



HAL
open science

Three applications of fragmentation and Poisson calculus to combinatorics

Adrien Joseph

► **To cite this version:**

Adrien Joseph. Three applications of fragmentation and Poisson calculus to combinatorics. Mathematics [math]. Université Pierre et Marie Curie - Paris VI, 2011. English. NNT: . tel-00605743

HAL Id: tel-00605743

<https://theses.hal.science/tel-00605743>

Submitted on 4 Jul 2011

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.

THÈSE DE DOCTORAT DE
L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité :
Mathématiques

présentée par
Adrien JOSEPH

pour obtenir le grade de
DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

**Trois applications de la fragmentation
et du calcul poissonien
à la combinatoire**

Directeur de thèse : M. Jean BERTOIN

Rapporteurs : M. Philippe CHASSAING
M. Svante JANSON

Soutenue le 30 juin 2011 devant le jury composé de

M. Jean BERTOIN	(Directeur de thèse)
M. Philippe CHASSAING	(Rapporteur)
Mme Brigitte CHAUVIN	(Examinatrice)
M. Thomas DUQUESNE	(Examineur)
M. Jean-François LE GALL	(Président du jury)
M. Grégory MIERMONT	(Examineur)
M. Alain ROUAULT	(Examineur)

REMERCIEMENTS

Je tiens à remercier très chaleureusement Jean Bertoin ; mes trois années de thèse sous sa direction constituent une expérience mathématique extrêmement enrichissante. Son écoute avenante, sa disponibilité, ses innombrables conseils m'ont permis de travailler dans un cadre idéal. Il m'a initié à la recherche mathématique en parvenant à m'en faire découvrir sa pluralité. Pour tout ce qu'il a fait pour moi, je lui suis très reconnaissant.

Je remercie Philippe Chassaing qui m'a fait l'honneur d'accepter les fonctions de rapporteur et de membre du jury. Un grand merci aussi à Svante Janson pour avoir rapporté ma thèse. Mes remerciements vont également à Jean-François Le Gall. Ses talents de pédagogue ainsi que son goût communicatif pour les mathématiques m'ont amené à choisir la voie des probabilités. C'est par ailleurs grâce à lui que j'ai eu l'opportunité de rencontrer Jean Bertoin. Je suis aussi honoré et heureux de la présence de Grégory Miermont et de Thomas Duquesne dans mon jury. Ils m'ont prouvé par l'exemple la richesse des probabilités. Enfin, j'aimerais remercier vivement Brigitte Chauvin et Alain Rouault qui m'ont accordé le privilège de faire partie de mon jury.

Je remercie Nicolas Curien avec qui j'ai partagé de formidables moments tout au long de ma thèse. Son optimisme et sa vivacité m'ont tout particulièrement impressionné lors de notre collaboration.

Chacune de mes années de thèse a été marquée par un lieu de travail différent, ce qui m'a donné l'occasion de faire de nombreuses rencontres. Je remercie tous mes compagnons de l'ancien bureau 3D1 de Chevaleret et du bureau 1626-201 de Jussieu pour leur bonne humeur. J'aimerais saluer plus généralement tous mes collègues et amis du LPMA, parmi lesquels Isabelle et Maria qui ont œuvré pour le bon déroulement de ma thèse. Merci également aux membres du CMLA qui m'ont fait vivre intensément cette dernière année de thèse à l'ENS de Cachan. Je souhaite encore remercier les membres du DMA, en particulier Zaina, qui m'ont permis de découvrir de lointaines contrées tant géographiques que mathématiques. Mes pensées vont enfin vers tous mes amis moins voire pas du tout matheux.

Je voudrais remercier mes parents pour leur soutien constant et pour la confiance qu'ils m'accordent. Je souhaite dire un immense merci à Anne-Lise, docteur Joseph en devenir. Je remercie toute ma famille, avec une pensée toute particulière pour mon grand-père Pierre. J'aimerais enfin remercier infiniment Marie, que j'ai connue au tout début de ma thèse. Elle a su m'aider et m'encourager à chaque instant de ces trois années. Je remercie aussi sa famille et termine en citant Gabin, sans cesse incroyable.

RÉSUMÉ

Cette thèse est consacrée à l'étude de trois modèles combinatoires intervenant dans la théorie des probabilités.

Nous nous intéresserons dans un premier temps aux arbres de fragmentation, et plus particulièrement à leur hauteur. À mesure de dislocation fixée, nous observerons l'apparition de deux régimes bien différents selon la capacité des sommets : au-delà d'une capacité critique, les hauteurs des arbres ont même asymptotique tandis que, en deçà de ce paramètre critique, les arbres sont de plus en plus hauts à mesure que le seuil de rupture diminue. Nous donnerons une interprétation de l'existence de ces deux phases grâce à la théorie des marches aléatoires branchantes.

Nous présenterons ensuite des résultats obtenus avec Nicolas Curien sur le quadtree. Nous expliciterons les comportements asymptotiques des coûts moyens des requêtes partielles. La théorie des fragmentations jouera encore un rôle clé.

La dernière partie sera consacrée à l'étude de grands graphes aléatoires, critiques pour le modèle de configuration. Sous certaines hypothèses, nous verrons que, correctement remises à l'échelle, les suites des tailles des composantes connexes de ces graphes convergent en un certain sens vers une suite aléatoire non triviale. Nous montrerons que cette dernière peut être vue comme la suite des longueurs d'excursion d'un processus à accroissements indépendants avec dérive polynomiale au-dessus de ses minima passés. Nous remarquerons que la situation est bien différente selon que la loi des degrés d'un sommet a un moment d'ordre 3 fini ou est une loi de puissance d'exposant compris entre 3 et 4.

ABSTRACT

This PhD thesis is devoted to the study of three combinatorial models occurring in probability theory.

We shall first be interested in fragmentation trees, and more particularly in their heights. Fixing the dislocation measure, we shall observe the appearance of two different regimes according to the vertex capacity: above a critical capacity the heights of the trees have the same asymptotics whereas below that critical parameter the trees get significantly higher as long as the rupture threshold decreases. The existence of those two phases will be explained via the theory of branching random walks.

We shall then present results obtained with Nicolas Curien on the quadtree. We shall explicit the asymptotic behaviors of the mean costs of the partial match queries. Here again, fragmentation theory will play a key role.

The last part will be devoted to large random graphs, critical for the configuration model. We shall see that, under some assumptions, the sequences of the component sizes of those graphs, once properly rescaled, converge in a certain sense to a non-trivial random sequence. We shall show that the latter may be viewed as the sequence of the excursion lengths of a process with independent increments and with polynomial drift above past minima. As will be seen, the situation is very different depending on whether the degree distribution has finite third moment or is a power law distribution with exponent between 3 and 4.

TABLE DES MATIÈRES

1	Introduction	1
1.1	Chaînes de fragmentation	2
1.2	Résultats généraux sur les fragmentations	5
1.3	Les arbres de fragmentation	10
1.4	Le quadtree	17
1.5	Graphes aléatoires	22
1.6	Les tailles des composantes d'un graphe aléatoire critique	26
2	A phase transition for the heights of a fragmentation tree	35
2.1	Introduction	36
2.2	Formulation of the main results	38
2.3	Preliminaries	43
2.4	Study of the heights	46
2.5	Study of the saturation levels	54
2.6	An explanation for the phase transition	59
3	Partial match queries in two-dimensional quadtrees	63
3.1	Introduction	64
3.2	Notation and first properties	65
3.3	Particular cases and fragmentation theory	68
3.4	The convergence at fixed $x \in (0, 1)$	74
3.5	Identifying the limit	78
3.6	Extensions and comments	79
4	The component sizes of a critical random graph	81
4.1	Introduction	82
4.2	Formulation of the results	84
4.3	The depth-first search	86
4.4	Convergence of the walk \bar{s}_n	89
4.5	Study of the cycle half-edges	94
4.6	Study of the random simple graph \mathcal{SG}_n	96
4.7	The power law distribution setting	98
	Bibliographie	107

CHAPITRE 1

INTRODUCTION

Cette thèse s'organise autour de quatre chapitres. Hormis le premier, introductif, ils proviennent chacun d'un article :

- Un article intitulé *A phase transition for the heights of a fragmentation tree*, à paraître dans *Random Structures and Algorithms*.
- Un article écrit en collaboration avec Nicolas Curien intitulé *Partial match queries in two-dimensional quadtrees: a probabilistic approach*, paru dans *Advances in Applied Probability* **43** (2011), 178-194.
- Un article intitulé *The component sizes of a critical random graph with given degree sequence*.

Cette introduction est quant à elle structurée de la manière suivante. La section 1.1 introduit la notion de chaîne de fragmentation. La section 1.2 dresse le tableau des principaux résultats sur ces processus. La section 1.3 présente les arbres de fragmentation. À la section 1.4, nous verrons une dernière application des fragmentations lors de l'étude du quadtree. Nous traiterons d'un autre sujet à la section 1.5. Elle sera en effet consacrée aux graphes aléatoires. Nous en présenterons plusieurs modèles, dont celui que nous décrirons plus largement à la section 1.6.

Sommaire

1.1	Chaînes de fragmentation	2
1.1.1	Marches aléatoires branchantes sur \mathbb{R}_+	2
1.1.2	Chaînes de fragmentation homogènes	4
1.1.3	Chaînes de fragmentation autosimilaires	4
1.2	Résultats généraux sur les fragmentations	5
1.2.1	Exposant malthusien	5

1.2.2	Extinction	6
1.2.3	Quelques martingales	6
1.2.4	Fragmentations homogènes	7
1.2.5	Fragmentations autosimilaires d'indice $\alpha > 0$	10
1.3	Les arbres de fragmentation	10
1.3.1	Le modèle	10
1.3.2	Transition de phase pour les hauteurs	11
1.3.3	Étude des niveaux de saturation	14
1.3.4	Exemple : les arbres fendus aléatoires	14
1.3.5	Présentation des techniques	17
1.4	Le quadtree	17
1.4.1	Le modèle et les principaux résultats	17
1.4.2	Lien avec les fragmentations	18
1.4.3	Techniques de preuves	21
1.4.4	Extensions des résultats	22
1.5	Graphes aléatoires	22
1.5.1	Modèle du graphe aléatoire d'Erdős et Rényi	22
1.5.2	Graphes aléatoires avec une suite de degrés fixée	23
1.5.3	Lien entre les deux modèles	24
1.5.4	Le modèle de configuration	24
1.5.5	Le modèle étudié	25
1.6	Les tailles des composantes d'un graphe aléatoire critique .	26
1.6.1	Présentation des résultats	26
1.6.2	Techniques de preuves	28

1.1 Chaînes de fragmentation

1.1.1 Marches aléatoires branchantes sur \mathbb{R}_+

Commençons par décrire un modèle de population sur \mathbb{R}_+ très simple. Initialement, un unique individu est présent. Il est situé à l'origine. Il vit pendant un temps exponentiel de paramètre 1. À sa mort, il est remplacé par des individus, appelés ses enfants, qui naissent tous aléatoirement à sa droite. Chacune de ces particules de première génération vit indépendamment de toutes les autres particules pendant un temps exponentiel de paramètre 1. À sa mort, elle est remplacée par des particules, ses enfants, qui naissent aléatoirement à sa droite. Les distances des enfants à leur particule mère suivent la même loi que celles des particules de la première génération à la particule initiale. Les nouveaux individus, qui appartiennent à la deuxième génération, évoluent de manière similaire, indépendamment de leur environnement. On procède de même pour tout individu créé. La Figure 1.1 illustre cette dynamique.

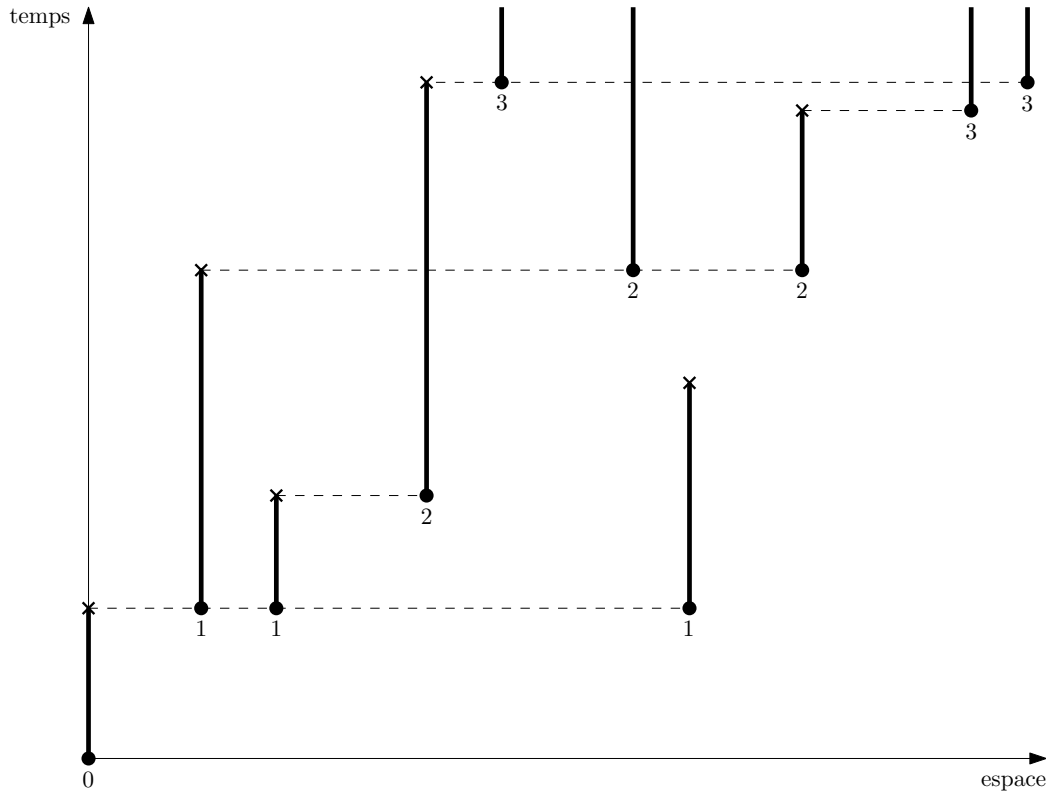


FIGURE 1.1 – Une réalisation de la marche aléatoire branchante. On a indiqué la génération de chaque individu.

Nous faisons l'hypothèse que les distances des enfants à leur particule mère peuvent être ordonnées en une suite croissante, cette dernière étant finie ou divergeant vers l'infini. Notons \mathcal{U}^\uparrow l'ensemble

$$\mathcal{U}^\uparrow = \left\{ \mathbf{u} = (u_1, u_2, \dots) : 0 \leq u_1 \leq u_2 \leq \dots \leq \infty \text{ et } \lim_{i \rightarrow \infty} u_i = \infty \right\}$$

muni de la distance d définie par $d(\mathbf{a}, \mathbf{b}) = \max\{|e^{-a_i} - e^{-b_i}| : i \geq 1\}$ et de la tribu borélienne associée. D'après la définition du modèle, la loi de la suite des distances d'une particule à ses enfants est la même quelle que soit la particule mère ; on la note ν_0 . *Dans tout ce qui suit, nous supposons que :*

$$\forall A > 0, \quad \int_{\mathcal{U}^\uparrow} \#\{i \geq 1 : u_i < A\} \nu_0(d\mathbf{u}) < \infty. \quad (1.1)$$

Il est alors possible d'ordonner la famille des positions des individus présents à l'instant t en une suite $(v_1(t), v_2(t), \dots)$ appartenant à \mathcal{U}^\uparrow . Notons $Z^{(t)}$ la mesure ponctuelle aléatoire sur \mathbb{R}_+ définie par

$$Z^{(t)}(dy) = \sum_{i \geq 1} \delta_{v_i(t)}(dy),$$

où δ_z est la mesure de Dirac au point z . On peut montrer que pour tous $t, s \geq 0$, $Z^{(t+s)}$ est obtenu en remplaçant chaque atome z de $Z^{(t)}$ par une famille $\{z + y : y \in \mathcal{Y}_z\}$, les

\mathcal{Y}_z étant des copies indépendantes de la famille des atomes de $Z^{(s)}$; $(Z^{(t)}, t \geq 0)$ est une marche aléatoire branchante.

1.1.2 Chaînes de fragmentation homogènes

Nous serons davantage intéressés par des processus branchants modélisant l'évolution de particules qui se brisent au cours du temps. Décrivons ce modèle à partir de la marche aléatoire branchante que nous venons de définir.

Posons ν la mesure image de ν_0 par l'application qui, à une suite (u_1, u_2, \dots) de \mathcal{U}^\uparrow , associe la suite $(e^{-u_1}, e^{-u_2}, \dots)$. La mesure de probabilité ν est portée par l'ensemble \mathcal{S}^\downarrow des suites décroissantes à termes appartenant à $[0, 1]$ qui tendent vers 0 :

$$\mathcal{S}^\downarrow = \left\{ \mathbf{s} = (s_1, s_2, \dots) : 1 \geq s_1 \geq s_2 \geq \dots \geq 0 \text{ et } \lim_{i \rightarrow \infty} s_i = 0 \right\}.$$

Nous n'imposons pas ici que $\nu(ds)$ -p.s., $\sum_{i \geq 1} s_i \leq 1$ ¹. L'hypothèse (1.1) se réécrit de la manière suivante :

$$\forall \varepsilon > 0, \quad \int_{\mathcal{S}^\downarrow} \#\{i \geq 1 : s_i > \varepsilon\} \nu(ds) < \infty. \quad (1.2)$$

Afin d'éliminer les cas triviaux, nous supposons par ailleurs qu'il existe $i \geq 1$ tel que $\nu(0 < s_i < 1) > 0$.

Le modèle qui nous intéresse se déduit simplement du modèle précédent en attribuant une masse aux particules plutôt qu'une position. Plus précisément, un individu qui vit pendant l'intervalle de temps $[d, g[$ à la position $v < \infty$ est vu dans ce nouveau modèle comme une particule de masse $e^{-v} > 0$ vivant pendant l'intervalle $[d, g[$. À sa mort, elle se brise en une suite de particules de masse $e^{-u_i} e^{-v}$, $i \geq 1$, où (u_1, u_2, \dots) est la suite des distances de l'individu à ses enfants dans l'ancien modèle de population.

Ce nouveau système de particules peut ainsi être vu comme une fragmentation. Nous la qualifierons d'*homogène*. Nous allons présenter dans le paragraphe suivant des chaînes de fragmentation plus générales.

1.1.3 Chaînes de fragmentation autosimilaires

À la différence des fragmentations homogènes pour lesquelles les particules vivent pendant un temps exponentiel dont le paramètre est toujours égal à 1, les fragmentations autosimilaires modélisent un système de particules pour lequel la durée de vie d'un objet de masse $m > 0$ suit une loi exponentielle dont le paramètre est fonction de m . Cette fonction est supposée de la forme $m \mapsto m^\alpha$, où α est un réel. En particulier, lorsque α est nul, on retrouve la définition d'une fragmentation homogène.

Présentons la dynamique d'une fragmentation autosimilaire. Ici, α est un réel et ν est une mesure de probabilité sur \mathcal{S}^\downarrow satisfaisant (1.2). Initialement, une unique particule est créée. Sa masse est égale à 1. On suit alors les règles suivantes. À chaque particule créée on attache une horloge exponentielle de paramètre m^α , $m > 0$ étant la masse de la particule. Tant que l'horloge ne sonne pas, la particule reste intacte, mais dès qu'elle

¹Cette relation sera toutefois satisfaite pour les modèles que nous étudierons (cf. sections 1.3 et 1.4).

retentit, la particule est remplacée par une suite d'objets de masse ms_i , $i \geq 1$, où la suite $\mathbf{s} = (s_i)_{i \geq 1} \in \mathcal{S}^\downarrow$ est tirée selon la loi ν . Chacun de ces fragments créés subit alors le même sort, son évolution étant supposée indépendante des autres particules.

On peut montrer qu'il est possible d'ordonner la famille des masses des particules présentes à l'instant t en une suite $F(t) = (F_1(t), F_2(t), \dots)$ appartenant à \mathcal{S}^\downarrow . Le processus F est appelé une *chaîne de fragmentation autosimilaire*. Il est caractérisé par le couple (α, ν) . Le réel α est appelé l'*indice d'autosimilarité* de F tandis que la mesure de probabilité ν est sa *mesure de dislocation*.

Quand $\alpha = 0$, on retrouve la définition d'une fragmentation homogène. Lorsque $\nu(\mathrm{d}\mathbf{s})$ -p.s., $\sum_{i \geq 1} s_i = 1$, on dit que la mesure de dislocation ν est *conservatrice*. Lorsque $\nu(\mathrm{d}\mathbf{s})$ -p.s., $\sum_{i \geq 1} s_i < 1$, on dit que ν est *dissipatrice*. Il y a alors perte de masse. Observons que l'hypothèse (1.2) est satisfaite dès que ν est conservatrice ou dissipatrice.

Précisons que lorsque ν est conservatrice, il se peut que la masse totale $\sum_{i \geq 1} F_i(t)$ ne soit pas constamment égale à 1. C'est par exemple le cas si l'indice d'autosimilarité α est strictement négatif et si $\int_{\mathcal{S}^\downarrow} \#\{i \geq 1 : s_i = 1\} \nu(\mathrm{d}\mathbf{s}) < 1$ (on peut même montrer qu'alors, p.s., $F(t) = (0, 0, \dots)$ pour tout t suffisamment grand, cf. [Ber06, Proposition 1.8]). Cependant, si α est positif, la masse totale est conservée. *Nous faisons dorénavant l'hypothèse que*

$$\alpha \geq 0.$$

Cela revient à supposer que plus un objet est gros, plus il est susceptible de se briser rapidement.

Il est possible de définir des processus de fragmentation pour lesquels les particules ne restent pas intactes pendant une durée strictement positive. Il peut aussi être intéressant d'étudier des systèmes dans lesquels les particules, au lieu de se fragmenter, s'agrègent entre elles. Nous recommandons au lecteur souhaitant découvrir ces processus l'ouvrage [Ber06] de Bertoin.

1.2 Résultats généraux sur les fragmentations

1.2.1 Exposant malthusien

Comme $\alpha \geq 0$, nous savons que, dans le cas particulier où ν est conservatrice, la masse totale du système est conservée. On peut se demander si, sans supposer ν conservatrice, une quantité simple est conservée. Pour cela, définissons la fonction strictement croissante et concave ψ définie par : pour tout réel θ ,

$$\psi(\theta) = 1 - \int_{\mathcal{S}^\downarrow} \sum_{i \geq 1} s_i^\theta \nu(\mathrm{d}\mathbf{s}) \in [-\infty, 1[,$$

où seuls les indices i pour lesquels $s_i > 0$ interviennent dans la somme². Posons

$$\underline{\theta} = \inf\{\theta \in \mathbb{R} : \psi(\theta) > -\infty\}.$$

²Cette convention pour les sommes sera appliquée dans tout ce qui suit.

La conservation d'une quantité simple nous est alors fournie par l'*hypothèse malthusienne* suivante : il existe $\beta^* > 0$, appelé *exposant malthusien*, tel que

$$\psi(\beta^*) = 0 \quad \text{et} \quad \int_{\mathcal{S}^\downarrow} \left(\sum_{i \geq 1} s_i^{\beta^*} \right)^\gamma \nu(ds) < \infty \quad \text{pour un certain } \gamma > 1.$$

Nous supposons dorénavant que l'*hypothèse malthusienne* est vérifiée. Il est clair que la condition (1.2) est alors satisfaite. Remarquons que $\beta^* = 1$ lorsque ν est conservatrice tandis que $\beta^* < 1$ lorsque ν est dissipatrice.

Nous allons voir que de nombreuses martingales peuvent désormais être construites. Mais auparavant, nous définissons quelques notions.

1.2.2 Extinction

Rappelons la définition des générations de particules (nous avons déjà vu cette notion pour les marches aléatoires branchantes).

Chaque fragment peut être vu comme descendant d'une lignée, dont l'ancêtre le plus vieux est la particule initiale de masse 1. Les particules peuvent ainsi être rassemblées en générations, chaque génération regroupant les fragments créés après un même nombre de brisures depuis l'instant initial. On procède de la manière itérative suivante. Il s'agit d'étiqueter les particules (et leur masse) par des k -uplets, l'entier naturel k correspondant à génération de la particule. La génération initiale $k = 0$ correspond à la particule initiale de masse 1 que l'on étiquette par \emptyset . On pose $m_\emptyset = 1$. À l'issue d'une brisure d'une particule d'étiquette $i \in \mathbb{N}^{*k}$ (appartenant donc à la génération k), les particules créées sont étiquetées par ij , $j \in \mathbb{N}^*$, et si $\mathbf{s}(i, \omega) \in \mathcal{S}^\downarrow$ désigne la réalisation de ν correspondant à la manière de briser la particule d'étiquette i , les masses m_{ij} , $j \in \mathbb{N}^*$, vérifient : $m_{ij} = m_i s_j(i, \omega)$.

Il est aisé de voir (voir par exemple [Ber06, Proposition 1.3]) que le processus qui à tout entier naturel k associe le nombre de particules présentes à la génération k est un processus de Galton-Watson dont la loi de reproduction est la loi de $\sup\{i \geq 1 : s_i > 0\}$ sous $\nu(ds)$. Puisque $\psi(\beta^*) = 0$ (et $\beta^* \geq 0$), ce processus de Galton-Watson est surcritique. On appelle *extinction* pour la fragmentation l'événement d'extinction du processus de Galton-Watson. La non-extinction a lieu avec probabilité non nulle, cette dernière étant égale à 1 si et seulement si $\nu(ds)$ -p.s., $s_1 > 0$ (ce qui est en particulier le cas lorsque ν est conservatrice).

1.2.3 Quelques martingales

Nous pouvons maintenant définir la première martingale importante, appelée *martingale intrinsèque*, définie en temps discret et notée $(\mathcal{M}_k, k \in \mathbb{N})$:

$$\mathcal{M}_k = \sum_{i \in \mathbb{N}^{*k}} m_i^{\beta^*}.$$

Bertoin prouve dans [Ber06, Théorème 1.1] qu'elle est bornée dans \mathbb{L}^γ pour un certain $\gamma > 1$, donc uniformément intégrable, et que sa valeur terminale \mathcal{M}_∞ est strictement

positive p.s. conditionnellement à l'événement de non-extinction. La propriété de branchement assure d'autre part que \mathcal{M}_∞ satisfait l'équation en loi

$$\mathcal{M}_\infty \stackrel{(\text{loi})}{=} \sum_{i \geq 1} s_i^{\beta^*} \mathcal{M}_\infty^{(i)},$$

où $\mathbf{s} = (s_i)_{i \geq 1}$ est tirée selon la loi ν et les $\mathcal{M}_\infty^{(i)}$ sont des copies indépendantes de \mathcal{M}_∞ et indépendantes de \mathbf{s} . Sous certaines hypothèses relativement faibles (voir [Rös92, Liu01]), cette équation permet de caractériser la loi de \mathcal{M}_∞ .

On peut définir une deuxième martingale, cette fois définie en temps continu. Elle prend en compte les masses des fragments présents à un instant donné $t \geq 0$. Rappelons que $F(t) = (F_i(t))_{i \geq 1}$ est la suite des masses des objets présents à l'instant t . Posons

$$\mathfrak{M}_t = \sum_{i \geq 1} F_i(t)^{\beta^*}.$$

On peut montrer (voir [Ber06, Proposition 1.5]) que $(\mathfrak{M}_t, t \geq 0)$ est une martingale relativement à la filtration naturelle de la fragmentation F . Plus précisément, on a

$$\mathfrak{M}_t = \mathbb{E}[\mathcal{M}_\infty | F(t)].$$

En particulier, \mathfrak{M}_t converge dans \mathbb{L}^γ vers \mathcal{M}_∞ pour un certain $\gamma > 1$.

1.2.4 Fragmentations homogènes

Résultats en temps continu

Voyons maintenant quelques résultats limites pour les fragmentations homogènes. Nous recommandons au lecteur intéressé de consulter l'article [Ber01] de Bertoin.

Nous supposons que $\beta^* > \underline{\theta}$. La plupart des masses des fragments présents à l'instant t décroissent alors comme $e^{-t\psi'(\beta^*)}$ (nous verrons que cette décroissance est bien différente dans le cas où $\alpha > 0$); d'après [Ber06, Théorème 1.2], pour toute fonction continue bornée $f : \mathbb{R}_- \rightarrow \mathbb{R}$, on a la convergence suivante dans \mathbb{L}^γ pour un certain $\gamma > 1$:

$$\sum_{i \geq 1} F_i(t)^{\beta^*} f(t^{-1} \ln F_i(t)) \xrightarrow[t \rightarrow \infty]{} \mathcal{M}_\infty f(-\psi'(\beta^*)). \quad (1.3)$$

Introduisons maintenant une dernière famille de martingales. Commençons par mentionner quelques résultats élémentaires. La fonction $\theta \mapsto \theta\psi'(\theta) - \psi(\theta)$ définie sur $]\underline{\theta}, \infty[$ est strictement croissante sur $]\underline{\theta}, \infty[\cap \mathbb{R}_-$ et strictement décroissante sur $]\underline{\theta}, \infty[\cap \mathbb{R}_+$. On peut montrer (voir [Ber06, Lemme 1.6]) qu'il existe un unique réel $\theta^* > \beta^*$ tel que

$$\theta^* \psi'(\theta^*) = \psi(\theta^*).$$

Ainsi l'ensemble des réels $\theta > \underline{\theta}$ tels que $\theta\psi'(\theta) > \psi(\theta)$ s'écrit $]\theta_*, \theta^*[$. Observons que le réel θ_* peut parfois être égal à $\underline{\theta}$ (par exemple lorsque $\underline{\theta} \geq 0$). Il peut également être strictement négatif. On a toutefois l'inégalité $\theta_* < \beta^*$.

