{R}$ , we write  $\mathbf{v}_r^*$  for the  $r$ -th core vertex met and  $\mathcal{C}_r^*$  to designate the part of the percolation cluster emanating from  $\mathbf{v}_r^*$ , contained in the current complement submap—that is the complement submap at the  $r$ -th refreshment time. Note that  $\mathcal{C}_1^* = \tilde{\mathcal{C}}_{v_1} \stackrel{(d)}{\sim} \tilde{\mathcal{C}}$ .

A straightforward consequence of the spatial Markov property is that for any  $1 \leq r \leq N_n$ :

$$\mathcal{C}_r^* \stackrel{(d)}{\sim} \tilde{\mathcal{C}} \quad \text{conditionally on the event } \mathcal{R} \geq r,$$

and  $\mathcal{C}_r^*$  is independent of the part of  $\tilde{\mathcal{C}}_{N_n}$  explored before the  $r$ -th refreshment time. In particular, the distribution of  $\mathcal{C}_r^*$  does not depend on the vertex  $\mathbf{v}_r^*$  of our sequence  $(v_k)_k$  where this refreshment occurs.

As said above, the intuition is that the clusters emanating from core vertices concentrate most of the mass of  $\tilde{\mathcal{C}}_{N_n}$ . Ultimately, we are going to prove that only one of them does, as the following proposition sums it up:

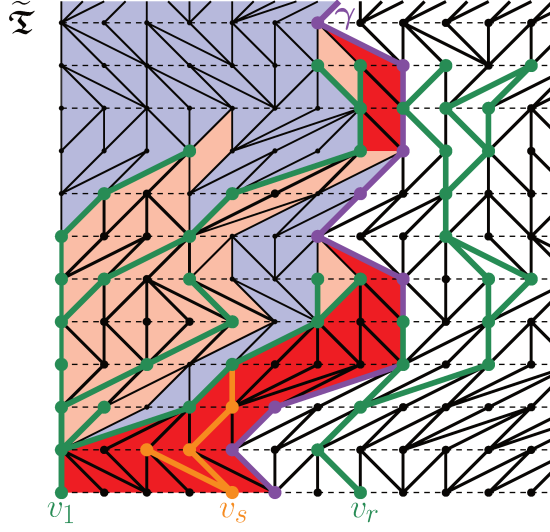


Figure 3.10: Illustration of our generalized peeling exploration. We start by exploring the tree skeleton of  $v_1$  (in green)—see Figure 3.8 for details. The revealed submap  $\Sigma_T$  is decomposed in blue and red areas corresponding to the successive upward/downward revelations. The former are not accessible (in a directed fashion) from  $v_1$ , whereas the latter might be. From the vertex  $v_s$  on the left of the purple path  $\gamma$  (the right boundary of  $\Sigma_T$ ), we can only reach vertices either contained in  $\tilde{\mathcal{C}}_{v_1}$  or in the bottom red area enclosed between the tree skeleton and  $\gamma$ . Some vertices in the latter category are here drawn in orange. The exploration then refreshes at  $v_r$ . We repeat the above procedure to reveal the unknown part of  $\tilde{\mathcal{C}}_{v_r}$  located on the right of  $\gamma$ . Some vertices of the interior of  $\Sigma_T$  might be accessible from  $v_r$ , but they are necessarily in red areas bordering  $\gamma$ .

**Proposition 3.5** *There exists a positive valued function  $\varphi$  satisfying  $\varphi(n) = o(n^{1/2})$ , such that if we define the events*

$$A_r := (\mathcal{R} \geq r) \cap (|\mathcal{C}_r^*| \geq n) \quad \text{and} \quad B_r := (\mathcal{R} \geq r) \cap (|\tilde{\mathcal{C}}_{N_n} \setminus \mathcal{C}_r^*| < \varphi(n))$$

for any  $1 \leq r \leq N_n$ , then it holds that:

$$\tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A_r \cap B_r^c \right) = o(n^{-1/2-\epsilon}) \quad (3.3.24)$$

and

$$\tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}_{N_n}| \geq n) \underset{n \rightarrow \infty}{\sim} \tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A_r \cap B_r \right). \quad (3.3.25)$$

The asymptotic equivalent (3.3.25) is the mathematical translation of what we have claimed on the mass spreading across  $\tilde{\mathcal{C}}_{N_n}$ , that only one cluster emerging from a core vertex has a macroscopic size. The proof of the above proposition is postponed to the end of the current section. Proposition 3.5 is all we need to prove both items of Theorem 3.7. We deal with them in separate paragraphs.

**The asymptotic of  $\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}_{N_n}| \geq n)$**  We go further than (3.3.25) by using (3.3.24):

$$\tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}_{N_n}| \geq n) \underset{n \rightarrow \infty}{\sim} \tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A_r \right). \quad (3.3.26)$$

The union  $\bigcup_{r=1}^{N_n} A_r = (\exists 1 \leq r \leq N_n : \mathcal{R} \geq r \text{ and } |\mathcal{C}_r^*| \geq n)$  can be written as a disjoint union of the events  $A_r \cap (\forall r' < r : |\mathcal{C}_{r'}^*| < n)$  for  $1 \leq r \leq N_n$ . By doing so and applying the spatial

Markov property, we derive that:

$$\begin{aligned} \tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A_r \right) &= \sum_{r=1}^{N_n} \tilde{\mathbb{P}}_{p_c} \left( A_r \cap (\forall r' < r : |\mathbf{c}_{r'}^*| < n - \varphi(n)) \right) \\ &= \tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}| \geq n) \cdot \sum_{r=1}^{N_n} \tilde{\mathbb{P}}_{p_c} \left( (\mathcal{R} \geq r) \cap (\forall r' < r : |\mathbf{c}_{r'}^*| < n) \right). \end{aligned} \quad (3.3.27)$$

From this equality, we directly deduce the upper bound:

$$\tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A_r \right) \leq N_n \cdot \tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}| \geq n).$$

Added to Theorem 3.5(ii), the fact that  $\varphi(n) = o(n)$  as  $n \rightarrow +\infty$  and the asymptotic (3.3.26), this leads to (3.3.22).

**Remark 3.3** *The asymptotic (3.3.26) combined with the equality (3.3.27) ensures that for any function  $\psi$  such that  $\psi(n) = o(n)$ :*

$$1 \leq \frac{\tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}_{N_n}| \geq n)}{\tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}_{N_n}| \geq n + \psi(n))} \leq \frac{\tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}| \geq n)}{\tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}| \geq n + \psi(n))} \stackrel{\text{Theorem 3.5(ii)}}{=} 1 + o(1).$$

Thus, we have  $\tilde{\mathbb{P}}_{p_c} (n \leq |\tilde{\mathcal{C}}_{N_n}| < n + \psi(n)) = o(\tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}_{N_n}| \geq n))$  as  $n \rightarrow +\infty$ .

**The scaling limit of  $\tilde{\mathcal{C}}_{N_n}$**  Let  $F$  be a measurable and bounded function on the set of compact metric spaces, which is moreover uniformly continuous with respect to the Gromov–Hausdorff distance. We get:

$$\begin{aligned} &\tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \tilde{\mathcal{C}}_{N_n}) \cdot \mathbf{1}_{|\tilde{\mathcal{C}}_{N_n}| \geq n}] \\ &= \tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \tilde{\mathcal{C}}_{N_n}) \cdot \mathbf{1}_{\bigcup_{r=1}^{N_n} A_r \cap B_r}] + o(\tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}_{N_n}| \geq n)), \end{aligned} \quad (3.3.28)$$

by using the fact that  $F$  is bounded, together with (3.3.25). Note that events  $A_r \cap B_r$  are disjoint for large  $n$  because  $\varphi(n) = (n^{1/2})$  as  $n \rightarrow +\infty$ . Thus:

$$\tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \tilde{\mathcal{C}}_{N_n}) \cdot \mathbf{1}_{\bigcup_{r=1}^{N_n} A_r \cap B_r}] = \sum_{r=1}^{N_n} \tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \tilde{\mathcal{C}}_{N_n}) \cdot \mathbf{1}_{A_r \cap B_r}],$$

for  $n$  sufficiently large. Given the definition of the shortest-path distance  $\mathbf{d}$  and since any  $\mathbf{c}_r^*$ —if it exists—is obviously a subset of  $\tilde{\mathcal{C}}_{N_n}$ , the Gromov-Hausdorff distance between them can be upper bounded as follows:

$$\mathbf{d}_{GH}(\tilde{\mathcal{C}}_{N_n}, \mathbf{c}_r^*) \leq \mathbf{d}_H(\tilde{\mathcal{C}}_{N_n}, \mathbf{c}_r^*) \leq |\tilde{\mathcal{C}}_{N_n} \setminus \mathbf{c}_r^*|,$$

where  $\mathbf{d}_H$  is the Hausdorff distance between compact sets of  $\tilde{\mathcal{C}}_{N_n}$ . It implies that on the event  $A_r \cap B_r$ :

$$\mathbf{d}_{GH}(n^{-1/2} \cdot \tilde{\mathcal{C}}_{N_n}, n^{-1/2} \cdot \mathbf{c}_r^*) \leq n^{-1/2} \cdot \varphi(n),$$

for any  $1 \leq r \leq N_n$ . If we denote by  $\omega(\cdot)$  the modulus of continuity of  $F$ , we deduce that:

$$\left| \tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \tilde{\mathcal{C}}_{N_n}) \cdot \mathbf{1}_{A_r \cap B_r}] - \tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \mathbf{c}_r^*) \cdot \mathbf{1}_{A_r \cap B_r}] \right| \leq \omega\left(\frac{\varphi(n)}{n^{1/2}}\right) \cdot \tilde{\mathbb{P}}_{p_c}(A_r \cap B_r),$$

again for any  $1 \leq r \leq N_n$ . By using (3.3.24) and the boundedness of  $F$ , we have:

$$\tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \mathbf{c}_r^*) \cdot \mathbf{1}_{A_r}] - \tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \mathbf{c}_r^*) \cdot \mathbf{1}_{A_r \cap B_r}] = o(n^{-1/2-\epsilon}),$$

uniformly with respect to  $r$ . Remark also that  $\lim_{n \rightarrow +\infty} \omega\left(\frac{\varphi(n)}{n^{1/2}}\right) = 0$  because  $\varphi(n) = o(n^{1/2})$  and  $F$  being uniformly continuous. Then, for  $n$  large enough:

$$\begin{aligned} & \tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \tilde{\mathcal{C}}_{N_n}) \cdot \mathbf{1}_{\cup_{r=1}^{N_n} A_r \cap B_r}] \\ &= \sum_{r=1}^{N_n} \tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \mathcal{C}_r^*) \cdot \mathbf{1}_{A_r}] + o\left(\tilde{\mathbb{P}}_{p_c}\left(\bigcup_{r=1}^{N_n} A_r \cap B_r\right)\right) + o(n^{-1/2}) \\ &= \sum_{r=1}^{N_n} \tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \mathcal{C}_r^*) \cdot \mathbf{1}_{A_r}] + o\left(\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}_{N_n}| \geq n)\right). \end{aligned} \quad (3.3.29)$$

The last estimate is due to (3.3.25). We state now that:

$$\sum_{r=1}^{N_n} \tilde{\mathbb{P}}_{p_c} \left( (\exists r' < r : |\mathcal{C}_{r'}^*| \geq n) \cap A_r \right) = \mathcal{O}(n^{-1+2\epsilon}). \quad (3.3.30)$$

Indeed, for any  $1 \leq r' < r \leq N_n$ :

$$\begin{aligned} \tilde{\mathbb{P}}_{p_c}(A_r \cap A_{r'}) &= \tilde{\mathbb{P}}_{p_c}(|\mathcal{C}_r^*| \geq n \mid \mathcal{R} \geq r) \cdot \tilde{\mathbb{P}}_{p_c}((|\mathcal{C}_{r'}^*| \geq n) \cap (\mathcal{R} \geq r)) \\ &\leq \tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| \geq n) \cdot \tilde{\mathbb{P}}_{p_c}(A_{r'}) \quad (\text{spatial Markov property}) \\ &\leq \tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| \geq n)^2, \end{aligned}$$

since  $\tilde{\mathbb{P}}_{p_c}(A_{r'}) \leq \tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| \geq n)$ , again thanks to the spatial Markov property. From the last inequality, we derive that:

$$\sum_{r=1}^{N_n} \tilde{\mathbb{P}}_{p_c} \left( (\exists r' < r : |\mathcal{C}_{r'}^*| \geq n) \cap A_r \right) \leq \sum_{r=1}^{N_n} \sum_{r'=1}^{r-1} \tilde{\mathbb{P}}_{p_c}(A_r \cap A_{r'}) \leq N_n^2 \cdot \tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| \geq n)^2.$$