Pour tout $\theta > \underline{\theta}$, posons $W(\theta, \cdot)$ le processus défini par

$$W(\theta, t) = e^{t\psi(\theta)} \sum_{i \geq 1} F_i(t)^\theta.$$

Notons que $(W(\beta^*, t), t \geq 0)$ correspond à la martingale $(\mathfrak{M}_t, t \geq 0)$. Bertoin prouve dans [Ber06, Corollaire 1.3] que plus généralement, pour tout $\theta > \underline{\theta}$, $W(\theta, \cdot)$ est une martingale relativement à la filtration naturelle de la fragmentation F . D'autre part, on peut aisément voir en adaptant la preuve de [Ber06, Théorème 1.4] que s'il existe $a, b > 0$ tels que

$$\sum_{i \geq 1} s_i^a \leq b \quad \nu(ds)\text{-p.s.}, \quad (1.4)$$

et si $\theta \in]\theta_*, \theta^*[$, alors la martingale $W(\theta, \cdot)$ est bornée dans \mathbb{L}^γ pour un certain $\gamma > 1$, donc uniformément intégrable, et sa valeur terminale $W(\theta, \infty)$ est strictement positive p.s. conditionnellement à l'événement de non-extinction. Nous utiliserons ce résultat pour des mesures de dislocation conservatrices, pour lesquelles (1.4) s'applique trivialement et la non-extinction a lieu p.s.

Quelques applications

Donnons maintenant des applications. On suppose ici encore que la condition (1.4) est vérifiée. Alors p.s. conditionnellement à l'événement de non-extinction,

$$t^{-1} \ln F_1(t) \xrightarrow[t \rightarrow \infty]{} -\psi'(\theta^*) \quad (1.5)$$

(voir [Ber06, Corollaire 1.4]). Ce résultat est à comparer avec (1.3), qui indique que le taux de décroissance exponentielle de la masse d'un fragment typique, égal à $\psi'(\beta^*)$, est strictement supérieur à $\psi'(\theta^*)$. On peut déduire de ce résultat que pour tout $\theta \geq \theta^*$, la martingale $W(\theta, \cdot)$ converge vers 0 p.s.

En s'inspirant de la preuve de (1.5), on peut démontrer un résultat analogue pour la masse du plus petit fragment. Pour que ceci ait un sens, il faut bien entendu au moins supposer qu'il n'y a qu'un nombre fini de fragments. Plus précisément, on se place sous les hypothèses suivantes : $-\infty < \theta_* < 0$ et $\theta_*\psi'(\theta_*) = \psi(\theta_*)$. En particulier $\psi(0) > -\infty$: l'espérance du nombre de particules apparaissant à l'issue d'une brisure est finie. On a alors, en notant $F_{\min}(t)$ la masse du plus petit fragment à l'instant t , p.s. conditionnellement à l'événement de non-extinction,

$$t^{-1} \ln F_{\min}(t) \xrightarrow[t \rightarrow \infty]{} -\psi'(\theta_*). \quad (1.6)$$

Comme auparavant, on en déduit que pour tout $\theta < \theta_*$, la martingale $W(\theta, \cdot)$ converge vers 0 p.s.

Terminons par mentionner un dernier résultat. On suppose que la mesure de dislocation ν n'est pas *géométrique*, c'est-à-dire qu'il n'existe aucun réel $r \in]0, 1[$ tel que $\nu(ds)$ -p.s., $\{s_i, i \geq 1\} \subset \{r^n, n \in \mathbb{N}\}$. On souhaite connaître le nombre de fragments présents à l'instant t de masse de l'ordre de $e^{-t\psi'(\theta)}$, où θ est, d'après (1.5) et (1.6), un réel compris strictement entre θ_* et θ^* . Il suffit pour cela d'appliquer les résultats de

Biggins sur les marches aléatoires branchantes (voir le corollaire 4 et la discussion de la page 150 de [Big92]), dont la précision est telle qu'on obtient exactement le nombre de fragments de masse $\Theta(e^{-t\psi'(\theta)})$: pour tout $\theta \in]\theta_*, \theta^*[$ et pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ à support compact directement Riemann intégrable, p.s. conditionnellement à l'événement de non-extinction,

$$\sqrt{t}e^{-t(\theta\psi'(\theta)-\psi(\theta))} \sum_{i \geq 1} f(\ln F_i(t) + t\psi'(\theta)) \xrightarrow{t \rightarrow \infty} \frac{W(\theta, \infty)}{\sqrt{2\pi|\psi''(\theta)|}} \int_{-\infty}^{\infty} f(-x)e^{\theta x} dx. \quad (1.7)$$

Cas particulier du temps discret

Nous présentons ici les résultats analogues pour les chaînes de fragmentation en temps discret. Le modèle est le suivant. Initialement, le système est constitué d'une unique particule de masse 1. Chaque particule reste intacte pendant une durée égale à 1, puis se fragmente selon la mesure de dislocation ν indépendamment de son environnement. Posons, pour tout $\theta \in \mathbb{R}$,

$$L(\theta) = \int_{\mathcal{S}^\downarrow} \sum_{i \geq 1} s_i^\theta \nu(ds) \in]0, \infty].$$

De manière informelle, à condition de transformer ψ en $-\ln L$, tous les résultats décrits auparavant dans le cas du temps continu peuvent se traduire dans ce cadre. La raison en est simple : pour tout $\theta > \underline{\theta}$, le processus

$$L(\theta)^{-k} \sum_{i \geq 1} F_i(k)^\theta, \quad k \in \mathbb{N}, \quad (1.8)$$

est une martingale relativement à la filtration naturelle de la fragmentation F . Cette famille de martingales est à comparer avec la famille des martingales $W(\theta, \cdot)$, $\theta > \underline{\theta}$.

Notons $]\theta_*, \theta^*[$ l'ensemble des réels $\theta > \underline{\theta}$ tels que $\ln L(\theta) > \theta L'(\theta)/L(\theta)$. Si l'hypothèse (1.4) est satisfaite, le résultat (1.5) se traduit ici en : p.s. conditionnellement à l'événement de non-extinction,

$$\frac{\ln F_1(k)}{k} \xrightarrow{k \rightarrow \infty} \frac{L'(\theta^*)}{L(\theta^*)}. \quad (1.9)$$

Remarquons que, pour une même mesure de dislocation ν , les comportements asymptotiques des modèles en temps continu et en temps discret sont bien différents (par exemple le taux de décroissance exponentielle de la masse du plus gros fragment de la fragmentation en temps continu est inférieur au taux de décroissance exponentielle de la masse du plus gros fragment de la fragmentation en temps discret).

Si ν n'est pas géométrique, un corollaire de (1.7) s'écrit ici : pour tout $\theta \in]\theta_*, \theta^*[$ et pour tous $b > a > 0$, p.s. conditionnellement à l'événement de non-extinction,

$$\sqrt{k}e^{-k\varphi(\theta)} \# \left\{ i \geq 1 : ae^{kL'(\theta)/L(\theta)} \leq F_i(k) \leq be^{kL'(\theta)/L(\theta)} \right\} \xrightarrow{k \rightarrow \infty} K_{\theta,a,b} W(\theta), \quad (1.10)$$

où $\varphi(\theta) = \ln L(\theta) - \theta L'(\theta)/L(\theta) > 0$, $W(\theta)$ est la valeur terminale de la martingale définie en (1.8) (qui est strictement positive p.s. conditionnellement à l'événement de non-extinction) et $K_{\theta,a,b}$ est un réel strictement positif.

1.2.5 Fragmentations autosimilaires d'indice $\alpha > 0$

Étudions le comportement asymptotique d'une fragmentation autosimilaire dont l'indice d'autosimilarité α est strictement positif. Nous recommandons la lecture de l'article [Ber02] de Bertoin.

Supposons $\beta^* > \underline{\theta}$ et ν non géométrique. On dispose alors d'un résultat analogue à (1.3). Cependant, ici, la plupart des masses des fragments présents à l'instant t décroissent comme $t^{-1/\alpha}$: d'après [Ber06, Théorème 1.3], pour toute fonction continue bornée $f : \mathbb{R}_+ \rightarrow \mathbb{R}$, on a la convergence suivante dans \mathbb{L}^1 :

$$\sum_{i \geq 1} F_i(t)^{\beta^*} f(t^{1/\alpha} F_i(t)) \xrightarrow[t \rightarrow \infty]{} \mathcal{M}_\infty \int_0^\infty f(x) \rho(dx),$$

où ρ est une mesure de probabilité (déterministe). On invite le lecteur intéressé à consulter [BG04] et [Ber06] pour connaître une expression de ρ .

Nous devons savoir comment se comporte l'espérance du nombre de fragments présents à l'instant t , étant entendu qu'un fragment n'est présent que si sa masse est strictement positive. Bertoin et Gnedin ont montré dans [BG04] que, sous les mêmes hypothèses mentionnées ci-dessus, pour tout complexes θ dont la partie réelle est strictement supérieure à $\underline{\theta}$,

$$t^{(\theta - \beta^*)/\alpha} \mathbb{E} \left[\sum_{i \geq 1} F_i(t)^\theta \right] \xrightarrow[t \rightarrow \infty]{} \int_0^\infty x^{\theta - \beta^*} \rho(dx). \quad (1.11)$$

La valeur précise de cette intégrale est fournie dans [BG04]. En particulier, pourvu que $\underline{\theta}$ est strictement négatif, on pourra prendre $\theta = 0$ et répondre à notre question.

1.3 Les arbres de fragmentation

Nous présentons ici des résultats obtenus dans [Jos11a] (cf. chapitre 2).

1.3.1 Le modèle

Décrivons le modèle d'urnes que nous avons étudié en premier lieu, appelé modèle d'occupation des cascades multiplicatives. On se donne une fragmentation homogène définie en temps discret de mesure de dislocation ν conservatrice (cf. Figure 1.2).

On jette sur l'unique particule présente initialement un nombre dénombrable de points, que l'on étiquète par \mathbb{N}^* . À l'issue de la première brisure, qui intervient à l'instant 1, des particules de masse m_i , $i \geq 1$, sont créées. Les points se répartissent alors sur ces fragments, chacun choisissant la particule d'étiquette i de masse $m_i > 0$ avec probabilité m_i . Tous ces choix sont supposés indépendants. On répète ce procédé à l'issue de chaque cassure : si un objet créé à l'instant k de masse $m_i > 0$, $i \in \mathbb{N}^{*k}$, se brise en une suite de particules de masse m_{ij} , $j \geq 1$, chacun des points qu'il abritait choisit la particule indicée par ij avec probabilité m_{ij}/m_i . Ici encore, tous les choix sont supposés indépendants.

Définissons le premier instant $H_{n,j}$, appelé *hauteur* de paramètre (n, j) , où tous les fragments contiennent strictement moins de $j \geq 2$ points parmi les n points d'étiquette appartenant à $\{1, \dots, n\}$. Définissons aussi le premier instant $G_{n,j}$, appelé *niveau de*

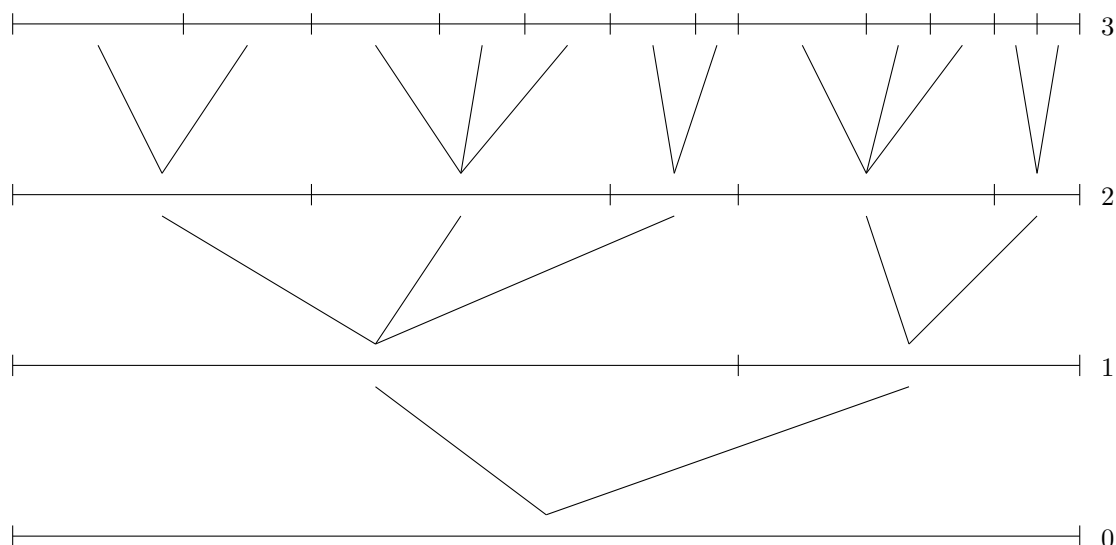


FIGURE 1.2 – Les particules correspondent ici à des segments de $[0, 1]$. On les a regroupées par génération.

saturation de paramètre (n, j) , où il existe un fragment contenant strictement moins de $j \geq 1$ points parmi les n points d'étiquette inférieure à n . On se demande comment se comportent $H_{n,j}$ et $G_{n,j}$ lorsque n tend vers l'infini (on étudiera successivement les cas où j est fixé et où j est fonction de n).

On a représenté dans la Figure 1.3 une réalisation du modèle d'occupation des cascades multiplicatives en reprenant l'exemple de la Figure 1.2.

1.3.2 Transition de phase pour les hauteurs

Dans [Ber08], Bertoin étudie le comportement asymptotique de la hauteur $H_{n,j}$ pour tout entier j fixé suffisamment grand. Il montre dans [Ber08, Proposition 2] que

$$\forall j \geq \theta^*, \quad \text{p.s.}, \quad \frac{H_{n,j}}{\ln n} \xrightarrow[n \rightarrow \infty]{} \frac{L(\theta^*)}{-L'(\theta^*)}.$$

Autrement dit, pour j suffisamment grand, toutes les hauteurs $H_{n,j}$ sont du même ordre de grandeur $\frac{L(\theta^*)}{-L'(\theta^*)} \ln n$. En utilisant les résultats présentés au paragraphe 1.2.4, nous exhiberons des constantes $C_2 > \dots > C_{\lceil \theta^* \rceil - 1} > -L(\theta^*)/L'(\theta^*)$ telles que

$$\forall j < \theta^*, \quad \text{p.s.}, \quad \frac{H_{n,j}}{\ln n} \xrightarrow[n \rightarrow \infty]{} C_j.$$

Pour j assez petit, la hauteur $H_{n,j}$ est sensiblement plus importante que $H_{n,j+1}$. On voit ainsi apparaître une transition de phase. Elle sera mieux comprise lorsque nous aborderons les preuves de ces résultats.

De manière informelle, nous verrons que, à l'instant $H_{n,j} - 1$, l'unique particule contenant encore au moins j points (parmi ceux d'étiquette inférieure à n) est le plus gros fragment lorsque $j \geq \theta^*$ tandis que, lorsque $2 \leq j < \theta^*$, il s'agit d'un fragment ayant une masse environ égale à $e^{H_{n,j}L'(j)/L(j)}$ (rappelons que, d'après (1.10), le nombre

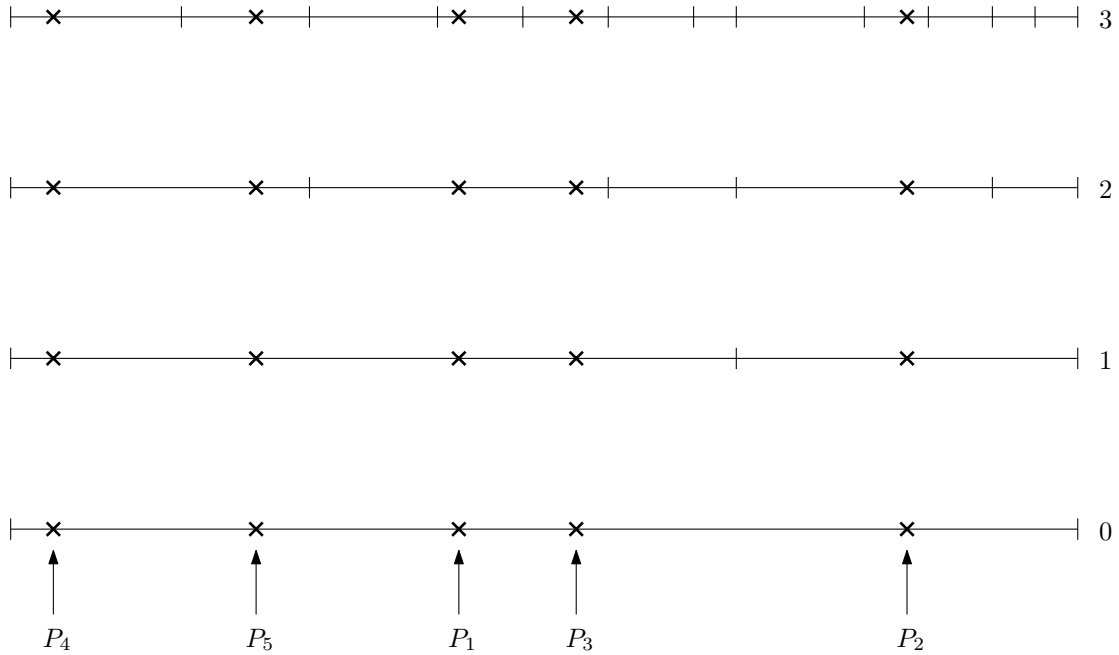


FIGURE 1.3 – On reprend la fragmentation de la Figure 1.2. On a jeté $n = 5$ points, représentés par des croix. Le premier instant où tous les points sont isolés des autres est $H_{5,2} = 3$. Le premier instant où il existe une particule ne contenant aucun point est $G_{5,1} = 2$. On a aussi par exemple $H_{5,3} = 2$ et $G_{5,2} = 1$.

de particules ayant une telle masse est environ $e^{H_{n,j}(\ln L(j) - jL'(j)/L(j))}$. On voit qu'il y a compétition entre la masse des particules et le nombre de particule ayant une masse donnée : pour $j \geq \theta^*$, c'est l'importance de la masse de l'unique plus gros fragment qui l'emporte, tandis que, pour $j < \theta^*$, c'est le nombre très élevé de petites particules qui fournit la principale contribution.

Déterminons de manière heuristique une minoration de C_j pour $j < \theta^*$. Nous traduisons pour cela le fait que, à l'instant $H_{n,j}$, toutes les particules de masse proche de $e^{H_{n,j}L'(j)/L(j)}$ ont toutes strictement moins de j points (parmi ceux d'étiquette appartenant à $\{1, \dots, n\}$). Observons tout d'abord que le nombre de points contenus dans une particule ayant une telle masse suit approximativement une loi binomiale de paramètres n et $e^{H_{n,j}L'(j)/L(j)}$. Si $ne^{H_{n,j}L'(j)/L(j)} \ll 1$, la probabilité qu'une telle particule contienne au moins j points est donc environ $n^j e^{jH_{n,j}L'(j)/L(j)}$. En admettant que les nombres de points que contiennent les particules sont presque indépendants, la probabilité que toutes les particules considérées aient strictement moins de j points à l'instant $H_{n,j}$ est environ

$$\left(1 - n^j e^{jH_{n,j}L'(j)/L(j)}\right)^{H_{n,j}^{-1/2} \exp\{H_{n,j}(\ln L(j) - jL'(j)/L(j))\}}$$

(le facteur $H_{n,j}^{-1/2}$ provient du facteur \sqrt{k} dans l'expression (1.10)). Cette quantité devant tendre vers 1 lorsque $n \rightarrow \infty$, la suite $(H_{n,j}^{-1/2} n^j e^{jH_{n,j}L'(j)/L(j)})$ converge vers 0. On en déduit que $C_j \geq -j / \ln L(j)^3$.

³Cette approche informelle permet même d'obtenir $H_{n,j} \geq \frac{j}{-\ln L(j)} \ln n + \frac{1}{2 \ln L(j)} \ln \ln n$.

On explicite une borne inférieure de $H_{n,j}/\ln n$ pour $j \geq \theta^*$ en procédant de la même manière, mais en considérant cette fois le plus gros fragment. Les bornes supérieures s'obtiendront en utilisant un argument de martingale et de fines estimées sur les lois de Poisson.

Nos principaux résultats peuvent se résumer ainsi :

Théorème 1.1. *Soit un entier $j \geq 2$. Si $j < \theta^*$, alors, presque sûrement,*

$$H_{n,j} = \frac{j}{-\ln L(j)} \ln n + O(\ln \ln n),$$

tandis que, si $j \geq \theta^*$, presque sûrement,

$$H_{n,j} = \frac{L(\theta^*)}{-L'(\theta^*)} \ln n + o(\ln n).$$

Si la condition suivante est vérifiée

$$\exists \delta > 0 \quad \text{tel que} \quad L(-\delta) < \infty \quad \text{et} \quad \int_{S^\downarrow} \left(\sum_{i \geq 1} \mathbf{1}_{s_i > 0} \right)^{1+\delta} \nu(ds) < \infty, \quad (1.12)$$

et si $j \geq \theta^*$, alors, presque sûrement,

$$H_{n,j} = \frac{L(\theta^*)}{-L'(\theta^*)} \ln n + O(\ln \ln n).$$

La précision du développement asymptotique de $H_{n,j}$ pour $j < \theta^*$ provient du résultat très fin (1.10). Le terme en $O(\ln \ln n)$ dans le développement asymptotique de $H_{n,j}$ pour $j \geq \theta^*$ s'obtient grâce aux récents résultats de Hu et Shi [HS09] sur les marches aléatoires branchantes en temps discret (il est alors crucial de supposer (1.12)). On peut en effet nettement améliorer l'asymptotique (1.9) en appliquant [HS09, Théorème 1.2 (1.7)] à une marche aléatoire branchante telle que la suite des distances d'un individu à enfants a même loi que $\mathbf{u} = (-\theta^* \ln s_i + \ln L(\theta^*))$ sous $\nu(ds)$.

Terminons ce tour d'horizon par mentionner un résultat où l'entier j est une fonction puissance de n :

Proposition 1.1. *Soit un réel $\alpha \in]0, 1[$. Presque sûrement,*

$$H_{n,n^\alpha} = (1 - \alpha) \frac{L(\theta^*)}{-L'(\theta^*)} \ln n + o(\ln n).$$

Si (1.12) est vérifiée et si $\theta^* < \infty$, alors, presque sûrement,

$$H_{n,n^\alpha} = (1 - \alpha) \frac{L(\theta^*)}{-L'(\theta^*)} \ln n + O(\ln \ln n).$$

De manière heuristique, le plus gros fragment est ici la dernière particule à ne plus contenir au moins n^α points parmi les n points d'étiquette minimale.

1.3.3 Étude des niveaux de saturation

Nous nous sommes aussi intéressés au comportement asymptotique du niveau de saturation $G_{n,j}$. Comme pour l'étude de la masse du plus petit fragment aboutissant à (1.6), nous faisons au moins l'hypothèse qu'il y a uniquement un nombre fini de particules créées à l'issue de chaque brisure. Nous verrons le résultat suivant :

Proposition 1.2. *Soit un entier $j \geq 1$. Si $\theta_* = -\infty$, ou si $-\infty < \theta_* < 0$ et $\ln L(\theta_*) = \theta_* L'(\theta_*)/L(\theta_*)$, alors, presque sûrement,*

$$G_{n,j} = \frac{L(\theta_*)}{-L'(\theta_*)} \ln n + o(\ln n).$$

D'autre part, si $-\infty < \theta_ < 0$, $\ln L(\theta_*) = \theta_* L'(\theta_*)/L(\theta_*)$ et (1.12) est vérifiée, alors, presque sûrement,*

$$G_{n,j} = \frac{L(\theta_*)}{-L'(\theta_*)} \ln n + O(\ln \ln n).$$

Il n'y a donc pas de transition de phase pour les niveaux de saturation. De façon informelle, le plus petit fragment est toujours la première particule qui contient strictement moins de j points.

Pour obtenir un terme en $O(\ln \ln n)$ dans le développement asymptotique de $G_{n,j}$ sous les hypothèses mentionnées dans la Proposition 1.2, on applique encore [HS09, Théorème 1.2 (1.7)] mais cette fois à une marche aléatoire branchante telle que la suite des distances d'un individu à ses enfants a même loi que $\mathbf{u} = (-\theta_* \ln s_i + \ln L(\theta_*))$ sous $\nu(ds)$.

Nous disposons ici aussi d'un résultat décrivant la situation lorsque l'entier j est une fonction puissance de n :

Proposition 1.3. *Soit un réel $\alpha \in]0, 1[$. Si $\theta_* = -\infty$, ou si $-\infty < \theta_* < 0$ et $\ln L(\theta_*) = \theta_* L'(\theta_*)/L(\theta_*)$, alors, presque sûrement,*

$$G_{n,n^\alpha} = (1 - \alpha) \frac{L(\theta_*)}{-L'(\theta_*)} \ln n + o(\ln n).$$

D'autre part, si $-\infty < \theta_ < 0$, $\ln L(\theta_*) = \theta_* L'(\theta_*)/L(\theta_*)$ et (1.12) est vérifiée, alors, presque sûrement,*

$$G_{n,n^\alpha} = (1 - \alpha) \frac{L(\theta_*)}{-L'(\theta_*)} \ln n + O(\ln \ln n).$$

Ici encore, le plus petit fragment est la première des particules à contenir moins de n^α points.

1.3.4 Exemple : les arbres fendus aléatoires

Nous supposons dans ce paragraphe qu'il existe un entier $b \geq 2$ tel que $\nu(ds)$ -p.s., $s_{b+1} = 0$, c'est-à-dire que, p.s., chaque dislocation d'une particule produit au plus b particules. Le modèle s'interprète alors en termes d'arbres fendus aléatoires (*random*

split trees en anglais). Il suffit pour cela de placer initialement les n points d'étiquette appartenant à $\{1, \dots, n\}$ dans un nœud racine puis de suivre la règle itérative suivante : chaque nœud \mathbf{n} contenant $p \geq j$ points (où $j \geq 2$ est un entier fixé) se scinde en b nœuds ; on tire une réalisation $\mathbf{s}(\mathbf{n}, \omega)$ de ν (indépendante des éventuelles réalisations de ν déjà tirées pour d'autres nœuds) ; les p points se répartissent dans les b nœuds aléatoirement selon $\mathbf{s}(\mathbf{n}, \omega)$ et indépendamment. L'objet obtenu à l'issue de cet algorithme est un arbre fini p.s. noté $T_{n,j}$ dont les feuilles contiennent strictement moins de j points et dont les nœuds internes ne contiennent aucun point. Il est aisé de voir que la hauteur de $T_{n,j}$ a même loi que $H_{n,j}$. La Figure 1.4 donne des exemples de tels arbres en reprenant la configuration des Figures 1.2 et 1.3.

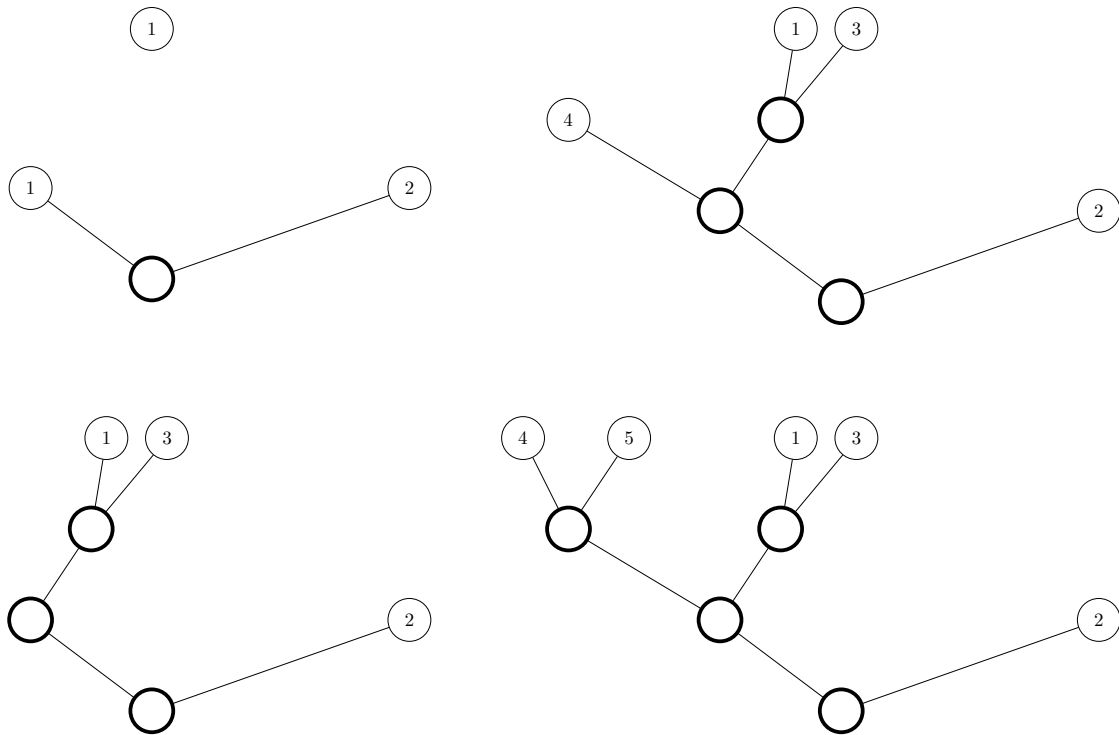


FIGURE 1.4 – Exemples d'arbres $T_{n,j}$ pour $n \leq 5$ et $j = 2$ (on n'a pas représenté les feuilles n'ayant aucun point). On a repris les réalisations des Figures 1.2 et 1.3. Le seuil de rupture étant ici $j = 2$, dès qu'un nœud contient plusieurs points, il se fragmente en donnant naissance à de nouveaux nœuds (il est dès lors représenté en gras). Il peut arriver que plusieurs nœuds gras apparaissent lorsqu'on passe de l'arbre $T_{n-1,j}$ à $T_{n,j}$, c'est-à-dire lorsqu'on ajoute le point d'étiquette n à l'arbre $T_{n-1,j}$. Il peut même arriver que les hauteurs de $T_{n-1,j}$ et $T_{n,j}$ diffèrent de plus de 1 (c'est le cas par exemple ici pour $n = 3$). La suite $(H_{n,2})$ pour $n \leq 5$ est $(0, 1, 3, 3, 3)$.

Les arbres $T_{n,j}$ sont des cas particuliers d'arbres fendus aléatoires étudiés par exemple par Devroye dans [Dev99]. Il faut toutefois bien noter qu'il y a par construction une probabilité non nulle⁴ pour que, à l'issue de la fragmentation d'un nœud, les points qui s'y trouvaient se retrouvent tous à nouveau dans un même nœud⁵. Ces phénomènes

⁴et non négligeable !

⁵C'est par exemple le cas à la Figure 1.4 pour l'arbre $T_{3,2}$ avec les points d'étiquette 1 et 3.

sont d'ailleurs à l'origine de l'existence de la transition de phase.

On peut coupler les arbres $T_{n,j}$ pour $n \geq 1$ et $j \geq 2$ en reprenant la construction du modèle donnée au début de cette section (cf. paragraphe 1.3.1). Par exemple la réalisation $\mathbf{s}(\mathbf{n}, \omega)$ peut se voir comme $(m_{ij}/m_i, j \leq b)$, où i est l'étiquette de la particule qui correspond au nœud \mathbf{n} . Ce couplage est d'ailleurs utilisé dans la Figure 1.4 puisqu'on s'est appuyé pour la dessiner sur la configuration fournie par les Figures 1.2 et 1.3.

Définissons maintenant une autre famille naturelle d'arbres fendus aléatoires, notée $(\tilde{T}_{n,j})$ et couplée avec la famille $(T_{n,j})$. Comme auparavant, initialement, les n points d'étiquette minimale sont placés dans un nœud racine. On suit ensuite la règle suivante : chaque nœud \mathbf{n} contenant $p \geq j$ points se scinde en b nœuds ; l'un des p points (par exemple celui d'étiquette minimale) reste dans \mathbf{n} ; les $p - 1$ autres points choisissent indépendamment l'un des b nœuds selon la réalisation $\mathbf{s}(\mathbf{n}, \omega)$. L'arbre obtenu est noté $\tilde{T}_{n,j}$. Il est fini p.s. Chacune de ses feuilles contient strictement moins de j points ; chacun de ses nœuds internes contient un point. Remarquons que la hauteur $\tilde{H}_{n,j}$ de $\tilde{T}_{n,j}$ est inférieure ou égale à $H_{n,j}$. La Figure 1.5 illustre la différence entre $T_{n,j}$ et $\tilde{T}_{n,j}$. Ici encore, on a repris le même exemple.

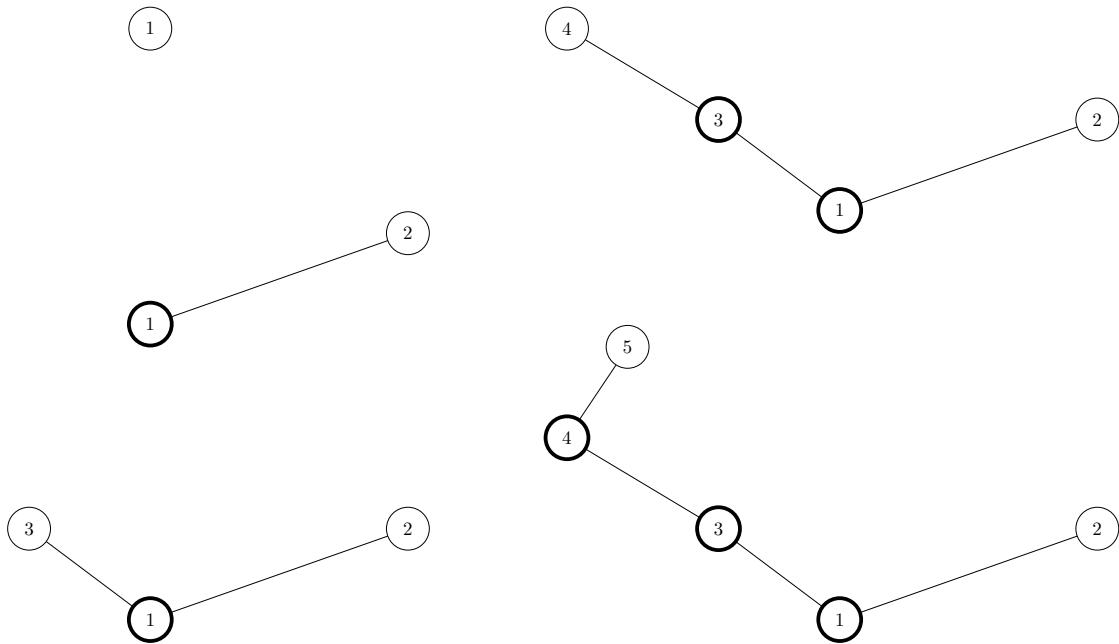


FIGURE 1.5 – Exemples d'arbres $\tilde{T}_{n,j}$ pour $n \leq 5$ et $j = 2$ (les feuilles ne contenant aucun point n'ont pas été représentées). Lorsqu'on passe de l'arbre $\tilde{T}_{n-1,j}$ à $\tilde{T}_{n,j}$, c'est-à-dire lorsqu'on ajoute le point d'étiquette n à l'arbre $\tilde{T}_{n-1,j}$, au plus un nœud interne (représenté en gras) est créé ; la différence entre les hauteurs de $\tilde{T}_{n-1,j}$ et $\tilde{T}_{n,j}$ est *a fortiori* au plus égale à 1. La suite $(\tilde{H}_{n,2})$ pour $n \leq 5$ est $(0, 1, 1, 2, 3)$.

Puisque tout nœud de $\tilde{T}_{n,j}$ visité par des points en conserve un, le phénomène décrit plus haut apparaissant pour les arbres $T_{n,j}$ ne se produit pas pour les arbres $\tilde{T}_{n,j}$. On peut montrer qu'il n'y a pas de transition de phase pour les hauteurs de cette famille d'arbres. Plus précisément, Devroye assure dans [Dev99, Théorème 1] que si $\theta^* < \infty$,

alors pour tout entier $j \geq 2$,

$$\frac{\tilde{H}_{n,j}}{\ln n} \underset{n \rightarrow \infty}{\overset{p}{\rightarrow}} \frac{L(\theta^*)}{-L'(\theta^*)}.$$

Pour finir, illustrons ce que nous venons de voir par un exemple simple. Supposons que ν soit la loi de la suite décroissante obtenue à partir de $(U, 1 - U, 0 \dots)$, U étant une variable aléatoire uniformément distribuée sur $[0, 1]$. On peut prouver que $\tilde{T}_{n,2}$ a la loi de l'arbre binaire de recherche aléatoire. On peut aussi montrer que $\lceil \theta^* \rceil = 4$. Ainsi les hauteurs des arbres $\tilde{T}_{n,j}$, $j \geq 2$, et celles des arbres $T_{n,j}$, $j \geq 4$, ont le même comportement asymptotique (au premier ordre) que la hauteur de l'arbre binaire de recherche aléatoire ; les arbres $T_{n,3}$ et $T_{n,2}$ sont quant à eux asymptotiquement bien plus hauts.

1.3.5 Présentation des techniques

Comme nous l'avons déjà souligné, la théorie des marches aléatoires branchantes est au cœur des preuves. Outre les résultats de Biggins déjà mentionnés, ceux de Hu et Shi nous seront très utiles. Un autre outil, plus technique, est la Poissonisation : au lieu d'étudier les points d'étiquette inférieure à n , on considérera les points d'étiquette inférieure à \mathcal{P}_n , où \mathcal{P}_n est une variable de Poisson de paramètre n indépendante de la fragmentation. Les calculs sont alors plus simples. On gagne d'autre part de l'indépendance puisque, conditionnellement aux masses des particules présentes à un instant donné, les nombres de points d'étiquette inférieure à \mathcal{P}_n contenus dans ces particules deviennent des variables aléatoires indépendantes ; l'étude sera nettement facilitée. Il restera à montrer via une dépoissonisation que les deux modèles sont proches.

1.4 Le quadtree

Nous présentons ici des résultats obtenus dans [CJ11] conjointement avec Nicolas Curien (cf. chapitre 3).

1.4.1 Le modèle et les principaux résultats

Le quadtree est une structure surtout étudiée par les informaticiens parce que, permettant de retrouver des données, elle a de nombreuses applications. L'enjeu est ici d'étudier de coût moyen de la *requête partielle* (*partial match query* en anglais). Cette question fut originellement posée par Flajolet.

Décrivons le modèle. Soit n points P_1, \dots, P_n indépendants et uniformément distribués sur le carré $[0, 1]^2$. On définit itérativement un recouvrement de p.s. $3n + 1$ rectangles fermés d'intérieur disjoint de la manière suivante. Le premier point P_1 découpe le carré $[0, 1]^2$ en quatre quadrants selon son abscisse et son ordonnée. Le point P_2 appartient p.s. à l'intérieur d'un unique quadrant, qu'il divise en quatre quadrants selon ses coordonnées. On procède de la même manière jusqu'au n -ème point. Pour tout $x \in [0, 1]$, notons alors $\mathcal{N}_n(x)$ le nombre de rectangles parmi les $3n + 1$ construits qui intersectent le segment $[(x, 0), (x, 1)]$. On s'intéresse au comportement asymptotique de

$\mathbb{E}[\mathcal{N}_n(x)]$ lorsque n tend vers l'infini, à $x \in [0, 1]$ fixé. La Figure 1.6 représente une réalisation du quadtree.

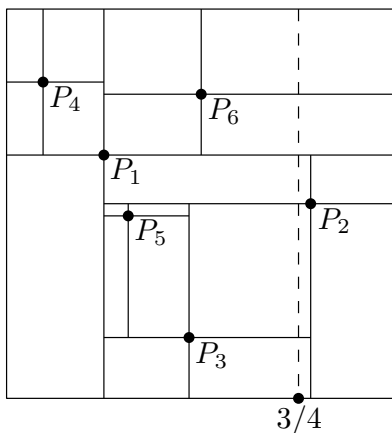


FIGURE 1.6 – Une réalisation du quadtree ayant $n = 6$ points. Ici, $\mathcal{N}_6(3/4) = 5$.

Nous verrons en particulier le résultat suivant.

Théorème 1.2. *Pour tout $x \in [0, 1]$,*

$$n^{-\beta^*} \mathbb{E}[\mathcal{N}_n(x)] \xrightarrow[n \rightarrow \infty]{} K_0 (x(1-x))^{\beta^*/2},$$

où $\beta^* = \frac{\sqrt{17}-3}{2}$ et $K_0 = \frac{\Gamma(2\beta^*+2)\Gamma(\beta^*+2)}{2\Gamma^3(\beta^*+1)\Gamma^2(\frac{\beta^*}{2}+1)}$.

Flajolet, Gonnet, Puech et Robson s'étaient intéressés dans [FGCR93] au comportement asymptotique de $\mathbb{E}[\mathcal{N}_n(x)]$ lorsque x est non pas fixé mais tiré uniformément au hasard dans $[0, 1]$ et indépendamment de P_1, \dots, P_n . Nous nous appuyerons d'ailleurs sur ce résultat, dont nous donnerons une nouvelle preuve grâce à la théorie des fragmentations, pour démontrer le Théorème 1.2.

1.4.2 Lien avec les fragmentations

Le modèle en temps continu

Comme pour les arbres de fragmentation, l'étude du quadtree est rendue plus simple lorsque, au lieu de jeter exactement n points, on en jette \mathcal{P}_n , où \mathcal{P}_n est une variable de Poisson de paramètre n . Cela revient à travailler en temps continu ; le modèle se reformule alors de la manière suivante. Initialement, le système est formé du carré $[0, 1]^2$ auquel on attache une horloge exponentielle de paramètre 1. Lorsque cette horloge sonne, on jette un point uniformément au hasard à l'intérieur du carré, le découpant en quatre quadrants. On attache alors à chacun de ces rectangles une horloge exponentielle dont le paramètre est égal à l'aire du rectangle, et on itère le procédé : dès qu'une horloge sonne, on divise le rectangle auquel elle est attachée en quatre rectangles selon un point jeté uniformément au hasard, puis on attache aux nouveaux rectangles des horloges exponentielles de paramètres idoines, etc.

Pour tout $x \in [0, 1]$ et pour tout $t \geq 0$, on note $N_t(x)$ le nombre de rectangles présents à l'instant t intersectant le segment $[(x, 0), (x, 1)]$. Modulo une dépoissonisation, notre problématique revient à connaître le comportement asymptotique de $\mathbb{E}[N_t(x)]$ lorsque t tend vers l'infini, x étant fixé.

Il est important de remarquer que, p.s., à l'issue de chaque brisure d'un rectangle en quatre sous-rectangles, au plus deux d'entre eux intersectent le segment $[(x, 0), (x, 1)]$. On peut par conséquent décréter que le système est constitué uniquement de rectangles intersectant $[(x, 0), (x, 1)]$. On peut alors voir $N_t(x)$ comme le nombre de rectangles présents à l'instant t . Nous adoptons cette approche à partir de maintenant.

Nous entrevoyons un lien avec les fragmentations autosimilaires d'indice d'autosimilarité égal à 1 et de mesure de dislocation dissipatrice.

Un modèle simplifié

Dans une première étape, nous nous intéressons à $\mathbb{E}[N_t(U)]$, où U est une variable aléatoire uniforme sur $[0, 1]$ indépendante des horloges exponentielles et des points jetés. Nous allons voir que cette quantité est aisée à étudier grâce à la théorie des fragmentations. Il faut toutefois bien noter que la suite décroissante des aires des rectangles intersectant $[(U, 0), (U, 1)]$ ne constitue pas une fragmentation ne serait-ce qu'à cause de la dépendance d'évolution des deux processus issus des deux premiers sous-rectangles créés (cela vient du fait que l'abscisse U , une fois tirée, est fixée pour toujours). Nous allons présenter ici une légère modification du modèle qui nous permettra de définir une véritable fragmentation que l'on saura étudier. De manière heuristique, il s'agira de tirer de nouvelles abscisses U à chaque fois que de nouveaux rectangles apparaissent.

Supposons que, au moment même où l'on attache à un rectangle R d'aire $\lambda(R)$ une horloge exponentielle de paramètre $\lambda(R)$, on choisisse une abscisse x_R uniformément au hasard dans le segment $[g_R, d_R]$, où g_R et d_R sont les abscisses des côtés gauche et droit de R . Ensuite, lorsque l'horloge sonne et qu'un point est jeté uniformément à l'intérieur du rectangle (et bien entendu indépendamment de x_R), divisant R en quatre sous-rectangles, on ne garde que les deux sous-rectangles intersectant le segment $[(x_R, 0), (x_R, 1)]$, les deux autres disparaissant du système. On attache à chacun d'eux, *indépendamment*, une horloge exponentielle de paramètre idoine et une abscisse choisie convenablement. Le procédé se répète. Nous avons représenté une réalisation de l'évolution de ce système dans la Figure 1.7.

Pour tout $t \geq 0$, définissons la suite décroissante $F(t)$ des aires des rectangles présents à l'instant t complétée par des zéros. On peut alors aisément montrer que F est une fragmentation autosimilaire d'indice d'autosimilarité 1 et de mesure de dislocation ν dissipatrice, supportée par $\{\mathbf{s} = (s_1, s_2, 0, \dots) : s_1 \geq s_2 > 0 \text{ et } s_1 + s_2 \leq 1\}$ et caractérisée par : pour fonction mesurable $f : [0, 1]^2 \rightarrow \mathbb{R}_+$,

$$\int_{\mathcal{S}^\downarrow} f(s_1, s_2) \nu(ds) = 4 \int_0^1 dx \int_{1/2}^1 dy x f(xy, x(1-y)).$$

Reprenant les notations du paragraphe 1.2.1, on a alors $\psi(\beta) = \frac{\beta^2 + 3\beta - 2}{(\beta+1)(\beta+2)}$, et l'exposant malthusien β^* de la fragmentation F est égal à $\frac{\sqrt{17}-3}{2}$.

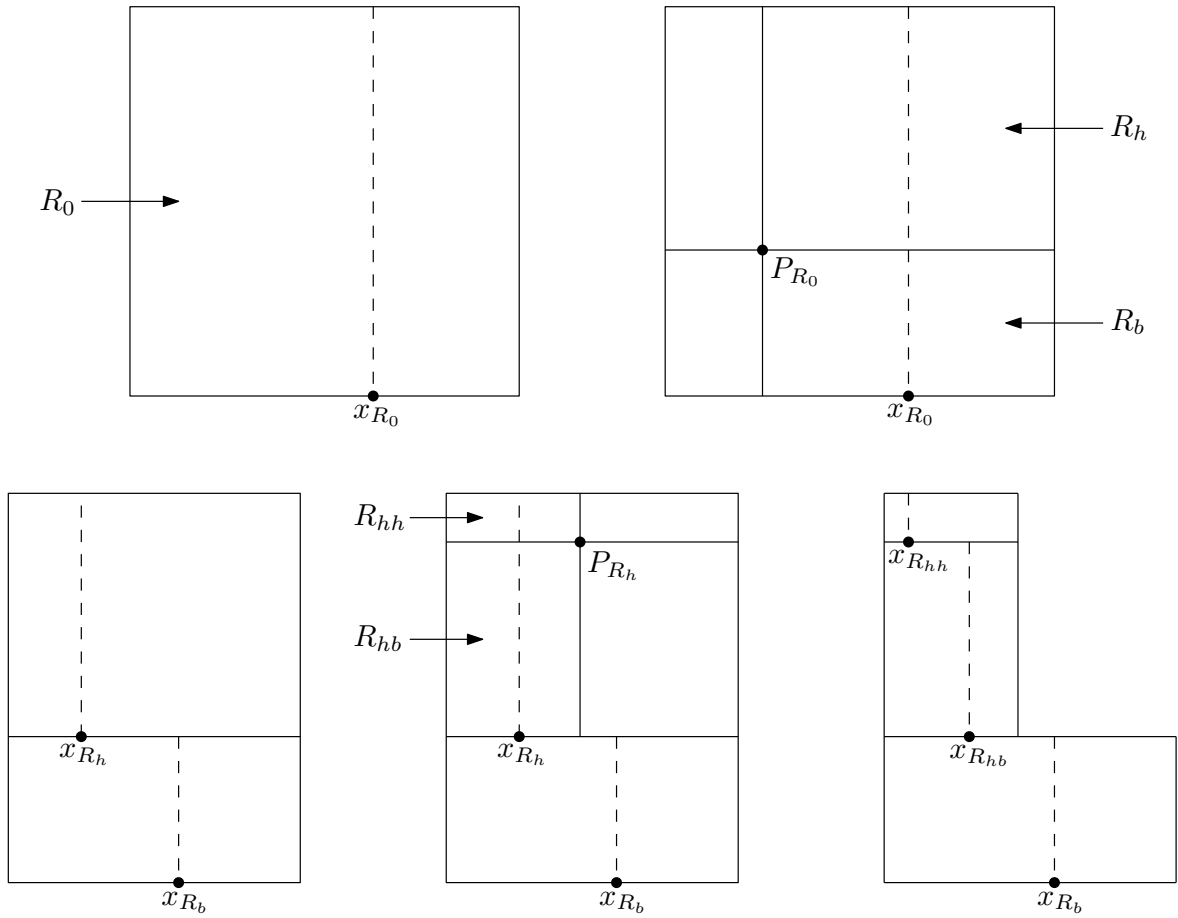


FIGURE 1.7 – Une réalisation du modèle simplifié jusqu’au retentissement de la deuxième horloge exponentielle.

Lien entre les deux modèles

Bien que le modèle que nous venons de présenter soit bien compris (en partie grâce aux résultats fournis par la théorie des fragmentations), il est bien différent du modèle initial, pour lequel l’abscisse x est fixée pour toujours. Cependant, en notant $\#F(t)$ le nombre de particules (qui sont ici des rectangles) de la fragmentation F présentes à l’instant t , nous verrons que $\mathbb{E}[N_t(U)]$ et $\mathbb{E}[\#F(t)]$ vérifient la même équation intégrale, de laquelle nous pourrions déduire que $\mathbb{E}[N_t(U)] = \mathbb{E}[\#F(t)]$. Il faut bien noter que $N_t(U)$ et $\#F(t)$ ne suivent toutefois pas la même loi.

En appliquant les résultats du paragraphe 1.2.5, et en particulier (1.11), nous en déduisons que

$$t^{-\beta^*} \mathbb{E}[N_t(U)] \xrightarrow[t \rightarrow \infty]{} \frac{\Gamma(2(\beta^* + 1))}{2\Gamma^3(\beta^* + 1)},$$

retrouvant ainsi le résultat de Flajolet, Gonnet, Puech et Robson dans [FGCR93].

1.4.3 Techniques de preuves

Voyons maintenant la stratégie adoptée pour obtenir les résultats à $x \in [0, 1]$ fixé. L'idée est de prouver tout d'abord que $t^{-\beta^*} \mathbb{E}[N_t(x)]$ admet une limite finie $n_\infty(x)$ en s'appuyant sur l'existence d'une limite finie pour $t^{-\beta^*} \mathbb{E}[N_t(U)]$. Le calcul de $n_\infty(x)$ se fera ensuite sans difficulté : nous remarquerons que la fonction n_∞ est continue et est solution d'une équation intégrale dont l'ensemble des solutions continues est une droite vectorielle que l'on peut déterminer explicitement.

Commençons par introduire quelques notations. Nous travaillons ici à x fixé. Nous allons suivre la place relative de l'abscisse x dans les rectangles successifs contenant le point $(x, 0)$. Plus précisément, on note $X_0 = x$ la place de l'abscisse x dans l'unique rectangle $R_0 = [0, 1]^2$ présent à l'origine. Notons R_1 le seul (p.s.) sous-rectangle de R_0 parmi ceux issus de brisure de R_0 qui contienne le point $(x, 0)$. Posons $X_1 \in [0, 1]$ la place qu'occupe le point $(x, 0)$ sur le côté inférieur de R_1 . On définit plus généralement la suite $(R_k, X_k)_{k \in \mathbb{N}}$ par récurrence : ayant construit R_k , on désigne par R_{k+1} le sous-rectangle issu de sa brisure contenant $(x, 0)$ et par $X_{k+1} \in [0, 1]$ la place relative de ce point sur le côté inférieur de R_{k+1} (voir la Figure 1.8 pour un exemple).

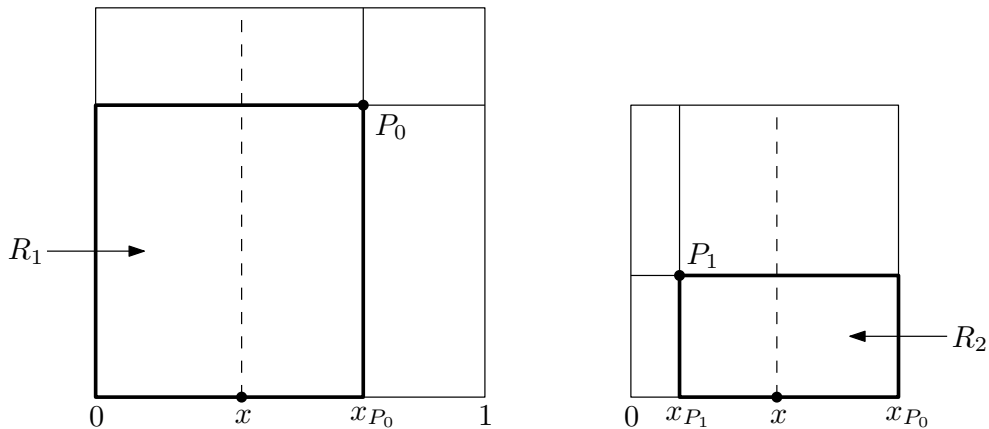


FIGURE 1.8 – Dans cet exemple, $X_1 = \frac{x}{x_{P_0}}$ et $X_2 = \frac{x - x_{P_1}}{x_{P_0} - x_{P_1}}$.

On peut aisément montrer que, pour $k \in \mathbb{N}$ fixé, notant $\lambda(R_k)$ l'aire du rectangle R_k , $t^{-\beta^*} \mathbb{E}[N_t(x)]$ peut être approché, pour t suffisamment grand, par

$$t^{-\beta^*} 2^k \mathbb{E} \left[\tilde{N}_{\lambda(R_k)t}(X_k) \right],$$

où $(\tilde{N}_s(y), (s, y))$ est une copie de $(N_s(y), (s, y))$ indépendante de $(\lambda(R_k), X_k)$. L'idée étant de se ramener à l'étude de $t^{-\beta^*} \mathbb{E}[N_t(U)]$, il s'agit de prouver que, pour k suffisamment grand, la loi de X_k est « proche » de l'uniforme sur $[0, 1]$ et les v.a. $\lambda(R_k)$ et X_k sont « presque » indépendantes⁶. Une majoration uniforme pour $s^{-\beta^*} \mathbb{E}[N_s(y)]$ permettra enfin d'appliquer le théorème de convergence dominée.

Les techniques mises en jeu pour quantifier ces phénomènes reposent sur l'étude d'une chaîne de Markov qui exploite le fait très important suivant : conditionnellement

⁶On peut montrer que, lorsque x est choisi uniformément sur $[0, 1]$, la v.a. X_1 est uniformément distribuée sur $[0, 1]$ et est indépendante de $\lambda(R_1)$.

à $\lambda(R_1) < x \wedge (1 - x)$, événement de probabilité strictement positive, la v.a. X_1 admet une densité par rapport à la mesure de Lebesgue minorée par un réel strictement positif dont l'expression ne fait pas intervenir $\lambda(R_1)$; de façon très informelle, cela montre que, avec probabilité $p_x > 0$, X_1 « devient » uniforme et indépendante de $\lambda(R_1)$. Si cet événement ne se produit pas (ce qui arrive avec probabilité $1 - p_x$), on poursuit avec X_2 , etc. Il s'agira d'estimer le premier instant où l'événement rendant X_k uniforme arrive. Nous utiliserons pour cela l'ouvrage [MT09].

1.4.4 Extensions des résultats

Les arguments présentés ici peuvent très certainement s'appliquer aussi pour étudier des quadrees en toute dimension.

Des questions intéressantes concernant le quadree 2-dimensionnel restent par ailleurs pour l'instant sans réponse. Nous avons vu que pour tout $x \in [0, 1]$, $t^{-\beta^*} N_t(x)$ converge en espérance. On peut se demander s'il y a convergence en loi, en probabilité, ou même presque sûre. Dans une autre direction, au lieu de fixer x dans $[0, 1]$, nous pourrions étudier le processus spatial $((t^{-\beta^*} N_t(x))_{x \in [0, 1]}, t \geq 0)$ et nous demander si ce processus converge au sens de Skorokhod vers un processus $(\mathcal{C}(x))_{x \in [0, 1]}$ (avec Nicolas Curien, nous avons abordé cette question avec Ralph Neininger et Henning Sulzbach).

1.5 Graphes aléatoires

1.5.1 Modèle du graphe aléatoire d'Erdős et Rényi

Parmi les modèles de graphes aléatoires, celui d'Erdős et Rényi est très certainement l'un des plus naturels et des plus étudiés. Il se définit très simplement de la manière suivante. Soit un réel p compris entre 0 et 1. Le graphe aléatoire $G(n, p)$ est un graphe à n sommets tel que chacune de ses $\binom{n}{2}$ paires de sommets est reliée par une arête avec probabilité p , indépendamment des autres paires de sommets.

Les questions qui nous intéressent ici sont liées aux tailles des composantes connexes du graphe⁷, et plus précisément aux tailles des grandes composantes, pour de grandes valeurs de n . Pour que l'étude soit intéressante, il faut supposer p de l'ordre de $1/n$. On peut résumer les résultats de la manière suivante. Pour tout $c \in]0, 1[$, a.p.s.⁸, la taille de la plus grande composante de $G(n, c/n)$ est $\Theta(\ln n)$, alors que pour tout $c > 1$, a.p.s., la taille de la plus grande composante de $G(n, c/n)$ est $\Theta(n)$ et la taille de la deuxième plus grande composante est $\Theta(\ln n)$. Enfin, les tailles de la plus grande et de la deuxième plus grande composantes de $G(n, 1/n)$ sont $\Theta(n^{2/3})$ a.p.s. On voit ainsi apparaître une transition de phase. Nous recommandons vivement au lecteur intéressé [ER60] d'Erdős et Rényi, ainsi que [Bol01] de Bollobás. Pour la phase critique $p \sim 1/n$, nous conseillons [Bol84] de Bollobás, [Łuc90] de Łuczak et [JKŁP93] de Janson, Knuth, Łuczak et Pittel.

Intéressons-nous plus particulièrement à la phase critique. Dans [Ald97], Aldous étudie la suite des tailles des composantes du graphe aléatoire d'Erdős et Rényi de

⁷La taille d'une composante est définie comme le cardinal de l'ensemble des sommets qu'elle contient.