Theorem 3.5(ii) then implies (3.3.30). Since the function  $F$  is bounded and  $\epsilon < 1/4$ , we can change (3.3.29) into:

$$\begin{aligned} & \tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \tilde{\mathcal{C}}_{N_n}) \cdot \mathbf{1}_{\cup_{r=1}^{N_n} A_r \cap B_r}] \\ &= \sum_{r=1}^{N_n} \tilde{\mathbb{E}}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}_r^*) \cdot \mathbf{1}_{A_r \cap (\forall r' < r: |\mathcal{C}_{r'}^*| < n)} \right] + o\left(\tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}_{N_n}| \geq n)\right). \end{aligned}$$

We apply the spatial Markov property like in (3.3.27) to deduce that:

$$\begin{aligned} & \sum_{r=1}^{N_n} \tilde{\mathbb{E}}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}_r^*) \cdot \mathbf{1}_{A_r \cap (\forall r' < r: |\mathcal{C}_{r'}^*| < n)} \right] \\ &= \tilde{\mathbb{E}}_{p_c} \left[ F(n^{-1/2} \cdot \tilde{\mathcal{C}}) \cdot \mathbf{1}_{|\tilde{\mathcal{C}}| \geq n} \right] \cdot \sum_{r=1}^{N_n} \tilde{\mathbb{P}}_{p_c} \left( (\mathcal{R} \geq r) \cap (\forall r' < r: |\mathcal{C}_{r'}^*| < n) \right). \end{aligned}$$

By using (3.3.28), together with (3.3.26) and (3.3.27), we get that:

$$\tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \tilde{\mathcal{C}}_{N_n}) \mid |\tilde{\mathcal{C}}_{N_n}| \geq n] = \tilde{\mathbb{E}}_{p_c} [F(n^{-1/2} \cdot \tilde{\mathcal{C}}) \mid |\tilde{\mathcal{C}}| \geq n] + o(1).$$

The scaling limit in Theorem 3.7 is then a consequence of Theorem 3.6.

**How the mass spreads across  $\tilde{\mathcal{C}}_{N_n}$**  Here we prove Proposition 3.5. Set  $\epsilon \in (0, 1/20)$ . We begin with a preliminary statement on the maximal time spent between two refreshment event. We write  $T_r$  for the length of the exploration of  $\mathcal{C}_r^*$ . By the spatial Markov property, as  $\mathcal{R} \geq r$ , the random variable  $T_r$  is distributed as the hitting time  $T$  of Proposition 3.4. Our claim is:

$$\tilde{\mathbb{P}}_{p_c} \left( \max_{1 \leq r \leq \mathcal{R}} T_r > n^{1+5\epsilon} \right) = o(n^{-1/2-2\epsilon}). \quad (3.3.31)$$

It stems directly from an union bound argument. With the help of this technical result, we gradually rule out the negligible contributions to the mass of  $\tilde{\mathcal{C}}_{N_n}$ .

We deal first with the specific contribution of non core vertices, that is the number of vertices of  $\tilde{\mathcal{C}}_{N_n}$  which are only accessible from them. We observe on Figure 3.10, that they are all located in the interior of bottom "red areas", those revealed right before a refreshment through some downward revelation. Note that there are as many as core vertices. We denote these sets by  $\mathfrak{A}_r$  for  $1 \leq r \leq \mathcal{R}$ . As  $\mathcal{R} \geq r$ , given the spatial Markov property, the volume of  $\mathfrak{A}_r$  is controlled by that of a subcritical Galton–Watson tree  $t_r \stackrel{(d)}{\sim} t \stackrel{(d)}{\sim} \mathbf{GW}_{1-\alpha}$ , whose distribution is exponentially-tailed. See the analysis of the peeling exploration algorithm in Section 3.3.1. Fix  $1 \leq r \leq N_n$  and  $t \leq n^{1+5\epsilon}$ . We have:

$$\begin{aligned} & \tilde{\mathbb{P}}_{p_c} \left( |\mathfrak{A}_r| > n^\delta; T_r = t \mid \mathcal{R} \geq r \right) \\ & \leq \tilde{\mathbb{P}}_{p_c} \left( |t_r| > n^\delta; T_r = t \mid \mathcal{R} \geq r \right) \leq \tilde{\mathbb{P}}_{p_c} \left( |t| > n^\delta \mid \mathcal{R} \geq r \right) = oe_\delta(n), \end{aligned}$$

for any  $\delta > 0$ . We recall that the notation  $oe$  for asymptotics was introduced in Definition 3.3. Now, we sum over  $t \leq n^{1+7\epsilon}$  and use (3.3.31) to deduce that

$$\tilde{\mathbb{P}}_{p_c} \left( |\mathfrak{A}_r| > n^\delta \mid \mathcal{R} \geq r \right) = o(n^{-1/2-2\epsilon}).$$

This holds again for any  $\delta > 0$ . Finally, by an union bound argument, we get that:

$$\forall \delta > 0, \quad \tilde{\mathbb{P}}_{p_c} \left( \sum_{r=1}^{\mathcal{R}} |\mathfrak{A}_r| > N_n \cdot n^\delta \right) = o(n^{-1/2-\epsilon}). \quad (3.3.32)$$

A roughly similar demonstration works for the (left-behind by the peeling exploration) contributions  $(\tilde{\mathcal{C}}_{\mathbf{v}_r} \setminus \mathcal{C}_r^*)_{\mathcal{R}}$  coming from clusters associated to core vertices. This time, vertices of  $\tilde{\mathcal{C}}_{\mathbf{v}_r} \setminus \mathcal{C}_r^*$  might be found in several distinct "red areas" generated during the exploration of  $\mathcal{C}_{r-1}^*$ . We witness on Figure 3.10 that the latter are necessarily bordering the left boundary of the complement submap obtained after the  $r$ -th refreshment. Hence, their number is bounded by the amount of vertices of  $\mathcal{C}_r^*$  tracked on the left boundary throughout the  $r$ -th exploration. Such quantity is geometrically distributed. Indeed, if at some point of the exploration of  $\mathcal{C}_r^*$ , an upward revelation occurs and an infinite part of the SCTHP is unveiled—because the underlying ascending tree is infinite, then we are sure that we will never meet again the left boundary in the future. This event has of course a non-zero chance to occur since ascending trees are supercritical. Thus, for any  $\delta > 0$ , the number of "red areas" containing vertices of  $\tilde{\mathcal{C}}_{\mathbf{v}_r} \setminus \mathcal{C}_r^*$  is lower than  $n^\delta$  with probability  $1 - oe_\delta(n)$  as  $n \rightarrow +\infty$ . Furthermore, the total number of "red areas" generated throughout the  $(r-1)$ -th exploration is bounded  $T_{r-1}$ , so is by  $n^{1+5\epsilon}$  with probability  $1 - o(n^{-1/2-2\epsilon})$ —see (3.3.31). Large deviations theory ensures that the probability for the total volume of  $n^\delta$  i.i.d. subcritical Galton–Watson trees to be greater than  $n^{2\delta}$  is  $oe_{2\delta}(n)$ . Since there are at most  $n^{(1+5\epsilon)n^\delta} = e^{(1+5\epsilon) \cdot \log(n) \cdot n^\delta}$  ways to choose  $n^\delta$  "red areas" among  $n^{1+5\epsilon}$ , we derive from an union bound argument that:

$$\forall \delta > 0, \quad \tilde{\mathbb{P}}_{p_c} \left( |\tilde{\mathcal{C}}_{\mathbf{v}_r} \setminus \mathcal{C}_r^*| > n^{2\delta}; T_r \leq n^{1+5\epsilon} \mid \mathcal{R} \geq r \right) = oe_\delta(n).$$

And as for (3.3.32), it implies that:

$$\forall \delta > 0, \quad \tilde{\mathbb{P}}_{p_c} \left( \sum_{r=1}^{\mathcal{R}} |\tilde{\mathcal{C}}_{\mathbf{v}_r} \setminus \mathcal{C}_r^*| > N_n \cdot n^\delta \right) = o(n^{-1/2-\epsilon}). \quad (3.3.33)$$

Take now  $\delta = 9\epsilon$ . It stems from (3.3.32) and (3.3.33) that the mass of  $\tilde{\mathcal{C}}_{N_n} \setminus \bigcup_{r=1}^{\mathcal{R}} \mathcal{C}_r^*$  does not exceed  $2N_n \cdot n^{9\epsilon}$  with probability  $1 - o(n^{-\frac{1}{2}-\epsilon})$ . Meanwhile, by using the spatial Markov property, combined with an union bound argument and Theorem 3.5(ii), we obtain that:

$$\begin{aligned} & \tilde{\mathbb{P}}_{p_c} \left( \exists 1 \leq r \neq r' \leq \mathcal{R} \text{ s.t. } |\mathcal{C}_r^*| \geq n - 4N_n \cdot n^{9\epsilon} \text{ and } |\mathcal{C}_{r'}^*| \geq n^{9\epsilon} \right) \\ & = \mathcal{O}(N_n^2 \cdot n^{-1/2-9\epsilon/2}) = \mathcal{O}(n^{-1/2-5\epsilon/2}) = o(n^{-1/2-\epsilon}). \end{aligned}$$



Remark that  $N_n \cdot n^{9\epsilon} = \mathcal{O}(n^{10\epsilon}) = o(n^{1/2})$  as  $n \rightarrow +\infty$  because  $10\epsilon < 1/2$ . See the proof of (3.3.30) for details. Hence:

$$\tilde{\mathbb{P}}_{p_c} \left( \exists 1 \leq r \leq \mathcal{R} \text{ s.t. } |\mathcal{C}_r^*| \geq n - 4N_n \cdot n^{9\epsilon} \text{ and } \left| \bigcup_{r' \neq r} \mathcal{C}_{r'}^* \right| \geq N_n \cdot n^{9\epsilon} \right) = o(n^{-1/2-\epsilon}). \quad (3.3.34)$$

For any  $1 \leq r \leq N_n$ , we define the event  $A'_r = (\mathcal{R} \geq r) \cap (|\mathcal{C}_r^*| \geq n - \varphi(n))$  where  $\varphi(n) = 4N_n \cdot n^{9\epsilon}$ . As already noticed before, the function  $\varphi$  is  $o(n^{1/2})$  as  $n \rightarrow +\infty$ . Recall the definition of the events  $(A_r)_r$  and  $(B_r)_r$  from Proposition 3.5. The straightforward consequence of the asymptotics (3.3.32), (3.3.33) and (3.3.34) is that

$$\tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A'_r \cap B_r^c \right) = o(n^{-1/2-\epsilon}). \quad (3.3.35)$$

This yields (3.3.24) because  $A_r \subset A'_r$  for any  $1 \leq r \leq N_n$ .

As  $|\tilde{\mathcal{C}}_{N_n}| \geq n$  and  $|\tilde{\mathcal{C}}_{N_n} \setminus \bigcup_{r=1}^{\mathcal{R}} \mathcal{C}_r^*| \leq 2N_n \cdot n^{9\epsilon}$ , we necessarily have  $|\bigcup_{r=1}^{\mathcal{R}} \mathcal{C}_r^*| \geq n - 2 \cdot N_n \cdot n^{8\epsilon}$ . Since  $\mathcal{R} \leq N_n$ , it means that there exists  $1 \leq r \leq \mathcal{R}$  such that

$$|\mathcal{C}_r^*| \geq n/N_n - 2 \cdot n^{9\epsilon} \geq C \cdot n^{1-\epsilon},$$

for some  $C > 0$ . In the same way as for (3.3.34), we can prove that:

$$\tilde{\mathbb{P}}_{p_c} \left( \exists 1 \leq r \leq \mathcal{R} \text{ s.t. } |\mathcal{C}_r^*| \geq C \cdot n^{1-\epsilon} \text{ and } \left| \bigcup_{r' \neq r} \mathcal{C}_{r'}^* \right| \geq N_n \cdot n^{9\epsilon} \right) = o(n^{-1/2-\epsilon}).$$

This implies that:

$$\tilde{\mathbb{P}}_{p_c} \left( |\tilde{\mathcal{C}}_{N_n}| \geq n; \bigcup_{r=1}^{N_n} A'_r \cap B_r \right) \underset{n \rightarrow +\infty}{\sim} \tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}_{N_n}| \geq n), \quad (3.3.36)$$

because  $\tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}_{N_n}| \geq n) = \Omega(n^{-1/2})$ .

Now, by using the fact that both  $\tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A'_r \right)$  and  $\tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A_r \right)$  are  $\Omega(n^{-1/2})$ , together with the asymptotic (3.3.35), we derive that:

$$\tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A'_r \cap B_r \right) \underset{n \rightarrow \infty}{\sim} \tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A'_r \right) \quad \text{and} \quad \tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A_r \cap B_r \right) \underset{n \rightarrow \infty}{\sim} \tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A_r \right). \quad (3.3.37)$$

Given (3.3.27), we know furthermore that:

$$1 \leq \frac{\tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A'_r \right)}{\tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A_r \right)} \leq \frac{\tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}| \geq n - \varphi(n))}{\tilde{\mathbb{P}}_{p_c} (|\tilde{\mathcal{C}}| \geq n)} \xrightarrow[n \rightarrow +\infty]{\text{Theorem 3.5(ii)}} 1. \quad (3.3.38)$$

We deduce that  $\tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A'_r \right) \underset{n \rightarrow \infty}{\sim} \tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A_r \right)$ . Then, by (3.3.37):

$$\tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A'_r \cap B_r \right) \underset{n \rightarrow \infty}{\sim} \tilde{\mathbb{P}}_{p_c} \left( \bigcup_{r=1}^{N_n} A_r \cap B_r \right).$$

Finally, we remark that  $\bigcup_{r=1}^{N_n} A_r \cap B_r \subset \{|\tilde{\mathcal{C}}_{N_n}| \geq n\}$ . So the latter asymptotic equivalent, combined with (3.3.36), completes the proof of (3.3.25) and of the proposition.

**Remark 3.4** *By reading carefully the section, we observe that Theorem 3.7 would still hold by taking random numbers  $N_n = \mathbf{N}_n$  satisfying  $\mathbb{P}(\mathbf{N}_n \leq Cn^\epsilon) = 1$  for some  $C > 0$  and any  $n \geq 1$ , provided that they are generated independently of the map and the percolation process.*

### 3.4 Percolation on SCT

We aim to transfer back to the SCT model the results on SCTHP obtained in Section 3.3. Far away from the root, the landscape around some fixed vertex in a SCT looks a lot like that in a SCTHP. There are however two significant differences. In the former, the rightmost path starting from some vertex  $v$  can rotate counterclockwise around the underlying Galton–Watson tree, and finally merge up at some point with the leftmost path. Such a thing cannot happen in a SCTHP. See Figure 3.11. The second important difference lies in the fact that we define a SCT from a supercritical Galton–Watson tree **conditioned to survive**. The conditioning kills the independence property between subtrees emerging from cousins, as it exists in the non conditioned version of the model or, to continue the comparison, between ascending trees in a SCTHP (see Proposition 3.1). In Section 3.4.1, we present a powerful trick to compensate such lack of independence and creates "hard boundaries" in SCT. We derive from it statements on the percolation probability function outlined in Theorem 3.1 and Theorem 3.2. In Section 3.4.2, we establish results on large critical clusters, namely Theorem 3.2(i) and the scaling limit of Theorem 3.3. Finally, in Section 3.4.3, we prove that the number of disjoint clusters in the SCT is infinite in the supercritical phase.

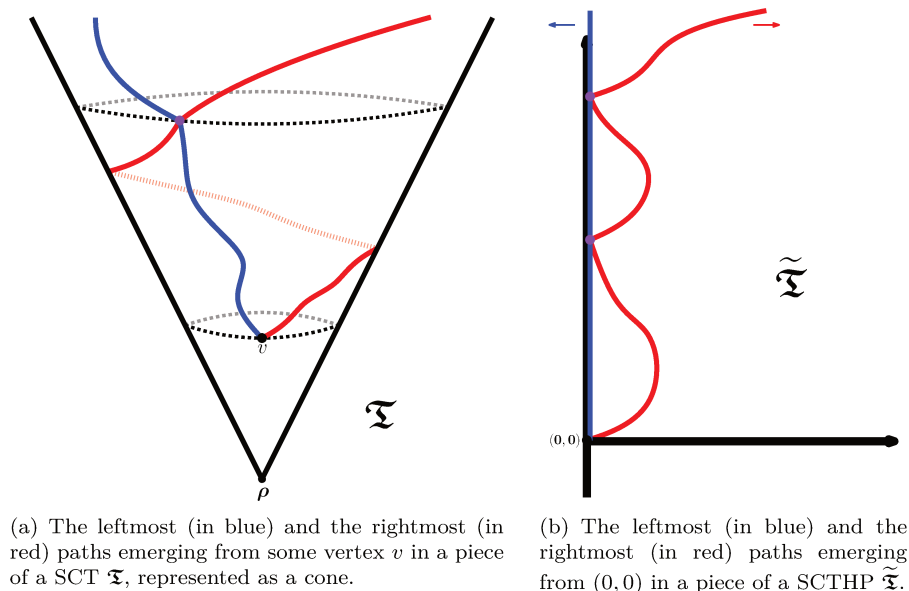


Figure 3.11: In a SCT, the leftmost path emerging from a vertex  $v$  can at some point touch on its left flank the rightmost path, after "rotating" around the cone. Such situation is excluded in the SCTHP model which is infinite in the horizontal direction.

#### 3.4.1 SCT and wide trees

In a SCT, far away from the root vertex, the bad situation depicted in Figure 3.11(a) does not occur in most cases. It is very likely that some middle subtree emerging from a cousin of  $v$  is "wide enough" to prevent it. What is at stake is to clarify the term "wide enough". Take any (finite or infinite) plane tree  $t$ . Start by removing—if possible—the leftmost children of the root, as well as the subtree emerging from it. Iterate indefinitely the procedure by consistently removing, at each step, the leftmost children in  $t$  of the vertices still present in the previous generation. If the subtree of  $t$  obtained after such pruning, is infinite, we say that  $t$  is *wide*. See Figure 3.12(a) for an illustration of the pruning operation. In this example, the tree is wide since the part made of the edges in black continuous line—that we call the *wide component*—is infinite.

When  $t = \mathbf{T}_\infty$  is a supercritical Galton–Watson tree conditioned to survive, with a geometric offspring distribution like in our model, it is wide with a positive chance:

**Lemma 3.2** *We have  $\mathbb{P}(\mathbf{T}_\infty \text{ is wide}) > 0$ .*

In fact, the result holds for any supercritical Galton–Watson tree, regardless of the offspring distribution. See [43, Lemma 3] for a proof of the statement. The author actually deals with a more general mechanism, in which are removed at each step a certain number of children—not necessarily a single one like us, nor even a constant number.

As argued in the foregoing, our interest for wide trees results from our wish to prevent events like that represented in Figure 3.11. Our observation is that wide trees fulfill their role by creating "hard boundaries" in the SCT as soon as they arise. Assume indeed that the subtree  $\mathbf{T}_v$  emerging from some vertex  $v \in \mathbf{T}_\infty$  is a wide tree. In the SCT derived from  $\mathbf{T}_\infty$ , any path starting from any cousin  $w$  of  $v$ —a vertex which is at same height as  $v$  in  $\mathbf{T}_\infty$ —can hit but cannot go through the infinite line made of the vertices bordering on its left the wide component of  $\mathbf{T}_v$  (in orange on Figure 3.12). Note that the latter line is by construction an existing path in the SCT.

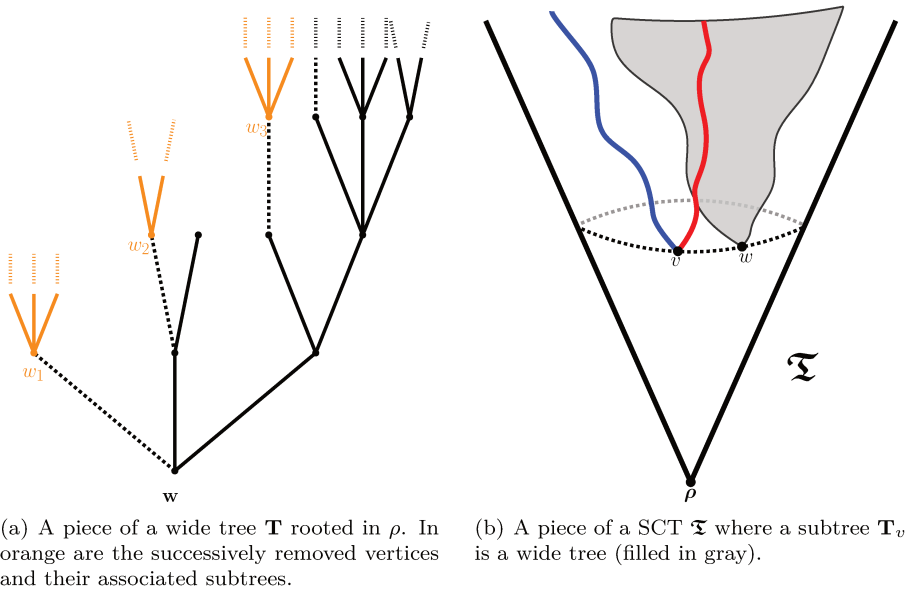


Figure 3.12: In a SCT, wide trees—like  $\mathbf{T}_v$  on the right figure—create hard boundaries. A path emerging from any cousin of  $v$  cannot overcome (but can reach) the orange line made of the successively removed vertices of  $\mathbf{T}_v$ , as we were checking whether the latter is wide or not. See for instance the rightmost path starting at  $w \neq v$ , coloured in red.

Furthermore, some stunning fact happens when is found—in a proper way, though—a wide tree in the SCT. The landscape around it is actually the same as that witnessed in a SCTHP. Indeed, if we condition for instance the underlying Galton–Watson tree  $\mathbf{T}_\infty$ —which is already conditioned to be infinite—so that at a given height  $h \geq 1$ , the subtree emanating from the leftmost vertex is wide, then the subtrees emanating from its cousins turn out to be i.i.d. with common distribution  $\mathbf{GW}_\alpha$ . This is simply because the forced condition on the former subtree is more binding than that on  $\mathbf{T}_\infty$  (to be infinite) and bring with it. In doing so, we consequently lift all constraints on the other subtrees emanating from height  $h$ . Meanwhile, we remark that requiring a Galton–Watson tree to be wide does not demand anything on the subtrees emerging from the "removed vertices"—those in orange on Figure 3.12(a)—since they are by construction outside the wide component. The latter subtrees are then distributed as usual  $\mathbf{GW}_\alpha$ .

To put it in a nutshell, as soon as we condition  $\mathbf{T}_\infty$  on the event that at height  $h$ , the subtree emanating from the leftmost vertex is wide, the environment seen in the SCT by its cousins, looking upward, is as if they were in a SCTHP. In particular, their associated (directed) percolation cluster is distributed as  $\tilde{\mathcal{C}}$ , the percolation cluster of the origin in a SCTHP. This is enough to prove the lower bound of Theorem 3.2(ii). Indeed, the event on which we condition occurs with a positive probability. Because the structure of  $\mathbf{T}_\infty$  below height  $h$  is independent of that above, it is still the case if in addition, we impose that there exists a second leftmost vertex at height  $h$  and that

the root of the SCT is connected to it. Given what it has just been argued, the probability that an infinite cluster emerges from the second leftmost vertex, conditionally on the previous event, is exactly  $\tilde{\Theta}(p)$ , where  $\tilde{\Theta}$  is the percolation probability function of the SCTHP model. We deduce that

$$\Theta(p) \geq a_1 \cdot \tilde{\Theta}(p), \tag{3.4.1}$$

for some  $a_1 \in (0, 1)$ . By using the third point of Theorem 3.5, we conclude that  $\Theta(p)$  is greater than a linear function in  $p - p_c$  for  $p > p_c$ . In particular, we have  $\Theta(p) > 0$  for such  $p$ , as stated in Theorem 3.1.