⁸abréviation d'asymptotiquement presque sûrement, qui signifie avec probabilité tendant vers 1 quand n tend vers l'infini

paramètres n et $p = n^{-1} + tn^{-4/3}$, où $t \in \mathbb{R}$ est fixé. Il montre que la suite décroissante des tailles des composantes de $G(n, n^{-1} + tn^{-4/3})$ divisées par $n^{2/3}$ converge en loi au sens d'une certaine topologie vers une suite décroissante aléatoire dont les termes sont tous non nuls p.s. Nous verrons dans la suite la formulation précise de ce résultat (cf. paragraphe 1.6.1).

1.5.2 Graphes aléatoires avec une suite de degrés fixée

Définissons un autre modèle de graphes aléatoires. Supposons disposer pour tout entier $n \geq 2$ d'une suite $\mathbf{d}^{(n)} = (d_i^{(n)})_{1 \leq i \leq n}$ d'entiers strictement positifs telle que $\sum_{i=1}^n d_i^{(n)}$ est paire. Notons $G(n, \mathbf{d}^{(n)})$ le graphe aléatoire à n sommets étiquetés de 1 à n dont la suite des degrés des sommets est $\mathbf{d}^{(n)}$, uniformément choisi parmi toutes les possibilités (on suppose qu'au moins un tel graphe existe). Nous souhaitons ici encore connaître le comportement des grandes composantes connexes de $G(n, \mathbf{d}^{(n)})$ lorsque n tend vers l'infini. Nous nous plaçons dans le cadre suivant : on suppose qu'il existe une mesure de probabilité sur \mathbb{N}^* , caractérisée par une suite $(\nu_k)_{k \geq 1}$, telle que $\#\{i : d_i^{(n)} = k\}/n \rightarrow \nu_k$ quand $n \rightarrow \infty$.

Molloy et Reed ont prouvé dans [MR95] que les comportements des grandes composantes connexes de $G(n, \mathbf{d}^{(n)})$ pour n grand dépendent du signe de

$$Q = \sum_{k=1}^{\infty} k(k-2)\nu_k.$$

De manière heuristique, la raison en est simple⁹. Supposons n grand et voyons comment on construit la composante connexe d'un sommet donné v . Prenons une demi-arête de v . Il s'agit de l'associer à une autre demi-arête. Elle en choisit une attachée à un sommet v' de degré k avec probabilité $k\nu_k/\mu$, où $\mu = \sum_i i\nu_i$ (par souci de simplicité, μ est supposé fini ici) ; il faut en effet ici considérer la loi de ν biaisée par la taille. On fait ensuite de même avec chacune des $k-1$ autres demi-arêtes de v' . On voit ainsi apparaître un simulacre de processus de Galton-Watson de loi de reproduction donnée par $((k+1)\nu_{k+1}/\mu, k \geq 0)$. Si $\sum k(k+1)\nu_{k+1}/\mu < 1$, c'est-à-dire si $Q < 0$, le processus de Galton-Watson est sous-critique ; les composantes connexes sont petites. En revanche, si $Q > 0$, le processus de Galton-Watson est surcritique ; il y a une probabilité non nulle pour qu'un sommet donné donne naissance à une composante géante.

Sans mentionner les résultats optimaux, notons seulement que les travaux de Molloy et Reed et ceux plus récents de Janson et Luczak [JL09] montrent que, sous certaines hypothèses, si $Q < 0$, alors a.p.s. la taille de la plus grande composante de $G(n, \mathbf{d}^{(n)})$ est $o(n)$, tandis que, si $Q > 0$, alors a.p.s. la taille de la plus grande composante est $\Theta(n)$.

Nous voyons apparaître ici aussi une transition de phase. Attachons-nous plus particulièrement au cas critique $Q = 0$. La structure de $G(n, \mathbf{d}^{(n)})$ dépend alors de la vitesse de convergence de

$$\alpha_n = \sum_{k=1}^{\infty} k(k-2) \frac{\#\{i : d_i^{(n)} = k\}}{n}$$

⁹L'interprétation qui suit sera mieux comprise après lecture du paragraphe 1.5.4. Il nous a semblé toutefois important de fournir tout de suite une explication informelle.

vers 0 (voir Kang et Seierstad [KS08]). Sous certaines hypothèses (parmi lesquelles $\beta = \sum_{k \geq 3} k(k-1)(k-2)\nu_k$ est fini et non nul), Janson et Luczak [JL09] ont prouvé que si $\alpha_n \rightarrow 0$ et $n^{1/3}\alpha_n \rightarrow \infty$, alors la taille de la plus grande composante de $G(n, \mathbf{d}^{(n)})$ divisée par $n\alpha_n$ converge en probabilité vers $2\mu/\beta$ et la taille de la deuxième plus grande composante divisée par $n\alpha_n$ converge en probabilité vers 0.

1.5.3 Lien entre les deux modèles

À première vue, les deux modèles que nous venons de présenter ne sont pas liés puisque, pour le modèle du graphe aléatoire d'Erdős et Rényi, à nombre de sommets fixé, la suite des degrés du graphe est aléatoire. Remarquant cependant que pour toute suite déterministe $\mathbf{d}^{(n)}$ de n entiers strictement positifs, le graphe $G(n, p)$ conditionné à avoir $\mathbf{d}^{(n)}$ comme suite de degrés est uniformément distribué selon toutes les possibilités, Janson et Luczak [JL09] ont montré que leurs résultats pouvaient s'adapter au modèle du graphe aléatoire d'Erdős et Rényi. Ainsi, pour étudier $G(n, n^{-1} + t_n n^{-4/3})$ avec (t_n) une suite de réels telle que $t_n n^{-1/3} \rightarrow 0$ et $t_n \rightarrow \infty$, les résultats de Janson et Luczak présentés plus haut s'appliquent en prenant pour ν la loi de Poisson de paramètre 1^{10} ; dans ce cas $\mu = \beta = 1$ et la taille de la plus grande composante de $G(n, n^{-1} + t_n n^{-4/3})$ divisée par $t_n n^{2/3}$ converge en probabilité vers 2 alors que la taille de la deuxième plus grande composante de $G(n, n^{-1} + t_n n^{-4/3})$ divisée par $t_n n^{2/3}$ converge en probabilité vers 0. Il est à noter que, comme nous l'avons entrevu dans le paragraphe 1.5.1, si (t_n) est une suite bornée, les tailles des deux plus grandes composantes connexes de $G(n, n^{-1} + t_n n^{-4/3})$ sont de l'ordre de $n^{2/3}$.

1.5.4 Le modèle de configuration

Notre étude portera sur la phase critique d'un modèle de (multi)graphes aléatoires proche de celui présenté dans le paragraphe 1.5.2. Présentons tout d'abord le *modèle de configuration* dans le cadre du modèle des graphes aléatoires avec une suite de degrés fixée. Il s'agit d'une méthode permettant de construire un *multigraphe* (c'est-à-dire un graphe pour lequel peuvent exister à la fois des arêtes dont les extrémités sont les mêmes, appelées *boucles*, et plusieurs mêmes arêtes, appelées *multiarêtes*) aléatoire tel que, conditionnellement à être *simple* (c'est-à-dire n'avoir ni boucle ni multiarête), il a même loi qu'un graphe aléatoire avec une suite de degrés fixée. Elle fut introduite par Bender et Canfield [BC78] puis étudiée par Bollobás [Bol80] et Wormald [Wor78].

Soit donc un entier $n \geq 2$ et une suite $\mathbf{d}^{(n)} = (d_i^{(n)})_{1 \leq i \leq n}$ d'entiers strictement positifs telle que $\sum_{i=1}^n d_i^{(n)}$ est paire. On étiquette les n sommets du multigraphe que l'on souhaite construire de 1 à n , et, pour tout $i \in \{1, \dots, n\}$, on attache $d_i^{(n)}$ demi-arêtes au sommet d'étiquette i . On apparie ensuite les demi-arêtes uniformément au hasard (rappelons que le nombre d'appariements possibles des $2s = \sum_{i=1}^n d_i^{(n)}$ demi-arêtes est $(2s-1)!! = \frac{(2s)!}{2^s s!}$; il s'agit donc de choisir l'un de ces appariements uniformément au hasard).

Toute réalisation de $G(n, \mathbf{d}^{(n)})$ pouvant être construite via ce procédé par exactement le même nombre d'appariements (à savoir $d_1^{(n)}! \cdots d_n^{(n)}!$), on a bien le graphe $G(n, \mathbf{d}^{(n)})$ si

¹⁰On sort légèrement de notre cadre puisque $\nu_0 > 0$, mais il est aisé de voir que l'on peut tout adapter.

on conditionne le multigraphe obtenu par le modèle de configuration à être simple. On peut démontrer d'autre part que, sous certaines conditions sur la suite $(\mathbf{d}^{(n)}, n \geq 2)$ (voir Bollobás [Bol01], Janson [Jan09]), la probabilité pour que le multigraphe soit simple est minorée par une constante strictement positive. Il est donc très commode de faire un travail préliminaire sur ces multigraphes, que l'on peut construire explicitement, pour obtenir via un conditionnement des résultats sur $G(n, \mathbf{d}^{(n)})$ (voir par exemple Pittel [Pit08], Janson [Jan08], Janson et Luczak [JL09]).

1.5.5 Le modèle étudié

Le modèle que l'on considère est une combinaison des modèles que nous venons de présenter. Il se situe dans un certain sens au centre de la fenêtre critique. Dans ce paragraphe, nous présentons ce modèle. Il sera vu en détail dans la section 1.6.

Soit $\nu = (\nu_k, k \geq 1)$ une mesure de probabilité sur \mathbb{N}^* ayant un moment d'ordre 2 fini telle que $\nu_2 < 1$ et $\sum_{k \geq 1} k(k-2)\nu_k = 0$. L'objet étudié, noté \mathcal{G}_n^ν , est le multigraphe aléatoire à n sommets étiquetés de 1 à n défini via le modèle de configuration de la manière suivante. Soit D_1, D_2, \dots, D_n n v.a. indépendantes de loi ν . On travaille conditionnellement à l'événement $\{\sum_{i=1}^n D_i \text{ est paire}\}$. Pour tout $i \in \{1, \dots, n\}$, on attache D_i demi-arêtes au sommet d'étiquette i . On apparie ensuite uniformément au hasard les demi-arêtes. Le multigraphe aléatoire obtenu est \mathcal{G}_n^ν .

Comme précédemment, nous nous demandons comment se comportent les tailles des plus grandes composantes connexes. Nous verrons que nous pourrions être très précis au sens où, comme dans [Ald97], nous obtiendrons le comportement asymptotique de la suite entière des tailles des composantes de \mathcal{G}_n^ν .

Posons \mathbf{C}_n^ν la suite décroissante des tailles des composantes de \mathcal{G}_n^ν . Nous nous intéresserons dans un premier temps au cas où ν a un moment d'ordre 3 fini ; nous verrons que $(n^{-2/3}\mathbf{C}_n^\nu)$ converge alors en loi selon une certaine topologie vers une suite décroissante aléatoire non triviale. Nous étudierons ensuite le cas où ν est une loi de puissance d'exposant $\gamma \in]3, 4[$, c'est-à-dire que la suite $(k^\gamma \nu_k)$ converge vers un réel strictement positif. Nous établirons cette fois que la suite $(n^{-(\gamma-2)/(\gamma-1)}\mathbf{C}_n^\nu)$ converge en loi selon une certaine topologie vers une suite décroissante aléatoire non triviale.

Nous nous intéresserons aussi au graphe *simple* aléatoire critique \mathcal{SG}_n^ν défini de la manière suivante : \mathcal{SG}_n^ν est le graphe aléatoire étiqueté à n sommets tel que, conditionnellement à la suite des degrés (D_1, \dots, D_n) ¹¹, il est uniformément choisi parmi tous les graphes simples ayant cette suite de degrés. Le graphe \mathcal{SG}_n^ν peut aussi être vu comme le multigraphe \mathcal{G}_n^ν conditionné à être simple. Nous verrons comment déduire de l'étude menée sur \mathcal{G}_n^ν des résultats sur les tailles des composantes de \mathcal{SG}_n^ν . Cette extension ne sera toutefois établie que dans le cas où ν a un moment d'ordre 3 fini (le cas général n'est pas traité à ce jour). Plus précisément, notant \mathbf{SC}_n^ν la suite décroissante des tailles des composantes du graphe \mathcal{SG}_n^ν , nous démontrerons que $(n^{-2/3}\mathbf{SC}_n^\nu)$ converge en loi vers la même limite que la suite $(n^{-2/3}\mathbf{C}_n^\nu)$.

¹¹qui est, rappelons-le, un n -échantillon de loi ν

1.6 Les tailles des composantes d'un graphe aléatoire critique

Nous nous intéressons ici au modèle présenté au paragraphe 1.5.5 dont nous reprenons les notations. L'étude que nous allons présenter a été conduite dans [Jos11b] (cf. chapitre 4).

1.6.1 Présentation des résultats

Le cas du moment d'ordre 3 fini

Nous supposons dans ce paragraphe que la mesure de ν vérifie :

$$\sum_{k=1}^{\infty} k(k-2)\nu_k = 0, \quad \sum_{k=1}^{\infty} k^3\nu_k < \infty \quad \text{et} \quad \nu_2 < 1. \quad (1.13)$$

Posons

$$\mu = \sum_{k=1}^{\infty} k\nu_k \quad \text{et} \quad \beta = \sum_{k=3}^{\infty} k(k-1)(k-2)\nu_k.$$

Observons que $\beta > 0$. Définissons les deux processus suivants : premièrement le mouvement brownien avec dérive parabolique

$$W^\nu : t \geq 0 \mapsto \sqrt{\frac{\beta}{\mu}} W(t) - \frac{\beta}{2\mu^2} t^2,$$

où W est un mouvement brownien standard, puis le processus réfléchi

$$R^\nu : t \geq 0 \mapsto W^\nu(t) - \min_{0 \leq s \leq t} W^\nu(s).$$

On appelle excursion de R^ν tout intervalle de temps $\gamma = [g(\gamma), d(\gamma)]$ tel que $R^\nu(g(\gamma)) = R^\nu(d(\gamma)) = 0$ et $R^\nu(t) > 0$ pour tout $t \in]g(\gamma), d(\gamma)[$. Une excursion γ a pour longueur $|\gamma| = d(\gamma) - g(\gamma)$. Aldous a observé dans [Ald97] que, p.s., l'ensemble des excursions de R^ν peut s'écrire $\{\gamma_j, j \geq 1\}$ de sorte que la suite des longueurs $(|\gamma_j|)$ est décroissante. Posons \mathcal{l}_\searrow^2 l'ensemble des suites $x = (x_1, x_2, \dots)$ telles que $x_1 \geq x_2 \geq \dots \geq 0$ et $\sum_i x_i^2 < \infty$, muni de la métrique euclidienne. Aldous a montré dans [Ald97, Lemme 25] que $\mathbb{E}[\sum_{j \geq 1} |\gamma_j|^2] < \infty$. En particulier la suite $(|\gamma_j|)$ appartient à \mathcal{l}_\searrow^2 p.s. D'autre part, on peut voir la suite décroissante \mathbf{C}_n^ν des tailles des composantes connexes de \mathcal{G}_n^ν comme une v.a. à valeurs dans \mathcal{l}_\searrow^2 en ajoutant des termes nuls à partir d'un certain rang.

Voici notre principal résultat dans le cas où ν admet un moment d'ordre 3 fini :

Théorème 1.3. *Supposons que la mesure de probabilité ν vérifie (1.13). Alors*

$$n^{-2/3} \mathbf{C}_n^\nu \xrightarrow[n \rightarrow \infty]{(d)} (|\gamma_j|, j \geq 1)$$

pour la topologie de \mathcal{l}_\searrow^2 .

Étude du graphe simple \mathcal{SG}_n^ν lorsque ν a un moment d'ordre 3 fini

Comme annoncé précédemment, lorsque ν a un moment d'ordre 3 fini, (\mathcal{SC}_n^ν) a le même comportement asymptotique que (\mathcal{C}_n^ν) . Nous établirons plus précisément le résultat suivant :

Théorème 1.4. *Supposons que la mesure de probabilité ν vérifie (1.13). Alors*

$$n^{-2/3} \mathcal{SC}_n^\nu \xrightarrow[n \rightarrow \infty]{(d)} (|\gamma_j|, j \geq 1)$$

pour la topologie de l_{\searrow}^2 .

Lien avec le résultat d'Aldous [Ald97, Corollaire 2]

Dans [Ald97], Aldous avait déjà démontré un résultat analogue concernant la suite \mathcal{C}_n^t des tailles des composantes du graphe aléatoire d'Erdős et Rényi de paramètres n et $p = n^{-1} + tn^{-4/3}$, le réel t étant fixé. En notant R^t le processus réfléchi du processus $s \geq 0 \mapsto W(s) + ts - s^2/2$, où W est un mouvement brownien standard, Aldous montre dans [Ald97, Lemme 25] que, p.s., l'ensemble des excursions de R^t peut s'écrire $\{\gamma_j, j \geq 1\}$ de sorte que la suite $(|\gamma_j|)$ appartient à l_{\searrow}^2 . [Ald97, Corollaire 2] assure que la suite $(n^{-2/3} \mathcal{C}_n^t)$ converge en distribution vers $(|\gamma_j|)$ pour la topologie de l_{\searrow}^2 .

Lorsque $t = 0$, centre de la fenêtre critique, ce résultat peut s'interpréter de manière heuristique en appliquant le Théorème 1.4 à la loi ν de Poisson de paramètre 1¹² : on voit alors que $G(n, 1/n)$ ressemble à \mathcal{SG}_n^ν . D'ailleurs, $\mu = \beta = 1$; R^ν et R^0 ont même loi.

Nous allons voir que la situation est bien différente lorsque ν a un moment d'ordre 3 infini.

Le cas d'une loi de puissance

Nous supposons ici que ν , au lieu de vérifier (1.13), satisfait :

$$\sum_{k=1}^{\infty} k(k-2)\nu_k = 0 \quad \text{et} \quad \nu_k \underset{k \rightarrow \infty}{\sim} ck^{-\gamma}, \quad \text{où } c > 0 \text{ et } \gamma \in]3, 4[. \quad (1.14)$$

Nous allons voir que, correctement remise à l'échelle, la suite (\mathcal{C}_n^ν) converge aussi en loi vers la suite réordonnée des longueurs des excursions du processus réfléchi d'un certain processus à accroissements indépendants. Ce processus n'est toutefois pas à accroissements stationnaires.

Posons tout d'abord

$$A^\nu : t \mapsto -\frac{c \Gamma(4-\gamma)}{(\gamma-3)(\gamma-2)\mu^{\gamma-2}} t^{\gamma-2}.$$

Soit d'autre part X^ν l'unique processus à accroissements indépendants tel que pour tous $t \geq 0$ et $u \in \mathbb{R}$,

$$\mathbb{E}[\exp\{iuX^\nu(t)\}] = \exp\left(\int_0^t ds \int_0^\infty dx (e^{iux} - 1 - iux) \frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xs/\mu}\right).$$

¹² ν n'est alors pas une mesure de probabilité sur \mathbb{N}^* puisque $\nu_0 > 0$. On peut toutefois adapter sans difficulté le Théorème 1.4 de sorte que les résultats énoncés restent valides.

Posons pour finir

$$W^\nu = X^\nu + A^\nu$$

et R^ν le processus réfléchi défini par

$$R^\nu : t \geq 0 \mapsto W^\nu(t) - \inf_{0 \leq s \leq t} W^\nu(s).$$

On définit les excursions et les longueurs d'intervalle comme pour le cas du moment d'ordre 3 fini.

Notre principal résultat dans le cas où ν est une loi de puissance s'énonce comme suit.

Théorème 1.5. *Supposons que la mesure de probabilité ν vérifie (1.14). Alors, p.s., l'ensemble des excursions de R^ν peut s'écrire $\{\gamma_j, j \geq 1\}$ de sorte que la suite des longueurs $(|\gamma_j|)$ appartient à l_{\searrow}^2 . D'autre part*

$$n^{-(\gamma-2)/(\gamma-1)} \mathbf{C}_n^\nu \xrightarrow[n \rightarrow \infty]{(d)} (|\gamma_j|, j \geq 1)$$

pour la topologie de l_{\searrow}^2 .

1.6.2 Techniques de preuves

Décrivons comment nous avons obtenu les Théorèmes 1.3 et 1.5. Que la mesure ν ait un moment d'ordre 3 fini ou qu'elle soit une loi de puissance, la méthode sera la même. Nous allons construire une suite de marches aléatoires $(\overline{W}_n^\nu)_{n \geq 2}$ telle que, pour tout $n \geq 2$, les longueurs des excursions de \overline{W}_n^ν au-dessus de ses minima passés correspondent aux tailles renormalisées des composantes de \mathcal{G}_n^ν . Nous verrons que cette suite $(\overline{W}_n^\nu)_{n \geq 2}$ converge en loi vers le processus aléatoire W^ν . Il s'agira enfin de montrer que les longueurs des excursions au-dessus des minima passés de W^ν (c'est-à-dire les longueurs des excursions de R^ν) correspondent à la limite des tailles renormalisées des composantes de \mathcal{G}_n^ν lorsque n tend vers l'infini.

Nous décrirons dans un dernier paragraphe la méthode ayant permis de démontrer le Théorème 1.4.

La recherche en profondeur

Commençons par donner une manière algorithmique pour construire le multigraphe \mathcal{G}_n^ν . Nous verrons ensuite que la marche aléatoire \overline{W}_n^ν se définit aisément à partir de cette construction.

On suppose que $\sum_{i=1}^n D_i$ est paire. On partitionne l'ensemble des demi-arêtes en trois parties : l'ensemble \mathcal{S} des demi-arêtes endormies, l'ensemble \mathcal{A} des demi-arêtes actives et l'ensemble \mathcal{D} des demi-arêtes mortes. La partie $\mathcal{S} \cup \mathcal{A}$ constitue l'ensemble des demi-arêtes vivantes. Initialement toutes les demi-arêtes sont endormies.

On tire alors une demi-arête endormie uniformément au hasard. On note v_1 le sommet auquel elle est attachée. On déclare actives toutes les demi-arêtes attachées à v_1 . Tant que \mathcal{A} n'est pas vide, on procède comme suit.

- On note i le plus grand entier k tel qu'il existe une demi-arête active attachée à v_k .

- On considère une demi-arête active l attachée à v_i .
- On tue la demi-arête l , c'est-dire qu'on la retire de l'ensemble $\mathcal{S} \cup \mathcal{A}$ et on la place dans l'ensemble \mathcal{D} .
- On apparie l à une demi-arête vivante r choisie uniformément au hasard.
- Si r est endormie, on note v_{j+1} le sommet auquel elle est attachée, où j est le nombre de sommets déjà découverts avant que la demi-arête r ne fût trouvée. On déclare actives toutes les demi-arêtes de v_{j+1} autres que r .
- On tue r .

On itère ce procédé jusqu'à ce que l'ensemble \mathcal{A} soit vide. La première composante connexe de \mathcal{G}_n^ν est alors totalement explorée. S'il reste des arêtes endormies, on réitère le même algorithme avec ces demi-arêtes. À la fin (c'est-à-dire lorsque toutes les demi-arêtes sont mortes), le multigraphe \mathcal{G}_n^ν est construit. La recherche en profondeur est illustrée dans le premier dessin de la Figure 1.9.

La marche aléatoire \overline{W}_n^ν

Voyons maintenant comment définir la marche aléatoire \overline{W}_n^ν . On commence par définir une première marche aléatoire non renormalisée W_n^ν définie sur $\{0, \dots, n\}$. Introduisons en premier lieu la notion de demi-arêtes cycles :

Définition 1.1. *On dit qu'une demi-arête l est une demi-arête cycle s'il existe une demi-arête r telle que :*

- l fut tuée avant r ,
- l fut appariée à r ,
- r était active au moment on l'apparia à l .

Notons, pour tout $i \in \{1, \dots, n\}$,

$$\hat{D}_i = \text{degré de } v_i.$$

On définit la marche aléatoire W_n^ν en posant, pour tout $i \in \{1, \dots, n\}$,

$$W_n^\nu(i) = \sum_{j=1}^i \left(\hat{D}_j - 2 - 2\# \{ \text{demi-arêtes cycles attachées à } v_j \} \right).$$

En notant $(\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_m)$ la suite des tailles des composantes de \mathcal{G}_n^ν ordonnées selon leur apparition au cours de la recherche en profondeur, on peut alors montrer qu'on a pour tout entier $j \in \{1, \dots, m\}$:

$$\mathcal{C}_j = \min\{i : W_n^\nu(i) = -2j\} - \min\{i : W_n^\nu(i) = -2(j-1)\}. \quad (1.15)$$

Aussi la connaissance de W_n^ν permet-elle de retrouver les tailles des composantes de \mathcal{G}_n^ν . Une réalisation de W_n^ν est donnée dans le second dessin de la Figure 1.9.

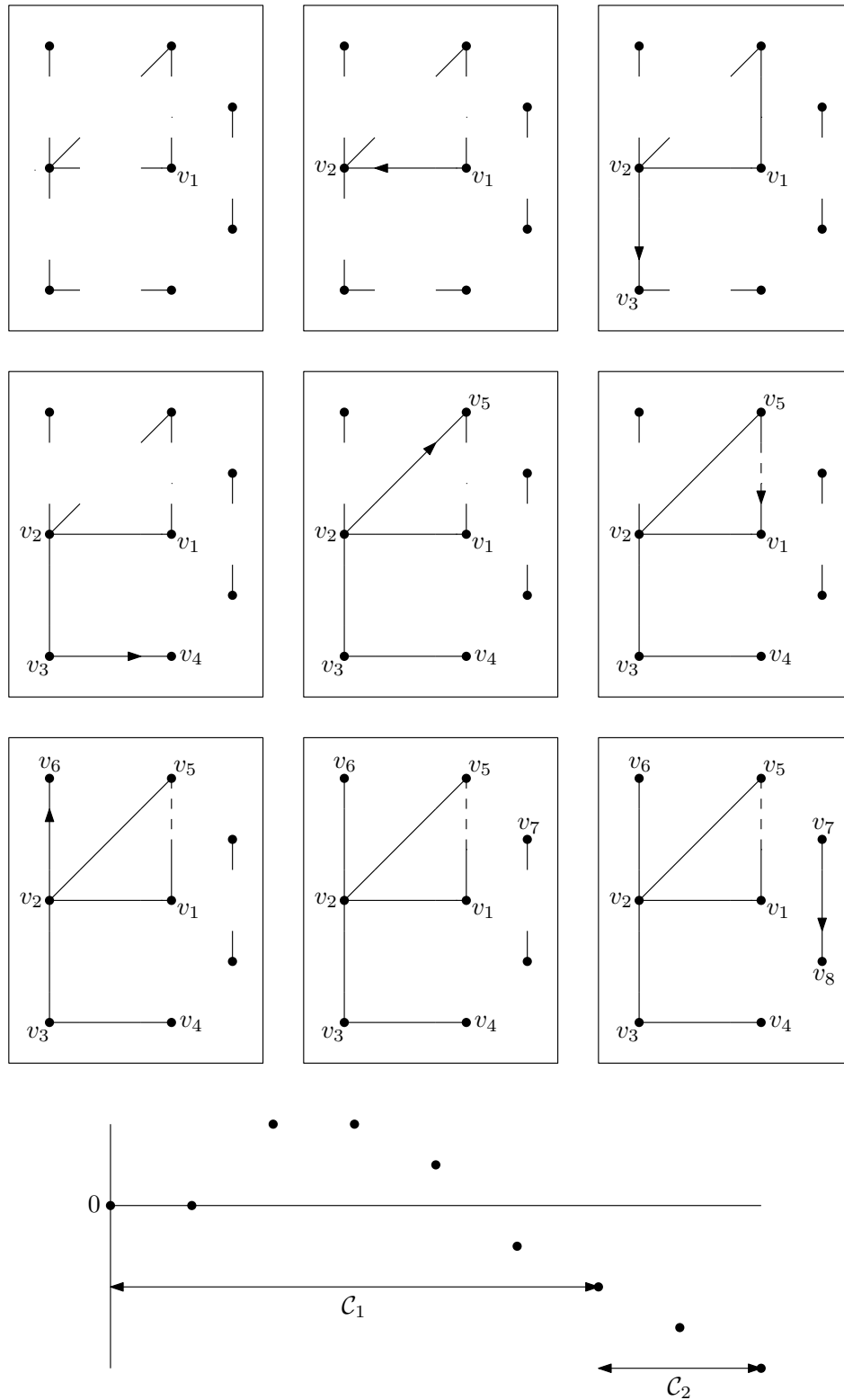


FIGURE 1.9 – Une réalisation de la recherche en profondeur pour $n = 8$, accompagnée de la marche aléatoire W_8^ν associée. L'une des demi-arêtes du sommet v_5 est une demi-arête cycle; on l'a représentée par un trait pointillé.

Un résultat clé

Si ν a un moment d'ordre 3 fini, posons $\gamma = 4$. Rappelons que si ν est une loi de puissance, $\gamma \in]3, 4[$ désigne son exposant.

Le résultat intermédiaire suivant constituera une étape clé dans les preuves des Théorèmes 1.3 et 1.5.

Proposition 1.4. *Supposons que la mesure de probabilité ν vérifie (1.13) ou (1.14). Renormalisant la marche aléatoire W_n^ν en définissant, pour tout $t \in [0, n^{1/(\gamma-1)}]$,*

$$\overline{W}_n^\nu(t) = n^{-1/(\gamma-1)} W_n^\nu(\lfloor tn^{(\gamma-2)/(\gamma-1)} \rfloor),$$

on a alors : $\overline{W}_n^\nu \xrightarrow[n \rightarrow \infty]{(d)} W^\nu$.