Obviously, the subtree emanating from the leftmost vertex is wide with a non zero probability, but not equal to one. So our method does not work as such to reverse the inequality 3.4.1 between  $\Theta(p)$  and  $\tilde{\Theta}(p)$ , which would be enough to establish the upper bound of Theorem 3.2(ii). Nevertheless, there exists a smart way to choose a vertex in  $\mathbf{T}_\infty$ —not just the leftmost at some given height—for which we roughly get in the same position:

**Lemma 3.3** *By some explicit procedure, we find almost surely a vertex  $\mathbf{v}^* \in \mathbf{T}_\infty$  such that:*

1. *the subtree emanating from  $\mathbf{v}^*$  is wide;*
2. *it is disconnected from the previous generation of vertices in  $\mathbf{T}_\infty$  after the percolation process is performed on the SCT;*
3. *the number of cousins of  $\mathbf{v}^*$  has a finite mean;*
4. *conditionally on  $\mathbf{v}^*$  and its number of cousins, the other subtrees in  $\mathbf{T}_\infty$  emanating from the latter are i.i.d. with common distribution  $\mathbf{GW}_\alpha$ .*

Admit temporarily the lemma to see how it helps us to get the desired result. Given the second point, an infinite path emerging from the root vertex of the SCT necessarily visits some cousin  $w$  of  $\mathbf{v}^*$ . The percolation cluster associated to  $w$  is then infinite. Thanks to the first and fourth points, it is distributed as the cluster of the origin in a SCTHP. By an union bound argument, we obtain:

$$\Theta(p) \leq \tilde{\Theta}(p) \cdot \mathbb{E}[\text{number of cousins of } \mathbf{v}^*].$$

Then we use the third point of Lemma 3.3, together with Theorem 3.5, to get as expected the upper bound of Theorem 3.2(ii). Note that the above inequality also implies that  $\Theta(p) = 0$  for  $p \leq p_c$ , which completes the proof of the first part of Theorem 3.1.

**Proof of Lemma 3.3** The broad idea of the proof is that a wide subtree can always be found in  $\mathbf{T}_\infty$  by going far enough from the root vertex. We proceed as follows. We start by choosing a vertex of the tree  $\mathbf{T}_\infty$ , which is—as required in Lemma 3.3—disconnected from the previous generation of vertices after the percolation process is performed on the SCT  $\mathfrak{T}$ . For instance, the leftmost one at minimal distance from the root vertex. Given the independance between the percolation process and the map, it is clear that such vertex always exists. We denote it by  $\mathbf{v}_1$ .

The number of cousins of  $\mathbf{v}_1$  has a finite mean. Indeed, for any  $h \geq 1$ , the total number of vertices of  $\mathbf{T}_\infty$  at height  $h$  has a finite mean and is negatively correlated with the event that the height of  $\mathbf{v}_1$  is greater than  $h$ , since more vertices gives more possibilities to disconnect one from the previous generation. Furthermore, this number is known to grow like  $m^h$  [89, 102], where  $m$  is the mean of the offspring distribution of  $\mathbf{T}_\infty$ . As  $h \rightarrow +\infty$ , according to a law of large numbers, a positive proportion of vertices at height  $h$  are linked in the map  $\mathfrak{T}$  to only one vertex from the previous generation. The probability that the percolation process does not disconnect any of them thus decreases super-exponentially fast when  $h \rightarrow +\infty$ . As a consequence, we have:

$$\sum_h m^h \cdot \mathbb{P}(\text{height of } \mathbf{v}_1 \geq h) < +\infty,$$

and this yields the expected conclusion.

We reveal now whether the subtree  $\mathbf{T}_{\mathbf{v}_1}$  emanating from  $\mathbf{v}_1$  is *wide* or not. It is with positive probability, and if so, we stop the algorithm and simply set  $\mathbf{v}^* = \mathbf{v}_1$ . Indeed, the first three conditions of Lemma 3.3 are obviously satisfied by  $\mathbf{v}_1$ . The fourth too, because the "wideness" of

the subtree ensures that the underlying Galton–Watson tree  $\mathbf{T}_\infty$  is indeed infinite. We already met this argument to prove the inequality (3.4.1).

When the subtree of  $\mathbf{v}_1$  is not wide, its *wide component*—that is the black part of the tree depicted on Figure 3.12(a)—is by definition finite. If its height is  $H \geq 0$  and if  $\mathbf{v}_1$  had height  $h \geq 1$ , we pursue our search for the vertex  $\mathbf{v}^*$  above height  $h + H + 1$ , by repeating the pattern that we have just described. More precisely, we look now for a vertex  $\mathbf{v}_2$ , at distance at least  $h + H + 1$  from the root vertex, disconnected from the previous generation of vertices after the percolation process is performed on the SCT. Like was  $\mathbf{v}_1$ . We choose again the leftmost one, with minimal height, and check whether the subtree  $\mathbf{T}_{\mathbf{v}_2}$  is wide or not. If it is, then  $\mathbf{v}^* = \mathbf{v}_2$ . Otherwise, we continue higher in the tree  $\mathbf{T}_\infty$ , and so on.

Why are we sure that we only need a finite number of steps before finding a vertex  $\mathbf{v}^*$  satisfying the four conditions of Lemma 3.3? Vertices above height  $h + H + 1$  all belong to subtrees emerging either from cousins of  $\mathbf{v}_1$  or from the orange "removed vertices" of  $\mathbf{T}_{\mathbf{v}_1}$  (see Figure 3.12(a)). Conditionally on the number of vertices belonging to one of these two sets, the latter subtrees are distributed as i.i.d. Galton–Watson trees with common distribution  $\mathbf{GW}_\alpha$ , but conditioned on the event that at least one of them is infinite, because we have to guarantee that the underlying tree  $\mathbf{T}_\infty$  is indeed infinite. Furthermore, the structure of these subtrees is (still conditionally) independent of the structure of  $\mathbf{T}_\infty$  below height  $h$  and of the finite wide component of  $\mathbf{T}_{\mathbf{v}_1}$ . The same things hold for subtrees emanating from vertices at height  $h + H + 1$ , conditionally on their number. As a consequence, the environment on which is carried out the second step of the algorithm consists of a finite collection of i.i.d. Galton–Watson trees with common distribution  $\mathbf{GW}_\alpha$ , conditioned on the event that at least one of them survives. The environment was the same during the initial step, except that there was only one subtree, that of the root vertex, namely  $\mathbf{T}_\infty$ . It is clear that the probability for the subtree  $\mathbf{T}_{\mathbf{v}_2}$  to be wide is lower bounded by a positive constant, uniformly in the number of trees forming the environment. So, by iterating the argument to the next steps—if  $\mathbf{T}_{\mathbf{v}_2}$  is not wide—and using Borel–Cantelli lemma, we conclude that the algorithm stops almost surely.

### 3.4.2 Large critical percolation clusters in SCT

In this section, we demonstrate the first item of Theorem 3.2 which provides an asymptotic on the tail distribution of the size of a critical percolation cluster (emerging from the root vertex) in a SCT, as well as the scaling limit of Theorem 3.3. An approximation (in distribution) of the cluster by that of the origin vertex in a SCTHP is mainly used.

We rely again on the emergence of wide trees in the underlying supercritical Galton–Watson tree  $\mathbf{T}_\infty$ , conditioned to survive. First, it is worth noting that we easily obtain some equivalent of (3.4.1) involving  $\mathbb{P}_{p_c}(|\mathcal{C}| \geq n)$ , that is:

$$\mathbb{P}_{p_c}(|\mathcal{C}| \geq n) \geq a_1 \cdot \tilde{\mathbb{P}}_{p_c}(|\tilde{\mathcal{C}}| \geq n), \quad (3.4.2)$$

where  $a_1 \in (0, 1)$ . Together with Theorem 3.5(ii), this implies that  $\mathbb{P}_{p_c}(|\mathcal{C}| \geq n) = \Omega(n^{-1/2})$ . As for (3.4.1), the idea is here to condition on a smartly chosen event having a non-zero chance to occur, under which the mass of  $\mathcal{C}$  is mostly supported by some sub-cluster distributed as  $\tilde{\mathcal{C}}$ . This can be achieved for instance, by conditioning on the joint events "the subtree emanating from the leftmost children of the root vertex of  $\mathbf{T}_\infty$  is *wide*" and "the root vertex is only connected to its second leftmost children". Indeed, the latter event ensures that the size of the percolation cluster emanating from the second leftmost children is at least  $n - 1$ , while the former event implies that its distribution is that of  $\tilde{\mathcal{C}}$ , as argued in the previous section. Hence (3.4.2).

The approach to get an upper bound on  $\mathbb{P}_{p_c}(|\mathcal{C}| \geq n)$  is roughly to say that wide trees quickly appear in  $\mathbf{T}_\infty$ , that is not so far from the root vertex, thus most of the mass of  $\mathcal{C}$  is always concentrated in a region of the SCT akin to a SCTHP. We are going to sum up such observation in a statement but we first need to introduce several notations.

In what follows, for any  $h \geq 1$ , we write  $\mathbf{B}_h$  to designate the ball of radius  $h$  of  $\mathbf{T}_\infty$  centered at the root vertex, which is by definition the set of vertices of  $\mathbf{T}_\infty$  whose height is at most  $h$ . We denote by  $\partial\mathbf{B}_r$  its boundary, that is the set of vertices whose height is exactly  $h$ . Now we set for

any  $N \geq 1$ :

$$\mathbf{H}_N := 1 + \inf\{h \geq 1 \text{ s.t. } |\partial\mathbf{B}_h| > N\}.$$

The random variable  $\mathbf{H}_N$  is almost surely finite because  $\partial\mathbf{B}_h \rightarrow +\infty$  as  $h \rightarrow +\infty$ . Finally, for any  $h, N \geq 1$ , we define the event  $W_h^N$  on which  $|\partial\mathbf{B}_h| > 2N$  and there exists two vertices  $v \neq w$  of  $\partial\mathbf{B}_h$  such that:

1. one is among the  $N$  leftmost vertices of  $\partial\mathbf{B}_h$ , the other among the  $N$  rightmost;
2. subtrees emanating from them in  $\mathbf{T}_\infty$  are both *wide*;
3. they are disconnected from  $\partial\mathbf{B}_{h-1}$  (in the SCT subject to the percolation process).

The heralded statement is then:

**Lemma 3.4** *For any  $\beta > 0$  large enough, we have:*

$$\mathbb{P}\left(m/2 \leq \frac{|\partial\mathbf{B}_{\mathbf{H}_{\beta \log n}}|}{\beta \log n} \leq 4m^2\right) = 1 - o_{n \rightarrow \infty}(n^{-1/2}), \quad (3.4.3)$$

and for any  $\delta > 0$ :

$$\mathbb{P}\left(|\mathbf{B}_{\mathbf{H}_{\beta \log n}}| \leq n^\delta\right) = 1 - o(n^{-1/2}). \quad (3.4.4)$$

Moreover, there exists some  $C > 0$  such that—again for  $\beta > 0$  large enough:

$$\mathbb{P}_{p_c}\left(W_{\mathbf{H}_{\beta \log n}}^C\right) = 1 - o_{n \rightarrow \infty}(n^{-1/2}). \quad (3.4.5)$$

We postpone the proof of the lemma at the end of the current section. The asymptotics (3.4.4) and (3.4.5) together say that as wide trees emerge in  $\mathbf{T}_\infty$  at the (random) height  $\mathbf{H}_{\beta \log n}$  with probability  $1 - o(n^{-1/2})$ , the volume of  $\mathbf{T}_\infty$  standing below is still subpolynomial in  $n$ . Under these conditions, most of the mass of a large cluster  $\mathcal{C}$ —that is with a mass larger than  $n$ —would necessarily be concentrated in a region of the SCT where some wide trees exists, making the part of the map there look like a SCTHP. Given our preliminary result stating that  $\mathbb{P}_{p_c}(|\mathcal{C}| \geq n) = \Omega(n^{-1/2})$ , the main consequence of Lemma 3.4 is that we can work under such conditions to study a large cluster. To be more precise, we have

$$\mathbb{P}_{p_c}(|\mathcal{C}| \geq n) \underset{n \rightarrow +\infty}{\sim} \mathbb{P}_{p_c}(|\mathcal{C}| \geq n \mid E_{\beta, n}), \quad (3.4.6)$$

if we denote by  $E = E_{\beta, n} = E_{\beta, C, \delta, n}$  the intersection of the three events whose the asymptotic probability is evaluated in Lemma 3.4.

On the event  $E$ , the tree  $\mathbf{T}_\infty$  contains at most  $n^\delta$  vertices up to height  $\mathbf{H}_{\beta \log n}$ . If the cluster  $\mathcal{C}$  itself contains more than  $n$  vertices, there will be at least  $n - n^\delta$  among them, located above or at the height  $\mathbf{H}_{\beta \log n}$ .

On the event  $E$  too, there exists two distinct vertices of  $\partial\mathbf{B}_{\mathbf{H}_{\beta \log n}}$  from which are emanating wide subtrees of  $\mathbf{T}_\infty$ . They are also disconnected from the previous generation in the percolated SCT. In particular, they cannot belong to  $\mathcal{C}$ . We write  $\mathbf{w}_\ell$  and  $\mathbf{w}_r$  for the leftmost and the rightmost such vertex, respectively. Conditionally on them, subtrees emanating from their cousins are mutually independent. The distribution of those being on the left of  $\mathbf{T}_{\mathbf{w}_\ell}$ , resp. on the right of  $\mathbf{T}_{\mathbf{w}_r}$ , depends on whether the root vertex is connected or not to  $\partial\mathbf{B}_{\mathbf{H}_{\beta \log n} - 1}$ . For those concerned, they are distributed as  $\mathbf{GW}_\alpha$  but conditioned to be *non wide*. For the others, the distribution is simply  $\mathbf{GW}_\alpha$ . Between  $\mathbf{T}_{\mathbf{w}_\ell}$  and  $\mathbf{T}_{\mathbf{w}_r}$ , subtrees are all distributed as  $\mathbf{GW}_\alpha$ .

The main consequence of the last point is that, as already explained in Section 3.4.1, the (upward directed) environment of the SCT seen by vertices flanked by  $\mathbf{w}_\ell$  and  $\mathbf{w}_r$  is like in a SCTHP. It is slightly more complex elsewhere given the *non wide* conditioning on some subtrees. That brings us to split the subgraph of  $\mathcal{C}$  starting at height  $\mathbf{H}_{\beta \log n}$  into two parts, denoted by  $\mathcal{C}_{mid}$  and  $\mathcal{C}_{out}$ . The former part gathers sub-clusters emanating from vertices of  $\mathcal{C}$  located in the *middle area* (between  $\mathbf{w}_\ell$  and  $\mathbf{w}_r$ ), while the latter does with those remaining outside. Conditionally on these vertices and on the event  $E$ , the graphs  $\mathcal{C}_{mid}$  and  $\mathcal{C}_{out}$  are independent, because the subtrees—on which they stand—are themselves.

The random number of vertices of  $\mathcal{C}$  in the middle area only depends on the structure of the percolated SCT up to height  $\mathbf{H}_{\beta \log n}$ . Given its value  $n' \geq 1$  and the event  $E$ , the graph  $\mathcal{C}_{mid}$  is distributed as  $\tilde{\mathcal{C}}_{n'}$ —see Section 3.3.4. Since we have furthermore  $|\partial \mathbf{B}_{\mathbf{H}_{\beta \log n}}| \leq 4m^2\beta \log n$  as the event  $E$  occurs, we get by using Theorem 3.7:

$$\mathbb{P}_{p_c}(|\mathcal{C}_{mid}| \geq n \mid E_{\beta,n}) = \mathcal{O}(\log(n) \cdot n^{-1/2}). \quad (3.4.7)$$

The same holds even more for  $\mathcal{C}_{out}$  although it could seem less obvious. Indeed, the potential *non wide* conditioning on some subtrees outside the middle area merely pushes down (in distribution) their volume, and consequently, the number of vertices which are accessible (in a directed fashion) from their root vertex in the SCT. So does once the map is percolated, on top of that.

Remember that the total mass  $|\mathcal{C}_{mid}| + |\mathcal{C}_{out}|$  is greater than  $n - n^\delta$  on the event  $\{|\mathcal{C}| \geq n\}$ . One of the two parts of course captures a half of it. From (3.4.7) and their conditional independence given  $E$ , we derive that the latter receives in fact, as  $n \rightarrow +\infty$ , an overwhelming proportion of the total mass. We mean that conditionally on both events  $E$  and  $\{|\mathcal{C}| \geq n\}$ , we have for instance either

$$|\mathcal{C}_{mid}| \geq n - 2n^\delta \quad \text{and} \quad |\mathcal{C}_{out}| \leq n^\delta,$$

or the converse (by switching roles of  $\mathcal{C}_{mid}$  and  $\mathcal{C}_{out}$ ), with probability tending to one. Hence, we get from (3.4.6):

$$\begin{aligned} & \mathbb{P}_{p_c}(|\mathcal{C}| \geq n \mid E_{\beta,n}) \\ & \underset{n \rightarrow +\infty}{\sim} \mathbb{P}_{p_c}(|\mathcal{C}_{mid}| \geq n - 2n^\delta; |\mathcal{C}| \geq n \mid E_{\beta,n}) + \mathbb{P}_{p_c}(|\mathcal{C}_{out}| \geq n - 2n^\delta; |\mathcal{C}| \geq n \mid E_{\beta,n}). \end{aligned} \quad (3.4.8)$$

However, chances that  $\mathcal{C}_{out}$  overtakes  $\mathcal{C}_{mid}$  are slim. Indeed, as the event  $E$  happens, the number of vertices on the left of  $\mathbf{w}_\ell$  or on the right of  $\mathbf{w}_r$  is at most  $2C \log n - 2$ , while there are more than  $m\beta/2 \log n - 2C \log n$  in the middle area. Remind you—by reading again through the proof of Theorem 3.7 in Section 3.3.4—that this is actually a single sub-cluster of  $\mathcal{C}_{mid} \cup \mathcal{C}_{out}$ —emanating from a vertex of the middle area or outside—which monopolizes most of the mass, and not  $\mathcal{C}_{mid}$  or  $\mathcal{C}_{out}$  as a whole. Given the rotationally invariant distribution of the SCT, it is clear that the probability that the sub-cluster emerges from the middle area increase up to one as  $\beta \rightarrow +\infty$ . Therefore:

$$\lim_{\beta \rightarrow +\infty} \limsup_{n \rightarrow +\infty} \frac{\mathbb{P}_{p_c}(|\mathcal{C}_{out}| \geq n - 2n^\delta; |\mathcal{C}| \geq n \mid E_{\beta,n})}{\mathbb{P}_{p_c}(|\mathcal{C}_{mid}| \geq n - 2n^\delta; |\mathcal{C}| \geq n \mid E_{\beta,n})} = 0. \quad (3.4.9)$$

With the help of Remark 3.3, and since  $\{|\mathcal{C}_{mid}| \geq n\} \subset \{|\mathcal{C}| \geq n\}$ , we remove the " $2n^\delta$ " of the event involving  $\mathcal{C}_{mid}$ . Then, from the asymptotic (3.4.8) and the limit (3.4.9), we derive that:

$$1 \leq \liminf_{n \rightarrow +\infty} \frac{\mathbb{P}_{p_c}(|\mathcal{C}| \geq n \mid E_{\beta,n})}{\mathbb{P}_{p_c}(|\mathcal{C}_{mid}| \geq n \mid E_{\beta,n})} \leq \limsup_{n \rightarrow +\infty} \frac{\mathbb{P}_{p_c}(|\mathcal{C}| \geq n \mid E_{\beta,n})}{\mathbb{P}_{p_c}(|\mathcal{C}_{mid}| \geq n \mid E_{\beta,n})} \leq 1 + M_\beta, \quad (3.4.10)$$

with  $M_\beta \xrightarrow{\beta \rightarrow +\infty} 0$ . Combined with (3.4.7), this is enough to complete the proof of Theorem 3.2(i).

We roughly adopt the same strategy to prove the scaling limit of Theorem 3.3. Take  $F$  a measurable and bounded function on the set of compact metric spaces, which is moreover uniformly continuous with respect to the Gromov–Hausdorff distance. On the event

$$E_{\beta,n} \cap \{|\mathcal{C}| \geq n\} \cap \{|\mathcal{C}_{mid}| \geq n - 2n^\delta\} \cap \{|\mathcal{C}_{out}| \leq n^\delta\},$$

we have

$$\mathbf{d}_{GH}\left(n^{-1/2} \cdot \mathcal{C}, n^{-1/2} \cdot \mathcal{C}_{mid}\right)_{\mathcal{C}_{mid} \subset \mathcal{C}} \leq n^{-1/2} \cdot 2n^\delta \xrightarrow{n \rightarrow +\infty} 0,$$

for  $\delta < 1/2$ . Since  $F$  is uniformly continuous, we deduce that:

$$\begin{aligned} & \mathbb{E}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}) \cdot \mathbf{1}_{|\mathcal{C}_{mid}| \geq n - 2n^\delta} \mathbf{1}_{|\mathcal{C}_{out}| \leq n^\delta} \mathbf{1}_{|\mathcal{C}| \geq n} \mid E_{\beta,n} \right] \\ & = \mathbb{E}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}_{mid}) \cdot \mathbf{1}_{|\mathcal{C}_{mid}| \geq n - 2n^\delta} \mathbf{1}_{|\mathcal{C}_{out}| \leq n^\delta} \mathbf{1}_{|\mathcal{C}| \geq n} \mid E_{\beta,n} \right] + o\left(\mathbb{P}_{p_c}(|\mathcal{C}| \geq n)\right). \end{aligned}$$

The same holds by switching roles of  $\mathcal{C}_{mid}$  and  $\mathcal{C}_{out}$ . Then, by using in addition the boundedness of  $F$  and the arguments developed up to (3.4.8), we get:

$$\begin{aligned} \mathbb{E}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}) \cdot \mathbf{1}_{|\mathcal{C}| \geq n} \right] &= \mathbb{E}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}_{mid}) \cdot \mathbf{1}_{|\mathcal{C}_{mid}| \geq n-2n^\delta} \cdot \mathbf{1}_{|\mathcal{C}| \geq n} \mid E_{\beta,n} \right] \\ &+ \mathbb{E}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}_{out}) \cdot \mathbf{1}_{|\mathcal{C}_{out}| \geq n-2n^\delta} \cdot \mathbf{1}_{|\mathcal{C}| \geq n} \mid E_{\beta,n} \right] + o\left(\mathbb{P}_{p_c}(|\mathcal{C}| \geq n)\right). \end{aligned} \quad (3.4.11)$$

The boundedness of  $F$  combined with (3.4.9) and Remark 3.3 ensures that:

$$\limsup_{n \rightarrow +\infty} \left| \frac{\mathbb{E}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}_{out}) \cdot \mathbf{1}_{|\mathcal{C}_{out}| \geq n-2n^\delta} \cdot \mathbf{1}_{|\mathcal{C}| \geq n} \mid E_{\beta,n} \right]}{\mathbb{P}_{p_c}(|\mathcal{C}_{mid}| \geq n \mid E_{\beta,n})} \right| \leq M_\beta \cdot \|F\|_\infty, \quad (3.4.12)$$

where  $M_\beta$  tends to 0 as  $\beta \rightarrow +\infty$ .