Le fait que cette proposition implique la convergence des suites des tailles des composantes paraît intuitivement clair d'après (1.15). La preuve rigoureuse de cette implication s'appuie sur les techniques développées par Aldous dans [Ald97] dans le cas où ν a un moment d'ordre 3 fini et par Aldous et Limic dans [AL98] dans le cas où ν est une loi de puissance.

Nous poursuivons notre discussion en mentionnant la ligne directrice permettant de démontrer la Proposition 1.4.

Outils pour prouver la Proposition 1.4

Observons que pour obtenir la Proposition 1.4, il suffit de connaître la marche W_n^ν jusqu'à l'instant $\lfloor t_0 n^{(\gamma-2)/(\gamma-1)} \rfloor$ pour tout $t_0 > 0$ fixé. Or, nous verrons que l'espérance du nombre de demi-arêtes cycles liées à l'un des $\lfloor t_0 n^{(\gamma-2)/(\gamma-1)} \rfloor$ premiers sommets découverts lors de la recherche en profondeur est bornée. La présence du facteur de renormalisation $n^{-1/(\gamma-1)}$ définissant \overline{W}_n^ν nous permet donc de réduire notre étude à celle, plus simple, de la marche aléatoire $(\sum_{j=1}^i (\hat{D}_j - 2), 0 \leq i \leq n)$ jusqu'à l'instant $\lfloor t_0 n^{(\gamma-2)/(\gamma-1)} \rfloor$.

Il est aisé de voir que la recherche en profondeur que nous avons décrite plus haut ordonne les sommets du multigraphe \mathcal{G}_n^ν selon un ordre biaisé par le degré. Puisque seules les suites du type $(\hat{D}_1, \dots, \hat{D}_{\lfloor t_0 n^{(\gamma-2)/(\gamma-1)} \rfloor})$ nous intéressent, il suffit de donner une méthode simple permettant de construire de telles suites. La technique que nous développerons s'appuie sur la remarque suivante. Pour ordonner un ensemble de réels strictement positifs $\{d_1, \dots, d_n\}$ selon un biais par la taille, il suffit de créer une suite d'horloges exponentielles indépendantes de paramètres d_i , $1 \leq i \leq n$. On ordonne ensuite les n réels selon les instants où les horloges qui leur sont associées retentissent.

Quitte à étudier des multigraphes ayant non plus exactement n sommets mais \mathcal{P}_n sommets, où \mathcal{P}_n est variable de Poisson de paramètre n , nous allons considérer un processus ponctuel de Poisson Π_n^0 sur $\mathbb{N}^* \times \mathbb{R}_+$ d'intensité $n\nu_k k e^{-ks} ds$. L'abscisse k d'un atome correspond au degré du sommet qu'il code tandis que son ordonnée s indique l'instant où l'horloge sonne. Il suffit ensuite d'ordonner la suite des sommets selon les dates d'apparition des atomes. Rappelons que seuls les $\lfloor t_0 n^{(\gamma-2)/(\gamma-1)} \rfloor$ premiers atomes nous intéressent, travailler avec le multigraphe issu de ce processus de Poisson nous convient.

Cependant, alors que la marche $(\sum_{j=1}^i (\hat{D}_j - 2), 0 \leq i \leq n)$ nous révèle la suite des degrés des sommets découverts $(\hat{D}_1, \dots, \hat{D}_n)$ à chaque instant entier jusqu'à l'instant n , les atomes de Π_n^0 ne sont pas découverts à taux 1 jusqu'à l'instant n ¹³. De manière informelle, il s'agit donc d'accélérer le temps; on remplace le temps s en un temps $\psi_n(s)$, où $\psi_n : [0, n[\rightarrow \mathbb{R}_+$ est une fonction strictement croissante et convexe telle que $\lim_{s \rightarrow n} \psi_n(s) = \infty$, de sorte que les atomes du processus ponctuel de Poisson Π_n sur $\mathbb{N}^* \times [0, n[$ d'intensité $\pi_n(k, ds) = n\nu_k k e^{-k\psi_n(s)} \psi_n'(s) ds$ apparaissent à taux 1 jusqu'à l'instant n . Cela revient à demander que pour tout $s \in [0, n[$, $\sum_{k \in \mathbb{N}^*} n\nu_k k e^{-k\psi_n(s)} \psi_n'(s) = 1$, c'est-à-dire, en posant $\phi : t \geq 0 \mapsto 1 - \sum_{k \in \mathbb{N}^*} \nu_k e^{-kt}$, que $\psi_n' \cdot \phi \circ \psi_n = 1/n$. Nous voyons que $\psi_n : s \in [0, n[\mapsto \phi^{-1}(s/n)$ convient.

Finalement, en notant Π_n le processus ponctuel de Poisson sur $\mathbb{N}^* \times [0, n[$ d'intensité $\pi_n(k, ds) = \nu_k k e^{-k\phi^{-1}(s/n)} (\phi^{-1})'(s/n) ds$, nous étudierons le processus $(S_n^\nu(t))_{t \geq 0}$ défini par :

$$S_n^\nu : t \mapsto \sum_{(k,s) \in \Pi_n} (k-2) \mathbf{1}_{s \leq t},$$

à comparer avec

$$i \mapsto \sum_{j \geq 1} (\hat{D}_j - 2) \mathbf{1}_{j \leq i}.$$

Il sera aisé de voir que, en posant $\bar{S}_n^\nu : t \geq 0 \mapsto n^{-1/(\gamma-1)} S_n^\nu(t n^{(\gamma-2)/(\gamma-1)})$, les marches aléatoires \bar{W}_n^ν et \bar{S}_n^ν sont très proches. Tout revient finalement à obtenir le résultat suivant.

Proposition 1.5. *Supposons que la mesure de probabilité ν vérifie (1.13) ou (1.14). On a alors la convergence suivante : $\bar{S}_n^\nu \xrightarrow[n \rightarrow \infty]{(d)} W^\nu$.*

Évoquons les outils mis en œuvre pour prouver la Proposition 1.5.

Esquisse de la preuve de la Proposition 1.5

Commençons par donner de nouvelles notations. On définit la fonction continue à variations bornées déterministe suivante

$$A_n^\nu : t \mapsto \int \pi_n(k, ds) (k-2) \mathbf{1}_{s \leq t},$$

de sorte que le processus M_n^ν défini par $M_n^\nu = S_n^\nu - A_n^\nu$ est une martingale. On définit comme précédemment \bar{A}_n^ν et \bar{M}_n^ν . Enfin, dans le cas où ν a un moment d'ordre 3 fini, on pose $A^\nu : t \geq 0 \mapsto -\frac{\beta}{2\mu^2} t^2$ et $X^\nu : t \geq 0 \mapsto \sqrt{\frac{\beta}{\mu}} W(t)$ ¹⁴.

La preuve de la Proposition 1.5 nécessite de prouver les deux résultats suivants : d'une part que la suite (\bar{A}_n^ν) converge uniformément sur tout compact vers A^ν (ce qui sera aisé), et d'autre part que la suite de martingales (\bar{M}_n^ν) converge en distribution vers X^ν . La preuve de ce dernier point constitue le cœur des démonstrations des Théorèmes 1.3 et 1.5.

¹³À cause de la décroissance de la densité de la loi exponentielle, de moins en moins d'atomes sont découverts au fur et à mesure que le temps s passe.

¹⁴Lorsque ν est une loi de puissance, les processus A^ν et X^ν ont été définis au paragraphe 1.6.1.

Dans le cas où ν a un moment d'ordre 3 fini, la convergence de (\overline{M}_n^ν) vers X^ν se justifie en reprenant les idées suivies par Aldous dans [Ald97]. Il suffit d'appliquer le théorème central limite pour les martingales en temps continu (voir par exemple [EK86]). Lorsque ν est une loi de puissance, on appliquera les outils présentés par Jacod et Shiryaev dans [JS03] permettant de prouver la convergence de processus à accroissements indépendants.

Éléments de preuve du Théorème 1.4

Terminons cette introduction en esquissant la preuve du Théorème 1.4. Nous supposons que ν a un moment d'ordre 3 fini. Nous verrons qu'alors, pour tout $t > 0$ fixé, a.p.s., aucun des $tn^{2/3}$ premiers sommets de \mathcal{G}_n^ν découverts au cours de la recherche en profondeur ne contient de boucle ni ne crée d'arête multiple. De manière heuristique, cela indique que savoir si le multigraphe \mathcal{G}_n^ν est simple ne dépend asymptotiquement pas de la connaissance de ces $tn^{2/3}$ sommets.

D'après la Proposition 1.4 et [Ald97], étudier la convergence en loi de $(n^{-2/3}\mathcal{C}_n^\nu)$ revient à connaître celle de (\overline{W}_n^ν) . La loi de \overline{W}_n^ν étant asymptotiquement caractérisée par les degrés des $tn^{2/3}$ premiers sommets (pour tout $t > 0$ fixé), elle ne dépend pas de l'événement $\{\mathcal{G}_n^\nu \text{ est simple}\}$, et la loi de \overline{W}_n^ν conditionnellement à l'événement $\{\mathcal{G}_n^\nu \text{ est simple}\}$ est, pour n grand, celle de \overline{W}_n^ν . Or, toujours d'après [Ald97], la convergence en loi de $(n^{-2/3}\mathcal{SC}_n^\nu)$ est donnée par celle de la suite des marches aléatoires associées, c'est-à-dire ici les marches \overline{W}_n^ν conditionnées à $\{\mathcal{G}_n^\nu \text{ est simple}\}$, dont les lois sont, comme nous venons de le voir, proches de celles des marches \overline{W}_n^ν . Nous en déduisons finalement que $(n^{-2/3}\mathcal{SC}_n^\nu)$ converge en loi vers la même limite que $(n^{-2/3}\mathcal{C}_n^\nu)$.

CHAPTER 2

A PHASE TRANSITION FOR THE HEIGHTS OF A FRAGMENTATION TREE

We provide information about the asymptotic regimes for a homogeneous fragmentation of a finite set. We establish a phase transition for the asymptotic behavior of the shattering times, defined as the first instants when all the blocks of the partition process have cardinality less than a fixed integer. Our results may be applied to the study of certain random split trees.

Contents

2.1	Introduction	36
2.2	Formulation of the main results	38
2.3	Preliminaries	43
2.3.1	Some results on branching random walks	43
2.3.2	Poissonization	45
2.4	Study of the heights	46
2.4.1	Upper bound	46
2.4.2	Lower bound	49
2.5	Study of the saturation levels	54
2.5.1	Lower bound	55
2.5.2	Upper bound	57
2.6	An explanation for the phase transition	59

2.1 Introduction

The occupancy scheme is a simple urn model which often occurs in probability, statistics, combinatorics and computer science. We can cite for example species sampling [BF93, HG00], analysis of algorithms [Gar02], learning theory [BG97], etc. The books by Johnson and Kotz [JK77] and by Kolchin *et al.* [KSC78] are standard references. The present work is partly motivated by the study of fragmentation trees.

Let us recall the occupancy scheme. Let \mathcal{I} be a countable set and $\mathbf{p} = (p_i : i \in \mathcal{I})$ be a probability measure on \mathcal{I} . The occupancy scheme on $(\mathcal{I}, \mathbf{p})$ is described as follows. For all $i \in \mathcal{I}$ such that $p_i \neq 0$, one places a box at i . One then throws successively and independently n balls in the boxes by assuming that each ball has probability p_i of falling into the box located at i (p_i should be seen as the size of that box). We may be interested in the number of boxes containing exactly j balls, or in the number of occupied boxes, etc.

We consider here a variant of the occupancy scheme which corresponds to a nested family of boxes. We introduce the infinite genealogical tree \mathcal{U} :

$$\mathcal{U} = \bigcup_{k=0}^{\infty} \mathbb{N}^k,$$

with $\mathbb{N} = \{1, 2, \dots\}$ and the convention $\mathbb{N}^0 = \{\emptyset\}$. The elements of \mathcal{U} are called individuals, and for all integers $k \in \mathbb{Z}_+$, the k -th generation of \mathcal{U} is formed by the individuals $i \in \mathbb{N}^k$ (we write $|i| = k$). For every individual $i = (i_1, \dots, i_k)$ of \mathcal{U} and $j \in \mathbb{N}$, the individual $ij = (i_1, \dots, i_k, j)$ is called the j -th child of i and i is the parent of ij . We suppose a real number $p_i \in [0, 1]$ is assigned to each individual i , such that $p_{\emptyset} = 1$ and $\sum_{j \in \mathbb{N}} p_{ij} = p_i$ for all $i \in \mathcal{U}$. Note that for all k , $\mathbf{p}(k) = (p_i : |i| = k)$ is a probability measure on \mathbb{N}^k . We can couple the occupancy schemes on $(\mathbb{N}^k, \mathbf{p}(k))$ as follows. Each individual i should be viewed as a box of size p_i (provided that $p_i \neq 0$), such that the box ij is contained into the box i (if $p_{ij} \neq 0$). Initially the n balls are thrown in the box located at \emptyset . Then one places successively and independently the n balls in the boxes of the first generation by assuming that each ball has probability p_i of falling into the box situated at $i \in \mathbb{N}$. Likewise, by iteration, a ball located in the box i is placed independently of the others into the sub-box ij with probability p_{ij}/p_i .

We denote the first generation at which all the boxes have less than $j \geq 2$ balls when n have been thrown by $H_{n,j}$ and the first level at which there exists a box containing less than $j \geq 1$ balls by $G_{n,j}$. $H_{n,j}$ is called a *height* and $G_{n,j}$ is called a *saturation level*. Our aim is to study the asymptotic behavior of $H_{n,j}$ and $G_{n,j}$ as n tends to infinity when we consider a certain *randomized* version of $(p_i, i \in \mathcal{U})$.

More precisely, in the present work, we shall assume that we are given a random probability measure $\rho = (\rho_1, \rho_2, \dots)$ on \mathbb{N} . We assign to each individual i an independent copy $\rho(i)$ of ρ . The *random* numbers p_i , $i \in \mathcal{U}$, are defined by induction: $p_{\emptyset} = 1$ and $p_{ij} = p_i \rho_j(i)$, for all $i \in \mathcal{U}$ and $j \in \mathbb{N}$. It means that $\rho(i)$ describes how the mass of the individual i is split to its children. This model is called the *occupancy scheme of multiplicative cascades*. We put emphasis on the fact that the numbers p_i are random. Once we are given a realization of the numbers $p_i(\omega)$, $i \in \mathcal{U}$, the balls are thrown in the boxes according to that realization. There are thus two levels of randomness in our model, namely the arrangement of boxes and the way one throws balls.

In the particular case when ρ is supported by a finite number of integers, in the sense that $\#\{j : \rho_j > 0\} \leq b$ a.s. for some integer $b \geq 2$, the height $H_{n,j}$ has a natural interpretation in terms of a special class of random split trees which have been considered, *e.g.*, by Devroye [Dev99]. Specifically, imagine that each box has a rupture threshold of j , in the sense that when a ball falls into some box i already containing $j-1$ balls, then this box is removed and the j balls are shared out amongst the children of i according to the random probability $\rho(i)$ (*i.e.*, conditionally on $\rho(i)$, each ball is put in the box ij with probability $\rho_j(i)$, independently of the other balls). This procedure yields a random tree where all balls are stored at leaves, and $H_{n,j}$ is the height of this tree when n balls have been thrown.

We further point out that the height $H_{n,j}$ also arises as a natural shattering time in homogeneous fragmentation chains, a class of partition-valued Markov chains. More precisely, the shattering time is defined as the first instant when all the blocks of the partition process have cardinality less than j . See [Ber06] for background and Section 3.2 in [Ber08] for a description which is closely related to the present work. In a different direction, we mention the work of Haas *et al.* [HMPW08] who associate another random tree to homogeneous fragmentation processes.

We shall show the following result: there exist an integer $\tau \in \{2, 3, \dots\} \cup \{\infty\}$ and a sequence of positive real numbers $C_2 > \dots > C_\tau = C_{\tau+1} = \dots =: C^*$ such that for all integers $j \geq 2$,

$$H_{n,j} \underset{n \rightarrow \infty}{\sim} C_j \ln n \quad \text{a.s.}$$

It should not be surprising that the heights $H_{n,j}$ have logarithmic asymptotics, as a large class of trees have such a behavior, like random split trees. On the other hand, it is remarkable that there exists a critical parameter $\tau \in \{2, 3, \dots\} \cup \{\infty\}$ at which a phase transition occurs (provided that $2 < \tau < \infty$). It is a peculiar behavior for a random split tree. Devroye indeed proved in [Dev99] that for a large class of random split trees such that their internal nodes contain at least one ball¹, then all the heights, regardless of the value of the rupture threshold j , have the same asymptotic behavior (in probability). In particular, no phase transition occurs.

It may also be interesting to consider the situation where the parameter j depends on the number of balls n . In the case of power functions, we shall prove that

$$H_{n,n^\alpha} \underset{n \rightarrow \infty}{\sim} (1 - \alpha)C^* \ln n \quad \text{a.s.},$$

for all $\alpha \in (0, 1)$, so that there is no phase transition in the asymptotics of H_{n,n^α} .

Concerning the asymptotics of the saturation level $G_{n,j}$, we shall prove the following result: under some technical conditions that will be detailed below, there exists a constant $C_* < C^*$ such that for all positive integers j ,

$$G_{n,j} \underset{n \rightarrow \infty}{\sim} C_* \ln n \quad \text{a.s.}$$

In particular, there is no phase transition in the asymptotics of $G_{n,j}$. We shall also show that under some further technical conditions, for all $\alpha \in (0, 1)$,

$$G_{n,n^\alpha} \underset{n \rightarrow \infty}{\sim} (1 - \alpha)C_* \ln n \quad \text{a.s.},$$

¹We stress that this assumption is crucial in the proof of Theorem 1 in [Dev99] and seems to have been overlooked in the statement.

so that there is also no phase transition in the asymptotic behavior of G_{n,n^α} .

Our approach essentially relies on the theory of branching random walks, and more precisely on their large deviations behaviors whose descriptions are due to Biggins [Big92]. The construction of the boxes given by the multiplicative cascades indeed enables us to define a branching random walk giving the sizes of the boxes at each generation. We shall see that the critical parameter τ and the real numbers C_2, \dots, C_τ and C_* are described by that branching random walk. Another key technique is Poissonization; instead of throwing exactly n balls, one throws \mathcal{P}_n balls, where \mathcal{P}_n is a Poisson variable with parameter n which is independent of $(\rho(i), i \in \mathcal{U})$. For all integers k , conditionally on the sizes of the boxes of the k -th generation, the numbers of balls per box of the k -th generation are thus independent Poisson variables.

The results will be stated in Section 2.2. We shall see the main techniques in Section 2.3. Section 2.4 will be devoted to the study of the heights. We shall first turn our attention to the upper bound. Due to the phase transition, we shall give two different proofs to show the lower bound of $H_{n,j}$, depending on whether the integer j is less than the critical parameter τ or not. The results on the saturation levels will be proved in Section 2.5. Finally, in Section 2.6, we shall explain heuristically why a phase transition may occur in the asymptotics of $H_{n,j}$, but not in those of H_{n,n^α} , $G_{n,j}$ and G_{n,n^α} .

2.2 Formulation of the main results

Recall that ρ is a random probability measure on \mathbb{N} . We denote its law by ν . If we denote by $\text{Prob}_{\mathbb{N}}$ the space of probability measures on \mathbb{N} , ν is a probability measure on $\text{Prob}_{\mathbb{N}}$, called the splitting law. We exclude the degenerate case when ρ is a Dirac point mass a.s.

As explained in the introduction, we consider a family $(\rho(i), i \in \mathcal{U})$ of independent copies of ρ labeled by the individuals of the genealogical tree \mathcal{U} . The multiplicative cascade construction defines for each generation k a probability measure $(p_i, |i| = k)$ on \mathbb{N}^k . Taking logarithm of masses, we may encode the latter by the following random point measure on \mathbb{R}_+

$$Z^{(k)}(\mathrm{d}y) = \sum_{i:|i|=k} \delta_{-\ln p_i}(\mathrm{d}y),$$

where δ_z stands for the Dirac point mass at z . Note that if $p_i = 0$, *i.e.*, if there is no box at i , the individual i is omitted in the sum defining $Z^{(k)}$. Likewise, in the sequel, we shall always consider individuals which have a positive mass. It follows immediately from the structure of the multiplicative cascades that $(Z^{(k)}, k \in \mathbb{Z}_+)$ is a branching random walk, in the sense that for all integers $k, k' \geq 0$, $Z^{(k+k')}$ is obtained from $Z^{(k)}$ by replacing each atom z of $Z^{(k)}$ by a family $\{z + y, y \in \mathcal{Y}_z\}$, where each \mathcal{Y}_z is an independent copy of the family of atoms of $Z^{(k')}$.

Let us introduce quantities defined via the splitting law ν . First, we define the Laplace transform of the intensity measure $Z^{(1)}$ by

$$L(\theta) = \mathbb{E} [\langle Z^{(1)}, e^{-\theta \cdot} \rangle]$$

for $\theta \in \mathbb{R}$. We can also write

$$L(\theta) = \mathbb{E} \left[\sum_{j \in \mathbb{N}} \rho_j^\theta \right] = \int_{\text{Prob}_{\mathbb{N}}} \left(\sum_{j \in \mathbb{N}} p_j^\theta \right) \nu(d\mathbf{p})$$

with the convention that $p^\theta = 0$ when $p = 0$ even when $\theta \leq 0$. Because ρ is not a Dirac point mass a.s., $L(0) > 1$. The function $L : \mathbb{R} \rightarrow (0, \infty]$ is decreasing with $L(1) = 1$. We define

$$\underline{\theta} = \inf\{\theta \in \mathbb{R} : L(\theta) < \infty\},$$

so that $L(\theta) < \infty$ when $\theta > \underline{\theta}$. Note that $\underline{\theta}$ may be negative. We can show from Hölder's inequality that $\ln L$ is a convex function, which implies the convexity of L and that the function

$$\varphi : \theta \mapsto \ln L(\theta) - \theta \frac{L'(\theta)}{L(\theta)}$$

is increasing on $(\underline{\theta}, 0)$ and decreasing on $(0, \infty)$. As $L(1) = 1$ and L decreases, we have $\varphi(1) = -L'(1) > 0$, and thus the set of $\theta \in (\underline{\theta}, \infty)$ such that $\varphi(\theta) > 0$ is a non-empty open interval (θ_*, θ^*) :

$$\theta_* = \inf\{\theta > \underline{\theta} : \varphi(\theta) > 0\} \quad \text{and} \quad \theta^* = \sup\{\theta > \underline{\theta} : \varphi(\theta) > 0\}.$$

Note that $\theta^* > 1$ and $\theta_* < 0$ if $\underline{\theta} < 0$. Let us consider the example of beta-splitting laws.

Example Let us describe the situation when $\rho = (B, 1 - B, 0, \dots)$, where B has a Beta(a, b) distribution. Easy calculations yield:

$$L(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)} \frac{\Gamma(a+\theta)}{\Gamma(a+b+\theta)} + \frac{\Gamma(a+b)}{\Gamma(b)} \frac{\Gamma(b+\theta)}{\Gamma(a+b+\theta)}.$$

Let $\theta^*(a, b)$ denote the parameter θ^* . Because this critical parameter is only defined implicitly as the solution of an equation involving the derivative of the gamma function, there is no simple explicit expression in terms of the parameter (a, b) . However numerical estimates can easily be obtained, and are represented in Figure 2.1 below in the special case $a = b$.

The convexity of $\ln L$ implies that $\frac{L}{-L'}$ is an increasing function. We define

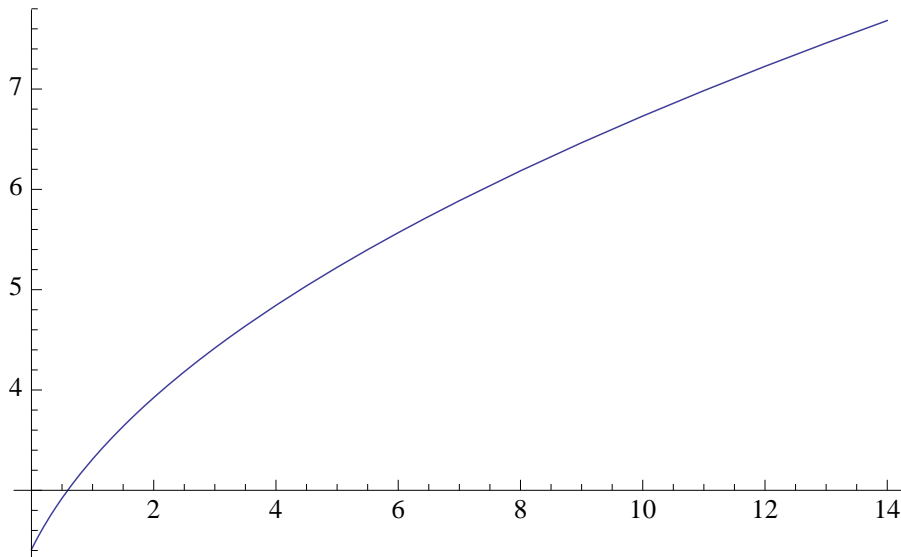
$$C_* = \lim_{\substack{\theta \rightarrow \theta_* \\ \theta > \theta_*}} \frac{L(\theta)}{-L'(\theta)} \quad \text{and} \quad C^* = \lim_{\substack{\theta \rightarrow \theta^* \\ \theta < \theta^*}} \frac{L(\theta)}{-L'(\theta)}.$$

Inspired by the article of Hu and Shi in [HS09] (see Lemma 2.4 below), we shall sometimes need the following assumption: there exists $\delta > 0$ such that

$$L(-\delta) < \infty \quad \text{and} \quad \int_{\text{Prob}_{\mathbb{N}}} \left(\sum_{i \in \mathbb{N}} \mathbf{1}_{p_i > 0} \right)^{1+\delta} \nu(d\mathbf{p}) < \infty. \quad (2.1)$$

To study the asymptotics of the saturation levels, we shall sometimes need the following hypothesis:

$$-\infty < \theta_* < 0 \quad \text{and} \quad \varphi(\theta_*) = 0. \quad (2.2)$$

Figure 2.1: graph of the function $a \mapsto \theta^*(a, a)$

We can now state the results that we shall prove. Concerning the asymptotic behavior of the heights $H_{n,j}$, we have the following results which complete and improve Proposition 2 in [Ber08].

Theorem 2.1. *Let $j \geq 2$ be an integer, set*

$$C_j = \begin{cases} \frac{j}{-\ln L(j)} & \text{if } j < \theta^*, \\ C^* & \text{if } j \geq \theta^*. \end{cases}$$

Then

$$H_{n,j} \sim C_j \ln n \quad \text{a.s.}$$

Furthermore, if $j < \theta^*$ or if (2.1) holds, then

$$H_{n,j} = C_j \ln n + O(\ln \ln n) \quad \text{a.s.}$$

In fact, the following result always holds:

$$H_{n,j} \leq C_j \ln n + O(\ln \ln n) \quad \text{a.s.}$$

We see that there is a phase transition in the asymptotic behavior of $H_{n,j}$ at the integer $\lceil \theta^* \rceil$ when $2 < \theta^* < \infty$ ($\lceil \theta^* \rceil$ is the cut-off parameter denoted by τ in the introduction). We point out that it may happen that $\theta^* \leq 2$ or $\theta^* = \infty$, in which case there is no phase transition. For instance, one can show that if $\rho = (\rho_1, \rho_2, \dots)$ with $\rho_1 = 1 - 0.75U$, $\rho_j = 0.05U$ for all $j \in 2, \dots, 16$ and $\rho_j = 0$ for all $j \geq 17$, where U is uniformly distributed on $(0, 1)$, then $\theta^* < 1.99$. On the other hand, for all $\alpha \in [\frac{1}{2}, 1)$, if ρ is equal to $(\frac{1}{2}, \frac{1}{2}, 0, \dots)$ with probability α and to $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, \dots)$ with probability $1 - \alpha$, then $\theta^* = \infty$.

Remark 2.1. The critical parameter θ^* is finite whenever

$$\left\| \max_{j \in \mathbb{N}} \rho_j \right\|_\infty = 1,$$

where $\rho = (\rho_j, j \in \mathbb{N})$ denotes a random probability measure on \mathbb{N} with law ν (recall that $\|\cdot\|_\infty$ is the essential supremum of the r. v. $\max_{j \in \mathbb{N}} \rho_j$). Indeed, it is easily seen that

$$\lim_{\theta \rightarrow \infty} L(\theta)^{\frac{1}{\theta}} = \|\max_{j \in \mathbb{N}} \rho_j\|_\infty.$$

If the right-hand side equals 1, then $g : \theta \mapsto \frac{-\ln L(\theta)}{\theta}$ has limit 0 at infinity. Now g is differentiable and $g(1) = 0$, so there exists $\theta_0 \in (1, \infty)$ such that $g'(\theta_0) = 0$. As $g'(\theta) = \theta^{-2}\varphi(\theta)$, we conclude that $\theta_0 = \theta^* < \infty$.

Remark 2.2. The parameter θ^* is infinite whenever ρ is not random, *i. e.*, $\rho = (\rho_1, \rho_2, \dots)$ where $\rho_j, j \in \mathbb{N}$, are deterministic real numbers (note that we excluded the case when only one ρ_j is nonzero). Indeed, in that setting, as for all $\theta > 1$, $L(\theta) = \sum_{j=1}^{\infty} \rho_j^\theta$, one can show that

$$\varphi(\theta) = \ln \left(\sum_{j=1}^{\infty} \rho_j^\theta \right) + \theta \frac{\sum_{j=1}^{\infty} -\ln \rho_j \rho_j^\theta}{\sum_{j=1}^{\infty} \rho_j^\theta}.$$

Let f denote the function $x \mapsto -x \ln x$. It is easily seen that $f(\sum_{j=1}^{\infty} \rho_j^\theta) < \sum_{j=1}^{\infty} f(\rho_j^\theta)$, which proves that $\varphi(\theta) > 0$ for all $\theta > 1$. As a result, $\theta^* = \infty$.