As explained above (3.4.7), conditionally on both events  $E_{n,\beta}$  and  $\{|\mathcal{C}_{mid}| \geq n\}$ , the graph  $\mathcal{C}_{mid}$  is distributed as  $\tilde{\mathcal{C}}_{\mathbf{N}}$  conditioned to have a mass larger than  $n$ , where  $\mathbf{N}$  is a random variable almost surely lower than  $4m^2\beta \log n$ , drawn independently of the underlying SCTHP and the associated percolation process. See Section 3.3.4. The quantity  $\mathbf{N}$  corresponds to the number of vertices of  $\mathcal{C}$  in the middle area at height  $\mathbf{H}_{\beta \log n}$ . Given the straightforward extension of Theorem 3.7 stated in Remark 3.4, we derive that:

$$\mathbb{E}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}_{mid}) \mid |\mathcal{C}_{mid}| \geq n; E_{\beta,n} \right] \xrightarrow[n \rightarrow +\infty]{} \mathbb{E} \left[ F(\kappa' \cdot \mathcal{T}_{\geq 1}) \right],$$

where  $\kappa' = \kappa'(\alpha)$  is the positive number defined in Theorem 3.6 and  $\mathcal{T}_{\geq 1}$  is the CRT of mass greater than 1. Then:

$$\begin{aligned} &\mathbb{E}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}_{mid}) \cdot \mathbf{1}_{|\mathcal{C}_{mid}| \geq n-2n^\delta} \cdot \mathbf{1}_{|\mathcal{C}| \geq n} \mid E_{\beta,n} \right] \cdot \mathbb{P}_{p_c}(|\mathcal{C}_{mid}| \geq n \mid E_{\beta,n})^{-1} \\ &= \underset{\text{Remark (3.3)}}{\mathbb{E}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}_{mid}) \mid |\mathcal{C}_{mid}| \geq n; E_{\beta,n} \right]} + o(1) \xrightarrow[n \rightarrow +\infty]{} \mathbb{E} \left[ F(\kappa' \cdot \mathcal{T}_{\geq 1}) \right]. \end{aligned} \quad (3.4.13)$$

The inequalities in (3.4.10), the asymptotic (3.4.11), the limit (3.4.13) and, the upper bound (3.4.12) together imply that:

$$\limsup_{n \rightarrow +\infty} \mathbb{E}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}) \mid |\mathcal{C}| \geq n \right] \leq \mathbb{E} \left[ F(\kappa' \cdot \mathcal{T}_{\geq 1}) \right] + M_\beta \cdot \|F\|_\infty$$

and

$$\liminf_{n \rightarrow +\infty} \mathbb{E}_{p_c} \left[ F(n^{-1/2} \cdot \mathcal{C}) \mid |\mathcal{C}| \geq n \right] \geq (1 + M_\beta)^{-1} \cdot \mathbb{E} \left[ F(\kappa' \cdot \mathcal{T}_{\geq 1}) \right] - M_\beta \cdot (1 + M_\beta)^{-1} \cdot \|F\|_\infty$$

We conclude by taking  $\beta \rightarrow +\infty$  to get  $M_\beta \rightarrow 0$ .

**Proof of Lemma 3.4** We start with the proof of (3.4.3). For sake of simplicity, we remove the conditioning on  $\mathbf{T}_\infty$  (imposing to be infinite). We will write  $\mathbb{P}_\infty$  instead of  $\mathbb{P}$  to indicate when the conditioning is back. By definition of  $\mathbf{H}_{\beta \log n}$ , we have  $|\partial \mathbf{B}_{\mathbf{H}_{\beta \log n}-1}| \geq \lceil \beta \log n \rceil$ . So:

$$\mathbb{P} \left( |\partial \mathbf{B}_{\mathbf{H}_{\beta \log n}}| < \frac{m}{2} \beta \log n \mid \mathbf{H}_{\beta \log n} < +\infty \right) \leq \mathbb{P} \left( \sum_{i=1}^{\lceil \beta \log n \rceil} \mathbf{X}_i < \frac{m}{2} \beta \log n \right),$$

where the  $(\mathbf{X}_i)_i$  are i.i.d. random variables with distribution  $\mu_\alpha$ . According to a classic result of the large deviations theory, the right-hand side of the inequality is  $o(n^{-1/2})$  when  $n \rightarrow +\infty$ , as  $\beta$  is large enough, because  $\mu_\alpha$  is an exponentially-tailed distribution. We deduce that:

$$\mathbb{P}_\infty \left( \frac{|\partial \mathbf{B}_{\mathbf{H}_{\beta \log n}}|}{\beta \log n} < \frac{m}{2} \right) = \underset{n \rightarrow \infty}{o} (n^{-1/2}). \quad (3.4.14)$$



Similarly, by using this times that  $\partial\mathbf{B}_{\mathbf{H}_{\beta \log n - 2}} \leq \lfloor \beta \log n \rfloor$ , we get for  $\beta$  large enough:

$$\mathbb{P}\left(|\partial\mathbf{B}_{\mathbf{H}_{\beta \log n - 1}}| > 2m\beta \log n \mid \mathbf{H}_{\beta \log n} < +\infty\right) \leq \mathbb{P}\left(\sum_{i=1}^{\lfloor \beta \log n \rfloor} \mathbf{X}_i > 2m\beta \log n\right) = \underset{n \rightarrow \infty}{o}\left(n^{-1/2}\right).$$

In the same way, again for  $\beta$  sufficiently large:

$$\begin{aligned} & \mathbb{P}\left(|\partial\mathbf{B}_{\mathbf{H}_{\beta \log n}}| > 4m^2\beta \log n \mid \mathbf{H}_{\beta \log n} < +\infty; |\partial\mathbf{B}_{\mathbf{H}_{\beta \log n - 1}}| \leq 2m\beta \log n\right) \\ & \leq \mathbb{P}\left(\sum_{i=1}^{\lfloor 2m\beta \log n \rfloor} \mathbf{X}_i > 4m^2\beta \log n\right) = \underset{n \rightarrow \infty}{o}\left(n^{-1/2}\right). \end{aligned} \quad (3.4.15)$$

We derive from the two last asymptotics that:

$$\mathbb{P}_{\infty}\left(\frac{|\partial\mathbf{B}_{\mathbf{H}_{\beta \log n}}|}{\beta \log n} > 4m^2\right) = \underset{n \rightarrow \infty}{o}\left(n^{-1/2}\right), \quad (3.4.16)$$

which yields the expected asymptotic (3.4.3), by combining with (3.4.14) and using an union bound argument.

We continue with the proof of (3.4.4). Since  $|\partial\mathbf{B}_h| \leq \beta \log n$  for every  $h < \mathbf{H}_{\beta \log n} - 1$ , the volume of  $\mathbf{B}_{\mathbf{H}_{\beta \log n}}$  is bounded by

$$\mathbf{H}_{\beta \log n} \cdot \beta \log n + \partial\mathbf{B}_{\mathbf{H}_{\beta \log n}}.$$

Hence, as we condition on the event  $\{|\partial\mathbf{B}_{\mathbf{H}_{\beta \log n}}| \leq 4m^2\beta \log n\}$ , if  $|\mathbf{B}_{\mathbf{H}_{\beta \log n}}| > n^{\delta}$  for some  $\delta > 0$ , it holds that:

$$\mathbf{H}_{\beta \log n} > n^{\delta/2},$$

for  $n$  large enough. In particular, this means that  $\partial\mathbf{B}_{n^{\delta/2-1}} \leq \beta \log n$ . Now we state that if a function  $\Phi : \mathbb{N}^* \mapsto ]0, +\infty[$  satisfies  $\Phi(n) = o(m^n)$ , then we have:

$$\mathbb{P}(1 \leq |\partial\mathbf{B}_n| \leq \Phi(n)) = \mathcal{O}(\Phi(n) \cdot m^{-n}).$$

The asymptotic can be proved via an explicit computation of the generative function of  $|\partial\mathbf{B}_n|$ , by mimicking the strategy employed in the proof of Proposition 3.2. Just set  $u_0 = s$  for  $s \in [0, 1]$  instead of  $u_0 = 0$ . Otherwise, see [67] where the authors deal with such asymptotic in a much wider context. In our situation, the above statement implies that:

$$\begin{aligned} & \mathbb{P}\left(|\mathbf{B}_{\mathbf{H}_{\beta \log n}}| > n^{\delta} \mid |\partial\mathbf{B}_{\mathbf{H}_{\beta \log n}}| \leq 4m^2\beta \log n; \mathbf{H}_{\beta \log n} < +\infty\right) \\ & \leq \underset{\text{for } n \text{ large enough}}{\mathbb{P}(\partial\mathbf{B}_{n^{\delta/2-1}} \leq \beta \log n)} = \mathcal{O}(\log n \cdot m^{-n^{\delta}}) = o(n^{-1/2}), \end{aligned}$$

for any  $\delta > 0$ . Given (3.4.16), we are allowed to conclude that  $\mathbb{P}_{\infty}(|\mathbf{B}_{\mathbf{H}_{\beta \log n}}| > n^{\delta}) = o(n^{-1/2})$ . This completes the proof of (3.4.3).

We finally turn our attention to the asymptotic (3.4.5). Again, we lighten the demonstration by temporarily removing the conditioning on  $\mathbf{T}_{\infty}$ . Assume that  $\partial\mathbf{B}_h > \beta \log n$  for some height  $h \geq 1$  and large  $n$ . Choose a vertex  $v$  in  $\mathbf{B}_h$  and assume that it gives birth to (at least) two children. Then, we force the edge connecting  $v$  to its rightmost child to be *closed*, after the percolation process has been performed. We also impose that the subtree emanating from the latter is *wide*. Since the closed edge was the only possible path in the SCT between  $\partial\mathbf{B}_h$  to  $v$ , we have yielded in  $\partial\mathbf{B}_{h+1}$  a vertex which is disconnected from the previous generation and whose the associated subtree is wide. By Lemma 3.2, it has of course a positive probability to occur, namely

$$q := \alpha^2(1-p) \cdot \mathbb{P}(\mathbf{T} \text{ is wide}),$$

where  $\mathbf{T} \stackrel{(d)}{\sim} \mathbf{GW}_{\alpha}$ . The probability that such event happens for one of the  $K \leq \beta \log n$  leftmost vertices of  $\partial\mathbf{B}_h$  is  $1 - (1-q)^K$ . Consequently, it does with probability  $1 - o(n^{-1/2})$  as  $n \rightarrow +\infty$ ,

when  $K = C \log n$  for sufficiently large  $0 < C < \beta$ . Conditionally on the latter event, the leftmost vertex for which the event occurs has, by definition, at most  $C \log n$  vertices on its left in  $\partial \mathbf{B}_h$ —including itself. The number of vertices on the left of its rightmost child in  $\partial \mathbf{B}_{h+1}$  is thus stochastically dominated by the sum of  $\lfloor C \log n \rfloor$  i.i.d. random variables with common distribution  $\mu_\alpha$ . By a large deviation argument—see (3.4.15), the sum does not exceed  $2m \cdot C \log n$  with probability  $1 - o(n^{-1/2})$  as  $n \rightarrow +\infty$ . It ensures that the rightmost child is among the  $2m \cdot C \log n$  leftmost vertices of  $\partial \mathbf{B}_{h+1}$ . A symmetric argument works to find another such vertex at the opposite side of the set. This puts an end to the demonstration.

### 3.4.3 The number of clusters in the supercritical phase

This paragraph is devoted to demonstrate the last part of Theorem 3.1, saying that there are almost surely infinitely many infinite *non-directed* clusters in the SCT, as  $p_c < p < 1$ . We insist on the fact that the statement holds for the *non directed* Bernoulli percolation process, contrary to the other results.

The proof is broadly similar to that of the asymptotic (3.4.5), outlined at the end of the previous section. We remove—for sake of simplicity—the conditioning (to be infinite) on  $\mathbf{T}_\infty$  and assume that  $\partial \mathbf{B}_h > \beta \log n$  for some  $h \geq 1$  and large  $n$ . This times, we choose a vertex  $v \in \partial \mathbf{B}_h$  which gives birth to (at least) four children. We force the three edges connecting  $v$  to its three rightmost children to be *closed*, once the percolation process has been performed. We impose that subtrees emanating from the rightmost child and the third are *wide*. Such operation isolates the second rightmost child of  $v$ , denoted by  $w$ , from the rest of the SCT. Furthermore, the (directed) cluster emanating from  $w$  is distributed as  $\tilde{\mathcal{C}}$  given the *wide* status of subtrees flanking  $w$ . See Section 3.4.1. The latter cluster is then infinite with probability  $\tilde{\Theta}(p)$  and it occurs independently of the previous steps of the operation. The probability of the whole operation is

$$q' := \alpha^4 \cdot (1 - p)^3 \cdot \mathbb{P}(\mathbf{T} \text{ is wide})^2 \cdot \tilde{\Theta}(p) > 0.$$

By construction, the *non directed* clusters emanating from several distinct vertices like  $w$  are disjoint. For any  $1 \leq K \leq \beta \log n$ , conditionally on the value of  $|\partial \mathbf{B}_h|$ , the probability to find less than  $K$  in  $\partial \mathbf{B}_{h+1}$  is

$$\sum_{k=0}^K \binom{|\partial \mathbf{B}_h|}{k} (q')^k \cdot (1 - q')^{|\partial \mathbf{B}_h| - k}.$$

It tends to zero as  $n \rightarrow +\infty$  because we have assumed that  $|\partial \mathbf{B}_h| > \beta \log n$ .