We may also be interested in the asymptotics of H_{n,n^α} , where $\alpha \in (0, 1)$. We shall prove the following result.

Proposition 2.1. *Let $\alpha \in (0, 1)$. Then*

$$H_{n,n^\alpha} \sim (1 - \alpha)C^* \ln n \quad a.s.$$

Moreover, if $\theta^* < \infty$ and (2.1) holds, then

$$H_{n,n^\alpha} = (1 - \alpha)C^* \ln n + O(\ln \ln n) \quad a.s.,$$

We shall see that if $\theta^* < \infty$ then

$$H_{n,n^\alpha} \leq (1 - \alpha)C^* \ln n + O(\ln \ln n) \quad a.s.$$

Concerning the asymptotic behavior of the saturation levels, we shall prove the following theorem.

Theorem 2.2. *Let $j \geq 1$ be an integer and $\alpha \in [0, 1)$. Suppose that $\theta_* = -\infty$ or (2.2) holds. Then*

$$G_{n,jn^\alpha} \sim (1 - \alpha)C_* \ln n \quad a.s.$$

Furthermore, if (2.2) and (2.1) hold, then

$$G_{n,jn^\alpha} = (1 - \alpha)C_* \ln n + O(\ln \ln n) \quad a.s.$$

In fact, if (2.2) holds, then

$$G_{n,jn^\alpha} \geq (1 - \alpha)C_* \ln n + O(\ln \ln n) \quad a.s.$$

Remark 2.3. A sufficient condition to guarantee that $-\infty < \theta_* < 0$ and $\varphi(\theta_*) = 0$ is:

$$-\infty < \underline{\theta} < 0 \quad \text{and} \quad \lim_{\theta \rightarrow \underline{\theta}} L(\theta) = \infty.$$

Indeed, suppose this condition fulfilled. Imagine that for all $\theta < 0$, $\varphi(\theta) > 0$. Let $\theta_0 \in (\underline{\theta}, 0)$. As $\theta \mapsto \frac{\ln L(\theta)}{\theta}$ is decreasing (its derivative is $-\theta^{-2}\varphi(\theta) < 0$), we have for all $\theta < \theta_0$:

$$\ln L(\theta) \leq \theta \frac{\ln L(\theta_0)}{\theta_0} \leq \underline{\theta} \frac{\ln L(\theta_0)}{\theta_0},$$

which contradicts $\lim_{\theta \rightarrow \underline{\theta}} L(\theta) = \infty$.

Remark 2.4. The final assertions in Theorem 2.1, Proposition 2.1 and Theorem 2.2 rely on the work of Hu and Shi [HS09]. McDiarmid's setting in [McD95] can however be considered; we can prove that the results stated in Theorem 2.1, Proposition 2.1 and Theorem 2.2 still hold if the assumption (2.1) is replaced by the following:

$$\int_{\text{Prob}_{\mathbb{N}}} \left(\sum_{i \in \mathbb{N}} \mathbf{1}_{p_i > 0} \right)^2 \nu(d\mathbf{p}) < \infty. \quad (2.3)$$

For instance, suppose that $\rho = (\rho_1, 1 - \rho_1, 0, 0, \dots)$, where ρ_1 is a random variable with density $\mathbf{1}_{0 < x < e^{-1}} x^{-1} \ln^{-2} x dx$. Then $\underline{\theta} = 0$, so (2.1) does not hold. Nonetheless, ρ is supported by two integers a.s., so (2.3) holds. As a result, for all $j \geq 2$, $H_{n,j} = C_j \ln n + O(\ln \ln n)$ a.s. Furthermore, as $\|1 - \rho_1\|_{\infty} = 1$, Remark 2.1 ensures that θ^* is finite, so for all $\alpha \in (0, 1)$, $H_{n,n^\alpha} = (1 - \alpha)C^* \ln n + O(\ln \ln n)$ a.s.

For the sake of simplicity, we shall show Theorem 2.1, Proposition 2.1 and Theorem 2.2 only in the setting of [HS09]. The general proof can however be easily carried out.

We now conclude this section by discussing an illustrative example. Consider the case $\rho = (U, 1 - U, 0, 0, \dots)$, where U is uniformly distributed on $[0, 1]$. Then $L(\theta) = \frac{2}{\theta+1}$, so $\underline{\theta} = -1$ and (2.1) holds. According to Remark 2.1 and Remark 2.3, $\theta^* < \infty$ and (2.2) hold. Hence, all our results may be applied. Easy calculations yield $\varphi(\theta) = \ln 2 - \ln(\theta + 1) + \frac{\theta}{\theta+1}$ and $[\theta^*] = 4$, so that there is a phase transition. We can show that $C_2 = \frac{2}{\ln 3 - \ln 2} \approx 4.93260\dots$, $C_3 = \frac{3}{\ln 2} \approx 4.32808\dots$ and that $C^* > C_*$ are the solutions of the equation

$$\ln \left(\frac{2}{c} \right) + \frac{c-1}{c} = 0.$$

Note that $C^* \approx 4.31107\dots$ and $C_* \approx 0.37336\dots$

As ρ is supported by two integers, our model may be interpreted in terms of random split trees; the procedure described in the introduction yields a random tree where all balls are stored at leaves. Let $T_{n,j}$ denote the tree obtained when n balls have been thrown and when the boxes have a rupture threshold of j . We now define another random split tree $\tilde{T}_{n,j}$ also related to our model. We imagine that each box has a rupture threshold of j , but when a ball falls into some box i already containing $j - 1$ balls, it *remains* in that box and the $j - 1$ other balls are shared out amongst the two children of i independently in such a way: each ball choses the left child of i with probability U_i , the right child of i with probability $1 - U_i$, where the r. v. U_i , $i \in \mathcal{U}$,

are independent and uniformly distributed on $[0, 1]$. No other ball is then allowed to be stored at the box i : when a ball falls into i , it is placed into one of the two children of i according to the probability vector $(U_i, 1 - U_i)$, and one has to consider whether the ball stays at that new box or not. Let $\tilde{T}_{n,j}$ denote the tree obtained when n balls have been thrown. Note that each internal node of $\tilde{T}_{n,j}$ has exactly one ball, whereas its leaves have less than j balls. We see that the height $\tilde{H}_{n,j}$ of the tree $\tilde{T}_{n,j}$ is less than or equal to the height $H_{n,j}$ of the tree $T_{n,j}$.

It is remarkable that $\tilde{T}_{n,2}$ has the law of the random binary search tree. Robson [Rob79] was the first to be interested in the height of the random binary search tree. In [Pit84], Pittel studied its height and its saturation level. Devroye proved in [Dev86, Dev87] that the saturation level is asymptotically equivalent to $C_* \ln n$ in probability as in our model, but that the height $\tilde{H}_{n,2}$ is equivalent to $C^* \ln n$ in probability, so that $\tilde{H}_{n,2}$ is not equivalent to $H_{n,2}$. Regarding the random binary search tree as a random split tree, he proved in [Dev99] that all the heights $\tilde{H}_{n,j}$ (regardless of the value of j) of the random binary search tree are asymptotically equivalent to $C^* \ln n$ in probability. In particular, there is no phase transition, contrary to our case. There is of course no contradiction, as the random binary search tree does not correspond to a model treated in this work. Indeed, in terms of random split trees, internal nodes of the binary search tree retain exactly one ball whereas all the balls are stored at leaves in our case.

2.3 Preliminaries

2.3.1 Some results on branching random walks

In this section, we recall some results on branching random walks that will be used in the proofs.

We begin by stating a key result obtained by Biggins in [Big92]. For all $\theta > \underline{\theta}$, we introduce

$$W^{(k)}(\theta) = L(\theta)^{-k} \langle Z^{(k)}, e^{-\theta \cdot} \rangle = L(\theta)^{-k} \sum_{i: |i|=k} p_i^\theta.$$

For all integers $k \in \mathbb{Z}_+ \cup \{\infty\}$, we denote by \mathcal{F}_k the σ -algebra generated by $(p_i, |i| \leq k)$.

Lemma 2.1. *For every $\theta > \underline{\theta}$, $(W^{(k)}(\theta), k \in \mathbb{Z}_+)$ is a martingale with respect to the filtration $(\mathcal{F}_k)_{k \in \mathbb{Z}_+}$. Moreover, if $\theta \in (\theta_*, \theta^*)$, it is bounded in $L^\gamma(\mathbb{P})$ for some $\gamma > 1$ and therefore uniformly integrable, and its terminal value*

$$W(\theta) = \lim_{k \rightarrow \infty} W^{(k)}(\theta)$$

is positive a.s.

In some proofs, we shall need a precise estimate of the number of boxes at generation k with size of order $\exp(k \frac{L'(\theta)}{L(\theta)})$. If the splitting law ν is not geometric, in the sense that there is no real number $r \in (0, 1)$ such that with probability one, all the masses of the atoms of ρ belong to $\{r^n, n \in \mathbb{Z}_+\}$, we may apply Corollary 4 in [Big92]:

Lemma 2.2. *Suppose that ν is not geometric. Then for all real numbers a and b such that $a < b$ and for all $\theta \in (\theta_*, \theta^*)$, we have with probability one that*

$$\begin{aligned} \lim_{k \rightarrow \infty} \sqrt{k} e^{-k\varphi(\theta)} \# \left\{ i \in \mathbb{N}^k : \exp \left(-b + k \frac{L'(\theta)}{L(\theta)} \right) \leq p_i \leq \exp \left(-a + k \frac{L'(\theta)}{L(\theta)} \right) \right\} \\ = \frac{1}{\sqrt{2\pi}} \frac{e^{\theta b} - e^{\theta a}}{\theta} \left(\frac{L''(\theta)}{L(\theta)} - \left(\frac{L'(\theta)}{L(\theta)} \right)^2 \right)^{-\frac{1}{2}} W(\theta). \end{aligned}$$

Note that Lemma 2.2 has a version in the geometric case (technically, replace Stone's local limit theorem by Gnedenko's local limit theorem). In our proofs, we shall always work with a non-geometric splitting law.

We next turn our interest to the asymptotic behavior of extreme sizes of boxes at generation k :

$$\underline{p}(k) = \inf \{ p_i : p_i > 0, |i| = k \} \quad \text{and} \quad \bar{p}(k) = \sup \{ p_i : |i| = k \}.$$

We have the general following results:

Lemma 2.3. *If $\theta^* < \infty$, then*

$$\lim_{k \rightarrow \infty} \left(-\ln \bar{p}(k) - \frac{k}{C^*} \right) = \infty \quad \text{and} \quad \lim_{k \rightarrow \infty} \frac{-\ln \bar{p}(k)}{k} = \frac{1}{C^*} \quad a.s.$$

Likewise, if (2.2) holds, then

$$\lim_{k \rightarrow \infty} \left(\ln \underline{p}(k) + \frac{k}{C_*} \right) = \infty \quad \text{and} \quad \lim_{k \rightarrow \infty} \frac{-\ln \underline{p}(k)}{k} = \frac{1}{C_*} \quad a.s.$$

To have precise estimates of the behavior of the heights and of the saturation levels, we shall sometimes need sharper results of the asymptotics of $\bar{p}(k)$ and $\underline{p}(k)$. The following result, proved by Hu and Shi in [HS09], will be very useful.

Lemma 2.4. *We assume that (2.1) holds.*

- *If $\theta^* < \infty$, then*

$$\limsup_{k \rightarrow \infty} \frac{-\ln \bar{p}(k) - \frac{k}{C^*}}{\ln k} = \frac{3}{2\theta^*} \quad a.s.$$

- *If (2.2) holds, then*

$$\limsup_{k \rightarrow \infty} \frac{\ln \underline{p}(k) + \frac{k}{C_*}}{\ln k} = -\frac{3}{2\theta_*} \quad a.s.$$

Addario-Berry and Reed in [ABR09] also studied the minima in branching random walks, but they require a stronger condition than (2.1). Their assumption is however fulfilled for random split trees.

2.3.2 Poissonization

Let us present the methods used in the proofs. We have to consider the number of balls belonging to each box at each generation. Now, conditionally on \mathcal{F}_k , when n balls have been thrown, the number of balls in the box i , $|i| = k$, follows a binomial law of parameter np_i . Furthermore, these random variables are not independent. A classical idea to circumvent those difficulties (see for instance Gnedin *et al.* in [GHP07] or Holst in [Hol86]) is to consider a randomized version of the total number of balls: instead of throwing initially n balls, one throws \mathcal{P}_n balls, where \mathcal{P}_n is a Poisson variable with parameter n which is independent of $(\rho(i), i \in \mathcal{U})$.

More precisely, we suppose we are given a standard Poisson process $(\mathcal{P}_x)_{x \geq 0}$ independent of \mathcal{F}_∞ . For all individuals $i \in \mathcal{U}$ and for all $x \in (0, \infty)$, we denote the number of balls at i when the first \mathcal{P}_x balls have been thrown by $\mathcal{C}(i; x)$. For all $x, y \in (0, \infty)$ and $k \in \mathbb{Z}_+$, we denote the number of boxes at generation k containing at least y balls when the first \mathcal{P}_x balls have been thrown by $\mathcal{N}_{x,y}(k)$:

$$\mathcal{N}_{x,y}(k) = \#\{i \in \mathbb{N}^k : \mathcal{C}(i; x) \geq y\}.$$

Conditionally on \mathcal{F}_k , the random variables $(\mathcal{C}(i; n))_{|i|=k}$ are independent Poisson variables with parameters np_i . That is why in our proofs, we shall first focus on $\mathcal{N}_{n,j}(k)$. We shall then show that $\mathcal{N}_{n,j}(k)$ is close to the number $N_{n,j}(k)$ of boxes at generation k containing at least j balls when exactly n balls have been thrown. Similarly, let $\mathcal{M}_{x,y}(k)$ denote the number of boxes at generation k containing less than y balls when the first \mathcal{P}_x balls have been thrown:

$$\mathcal{M}_{x,y}(k) = \#\{i \in \mathbb{N}^k : \mathcal{C}(i; x) < y\}$$

and let $M_{n,j}$ denote the number of boxes at generation k containing less than j balls when n balls have been thrown.

We now prove two estimates using the technique of Poissonization that will then be used in the sequel.

Lemma 2.5. *Let $p \in (0, \infty)$. There exist two finite constants $c(p)$ and $d(p)$ such that*

$$\sup_{\substack{j \geq p \\ j \in \mathbb{N}}} \sup_{x > 0} \mathbb{P}(\mathcal{P}_x \geq j) j^p x^{-p} \leq c(p), \quad (2.4)$$

and

$$\sup_{j \in \mathbb{N}} \sup_{x > 0} \mathbb{P}(\mathcal{P}_x < j) j^{-p} x^p \leq d(p). \quad (2.5)$$

Proof. We begin by showing (2.4). Let $j \geq p$ be an integer. Let $x > 0$. By Markov's inequality,

$$\mathbb{P}(\mathcal{P}_x \geq j) = \mathbb{P}(\mathcal{P}_x^p \geq j^p) \leq j^{-p} \mathbb{E}[\mathcal{P}_x^p \mathbf{1}_{\mathcal{P}_x \geq j}].$$

Therefore we only have to bound from above

$$x^{-p} \mathbb{E}[\mathcal{P}_x^p \mathbf{1}_{\mathcal{P}_x \geq j}] = x^{-p} \sum_{k=j}^{\infty} k^p e^{-x} \frac{x^k}{k!}.$$

As $\frac{\Gamma(k-p+1)k^p}{k!} \rightarrow 1$ as k tends to infinity, there exists a finite constant $c(p) \geq 1$ such that for all $k > p - 1$,

$$\frac{k^p}{k!} \leq \frac{c(p)}{\Gamma(k-p+1)}.$$

As $j \geq p > p - 1$, we thus have

$$x^{-p} \sum_{k=j}^{\infty} k^p e^{-x} \frac{x^k}{k!} \leq c(p) e^{-x} \sum_{k=j}^{\infty} \frac{x^{k-p}}{\Gamma(k-p+1)} = c(p) e^{-x} \sum_{k=0}^{\infty} \frac{x^{k+u}}{\Gamma(k+u+1)},$$

where $u = j - p \geq 0$. Applying the formulae 6.5.1, 6.5.4 and 6.5.29 in [AS64], we get that

$$\sum_{k=0}^{\infty} \frac{x^{k+u}}{\Gamma(k+u+1)} \leq e^x,$$

which proves (2.4).

To show (2.5), we write by Markov's inequality:

$$\mathbb{P}(\mathcal{P}_x < j) = \mathbb{P}((\mathcal{P}_x + 1)^{-p} \geq j^{-p}) \leq j^p \mathbb{E}[(\mathcal{P}_x + 1)^{-p}] = j^p \sum_{k=0}^{\infty} (k+1)^{-p} e^{-x} \frac{x^k}{k!}.$$

As $\frac{\Gamma(k+p+1)}{(k+1)^p k!} \rightarrow 1$ as k tends to infinity, there exists a finite constant $d(p) \geq 1$ such that for all $k \in \mathbb{Z}_+$, $\frac{\Gamma(k+p+1)}{(k+1)^p k!} \leq d(p)$. We thus have

$$\mathbb{P}(\mathcal{P}_x < j) \leq d(p) j^p x^{-p} e^{-x} \sum_{k=0}^{\infty} \frac{x^{k+p}}{\Gamma(k+p+1)}.$$

We conclude as before. \square

Remark 2.5. We stress that in (2.4), the integer j cannot be less than p . This restriction lies in the heart of the existence of the phase transition in the asymptotics of $H_{n,j}$ (see Proposition 2.2 below, and more precisely Lemma 2.6). On the other hand, there is no restriction in (2.5); no phase transition appears in the asymptotics of $G_{n,j}$ (see Proposition 2.6 below).

2.4 Study of the heights

In this section, we prove Theorem 2.1 and Proposition 2.1. We are first interested in the upper bound. We shall then focus on the lower bound. We stress that the phase transition is glimpsed at the beginning of the study of the upper bound (see Remark 2.6 below). It is however proved in the paragraph dealing with the lower bound.

2.4.1 Upper bound

Equation (2.4) will enable us to have a uniform upper bound of $N_{n,j}(k)$ independent of the sizes of the boxes that will eventually lead to the upper bounds stated in Theorem 2.1 and in Proposition 2.1.

Proposition 2.2. *Let $j \geq 1$ be an integer and $\alpha \in [0, 1)$ such that $(j, \alpha) \neq (1, 0)$. Then*

$$\limsup_{n \rightarrow \infty} \frac{H_{n, jn^\alpha} - (1 - \alpha) \frac{\theta}{-\ln L(\theta)} \ln n}{\ln \ln n} \leq \frac{1}{-\ln L(\theta)} \quad a.s.$$

for all $\theta > 1$ if $\alpha > 0$, and for all $\theta \in (1, j]$ if $\alpha = 0$.

Remark 2.6. Suppose that $\alpha = 0$. We deduce from Proposition 2.2 that:

$$H_{n, j} \leq \min_{\theta \in (1, j]} h(\theta) \ln n + O(\ln \ln n) \quad a.s.,$$

where h is the function $\theta \mapsto \frac{\theta}{-\ln L(\theta)}$. Now, h is differentiable and $h'(\theta) = -\varphi(\theta) \ln^{-2} L(\theta)$, so h is decreasing on $(1, \theta^*]$ and increasing on $[\theta^*, \infty)$. The minimum of h is therefore

- $\frac{j}{-\ln L(j)}$ if $j < \theta^*$,
- $\frac{\theta^*}{-\ln L(\theta^*)} = C^*$ if $j \geq \theta^*$.

The phase transition is shown. It will be proved in the study of the lower bound.

Recall that $H_{n, jn^\alpha} \leq k$ if and only if at generation k , every box contains less than jn^α balls when n balls have been thrown. Because we want to show that H_{n, jn^α} is bounded from above by $((1 - \alpha) \frac{\theta}{-\ln L(\theta)}) \ln(n) + r \ln \ln n$, we take $k = ((1 - \alpha) \frac{\theta}{-\ln L(\theta)}) \ln(n) + r \ln \ln n$. In other words, one initially throws $n = \exp(k \frac{1}{1 - \alpha} \frac{-\ln L(\theta)}{\theta}) k^{-s}$ balls and we show that every box of the k -th generation contains less than jn^α balls.

Let u and u' be two real numbers such that

$$u > \frac{1}{(1 - \alpha)\theta} \quad \text{and} \quad u < u' < \frac{1}{\alpha} \left(u - \frac{1}{\theta} \right),$$

where for $\alpha = 0$, the second condition reduces to $u' > u$. Define for all $k \geq 1$

$$x_k = k^{-u} \exp \left(k \frac{1}{1 - \alpha} \frac{-\ln L(\theta)}{\theta} \right) \quad \text{and} \quad \phi_k = k^{-u'} \exp \left(k \frac{1}{1 - \alpha} \frac{-\ln L(\theta)}{\theta} \right).$$

Informally, x_k corresponds to the number of balls thrown when we consider boxes at generation k . As mentioned in the preliminaries, the argument relies on Poissonization. We are first interested in $\mathcal{N}_{x, y}(k)$ and we shall see how to depoissonize.

Lemma 2.6. *For almost all ω , there exists $k_0(\omega)$ such that*

$$\mathcal{N}_{x_k, j\phi_k^\alpha}(k) = 0, \quad \text{for all } k \geq k_0(\omega).$$

Proof. Let $x > 0$ and $y \geq \theta$. We calculate $\mathbb{E}[\mathcal{N}_{x, y}(k) | \mathcal{F}_k]$. We write

$$\mathcal{N}_{x, y}(k) = \sum_{i: |i|=k} \mathbf{1}_{\mathcal{C}(i; x) \geq y} = \sum_{i: |i|=k} \mathbf{1}_{\mathcal{C}(i; x) \geq \lceil y \rceil}.$$

Conditionally on \mathcal{F}_k , $(\mathcal{C}(i; x))_{|i|=k}$ are Poisson variables with parameters $x p_i$, so

$$\mathbb{E}[\mathcal{N}_{x, y}(k) | \mathcal{F}_k] = \sum_{i: |i|=k} \mathbb{P}(\mathcal{P}_{x p_i} \geq \lceil y \rceil).$$

As $\lceil y \rceil \geq \theta$, (2.4) ensures that

$$\mathbb{E}[\mathcal{N}_{x,y}(k)|\mathcal{F}_k] \leq \sum_{i:|i|=k} c(\theta)[y]^{-\theta}(xp_i)^\theta.$$

Now, as $y \geq \theta$,

$$\lceil y \rceil \geq \frac{y-1}{y}y \geq \frac{\theta-1}{\theta}y.$$

In the notation of Lemma 2.1, we have:

$$\mathbb{E}[\mathcal{N}_{x,y}(k)|\mathcal{F}_k] \leq c'(\theta)y^{-\theta}x^\theta L(\theta)^k W^{(k)}(\theta),$$

where $c'(\theta) = c(\theta)(1 - \frac{1}{\theta})^{-\theta}$. We finally get

$$\mathbb{E}[\mathcal{N}_{x,y}(k)] \leq c'(\theta)y^{-\theta}x^\theta L(\theta)^k.$$

Taking $x = x_k$ and $y = j\phi_k^\alpha$, we get for all k sufficiently large so that $j\phi_k^\alpha \geq \theta$ (note that if $\alpha = 0$, then for all k , $j\phi_k^\alpha \geq \theta$):

$$\mathbb{E}[\mathcal{N}_{x_k, j\phi_k^\alpha}(k)] \leq c'(\theta)j^{-\theta}k^{\theta(u'\alpha - u)}.$$

Now $\theta(u'\alpha - u) < -1$. As a consequence

$$\mathbb{E}\left[\sum_{k \in \mathbb{N}} \mathbf{1}_{\mathcal{N}_{x_k, j\phi_k^\alpha}(k) \geq 1}\right] \leq \mathbb{E}\left[\sum_{k \in \mathbb{N}} \mathcal{N}_{x_k, j\phi_k^\alpha}(k)\right] < \infty.$$

In particular, there is an a.s. finite number of integers k such that $\mathcal{N}_{x_k, j\phi_k^\alpha}(k) \geq 1$. \square

We now show how to have information on $N_{n, jn^\alpha}(k)$ itself.

Proof of Proposition 2.2. Let E_k denote the event $\{\mathcal{P}_{x_k} \leq \phi_{k+1}\}$. Because $u' > u$, $\frac{\phi_{k+1}}{x_k}$ tends to 0 as k tends to infinity, so there exists an integer k_1 such that for all $k \geq k_1$, $\phi_{k+1} \leq \frac{x_k}{2}$. Consequently,

$$\mathbb{P}(E_k) \leq \mathbb{P}\left(\mathcal{P}_{x_k} \leq \frac{x_k}{2}\right) = \mathbb{P}\left(\mathcal{P}_{x_k} - x_k \leq -\frac{x_k}{2}\right) \leq \mathbb{P}\left(|\mathcal{P}_{x_k} - x_k| \geq \frac{x_k}{2}\right).$$

The variance of the Poisson variable \mathcal{P}_{x_k} being x_k , we get by Chebichev's inequality: $\mathbb{P}(E_k) \leq 4x_k^{-1}$ for all $k \geq k_1$. As $\sum_{k \in \mathbb{N}} x_k^{-1} < \infty$, the Borel-Cantelli lemma ensures that for almost all ω , there exists $k_2(\omega)$ such that $\mathcal{P}_{x_k} > \lfloor \phi_{k+1} \rfloor$ for all $k \geq k_2(\omega)$.

Applying Lemma 2.6, we deduce that, if we define the event Ω_0 by

$$\Omega_0 = \left\{ \omega : \text{there exists } k_3(\omega) \text{ such that } N_{\lfloor \phi_{k+1} \rfloor, j\phi_k^\alpha}(k) = 0 \text{ for all } k \geq k_3(\omega) \right\},$$

then $\mathbb{P}(\Omega_0) = 1$. Notice that there exists an integer k_4 such that for all $k \geq k_4$,

$$k^{-u'} \geq \exp\left(-k \frac{1}{2(1-\alpha)} \frac{-\ln L(\theta)}{\theta}\right).$$

There exists a rank k_5 from which the sequence $(\phi_k)_{k \geq k_5}$ is increasing. Furthermore, that sequence tends to infinity. Let $\omega \in \Omega_0$. Let n be an integer greater than ϕ_{k_5} large

enough so that the unique $k \geq k_5$ satisfying $\phi_k \leq n < \phi_{k+1}$ is greater than $k_3(\omega)$ and k_4 . Because $N_{[\phi_{k+1}], j\phi_k^\alpha}(k) = 0$ and $n \leq \lfloor \phi_{k+1} \rfloor$, $N_{n, j\phi_k^\alpha}(k) = 0$. Now, $j n^\alpha \geq j\phi_k^\alpha$, so $N_{n, j n^\alpha}(k) = 0$ and $H_{n, j n^\alpha} \leq k$. Moreover, as $n \geq \phi_k$,

$$k \leq (1 - \alpha) \frac{\theta}{-\ln L(\theta)} \ln n + u'(1 - \alpha) \frac{\theta}{-\ln L(\theta)} \ln k.$$

Further, as $k \geq k_4$, we have:

$$\exp\left(k \frac{1}{2(1 - \alpha)} \frac{-\ln L(\theta)}{\theta}\right) \leq k^{-u'} \exp\left(k \frac{1}{1 - \alpha} \frac{-\ln L(\theta)}{\theta}\right) \leq n,$$

so $k \leq 2(1 - \alpha) \frac{\theta}{-\ln L(\theta)} \ln n$. Therefore

$$H_{n, j n^\alpha} \leq k \leq (1 - \alpha) \frac{\theta}{-\ln L(\theta)} \ln n + u'(1 - \alpha) \frac{\theta}{-\ln L(\theta)} \ln\left(2(1 - \alpha) \frac{\theta}{-\ln L(\theta)} \ln n\right).$$

We conclude that

$$\limsup_{n \rightarrow \infty} \frac{H_{n, j n^\alpha} - (1 - \alpha) \frac{\theta}{-\ln L(\theta)} \ln n}{\ln \ln n} \leq u'(1 - \alpha) \frac{\theta}{-\ln L(\theta)} \quad \text{a.s.}$$

The fact that u' can be chosen arbitrarily close to $\frac{1}{(1 - \alpha)\theta}$ completes the proof. \square

Remark 2.7. Note that to deduce from Proposition 2.2 that $H_{n, n^\alpha} \leq (1 - \alpha)C^*$ a.s. in the case $\theta^* = \infty$, one has to show that

$$\lim_{\theta \rightarrow \infty} \frac{\theta}{-\ln L(\theta)} = C^*.$$

Proof. Let $\psi = \ln L$. The condition $\theta^* = \infty$ implies that $C^* = \lim_{\theta \rightarrow \infty} -\frac{1}{\psi'(\theta)}$. As ψ is convex, it is known that $\frac{\psi(\theta)}{\theta}$ tends to $l \in]-\infty, \infty]$ as θ tends to ∞ . Let f be the function $\theta \mapsto \psi(\theta) - l\theta$. f is convex and $\frac{f(\theta)}{\theta}$ tends to 0 as θ tends to ∞ . One can easily deduce that f is non-increasing. As a result, $\psi' \leq l$. If $l \neq -\frac{1}{C^*}$, then $\psi'(\theta) \rightarrow l$ as θ tends to ∞ . Now, ψ' is non-decreasing. So there exists $\varepsilon > 0$ such that $\psi' \leq l - \varepsilon$. Let $\theta > 1$. Then $\int_1^\theta \psi' \leq (l - \varepsilon)(\theta - 1)$, i.e., $\psi(\theta) - l\theta \leq -\varepsilon\theta + \varepsilon - l$. Dividing by θ and taking the limit, we get $\varepsilon \leq 0$, which is absurd. We deduce that $l = -\frac{1}{C^*}$, which means that $\frac{\theta}{-\ln L(\theta)} \rightarrow C^*$ as θ tends to ∞ . \square

2.4.2 Lower bound

We now study the lower bound. Combined with the upper bound supplied by Proposition 2.2, Theorem 2.1 and Proposition 2.1 will be proved.

The case $j \geq \theta^*$ or $\alpha \in (0, 1)$ and $\theta^* < \infty$

In this section, we prove that $H_{n, j} \geq C^* \ln n + O(\ln \ln n)$ a.s. if $j \geq \theta^*$ and (2.1) holds. We also prove that $H_{n, n^\alpha} \geq (1 - \alpha)C^* \ln n + O(\ln \ln n)$ a.s. if $\theta^* < \infty$ and (2.1) holds. We shall see that the largest box plays a key role.

Proposition 2.3. *We suppose that $\theta^* < \infty$. Let $j \geq 1$ be an integer and $\alpha \in [0, 1)$ such that $(j, \alpha) \neq (1, 0)$. Under the assumption (2.1), we have:*

$$\liminf_{n \rightarrow \infty} \frac{H_{n, jn^\alpha} - (1 - \alpha)C^* \ln n}{\ln \ln n} \geq \frac{3}{2 \ln L(\theta^*)} \quad a.s.$$

By definition, $H_{n, jn^\alpha} > k$ if and only if at generation k , there exists a box containing at least jn^α balls when n balls have been thrown. We shall see that in our setting, it suffices to consider the largest box. As we intend to show that H_{n, jn^α} is bounded from below by $((1 - \alpha)C^*) \ln n - r \ln \ln n$, we take $k = ((1 - \alpha)C^*) \ln n - r \ln \ln n$, i.e., one initially throws $n = \exp(k \frac{1}{(1 - \alpha)C^*}) k^s$ balls and we show that the largest box of the k -th generation contains at least jn^α balls.

Let γ, γ' and γ'' be three real numbers such that

$$\gamma > \frac{3}{2\theta^*}, \quad \gamma' > \frac{\gamma}{1 - \alpha} \quad \text{and} \quad \gamma' < \gamma'' < \frac{\gamma' - \gamma}{\alpha}.$$

Note that if $\alpha = 0$, the third condition is simply $\gamma'' > \gamma'$. Define for all $k \in \mathbb{Z}_+$

$$x_k = k^{\gamma'} \exp\left(k \frac{1}{(1 - \alpha)C^*}\right) \quad \text{and} \quad \phi_k = k^{\gamma''} \exp\left(k \frac{1}{(1 - \alpha)C^*}\right).$$

We first show that a.s., $\mathcal{N}_{x_k, j\phi_k^\alpha}(k) \geq 1$ for all integers k sufficiently large. To do so, we simply consider the largest box.

Lemma 2.7. *For almost all ω , there exists $k_0(\omega)$ such that*

$$\mathcal{N}_{x_k, j\phi_k^\alpha}(k) \geq 1, \quad \text{for all } k \geq k_0(\omega).$$

Proof. For every generation $k \in \mathbb{N}$, we consider an imaginary box $\mathfrak{b}(k)$ contained in the largest box $\bar{\mathfrak{b}}(k)$ of size $\bar{p}(k)$ such that a ball fallen in $\bar{\mathfrak{b}}(k)$ is thrown in $\mathfrak{b}(k)$ with probability

$$\frac{\mathfrak{p}(k)}{\bar{p}(k)} \wedge 1, \quad \text{where} \quad \mathfrak{p}(k) = k^{-\gamma} e^{-\frac{k}{C^*}}.$$

Informally, the box $\mathfrak{b}(k)$ has size $\mathfrak{p}(k) \wedge \bar{p}(k)$. We denote by A_k the event $\{\mathfrak{p}(k) < \bar{p}(k)\}$ and by B_k the event defined by

$$B_k = \{\text{the box } \mathfrak{b}(k) \text{ contains less than } j\phi_k^\alpha \text{ balls when the first } \mathcal{P}_{x_k} \text{ have been thrown}\}.$$

Conditionally on A_k , the box $\mathfrak{b}(k)$ has size $\mathfrak{p}(k)$ so the number of balls contained in $\mathfrak{b}(k)$ when \mathcal{P}_{x_k} balls have been thrown is a Poisson variable with parameter $x_k \mathfrak{p}(k)$. As a result

$$\mathbb{P}(A_k \cap B_k) \leq \mathbb{P}(B_k | A_k) = \mathbb{P}(\mathcal{P}_{x_k \mathfrak{p}(k)} < j\phi_k^\alpha).$$

Applying (2.5), we get:

$$\mathbb{P}(A_k \cap B_k) \leq d(p) j^p k^{-2},$$

where $p = \frac{2}{\gamma' - \gamma - \alpha\gamma''} > 0$, and hence $\sum \mathbb{P}(A_k \cap B_k) < \infty$. By the Borel-Cantelli lemma, we deduce that a.s., for all k sufficiently large, $\mathfrak{p}(k) \geq \bar{p}(k)$ or the box $\mathfrak{b}(k)$, which is contained in the box $\bar{\mathfrak{b}}(k)$, has at least $j\phi_k^\alpha$ balls when the first \mathcal{P}_{x_k} balls have been thrown. Now, by Lemma 2.4, we know that a.s., for all integers k sufficiently large, $\mathfrak{p}(k) < \bar{p}(k)$. Lemma 2.7 is therefore proved. \square

We now deduce from Lemma 2.7 the lower bound of H_{n,jn^α} .

Proof of Proposition 2.3. The same calculations performed at the beginning of the proof of Proposition 2.2 show that for almost all ω , there exists $k_1(\omega)$ such that $\mathcal{P}_{x_k} < \lceil \phi_{k-1} \rceil$ for all $k \geq k_1(\omega)$. Applying Lemma 2.7, we deduce that, if we define the event Ω_0 by

$$\Omega_0 = \left\{ \omega : \text{there exists } k_2(\omega) \text{ such that } N_{\lceil \phi_{k-1} \rceil, j\phi_k^\alpha}(k) \geq 1 \text{ for all } k \geq k_2(\omega) \right\},$$

then $\mathbb{P}(\Omega_0) = 1$. Let $\omega \in \Omega_0$. The sequence (ϕ_k) is increasing and tends to infinity. Let n be an integer large enough so that the unique integer k satisfying $\phi_{k-1} < n \leq \phi_k$ is greater than $k_2(\omega)$. Because $N_{\lceil \phi_{k-1} \rceil, j\phi_k^\alpha}(k) \geq 1$ and $\lceil \phi_{k-1} \rceil \leq n$, $N_{n, j\phi_k^\alpha}(k) \geq 1$. Now, $jn^\alpha \leq j\phi_k^\alpha$, so $N_{n, jn^\alpha}(k) \geq 1$ and $H_{n, jn^\alpha} > k$. Further, as $n \leq \phi_k$,

$$k \geq (1 - \alpha)C^* \ln n - \gamma''(1 - \alpha)C^* \ln k.$$

As $\exp((k-1)\frac{1}{(1-\alpha)C^*}) \leq \phi_{k-1} \leq n$, we have: $k \leq (1 - \alpha)C^* \ln n + 1$. Thus

$$H_{n, jn^\alpha} > k \geq (1 - \alpha)C^* \ln n - \gamma''(1 - \alpha)C^* \ln((1 - \alpha)C^* \ln n + 1)$$

and

$$\liminf_{n \rightarrow \infty} \frac{H_{n, jn^\alpha} - (1 - \alpha)C^* \ln n}{\ln \ln n} \geq -\gamma''(1 - \alpha)C^* \quad \text{a.s.}$$

The fact that γ'' can be chosen arbitrarily close to $\frac{3}{2(1-\alpha)\theta^*}$ completes the proof. \square

Remark 2.8. In the case $\theta^* < \infty$, if the assumption (2.1) does not hold, one can still show that for all integers $j \geq 1$ and for all real numbers $\alpha \in [0, 1[$ such that $(j, \alpha) \neq (1, 0)$, with probability one, $H_{n, jn^\alpha} \geq (1 - \alpha)C^* \ln n + o(\ln n)$. To do so, recall that Lemma 2.3 ensures that $\frac{-\ln \bar{p}(k)}{k}$ tends to $\frac{1}{C^*}$ a.s., and use an argument similar to that in Proposition 2.3.

The case $2 \leq j < \theta^*$ and $\alpha = 0$

We now prove that $H_{n, j} \geq C_j \ln n + O(\ln \ln n)$ a.s. if $j < \theta^*$. We shall notice that contrary to the case $j \geq \theta^*$, the number of boxes matters more than their sizes.

Proposition 2.4. *Let $j < \theta^*$ be an integer greater than 1. Then*

$$\liminf_{n \rightarrow \infty} \frac{H_{n, j} - \frac{j}{-\ln L(j)} \ln n}{\ln \ln n} \geq \frac{1}{2 \ln L(j)} \quad \text{a.s.}$$

The main difference with the case $j \geq \theta^*$ is that we cannot consider the largest box any more. We shall rather focus on some other boxes which are smaller but sufficiently numerous so that it is very unlikely that all of them contain less than j balls when n balls have been thrown. As we want to prove that $H_{n, j}$ is bounded from below by $\frac{j}{-\ln L(j)} \ln n - r \ln \ln n$, we consider the situation at the k -th generation when one initially throws $\exp(k\frac{-\ln L(j)}{j})k^s$ balls.

The boxes that will play a key role are those appearing in Lemma 2.2 (recall that $j < \theta^*$): with probability one,

$$\lim_{k \rightarrow \infty} \sqrt{k} e^{-k\varphi(j)} \# \{i \in \mathbb{N}^k : \mathfrak{s}(k) \leq p_i \leq 2\mathfrak{s}(k)\} = Q(j),$$

where

$$\mathfrak{s}(k) = \exp\left(-k \frac{-L'(j)}{L(j)}\right)$$

and

$$Q(j) = \frac{1}{\sqrt{2\pi}} \frac{1-2^{-j}}{j} \left(\frac{L''(j)}{L(j)} - \left(\frac{L'(j)}{L(j)} \right)^2 \right)^{-\frac{1}{2}} W(j).$$

Notice that, by Lemma 2.1, as $j < \theta^*$, $Q(j) > 0$ a.s. Define

$$\nu(k) = \left\lfloor \frac{Q(j)k^{-\frac{1}{2}}e^{k\varphi(j)}}{2} \right\rfloor.$$

Then (recall that $\varphi(j) > 0$), for almost all ω , there exists $k_0(\omega)$ such that

$$\#\{i \in \mathbb{N}^k : p_i \geq \mathfrak{s}(k)\} \geq \nu(k) \geq 1, \quad \text{for all } k \geq k_0(\omega). \quad (2.6)$$

We can now prove the following result.

Lemma 2.8. *Define for all $k \in \mathbb{Z}_+$*

$$x_k = k^u \exp\left(k \frac{-\ln L(j)}{j}\right), \quad \text{where } u > \frac{1}{2j}.$$

Then for almost all ω , there exists $k_1(\omega)$ such that

$$\mathcal{N}_{x_k, j}(k) \geq 1, \quad \text{for all } k \geq k_1(\omega).$$

Proof. As \mathbb{N}^k is countable, we can order the boxes of the k -th generation in decreasing order. Denote the family of boxes having a size at least $\mathfrak{s}(k)$ by \mathcal{G}_k . We create $\nu(k)$ new boxes: if $\#\mathcal{G}_k \geq \nu(k)$, consider the first $\nu(k)$ boxes belonging to the family \mathcal{G}_k , denoted by $b_1(k), \dots, b_{\nu(k)}(k)$. For all $l \leq \nu(k)$, we place an imaginary box $\mathfrak{b}_l(k)$ inside the box $b_l(k)$ such that a ball fallen in $b_l(k)$ is thrown in $\mathfrak{b}_l(k)$ with probability $\frac{\mathfrak{s}(k)}{s_l(k)}$, where $s_l(k)$ is the size of $b_l(k)$. In particular, every imaginary box has size $\mathfrak{s}(k)$. If $\#\mathcal{G}_k < \nu(k)$, we denote the first $\nu(k)$ boxes by $\mathfrak{b}_1(k), \dots, \mathfrak{b}_{\nu(k)}(k)$. Introduce the events $A_k = \{\#\mathcal{G}_k \geq \nu(k)\}$ and

$$B_k = \{\forall l \leq \nu(k), \mathfrak{b}_l(k) \text{ contains less than } j \text{ balls when the first } \mathcal{P}_{x_k} \text{ have been thrown}\}.$$

Conditionally on \mathcal{F}_∞ and A_k , the boxes $\mathfrak{b}_l(k)$ have size $\mathfrak{s}(k)$ so the numbers of balls contained in $\mathfrak{b}_l(k)$, $l \leq \nu(k)$, when \mathcal{P}_{x_k} balls have been thrown are independent Poisson variables with parameters $x_k \mathfrak{s}(k)$. Therefore

$$\mathbb{P}(A_k \cap B_k | \mathcal{F}_\infty) \leq \mathbb{P}(\mathcal{P}_{x_k \mathfrak{s}(k)} < j)^{\nu(k)}.$$

Thus

$$\ln \mathbb{P}(A_k \cap B_k | \mathcal{F}_\infty) \leq \frac{Q(j)}{2} k^{-\frac{1}{2}} e^{k\varphi(j)} \ln \mathbb{P}(\mathcal{P}_{x_k \mathfrak{s}(k)} < j).$$

It can be easily seen that $x_k \mathfrak{s}(k)$ tends to 0, and that, as a result,

$$\ln \mathbb{P}(\mathcal{P}_{x_k \mathfrak{s}(k)} < j) \sim -\frac{x_k^j \mathfrak{s}(k)^j}{j!}.$$

Thus there exists a constant $c > 0$ such that for all $k \in \mathbb{Z}_+$,

$$\ln \mathbb{P}(\mathcal{P}_{x_k \mathbf{s}(k)} < j) \leq -2cx_k^j \mathbf{s}(k)^j.$$

Finally, we get

$$\mathbb{P}(A_k \cap B_k | \mathcal{F}_\infty) \leq \exp\left(-cQ(j)k^{-\frac{1}{2}}e^{k\varphi(k)}x_k^j \mathbf{s}(k)^j\right) = \exp\left(-cQ(j)k^{ju-\frac{1}{2}}\right).$$

As $u > \frac{1}{2j}$ and $Q(j) > 0$ a.s., we get $\mathbb{E}[\sum \mathbf{1}_{A_k \cap B_k} | \mathcal{F}_\infty] < \infty$ a.s., so $\sum \mathbf{1}_{A_k \cap B_k} < \infty$ a.s. Combined with (2.6), this proves that a.s., for all integers k sufficiently large, there exists an imaginary box containing at least j balls when \mathcal{P}_{x_k} have been thrown. As every imaginary box is contained in a real box, Lemma 2.8 is proved. \square

We now deduce from Lemma 2.8 the lower bound of $H_{n,j}$.

Proof of Proposition 2.4. Let u and u' be two real numbers such that $\frac{1}{2j} < u < u'$. Define $x_k = k^u \exp(k \frac{-\ln L(j)}{j})$ and $\phi_k = k^{u'} \exp(k \frac{-\ln L(j)}{j})$. One can show that for almost all ω , there exists $k_2(\omega)$ such that $\mathcal{P}_{x_k} < \lceil \phi_{k-1} \rceil$ for all $k \geq k_2(\omega)$. Applying Lemma 2.8, we deduce that if we define the event Ω_0 by

$$\Omega_0 = \left\{ \omega : \text{there exists } k_3(\omega) \text{ such that } N_{\lceil \phi_{k-1} \rceil, j}(k) \geq 1 \text{ for all } k \geq k_3(\omega) \right\},$$

then $\mathbb{P}(\Omega_0) = 1$. The sequence (ϕ_k) is increasing and tends to infinity. Let $\omega \in \Omega_0$. Let n be an integer large enough so that the unique integer k satisfying $\phi_{k-1} < n \leq \phi_k$ is greater than $k_3(\omega)$. Because $N_{\lceil \phi_{k-1} \rceil, j}(k) \geq 1$ and $n \geq \lceil \phi_{k-1} \rceil$, $N_{n,j}(k) \geq 1$, so $H_{n,j} > k$. Moreover, as $n \leq \phi_k$,

$$k \geq \frac{j}{-\ln L(j)} \ln n - u' \frac{j}{-\ln L(j)} \ln k.$$

As $n \geq \phi_{k-1} \geq \exp((k-1) \frac{-\ln L(j)}{j})$, we have: $k \leq \frac{j}{-\ln L(j)} \ln n + 1$. Therefore

$$H_{n,j} > k \geq \frac{j}{-\ln L(j)} \ln n - u' \frac{j}{-\ln L(j)} \ln \left(\frac{j}{-\ln L(j)} \ln n + 1 \right).$$

We conclude that

$$\liminf_{n \rightarrow \infty} \frac{H_{n,j} - \frac{j}{-\ln L(j)} \ln n}{\ln \ln(n)} \geq -u' \frac{j}{-\ln L(j)} \quad \text{a.s.}$$

The fact that u' can be chosen arbitrarily close to $\frac{1}{2j}$ completes the proof. \square

The case $\theta^* = \infty$ and $\alpha \in (0, 1)$

In this section, we prove that $H_{n,n^\alpha} \geq (1 - \alpha)C^* \ln n + o(\ln n)$ a.s. if $\theta^* = \infty$ and $\alpha \in (0, 1)$. We shall see that, like the case $\theta^* < \infty$, the largest box plays a key role.

Proposition 2.5. *We suppose that $\theta^* = \infty$. Let $\alpha \in (0, 1)$. With probability one, we have:*

$$\liminf_{n \rightarrow \infty} \frac{H_{n,n^\alpha}}{\ln n} \geq (1 - \alpha)C^*.$$

We are inspired by the proofs of Proposition 2.3 and Proposition 2.4.

Proof. Let θ , ε and ε' be three real numbers such that

$$\theta > 1 \quad \text{and} \quad 0 < \varepsilon < \varepsilon' < \frac{\varepsilon}{\alpha}.$$

Define for all $k \in \mathbb{Z}_+$

$$x_k = \exp(k\varepsilon) \exp\left(\frac{k}{1-\alpha} \frac{-L'(\theta)}{L(\theta)}\right) \quad \text{and} \quad \phi_k = \exp(k\varepsilon') \exp\left(\frac{k}{1-\alpha} \frac{-L'(\theta)}{L(\theta)}\right).$$

We first show that a.s., $\mathcal{N}_{x_k, \phi_k^\alpha}(k) \geq 1$ for all integers k sufficiently large. To do so, we consider the largest box. Even though we cannot apply Lemma 2.3, Lemma 2.2 provide information about the size $\bar{p}(k)$ of the largest box $\bar{b}(k)$.

Let us prove the following key result: for almost all ω , there exists $k_0(\omega)$ such that

$$\mathcal{N}_{x_k, \phi_k^\alpha}(k) \geq 1, \quad \text{for all } k \geq k_0(\omega). \quad (2.7)$$

We denote by A_k the event $\{\bar{p}(k) \geq \exp(k \frac{L'(\theta)}{L(\theta)})\}$ and by B_k the event $\{\text{the box } \bar{b}(k) \text{ contains less than } \phi_k^\alpha \text{ balls when the first } \mathcal{P}_{x_k} \text{ have been thrown}\}$. Then:

$$\mathbb{P}(A_k \cap B_k) \leq \mathbb{P}(B_k | A_k) \leq \mathbb{P}\left(\mathcal{P}_{x_k \exp(k \frac{L'(\theta)}{L(\theta)})} < \phi_k^\alpha\right).$$

Applying (2.5), we get:

$$\mathbb{P}(A_k \cap B_k) \leq d(1)e^{-k(\varepsilon - \alpha\varepsilon')}.$$

Hence $\sum \mathbb{P}(A_k \cap B_k) < \infty$. By the Borel-Cantelli lemma, we deduce that a.s., for all k sufficiently large, $\bar{p}(k) < \exp(k \frac{L'(\theta)}{L(\theta)})$ or the box $\bar{b}(k)$ has at least ϕ_k^α balls when the first \mathcal{P}_{x_k} balls have been thrown. Now, by Lemma 2.2 (recall that $\theta \in (\theta_*, \theta^*)$ since $\theta^* = \infty$), we know that a.s., for all integers k sufficiently large, $\bar{p}(k) \geq \exp(k \frac{L'(\theta)}{L(\theta)})$. (2.7) is therefore proved.

Performing the same calculations as at the end of the proof of Proposition 2.3, one finally gets:

$$\liminf_{n \rightarrow \infty} \frac{H_{n, n^\alpha}}{\ln n} \geq \frac{1}{\varepsilon' + \frac{1}{1-\alpha} \frac{-L'(\theta)}{L(\theta)}} \quad \text{a.s.}$$

The fact that ε' can be chose arbitrarily close to 0 ensures that:

$$\liminf_{n \rightarrow \infty} \frac{H_{n, n^\alpha}}{\ln n} \geq (1-\alpha) \frac{L(\theta)}{-L'(\theta)} \quad \text{a.s.}$$

Taking the limit as θ goes to $\infty = \theta^*$, we get the result. \square

2.5 Study of the saturation levels

In this section, we prove Theorem 2.2. We shall assume that $\theta_* = -\infty$ or (2.2) holds. We first deal with the lower bound. We shall then focus on the upper bound.

2.5.1 Lower bound

We are inspired by the techniques developed in the previous section. Instead of applying (2.4) as in the study of the heights, we will apply (2.5). It may explain why there is no phase transition in the asymptotics of $G_{n,j}$ (see Remark 2.5).

Proposition 2.6. *Suppose that $\theta_* = -\infty$ or (2.2) holds. Let $j \geq 1$ be an integer and $\alpha \in [0, 1)$. Then*

$$\liminf_{n \rightarrow \infty} \frac{G_{n,jn^\alpha} - (1-\alpha) \frac{-\theta}{\ln L(\theta)} \ln(n)}{\ln \ln(n)} \geq -\frac{1}{\ln L(\theta)} \quad a.s.$$

for every $\theta \in [\theta^*, 0)$ if (2.2) holds, for every $\theta \in (-\infty, 0)$ if $\theta_* = -\infty$.

Recall that $G_{n,jn^\alpha} > k$ if and only if at generation k , every box contains at least jn^α balls when n balls have been thrown. Because we want to show that G_{n,jn^α} is bounded from below by $(1-\alpha) \frac{-\theta}{\ln L(\theta)} \ln n - r \ln \ln n$, we take $k = (1-\alpha) \frac{-\theta}{\ln L(\theta)} \ln n - r \ln \ln n$, i.e., $n = \exp(k \frac{1}{1-\alpha} \frac{\ln L(\theta)}{-\theta}) k^s$.

Proof. Let u and u' be two real numbers such that

$$u > -\frac{1}{(1-\alpha)\theta} \quad \text{and} \quad u < u' < \frac{1}{\alpha} \left(u + \frac{1}{\theta} \right),$$

where the second condition reduces to $u' > u$ for $\alpha = 0$. Define for all $k \in \mathbb{N}$:

$$x_k = k^u \exp\left(k \frac{1}{1-\alpha} \frac{\ln L(\theta)}{-\theta}\right) \quad \text{and} \quad \phi_k = k^{u'} \exp\left(k \frac{1}{1-\alpha} \frac{\ln L(\theta)}{-\theta}\right).$$

Proceeding as usual, let us prove the following key result: for almost all ω , there exists $k_0(\omega)$ such that

$$\mathcal{M}_{x_k, j\phi_k^\alpha}(k) = 0, \quad \text{for all } k \geq k_0(\omega). \quad (2.8)$$

Let x and y be two real numbers such that $y \geq j$ and $k \in \mathbb{N}$. We calculate $\mathbb{E}[\mathcal{M}_{x,y}(k) | \mathcal{F}_k]$. We write

$$\mathcal{M}_{x,y}(k) = \sum_{i:|i|=k} \mathbf{1}_{\mathcal{C}(i;x) < y} = \sum_{i:|i|=k} \mathbf{1}_{\mathcal{C}(i;x) < [y]}$$

Now, conditionally on \mathcal{F}_k , $(\mathcal{C}(i;x))_{|i|=k}$ are Poisson variables with parameters $x p_i$, so

$$\mathbb{E}[\mathcal{M}_{x,y}(k) | \mathcal{F}_k] = \sum_{i:|i|=k} \mathbb{P}(\mathcal{P}_{x p_i} < [y]).$$

Applying (2.5), we get

$$\mathbb{E}[\mathcal{M}_{x,y}(k) | \mathcal{F}_k] \leq d(-\theta) [y]^{-\theta} \sum_{i:|i|=k} (x p_i)^\theta.$$

Now, as $y \geq j$,

$$[y] \leq \frac{y+1}{y} y \leq \frac{j+1}{j} y.$$

In the notation of Lemma 2.1, we have:

$$\mathbb{E} [\mathcal{M}_{x,y}(k) | \mathcal{F}_k] \leq d' y^{-\theta} x^\theta L(\theta)^k W^{(k)}(\theta),$$

where $d' = d(-\theta)(1 + \frac{1}{j})^{-\theta}$. We finally get

$$\mathbb{E} [\mathcal{M}_{x,y}(k)] \leq d' y^{-\theta} x^\theta L(\theta)^k.$$

For $x = x_k$ and $y = j\phi_k^\alpha \geq j$, we obtain for all integers $k \in \mathbb{N}$:

$$\mathbb{E} [\mathcal{M}_{x_k, j\phi_k^\alpha}(k)] \leq d' j^{-\theta} k^{\theta(u-u'\alpha)}.$$

Now $\theta(u - u'\alpha) < -1$. As a consequence

$$\mathbb{E} \left[\sum_{k \in \mathbb{Z}_+} \mathbf{1}_{\mathcal{M}_{x_k, j\phi_k^\alpha}(k) \geq 1} \right] \leq \mathbb{E} \left[\sum_{k \in \mathbb{Z}_+} \mathcal{M}_{x_k, j\phi_k^\alpha}(k) \right] < \infty.$$

In particular, there is an a.s. finite number of integers k such that $\mathcal{M}_{x_k, j\phi_k^\alpha}(k) \geq 1$, which proves (2.8).

Performing the same calculations as at the end of the proof of Proposition 2.3, one finally gets:

$$\liminf_{n \rightarrow \infty} \frac{G_{n, jn^\alpha} - (1 - \alpha) \frac{-\theta}{\ln L(\theta)} \ln(n)}{\ln \ln(n)} \geq -u'(1 - \alpha) \frac{-\theta}{\ln L(\theta)} \quad \text{a.s.}$$

The fact that u' can be chose arbitrarily close to $-\frac{1}{(1-\alpha)\theta}$ completes the proof. \square

Remark 2.9. Note that to deduce from Proposition 2.6 that $G_{n, jn^\alpha} \geq (1 - \alpha)C_* \ln n$ a.s. in the case $\theta_* = -\infty$, one has to show that

$$\lim_{\theta \rightarrow -\infty} \frac{-\theta}{\ln L(\theta)} = C_*.$$

Proof. Let $\psi = \ln L$. The condition $\theta_* = -\infty$ implies that $C_* = \lim_{\theta \rightarrow -\infty} -\frac{1}{\psi'(\theta)}$. As ψ is convex, it is known that $\frac{\psi(\theta)}{-\theta}$ tends to $l \in]-\infty, \infty]$ as θ tends to $-\infty$. We distinguish two cases.

Either l is finite. Let $g : \theta \mapsto \psi(\theta) + l\theta$. g is a convex function and $\frac{g(\theta)}{\theta}$ tends to 0 as θ tends to $-\infty$. One can easily deduce that g is non-decreasing. As a result, $\psi' \geq -l$. If $l \neq \frac{1}{C_*}$, then $\psi'(\theta) \not\rightarrow -l$ as θ tends to $-\infty$. Now, ψ' is non-decreasing. So there exists $\varepsilon > 0$ such that $\psi' \geq \varepsilon - l$. Let $\theta < 0$. Then $\int_\theta^0 \psi' \geq (\varepsilon - l)(-\theta)$, i.e., $\psi(\theta) + l\theta \leq \psi(0) + \varepsilon\theta$. Dividing by $-\theta$ and taking the limit, we get $\varepsilon \leq 0$, which is absurd. We deduce that $l = \frac{1}{C_*}$, which means that $\frac{-\theta}{\ln L(\theta)} \rightarrow C_*$ as θ tends to $-\infty$.

Or l is infinite. As $\theta_* = -\infty$, $\varphi(\theta) > 0$ for all $\theta < 0$, so the function $\theta \in (-\infty, 0) \mapsto \frac{\theta}{\psi(\theta)} - \frac{1}{\psi'(\theta)}$ is increasing. It is also positive. If it does not tend to 0 at $-\infty$, then it is bounded from below by some $\varepsilon > 0$. Multiplying by $-\psi'(\theta) > 0$, we get: $1 \geq 1 - \frac{\theta\psi'(\theta)}{\psi(\theta)} \geq -\varepsilon\psi'(\theta)$. Consequently $\psi'(\theta) \geq -\frac{1}{\varepsilon}$ and $\psi(0) - \psi(\theta) \geq \frac{\theta}{\varepsilon}$. Dividing by $\theta < 0$, we obtain: $-\frac{\psi(\theta)}{\theta} \leq \frac{1}{\varepsilon} - \frac{\psi(0)}{\theta}$. Taking the limit, we have $\infty \leq \frac{1}{\varepsilon}$, which is absurd. Finally, $\frac{\theta}{\psi(\theta)} - \frac{1}{\psi'(\theta)}$ tends to 0 as θ tends to $-\infty$. Now, by definition of C_* , $\psi'(\theta)$ tends to $-\frac{1}{C_*}$. As a result, $\frac{-\theta}{\psi(\theta)}$ tends to C_* as θ tends to $-\infty$. \square

2.5.2 Upper bound

In this section, we are interested in the upper bound of the saturation levels.