The straightforward consequence is that the probability to get at least  $K$  such vertices at height  $\mathbf{H}_{\beta \log n}$ —if it exists—tends to one as  $n \rightarrow +\infty$ . Since  $\mathbf{H}_{\beta \log n} < +\infty$  for any  $n \geq 1$  in a supercritical Galton–Watson tree conditioned to survive, there exists almost surely some height at which we will find them. In particular, this implies that the SCT contains at least  $K$  disjoint infinite non-directed clusters with probability one, as  $p_c < p < 1$ . The latter being true for any  $K \geq 1$ , there are in fact infinitely many in the map.



# Bibliography

- [1] Romain Abraham and Jean-François Delmas. Local limits of conditioned Galton-Watson trees: the condensation case. *Electronic Journal of Probability*, 19:1–29, 2014.
- [2] Romain Abraham and Jean-François Delmas. An introduction to Galton-Watson trees and their local limits. *arXiv preprint arXiv:1506.05571*, 2015.
- [3] Louigi Addario-Berry, Nicolas Broutin, and Christina Goldschmidt. The continuum limit of critical random graphs. *Probability Theory and Related Fields*, 152(3-4):367–406, 2012.
- [4] Luigi Addario-Berry and Bruce A Reed. Ballot theorems, old and new. In *Horizons of combinatorics*, pages 9–35. Springer, 2008.
- [5] Michael Aizenman, Harry Kesten, and Charles M Newman. Uniqueness of the infinite cluster and continuity of connectivity functions for short and long range percolation. *Communications in Mathematical Physics*, 111(4):505–531, 1987.
- [6] David Aldous. The continuum random tree. I. *The Annals of Probability*, pages 1–28, 1991.
- [7] David Aldous. The continuum random tree III. *The Annals of Probability*, pages 248–289, 1993.
- [8] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *The Annals of Probability*, pages 812–854, 1997.
- [9] David Aldous and Russell Lyons. Processes on unimodular random networks. *Electronic Journal of Probability*, 12:1454–1508, 2007.
- [10] Jan Ambjørn, Bergfinnur Durhuus, Thordur Jonsson, and Orur Jonsson. *Quantum geometry: a statistical field theory approach*. Cambridge University Press, 1997.
- [11] Jan Ambjørn, Andrzej Görlich, Jerzy Jurkiewicz, and Renate Loll. Nonperturbative quantum gravity. *Physics Reports*, 519(4-5):127–210, 2012.
- [12] Jan Ambjørn, Jerzy Jurkiewicz, and Renate Loll. Reconstructing the universe. *Physical Review D*, 72(6):064014, 2005.
- [13] Jan Ambjørn and Renate Loll. Non-perturbative Lorentzian quantum gravity, causality and topology change. *Nuclear Physics B*, 536(1-2):407–434, 1998.
- [14] Jan Ambjørn and Y Watabiki. Scaling in quantum gravity. *Nuclear Physics B*, 445(1):129–142, 1995.
- [15] Omer Angel. Growth and percolation on the uniform infinite planar triangulation. *Geometric And Functional Analysis*, 13(5):935–974, 2003.
- [16] Omer Angel. Scaling of percolation on infinite planar maps, I. *arXiv preprint math/0501006*, 2005.
- [17] Omer Angel and Nicolas Curien. Percolations on random maps I: Half-plane models. In *Annales de l’IHP Probabilités et statistiques*, volume 51, pages 405–431, 2015.
- [18] Omer Angel, Tom Hutchcroft, Asaf Nachmias, and Gourab Ray. Hyperbolic and parabolic unimodular random maps. *Geometric and Functional Analysis*, 28(4):879–942, 2018.
- [19] Omer Angel and Gourab Ray. Classification of half-planar maps. *The Annals of Probability*, 43(3):1315–1349, 2015.
- [20] Omer Angel and Oded Schramm. Uniform Infinite Planar Triangulations. *Communications in mathematical physics*, 241(2):191–213, 2003.

- [21] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [22] Itai Benjamini and Nicolas Curien. Ergodic theory on stationary random graphs. *Electronic Journal of Probability*, 17:1–20, 2012.
- [23] Itai Benjamini and Oded Schramm. Percolation beyond  $Z^d$ , many questions and a few answers. *Electronic Communications in Probability*, 1:71–82, 1996.
- [24] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. In *Selected Works of Oded Schramm*, pages 533–545. Springer, 2011.
- [25] Noam Berger, Christian Borgs, Jennifer T Chayes, and Amin Saberi. On the spread of viruses on the internet. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 301–310. Society for Industrial and Applied Mathematics, 2005.
- [26] Noam Berger, Christian Borgs, Jennifer T Chayes, and Amin Saberi. Asymptotic behavior and distributional limits of preferential attachment graphs. *The Annals of Probability*, 42(1):1–40, 2014.
- [27] Jean Bertoin and Gerónimo Uribe Bravo. Supercritical percolation on large scale-free random trees. *The Annals of Applied Probability*, 25(1):81–103, 2015.
- [28] Jérémie Bettinelli. Scaling limit of random planar quadrangulations with a boundary. In *Annales de l’IHP Probabilités et statistiques*, volume 51, pages 432–477, 2015.
- [29] Shankar Bhamidi. Universal techniques to analyze preferential attachment trees: Global and Local analysis. 2007. preprint.
- [30] Béla Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316, 1980.
- [31] Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.
- [32] Béla Bollobás and Oliver Riordan. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics*, 1(1):1–35, 2004.
- [33] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.
- [34] J. Bouttier, P. Francesco, and E. Guitter. Planar maps as labeled mobiles. *Electron. J. Comb.*, 11, 2004.
- [35] Tom Britton, Maria Deijfen, and Anders Martin-Löf. Generating simple random graphs with prescribed degree distribution. *Journal of statistical physics*, 124(6):1377–1397, 2006.
- [36] Simon R Broadbent and John M Hammersley. Percolation processes: I. crystals and mazes. In *Mathematical proceedings of the Cambridge philosophical society*, volume 53, pages 629–641. Cambridge University Press, 1957.
- [37] Sébastien Bubeck, Luc Devroye, and Gábor Lugosi. Finding Adam in random growing trees. *Random Structures & Algorithms*, 50(2):158–172, 2017.
- [38] Sébastien Bubeck, Ronen Eldan, Elchanan Mossel, and Miklós Z. Rácz. From trees to seeds: On the inference of the seed from large trees in the uniform attachment model. *Bernoulli*, 23(4A):2887–2916, 11 2017.
- [39] Sébastien Bubeck, Elchanan Mossel, and Miklós Z Rácz. On the influence of the seed graph in the preferential attachment model. *IEEE Transactions on Network Science and Engineering*, 2(1):30–39, 2015.
- [40] Timothy George Budd. *Non-perturbative quantum gravity: a conformal perspective*. Utrecht University, 2012.
- [41] Thomas Budzinski. *Cartes aléatoires hyperboliques*. PhD thesis, Université Paris-Saclay, 2018.
- [42] Thomas Budzinski. The hyperbolic Brownian plane. *Probability Theory and Related Fields*, 171(1):503–541, 2018.
- [43] Thomas Budzinski. Supercritical causal maps: geodesics and simple random walk. *Electronic Journal of Probability*, 24, 2019.

- [44] Thomas Budzinski. Infinite geodesics in hyperbolic random triangulations. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 1129–1161. Institut Henri Poincaré, 2020.
- [45] Thomas Budzinski and Baptiste Louf. Local limits of uniform triangulations in high genus. *Inventiones mathematicae*, 223(1):1–47, 2021.
- [46] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33. American Mathematical Soc., 2001.
- [47] Robert M Burton and Michael Keane. Density and uniqueness in percolation. *Communications in mathematical physics*, 121(3):501–505, 1989.
- [48] Philippe Chassaing and Gilles Schaeffer. Random planar lattices and integrated superBrownian excursion. *Probability Theory and Related Fields*, 128(2):161–212, 2004.
- [49] Xinxin Chen and Grégory Miermont. Long Brownian bridges in hyperbolic spaces converge to Brownian trees. *Electronic Journal of Probability*, 22:1–15, 2017.
- [50] Fan Chung and Linyuan Lu. The diameter of sparse random graphs. *Advances in Applied Mathematics*, 26(4):257–279, 2001.
- [51] Robert Cori and Bernard Vauquelin. Planar maps are well labeled trees. *Canadian Journal of Mathematics*, 33(5):1023–1042, 1981.
- [52] Nicolas Curien. Planar stochastic hyperbolic triangulations. *Probability Theory and Related Fields*, 165(3):509–540, 2016.
- [53] Nicolas Curien. Peeling random planar maps. *Saint-Flour lecture notes*, 2019.
- [54] Nicolas Curien, Thomas Duquesne, Igor Kortchemski, and Ioan Manolescu. Scaling limits and influence of the seed graph in preferential attachment trees (limites d'échelle et ontogénèse des arbres construits par attachement préférentiel). *Journal de l'École polytechnique—Mathématiques*, 2:1–34, 2015.
- [55] Nicolas Curien, Bénédicte Haas, and Igor Kortchemski. The CRT is the scaling limit of random dissections. *Random Structures & Algorithms*, 47(2):304–327, 2015.
- [56] Nicolas Curien, Tom Hutchcroft, and Asaf Nachmias. Geometric and spectral properties of causal maps. *Journal of the European Mathematical Society*, 22(12):3997–4024, 2020.
- [57] Nicolas Curien and Igor Kortchemski. Percolation on random triangulations and stable looptrees. *Probability Theory and Related Fields*, 163(1):303–337, 2015.
- [58] Nicolas Curien and Jean-François Le Gall. Scaling limits for the peeling process on random maps. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 53, pages 322–357. Institut Henri Poincaré, 2017.
- [59] Nicolas Curien and Grégory Miermont. Uniform infinite planar quadrangulations with a boundary. *Random Structures & Algorithms*, 47(1):30–58, 2015.
- [60] Pierre de La Harpe. *Topics in geometric group theory*. University of Chicago Press, 2000.
- [61] Luc Devroye and Tommy Reddad. On the discovery of the seed in uniform attachment trees. *arXiv preprint arXiv:1810.00969*, 2018.
- [62] Thomas Duquesne. A limit theorem for the contour process of conditioned Galton–Watson trees. *The Annals of Probability*, 31(2):996–1027, 2003.
- [63] Thomas Duquesne and Jean-François Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probability Theory and Related Fields*, 131(4):553–603, 2005.
- [64] Bergfinnur Durhuus, Thordur Jonsson, and John F Wheeler. On the spectral dimension of causal triangulations. *Journal of Statistical Physics*, 139(5):859–881, 2010.
- [65] Paul Erdos and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5(1):17–60, 1960.
- [66] Robert Fitzner and Remco van der Hofstad. Mean-field behavior for nearest-neighbor percolation in  $d > 10$ . *Electronic Journal of Probability*, 22:1–65, 2017.
- [67] Klaus Fleischmann and Vitali Wachtel. Lower deviation probabilities for supercritical Galton–Watson processes. In *Annales de l'IHP Probabilités et statistiques*, volume 43, pages 233–255, 2007.

- [68] Cornelius Marius Fortuin and Piet W Kasteleyn. On the random-cluster model: I. introduction and relation to other models. *Physica*, 57(4):536–564, 1972.
- [69] Jean-François Le Gall and Grégory Miermont. Scaling limits of random planar maps with large faces. *The Annals of Probability*, 39(1):1 – 69, 2011.
- [70] Edgar N Gilbert. Random graphs. *The Annals of Mathematical Statistics*, 30(4):1141–1144, 1959.
- [71] Matthias Gorny, Édouard Maurel-Segala, and Arvind Singh. The geometry of a critical percolation cluster on the UIPT. In *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, volume 54, pages 2203–2238. Institut Henri Poincaré, 2018.
- [72] G.R. Grimmett. *Percolation*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2013.
- [73] Bénédicte Haas and Grégory Miermont. Scaling limits of Markov branching trees with applications to Galton–Watson and random unordered trees. *The Annals of Probability*, 40(6):2589–2666, 2012.
- [74] Takashi Hara and Gordon Slade. Mean-field critical behaviour for percolation in high dimensions. *Communications in Mathematical Physics*, 128(2):333–391, 1990.
- [75] Takashi Hara and Gordon Slade. Mean-field behaviour and the lace expansion. In *Probability and phase transition*, pages 87–122. Springer, 1994.
- [76] Theodore Edward Harris. *The theory of branching processes*, volume 6. Springer Berlin, 1963.
- [77] Remco van der Hofstad. *Random Graphs and Complex Networks*, volume 1 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 2016.
- [78] Remco van der Hofstad. *Random Graphs and Complex Networks. Volume 2*. To appear in Cambridge Series in Statistical and Probabilistic Mathematics. A preliminary version is available at <https://www.win.tue.nl/~rhofstad/>. 2021+.
- [79] Fereydoun Hormozdiari, Petra Berenbrink, Nataša Pržulj, and S Cenk Sahinalp. Not all scale-free networks are born equal: the role of the seed graph in ppi network evolution. *PLoS computational biology*, 3(7):e118, 2007.
- [80] Tom Hutchcroft. Percolation on hyperbolic graphs. *Geometric and Functional Analysis*, 29(3):766–810, 2019.
- [81] Peter Jagers. Some notes on the history of branching processes, from my perspective. 2009.
- [82] Svante Janson. Monotonicity, asymptotic normality and vertex degrees in random graphs. *Bernoulli*, 13(4):952–965, 2007.
- [83] Varun Jog and Po-Ling Loh. Analysis of centrality in sublinear preferential attachment trees via the crump-mode-jagers branching process. *IEEE Transactions on Network Science and Engineering*, 4(1):1–12, 2016.
- [84] David G Kendall. The genealogy of genealogy branching processes before (and after) 1873. *Bulletin of the London Mathematical Society*, 7(3):225–253, 1975.
- [85] Harry Kesten. The critical probability of bond percolation on the square lattice equals  $1/2$ . *Communications in mathematical physics*, 74(1):41–59, 1980.
- [86] Harry Kesten. Subdiffusive behavior of random walk on a random cluster. In *Annales de l’IHP Probabilités et statistiques*, volume 22, pages 425–487, 1986.
- [87] Harry Kesten. Scaling relations for 2D-percolation. *Communications in Mathematical Physics*, 109(1):109–156, 1987.
- [88] Harry Kesten, P Ney, and F Spitzer. The Galton-Watson process with mean one and finite variance. *Theory of Probability & Its Applications*, 11(4):513–540, 1966.
- [89] Harry Kesten and Bernt P Stigum. A limit theorem for multidimensional Galton-Watson processes. *The Annals of Mathematical Statistics*, 37(5):1211–1223, 1966.
- [90] Robert D Kleinberg and Jon M Kleinberg. Isomorphism and embedding problems for infinite limits of scale-free graphs. In *SODA*, pages 277–286, 2005.

- [91] Vadim G Knizhnik, Alexander M Polyakov, and Alexander B Zamolodchikov. Fractal structure of 2D-quantum gravity. *Modern Physics Letters A*, 3(08):819–826, 1988.
- [92] Igor Kortchemski. A simple proof of Duquesne’s theorem on contour processes of conditioned Galton–Watson trees. In *Séminaire de Probabilités XLV*, pages 537–558. Springer, 2013.
- [93] Bernhard H Korte, Jens Vygen, B Korte, and J Vygen. *Combinatorial optimization*, volume 1. Springer, 2011.
- [94] Paul L Krapivsky, Sidney Redner, and Francois Leyvraz. Connectivity of growing random networks. *Physical review letters*, 85(21):4629, 2000.
- [95] Maxim Krikun. Local structure of random quadrangulations. *arXiv preprint math/0512304*, 2005.
- [96] Jean-François Le Gall. *Spatial branching processes, random snakes and partial differential equations*. Springer Science & Business Media, 1999.
- [97] Jean-François Le Gall. Random trees and applications. *Probability surveys*, 2:245–311, 2005.
- [98] Jean-François Le Gall. Uniqueness and universality of the Brownian map. *The Annals of Probability*, 41(4):2880–2960, 2013.
- [99] Jean-François Le Gall and Frédéric Paulin. Scaling limits of bipartite planar maps are homeomorphic to the 2-sphere. *Geometric and Functional Analysis*, 18(3):893–918, 2008.
- [100] Gabor Lugosi and Alan S Pereira. Finding the seed of uniform attachment trees. *arXiv preprint arXiv:1801.01816*, 2018.
- [101] Russell Lyons. Random walks and percolation on trees. *The annals of Probability*, 18(3):931–958, 1990.
- [102] Russell Lyons, Robin Pemantle, and Yuval Peres. Conceptual proofs of L log L criteria for mean behavior of branching processes. *The Annals of Probability*, pages 1125–1138, 1995.
- [103] Russell Lyons, Robin Pemantle, and Yuval Peres. Ergodic theory on Galton–Watson trees: speed of random walk and dimension of harmonic measure. *Ergodic Theory and Dynamical Systems*, 15(3):593–619, 1995.
- [104] Russell Lyons and Yuval Peres. *Probability on trees and networks*, volume 42. Cambridge University Press, 2017.
- [105] David Corlin Marchand. Percolation on supercritical causal triangulations. *in preparation*, 2021.
- [106] David Corlin Marchand and Ioan Manolescu. Influence of the seed in affine preferential attachment trees. *Bernoulli*, 26(3):1665–1705, 2020.
- [107] Jean-François Marckert and Grégory Miermont. Invariance principles for random bipartite planar maps. *The Annals of Probability*, 35(5):1642–1705, 2007.
- [108] Cyril Marzouk. Scaling limits of random bipartite planar maps with a prescribed degree sequence. *Random Structures & Algorithms*, 53(3):448–503, 2018.
- [109] Laurent Ménard and Pierre Nolin. Percolation on uniform infinite planar maps. *Electronic Journal of Probability*, 19:1–27, 2014.
- [110] Marcus Michelen, Robin Pemantle, and Josh Rosenberg. Quenched survival of Bernoulli percolation on Galton–Watson trees. *Journal of Statistical Physics*, 181(4):1323–1364, 2020.
- [111] Manuel Middendorf, Etay Ziv, and Chris H Wiggins. Inferring network mechanisms: the drosophila melanogaster protein interaction network. *Proceedings of the National Academy of Sciences of the United States of America*, 102(9):3192–3197, 2005.
- [112] Grégory Miermont. On the sphericity of scaling limits of random planar quadrangulations. *Electronic Communications in Probability*, 13:248–257, 2008.
- [113] Grégory Miermont. The Brownian map is the scaling limit of uniform random plane quadrangulations. *Acta mathematica*, 210(2):319–401, 2013.
- [114] Grégory Miermont. Aspects of random maps. *Saint-Flour lecture notes*, 2014.
- [115] T Móri. On random trees. *Studia Scientiarum Mathematicarum Hungarica*, 39(1-2):143–155, 2002.



- [116] Tamás F Móri. The maximum degree of the barabási–albert random tree. *Combinatorics, Probability and Computing*, 14(03):339–348, 2005.
- [117] Mark Newman. *Networks*. Oxford university press, 2018.
- [118] Roberto Oliveira and Joel Spencer. Connectivity transitions in networks with super-linear preferential attachment. *Internet Math.*, 2(2):121–163, 2005.
- [119] Erol Peköz, Adrian Röllin, and Nathan Ross. Joint degree distributions of preferential attachment random graphs. *Advances in Applied Probability*, 49(2):368–387, 2017.
- [120] Erol A Peköz, Adrian Röllin, and Nathan Ross. Degree asymptotics with rates for preferential attachment random graphs. *The Annals of Applied Probability*, 23(3):1188–1218, 2013.
- [121] Gourab Ray. Geometry and percolation on half planar triangulations. *Electronic Journal of Probability*, 19:1–28, 2014.
- [122] Loïc Richier. *Géométrie et percolation sur des cartes à bord aléatoires*. PhD thesis, ENS Lyon, 2017.
- [123] Oliver Riordan and Nicholas Wormald. The diameter of sparse random graphs. *Combinatorics, Probability and Computing*, 19(5-6):835–926, 2010.
- [124] Anna Rudas, Bálint Tóth, and Benedek Valkó. Random trees and general branching processes. *Random Structures & Algorithms*, 31(2):186–202, 2007.
- [125] Roberto H Schonmann. Mean-field criticality for percolation on planar non-amenable graphs. *Communications in mathematical physics*, 225(3):453–463, 2002.
- [126] Oded Schramm. Scaling limits of loop-erased random walks and uniform spanning trees. *Israel Journal of Mathematics*, 118(1):221–288, 2000.
- [127] Regev Schweiger, Michal Linial, and Nathan Linial. Generative probabilistic models for protein–protein interaction networks—the biclique perspective. *Bioinformatics*, 27(13):i142–i148, 2011.
- [128] Delphin Sénizergues. Growing random graphs with a preferential attachment structure. *arXiv preprint arXiv:2002.09999*, 2020.
- [129] Stanislav Smirnov. Critical percolation in the plane: conformal invariance, Cardy’s formula, scaling limits. *Comptes Rendus de l’Académie des Sciences-Series I-Mathematics*, 333(3):239–244, 2001.
- [130] Stanislav Smirnov and Wendelin Werner. Critical exponents for two-dimensional percolation. *arXiv preprint math/0109120*, 2001.
- [131] Robin Stephenson. Local convergence of large critical multi-type Galton–Watson trees and applications to random maps. *Journal of Theoretical Probability*, 31(1):159–205, 2018.
- [132] Gerard ’t Hooft. A planar diagram theory for strong interactions. In *The Large N Expansion In Quantum Field Theory And Statistical Physics: From Spin Systems to 2-Dimensional Gravity*, pages 80–92. World Scientific, 1993.
- [133] William Thomas Tutte. A census of hamiltonian polygons. *Canadian Journal of Mathematics*, 14:402–417, 1962.
- [134] William Thomas Tutte. A census of planar triangulations. *Canadian Journal of Mathematics*, 14:21–38, 1962.
- [135] William Thomas Tutte. A census of slicings. *Canadian Journal of Mathematics*, 14:708–722, 1962.
- [136] William Thomas Tutte. A census of planar maps. *Canadian Journal of Mathematics*, 15:249–271, 1963.
- [137] William Thomas Tutte. On the enumeration of four-colored maps. *SIAM Journal on Applied Mathematics*, 17(2):454–460, 1969.
- [138] Remco van der Hofstad, Gerard Hooghiemstra, and Piet Van Mieghem. Distances in random graphs with finite variance degrees. *Random Structures & Algorithms*, 27(1):76–123, 2005.
- [139] Remco van der Hofstad, Gerard Hooghiemstra, and Dmitri Znamenski. Distances in random graphs with finite mean and infinite variance degrees. *Electronic Journal of Probability*, 12:703–766, 2007.

- [140] Vladimir A Vatutin and Vitali Wachtel. Local probabilities for random walks conditioned to stay positive. *Probability Theory and Related Fields*, 143(1-2):177–217, 2009.
- [141] Balint Virág. Anchored expansion and random walk. *Geometric & Functional Analysis GAFA*, 10(6):1588–1605, 2000.
- [142] Mingyang Wang, Guang Yu, and Daren Yu. Measuring the preferential attachment mechanism in citation networks. *Physica A: Statistical Mechanics and its Applications*, 387(18):4692 – 4698, 2008.
- [143] Yoshiyuki Watabiki. Construction of non-critical string field theory by transfer matrix formalism in dynamical triangulation. *Nuclear Physics B*, 441(1-2):119–163, 1995.
- [144] John C Wierman. Bond percolation on honeycomb and triangular lattices. *Advances in Applied Probability*, 13(2):298–313, 1981.