The case $-\infty < \theta_* < 0$ **and** $\varphi(\theta_*) = 0$

We prove that $G_{n,jn^\alpha} \leq (1 - \alpha)C_* \ln n + O(\ln \ln n)$ a.s. if (2.2) and (2.1) hold by studying the smallest box, regardless of the value of j and α ; there is no phase transition.

Proposition 2.7. *Suppose that (2.2) and (2.1) hold. Let $j \geq 1$ be an integer and $\alpha \in (0, 1)$. Then*

$$\limsup_{n \rightarrow \infty} \frac{G_{n,j} - C_* \ln n}{\ln \ln n} \leq \frac{3}{2 \ln L(\theta_*)} + \frac{C_*}{j} \quad a.s.$$

and

$$\limsup_{n \rightarrow \infty} \frac{G_{n,jn^\alpha} - (1 - \alpha)C_* \ln n}{\ln \ln n} \leq \frac{3}{2 \ln L(\theta_*)} \quad a.s.$$

In order to prove both inequalities simultaneously, we may suppose that $\alpha \in [0, 1)$, the case $G_{n,j}$ corresponding to G_{n,jn^α} with $\alpha = 0$.

Proof. Let γ , γ' and γ'' be three real numbers such that:

- $\gamma > -\frac{3}{2\theta_*}$, $\gamma' > \gamma + \frac{1}{j}$ and $\gamma'' > \gamma'$ if $\alpha = 0$,
- $\gamma > -\frac{3}{2\theta_*}$, $\gamma' > \frac{\gamma}{1-\alpha}$ and $\gamma' < \gamma'' < \frac{\gamma' - \gamma}{\alpha}$ if $\alpha > 0$.

Define

$$x_k = k^{-\gamma'} \exp\left(k \frac{1}{(1-\alpha)C_*}\right) \quad \text{and} \quad \phi_k = k^{-\gamma''} \exp\left(k \frac{1}{(1-\alpha)C_*}\right).$$

Let us prove the following result: for almost all ω , there exists $k_0(\omega)$ such that

$$\mathcal{M}_{x_k, j\phi_k^\alpha}(k) \geq 1, \quad \text{for all } k \geq k_0(\omega). \quad (2.9)$$

Define $\mathbf{p}(k) = k^\gamma e^{-\frac{k}{C_*}}$. We consider an imaginary box $\mathbf{b}(k)$ at generation k such that

1. if $1 \leq \mathbf{p}(k)$: every ball thrown is placed in $\mathbf{b}(k)$.
2. if $\underline{\mathbf{p}}(k) < \mathbf{p}(k) < 1$, where $\underline{\mathbf{p}}(k)$ is the size of the smallest box $\underline{\mathbf{b}}(k)$: every ball fallen in $\underline{\mathbf{b}}(k)$ is also placed in the imaginary box and every other ball is placed in $\mathbf{b}(k)$ with probability $\frac{\mathbf{p}(k) - \underline{\mathbf{p}}(k)}{1 - \underline{\mathbf{p}}(k)}$. Hence, the imaginary box has size $\mathbf{p}(k)$.
3. if $\mathbf{p}(k) \leq \underline{\mathbf{p}}(k)$: every ball fallen in the smallest box is placed in the imaginary box with probability $\frac{\mathbf{p}(k)}{\underline{\mathbf{p}}(k)}$, and no other ball is placed in $\mathbf{b}(k)$.

The box $\mathbf{b}(k)$ has thus size $\mathbf{p}(k) \wedge 1$, and whenever $\mathbf{p}(k) > \underline{p}(k)$, it contains the smallest box. From Lemma 2.4, we know that a.s., for all integers k sufficiently large, $\mathbf{p}(k) > \underline{p}(k)$. To prove (2.9), all we have to do is therefore to prove that $\mathbb{P}(\limsup A_k) = 0$, where the event A_k is defined by

$$A_k = \{\mathbf{b}(k) \text{ contains at least } j\phi_k^\alpha \text{ balls when the first } \mathcal{P}_{x_k} \text{ balls have been thrown}\}.$$

By the Borel-Cantelli lemma, it suffices to show that $\sum \mathbb{P}(A_k) < \infty$. Let k be an integer sufficiently large so that $\mathbf{p}(k) < 1$. The imaginary box has then size $\mathbf{p}(k)$, so we have:

$$\mathbb{P}(A_k) = \mathbb{P}(\mathcal{P}_{x_k \mathbf{p}(k)} \geq j\phi_k^\alpha).$$

We would like to apply (2.4). To do so, let $p = j$ if $\alpha = 0$ and $p = \frac{2}{\gamma' - \gamma - \alpha\gamma''} > 0$ if $\alpha > 0$. Note that for all integers k sufficiently large, $j\phi_k^\alpha \geq p$. (2.4) then ensures that:

$$\mathbb{P}(A_k) \leq c(p)j^{-p}k^{p(\alpha\gamma'' + \gamma - \gamma')}.$$

As $p(\alpha\gamma'' + \gamma - \gamma') < -1$, $\sum \mathbb{P}(A_k) < \infty$, which proves (2.9).

Performing the same calculations as at the end of the proof of Proposition 2.2, one finally gets:

$$\limsup_{n \rightarrow \infty} \frac{G_{n, jn^\alpha} - (1 - \alpha)C_* \ln n}{\ln \ln n} \leq \gamma''(1 - \alpha)C_* \quad \text{a.s.}$$

The fact that γ'' can be chosen arbitrarily close either to $-\frac{3}{2\theta_*} + \frac{1}{j}$ if $\alpha = 0$ or to $-\frac{3}{2(1-\alpha)\theta_*}$ if $\alpha > 0$ completes the proof. \square

Remark 2.10. If the assumption (2.2) holds but (2.1) does not, one can still show that for all integers $j \geq 1$ and for all real numbers $\alpha \in [0, 1[$, with probability one, $G_{n, jn^\alpha} \leq (1 - \alpha)C_* \ln n + o(\ln n)$. To do so, recall that Lemma 2.3 ensures that $\frac{-\ln \mathbf{p}(k)}{k}$ tends to $\frac{1}{C_*}$ a.s., and use an argument similar to that in Proposition 2.7.

The case $\theta_* = -\infty$

In this section, we prove that $G_{n, jn^\alpha} \leq (1 - \alpha)C_* \ln n + o(\ln n)$ a.s. if $\theta_* = -\infty$. We shall see that, once again, the smallest box plays a key role.

Proposition 2.8. *We suppose that $\theta_* = -\infty$. Let j be a positive integer and $\alpha \in [0, 1)$. Then*

$$\limsup_{n \rightarrow \infty} \frac{G_{n, jn^\alpha}}{\ln n} \leq (1 - \alpha)C_* \quad \text{a.s.}$$

We are inspired by the proof of Proposition 2.5.

Proof. Let θ , ε and ε' be three real numbers such that

$$\theta < 0, \quad 0 < \varepsilon < \varepsilon' < \frac{\varepsilon}{\alpha} \quad \text{and} \quad \varepsilon' < \frac{1}{1 - \alpha} \frac{-L'(\theta)}{L(\theta)}.$$

Define for all $k \in \mathbb{Z}_+$

$$x_k = \exp(-k\varepsilon) \exp\left(\frac{k}{1 - \alpha} \frac{-L'(\theta)}{L(\theta)}\right) \quad \text{and} \quad \phi_k = \exp(-k\varepsilon') \exp\left(\frac{k}{1 - \alpha} \frac{-L'(\theta)}{L(\theta)}\right).$$

Note that we suppose that $\varepsilon' < \frac{1}{1-\alpha} \frac{-L'(\theta)}{L(\theta)}$ so that x_k and ϕ_k tend to ∞ as k tends to ∞ . We first show that a.s., $\mathcal{M}_{x_k, j\phi_k^\alpha}(k) \geq 1$ for all integers k sufficiently large. To do so, we consider the smallest box. Even though we cannot apply Lemma 2.3, Lemma 2.2 provide information about the size $\underline{p}(k)$ of the smallest box $\underline{b}(k)$.

Let us prove the following key result: for almost all ω , there exists $k_0(\omega)$ such that

$$\mathcal{M}_{x_k, j\phi_k^\alpha}(k) \geq 1, \quad \text{for all } k \geq k_0(\omega). \quad (2.10)$$

We denote by A_k the event $\{\underline{p}(k) \leq \exp(k \frac{L'(\theta)}{L(\theta)})\}$ and by B_k the event $\{\text{the box } \underline{b}(k) \text{ contains at least } j\phi_k^\alpha \text{ balls when the first } \mathcal{P}_{x_k} \text{ have been thrown}\}$. Then:

$$\mathbb{P}(A_k \cap B_k) \leq \mathbb{P}(B_k | A_k) \leq \mathbb{P}\left(\mathcal{P}_{x_k \exp(k \frac{L'(\theta)}{L(\theta)})} \geq j\phi_k^\alpha\right).$$

By Markov's inequality, we get:

$$\mathbb{P}(A_k \cap B_k) \leq j^{-1} e^{-k(\varepsilon - \alpha\varepsilon')}.$$

Hence $\sum \mathbb{P}(A_k \cap B_k) < \infty$. By the Borel-Cantelli lemma, we deduce that a.s., for all k sufficiently large, $\underline{p}(k) > \exp(k \frac{L'(\theta)}{L(\theta)})$ or the box $\underline{b}(k)$ has less than $j\phi_k^\alpha$ balls when the first \mathcal{P}_{x_k} balls have been thrown. Now, by Lemma 2.2 (recall that $\theta \in (\theta_*, \theta^*)$ since $\theta_* = -\infty$), we know that a.s., for all integers k sufficiently large, $\underline{p}(k) \leq \exp(k \frac{L'(\theta)}{L(\theta)})$. Finally, (2.10) is proved.

The usual calculations yield:

$$\limsup_{n \rightarrow \infty} \frac{G_{n, jn^\alpha}}{\ln n} \leq \frac{1}{-\varepsilon' + \frac{1}{1-\alpha} \frac{-L'(\theta)}{L(\theta)}} \quad \text{a.s.}$$

The fact that ε' can be chose arbitrarily close to 0 ensures that:

$$\limsup_{n \rightarrow \infty} \frac{G_{n, jn^\alpha}}{\ln n} \leq (1 - \alpha) \frac{L(\theta)}{-L'(\theta)} \quad \text{a.s.}$$

Taking the limit as θ goes to $-\infty = \theta^*$, we get the result. \square

2.6 An explanation for the phase transition

In this section, we explain heuristically why a phase transition may occur in the asymptotics of $H_{n, j}$ but not in those of H_{n, n^α} , $G_{n, j}$ and G_{n, n^α} . To do so, we rephrase the setting in terms of interval-fragmentations. One can easily construct a family $(F(k), k \in \mathbb{Z}_+)$ of random open subsets of $(0, 1)$ with $F(0) = (0, 1)$, which is nested in the sense that $F(k') \subseteq F(k)$ for all $k \leq k'$, and such that the set of the lengths of the interval components of $F(k)$ is $\{p_i : |i| = k\}$ for all integers k . The boxes correspond to the interval components and the balls to a sequence $(U_i)_{i \in \mathbb{N}}$ of independent random variables uniformly distributed on $(0, 1)$ and independent of the fragmentation F . The height $H_{n, j}$ corresponds to the least integer k such that all interval components of $F(k)$ contain less than j elements of the set $\{U_1, \dots, U_n\}$, and the saturation level $G_{n, j}$ to the least

integer k such that there exists an interval component of $F(k)$ containing less than j elements of the set $\{U_1, \dots, U_n\}$. Roughly speaking, the height $H_{n,j}$ depends crucially on the minimal length $\underline{m}_{n,j}$ of the intervals $[\hat{U}_i, \hat{U}_{i+j-1}]$ for $1 \leq i \leq n-j+1$, where $0 < \hat{U}_1 < \dots < \hat{U}_n < 1$ are the ordered statistics of the family (U_1, \dots, U_n) , whereas the saturation level $G_{n,j}$ is related to the maximal length $\overline{m}_{n,j}$ of the intervals $[\hat{U}_i, \hat{U}_{i+j}]$ for $0 \leq i \leq n-j+1$, where $\hat{U}_0 = 0$ and $\hat{U}_{n+1} = 1$. Indeed, the height $H_{n,j}$ is the first time k when no cluster $[\hat{U}_i, \hat{U}_{i+j-1}]$, $1 \leq i \leq n-j+1$, of size j (in particular the smallest one) is included in an interval component of $F(k)$, and the saturation level $G_{n,j}$ is the first time k when there exists a cluster $[\hat{U}_i, \hat{U}_{i+j}]$, $0 \leq i \leq n-j+1$, of size $j+1$ (possibly the largest one) containing an interval component of $F(k)$.

It is easy to show from Lemma 2.3 that if we used equidistributed points $\{\frac{j}{n+1} : 1 \leq j \leq n\}$ instead of i.i.d. uniform points, then no phase transition would occur. More precisely, all the heights $H_{n,j}$ would be equivalent to $C^* \ln n$ a.s. Further, the heights H_{n,n^α} would be equivalent to $(1-\alpha)C^* \ln n$ a.s. and the saturation levels G_{n,jn^α} to $(1-\alpha)C_* \ln n$ a.s.

We first explain why the clusters of size n^α behave as if the points U_i , $1 \leq i \leq n$, were equidistributed on $(0,1)$. It will follow that no phase transition occurs in the asymptotics of H_{n,n^α} and G_{n,n^α} . To do so, let us first prove that $\underline{m}_{n,n^\alpha} \geq n^{-1+\alpha+o(1)}$ a.s. Let $\varepsilon > 0$. We have:

$$\mathbb{P}(\underline{m}_{n,n^\alpha} \leq n^{-1+\alpha-2\varepsilon}) \leq n\mathbb{P}(\hat{U}_{n^\alpha} - \hat{U}_1 \leq n^{-1+\alpha-2\varepsilon}).$$

Let $(e_i)_{i \in \mathbb{N}}$ be a sequence of independent exponential variables with parameters 1. Define $\gamma_k = \sum_{1 \leq i \leq k} e_i$, for all $k \geq 1$. Then we have:

$$(\hat{U}_1, \dots, \hat{U}_n) \stackrel{(d)}{=} \left(\frac{\gamma_1}{\gamma_{n+1}}, \dots, \frac{\gamma_n}{\gamma_{n+1}} \right), \quad (2.11)$$

so

$$\begin{aligned} \mathbb{P}(\hat{U}_{n^\alpha} - \hat{U}_1 \leq n^{-1+\alpha-2\varepsilon}) &= \mathbb{P}\left(\frac{\gamma_{n^\alpha} - \gamma_1}{\gamma_{n+1}} \leq n^{-1+\alpha-2\varepsilon} \right) \\ &\leq \mathbb{P}(\gamma_{n^\alpha} - \gamma_1 < n^{1+\varepsilon} n^{-1+\alpha-2\varepsilon}) + \mathbb{P}(\gamma_{n+1} \geq n^{1+\varepsilon}) \\ &= \mathbb{P}(\mathcal{P}_{n^\alpha-\varepsilon} \geq n^\alpha) + \mathbb{P}(\mathcal{P}_{n^{1+\varepsilon}} \leq n+1) \end{aligned}$$

since the sequence $(\gamma_i)_{i \in \mathbb{N}}$ has the same law as the ordered sequence of points of a standard Poisson process. Applying Lemma 2.5 with $p = \frac{3}{\varepsilon}$, we have for all integers n sufficiently large:

$$\mathbb{P}(\mathcal{P}_{n^\alpha-\varepsilon} \geq n^\alpha) + \mathbb{P}(\mathcal{P}_{n^{1+\varepsilon}} \leq n+1) \leq (c(p) + 2^p d(p)) n^{-3}.$$

We deduce that that for all integers n sufficiently large, we have:

$$\mathbb{P}(\underline{m}_{n,n^\alpha} \leq n^{-1+\alpha-2\varepsilon}) \leq (c(p) + 2^p d(p)) n^{-2}.$$

Applying the Borel-Cantelli lemma and letting ε tend to 0, we easily conclude that $\underline{m}_{n,n^\alpha} \geq n^{-1+\alpha+o(1)}$ a.s. Similar calculations yield: $\overline{m}_{n,n^\alpha} \leq n^{-1+\alpha+o(1)}$ a.s. We conclude that the clusters of size n^α essentially behave as if the points U_i , $1 \leq i \leq n$, were equidistributed on $(0,1)$, which explains why no phase transition occur in the asymptotics of H_{n,n^α} and G_{n,n^α} .

On the contrary, the clusters of size j behave very differently from the equidistribution case. For instance, thanks to the identification (2.11), we may apply well-known results on extreme values (see, *e.g.*, [Res08]). One finds that $\underline{m}_{n,j} \approx n^{-\frac{j}{j-1}}$, so that $\underline{m}_{n,j}$ is much smaller than $\frac{j}{n}$, even at a logarithmic scale. Furthermore, the larger the integer j is, the closer these two quantities are, so that $H_{n,j}$ may eventually be equivalent to $C^* \ln n$ a.s. if j is sufficiently large. On the other hand, if the integer j is too small, then $H_{n,j}$ may be greater than $(1 + \varepsilon)C^* \ln n + o(\ln n)$ a.s. for some $\varepsilon > 0$: one has to wait for a long time before the smallest clusters of size j are no longer included in any interval component of the fragmentation; in that case, there is a phase transition.

Concerning the saturation level $G_{n,j}$, one gets $\overline{m}_{n,j} \approx \frac{\ln n}{n}$, which differs from the equidistribution case only up to a logarithmic factor, and the impact of the latter is asymptotically negligible. It explains why no phase transition occurs in the asymptotics of $G_{n,j}$.

CHAPTER 3

PARTIAL MATCH QUERIES IN TWO-DIMENSIONAL QUADTREES: A PROBABILISTIC APPROACH

We analyze the mean cost of the partial match queries in random two-dimensional quadtrees. The method is based on fragmentation theory. The convergence is guaranteed by a coupling argument of Markov chains, whereas the value of the limit is computed as the fixed point of an integral equation.

Contents

3.1	Introduction	64
3.2	Notation and first properties	65
3.2.1	The continuous-time model	65
3.2.2	Main equations	66
3.2.3	Depoissonization	67
3.3	Particular cases and fragmentation theory	68
3.3.1	The uniform case	68
3.3.2	Case $x = 0$	71
3.3.3	An <i>a priori</i> uniform bound	72
3.4	The convergence at fixed $x \in (0, 1)$	74
3.4.1	A key Markov chain	74
3.4.2	Study of $t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\overline{M}_k t - F_k}(X_k) \mathbf{1}_{T > k}]$	76
3.4.3	Conclusion	77
3.5	Identifying the limit	78
3.6	Extensions and comments	79

3.6.1	Various convergences	79
3.6.2	Multidimensional case	80
3.6.3	Quadtree as a model of random geometry	80

3.1 Introduction

Introduced by Finkel and Bentley [FB74], the quadtree structure is a comparison based algorithm designed for retrieving multidimensional data. It is often studied in computer science because of its numerous applications. The aim of this paper is to study the mean cost of the so-called *partial match queries* in random quadtrees. This problem was first analyzed by Flajolet *et al.* [FGCR93].

Let us briefly describe the discrete model. We choose to focus only on the two-dimensional case. Let P_1, \dots, P_n be n independent random variables uniformly distributed over $(0, 1)^2$. We shall assume that the points have different x and y coordinates, an event that has probability 1. We construct iteratively a finite covering of $[0, 1]^2$ composed of rectangles with disjoint interiors as follows. The first point P_1 divides the original square $[0, 1]^2$ into four closed quadrants according to the vertical and horizontal positions of P_1 . By induction, a point P_k divides the quadrant in which it falls into four quadrants according to its position in this quadrant (see Figure 3.1). Hence the n points P_1, \dots, P_n give rise to a covering of $[0, 1]^2$ into $3n + 1$ closed rectangles with disjoint interiors that we denote by $\text{Quad}(P_1, \dots, P_n)$.

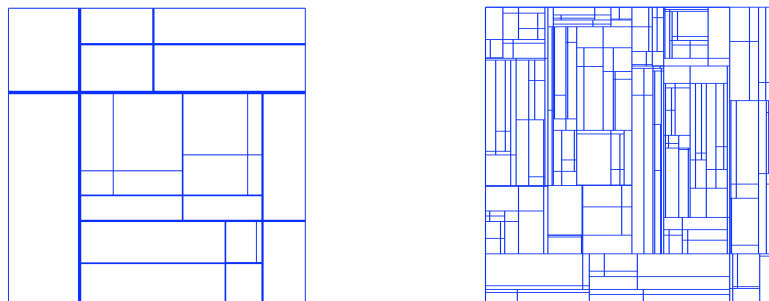


Figure 3.1: Two splittings of $[0, 1]^2$ with resp. 8 and 100 points.

We are interested in the partial match query. As explained by Flajolet and Sedgewick [FS09, Example VII.23.], given $x_0 \in [0, 1]$, it determines the set of points P_i , $i \in \{1, \dots, n\}$, with x coordinates equal to x_0 , regardless of the y coordinates (that set is either empty or a singleton). Denoting the vertical segment $[(x, 0), (x, 1)]$ by S_x , the cost of this partial match query is measured by the number $\mathcal{N}_n(x)$ of rectangles of $\text{Quad}(P_1, \dots, P_n)$ intersecting S_x minus 1 ($\mathcal{N}_0(x) = 0$ by convention). We shall study the cost of a *fixed* query. Our main result is:

Theorem 3.1. *For every $x \in [0, 1]$, we have the following convergence:*

$$n^{-\beta^*} \mathbb{E}[\mathcal{N}_n(x)] \xrightarrow{n \rightarrow \infty} K_0(x(1-x))^{\beta^*/2},$$

where $\beta^* = \frac{\sqrt{17} - 3}{2}$ and $K_0 = \frac{\Gamma(2\beta^* + 2)\Gamma(\beta^* + 2)}{2\Gamma^3(\beta^* + 1)\Gamma^2(\frac{\beta^*}{2} + 1)}$.

Flajolet *et al.* [FGCR93] proved the convergence in mean of the properly rescaled cost of partial match queries when x is *random* with the uniform law on $[0, 1]$ and independent of P_1, \dots, P_n . See also Chern and Hwang [CH03] for a more precise asymptotic behavior. We shall give another proof of this result using *fragmentation theory* (see Corollary 3.1 below). As a by-product of our techniques, we shall also prove in Corollary 3.2 below that, when rescaled by $n^{1-\sqrt{2}}$, $\mathcal{N}_n(0)$ converges in \mathbb{L}^2 (its convergence in mean was obtained in [FGCR93]).

The paper is organized as follows. Section 3.2 introduces the model embedded in continuous-time and presents the first properties. Section 3.3 is devoted to the link between quadtrees and fragmentation theory. Section 3.4, the most technical one, contains the proof of the convergence at a fixed point x without knowing the limit. The identification of the limit is done in Section 3.5 using a fixed point argument for integral equation.

3.2 Notation and first properties

In order to apply probabilistic techniques, we first introduce a continuous-time version of the quadtree: the points P_1, \dots, P_n are replaced by the arrival points of a Poisson point process over $\mathbb{R}_+ \times [0, 1]^2$ with intensity $dt \otimes dx dy$. All the results obtained in this model can easily be translated into results for the discrete-time model.

3.2.1 The continuous-time model

Let Π be a Poisson point process on $\mathbb{R}_+ \times [0, 1]^2$ with intensity $dt \otimes dx dy$. Let $((\tau_i, x_i, y_i), i \geq 1)$ be the atoms of Π ranked in the increasing order of their τ -component. We define a process $(Q(t))_{t \geq 0}$ with values in finite covering of $[0, 1]^2$ by closed rectangles with disjoint interiors as follows. We first introduce the operation **SPLIT**: for every subset R of $[0, 1]^2$ and for every $(x, y) \in [0, 1]^2$,

$$\text{SPLIT}(R, x, y) = \{R \cap [0, x] \times [0, y], R \cap [0, x] \times [y, 1], R \cap [x, 1] \times [0, y], R \cap [x, 1] \times [y, 1]\}.$$

In other words, if R is a rectangle with sides parallel to the x and y axes, then $\text{SPLIT}(R, x, y)$ is the set of the four quadrants in R determined by the point (x, y) . We may now recursively define the process $(Q(t))_{t \geq 0}$. Let $\tau_0 = 0$. For every $t \in [0, \tau_1)$, define $Q(t) = \{[0, 1]^2\}$, and for every $t \in [\tau_i, \tau_{i+1})$, denoting by R the only element (if any) of $Q(\tau_{i-1})$ such that (x_i, y_i) is in the interior of the rectangle R , let

$$Q(t) = \text{SPLIT}(R, x_i, y_i) \cup Q(\tau_{i-1}) \setminus \{R\}.$$

Observe that a.s., for every $i \in \mathbb{Z}_+$, there indeed exists a unique rectangle of $Q(\tau_i)$ such that (x_{i+1}, y_{i+1}) is in its interior, hence the process $(Q(t))_{t \geq 0}$ is well defined up to an event of zero probability. In the sequel we shall assume that the points of Π always fall in the interior of some rectangle of $(Q(t))_{t \geq 0}$. As explained in the introduction, we are

interested in the number of rectangles of $Q(t)$ intersecting the segment S_x , specifically we set:

$$N_t(x) = \#\{R \in Q(t) : R \cap S_x \neq \emptyset\} - 1,$$

so that $N_t(x) = 0$ for every $0 \leq t < \tau_1$. Recalling that τ_n is the arrival time of the n -th point of Π , $Q(\tau_n)$ has the same distribution as the random variable $\text{Quad}(P_1, \dots, P_n)$ of the introduction. In particular, for every $(n, x) \in \mathbb{N} \times [0, 1]$, we have $N_{\tau_n}(x) = \mathcal{N}_n(x)$ in distribution.

3.2.2 Main equations

Let $x \in [0, 1]$. We denote by \mathcal{A} the set of words over the alphabet $\{0, 1\}$,

$$\mathcal{A} = \bigcup_{n \geq 0} \{0, 1\}^n,$$

where by convention $\{0, 1\}^0 = \{\emptyset\}$. Thus, if $u \in \mathcal{A}$, u is either \emptyset or a finite sequence of 0 and 1. If u and v are elements of \mathcal{A} then uv denotes the concatenation of the two words u and v . We label the rectangles appearing in $(Q(t))_{t \geq 0}$ whose intersection with the segment S_x is non-empty by elements of \mathcal{A} according to the following rule. By convention $R_\emptyset(x)$ is the unit square $[0, 1]^2$. The first point (τ_1, x_1, y_1) of Π splits $[0, 1]^2$ into four rectangles, a.s. only two of them intersect S_x , we denote the bottom rectangle by $R_0(x)$ and the top one by $R_1(x)$. Inductively, for every $u \in \mathcal{A}$, a point of Π eventually falls into $R_u(x)$, dividing it into four rectangles. Almost surely, only two of them intersect S_x , denote the bottom one by $R_{u0}(x)$ and the top one by $R_{u1}(x)$. For $u \in \mathcal{A}$, we denote the minimal (resp. maximal) horizontal coordinate of $R_u(x)$ by $G_u(x)$ (resp. $D_u(x)$), and define the *place* of x in $R_u(x)$ to be

$$X_u(x) = \frac{x - G_u(x)}{D_u(x) - G_u(x)}.$$

If $u \neq \emptyset$, we denote the parent of u by \overleftarrow{u} which is the word u without its last letter. We write $M_u(x)$ for the ratio of the (two-dimensional) Lebesgue measure $\text{Leb}(R_u(x))$ of $R_u(x)$ by the measure of $R_{\overleftarrow{u}}(x)$,

$$M_u(x) = \frac{\text{Leb}(R_u(x))}{\text{Leb}(R_{\overleftarrow{u}}(x))}.$$

We also set for all $x \in [0, 1]$, $M_\emptyset(x) = 1$. For $u \in \{0, 1\}$ and $t \geq 0$, we introduce the “subquadtrees” $Q_{u,x}(t) = \{R \in Q(t + \tau_1) : R \subset R_u(x)\}$. Then, for every $t \geq 0$, one has:

$$N_t(x) = \mathbf{1}_{t \geq \tau_1} + \mathbf{1}_{t \geq \tau_1} \sum_{u \in \{0, 1\}} \left(\#\{R \in Q_{u,x}(t - \tau_1) : R \cap S_x \neq \emptyset\} - 1 \right). \quad (3.1)$$

If R is a rectangle with sides parallel to the x and y axes, we denote by $\Phi_R : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ the only affine transformation that maps the bottom left vertex of R to $(0, 0)$, the bottom right vertex of R to $(1, 0)$ and the up left vertex of R to $(0, 1)$. It should be plain from properties of Poisson point measures that, conditionally on $(M_u(x), X_u(x), R_u(x))$, the

process $(\Phi_{R_u(x)}(Q_{u,x}(t)))_{t \geq 0}$ has the same distribution as the process $(\tilde{Q}(M_u(x)t))_{t \geq 0}$, where \tilde{Q} is an independent copy of Q . In particular, conditionally on $(M_u(x), X_u(x))$, the number of rectangles in $Q_{u,x}$ that intersect S_x (minus 1), viewed as a process of t , has the same distribution as the process $(\tilde{N}_{M_u(x)t}(X_u(x)))_{t \geq 0}$ where \tilde{N} is defined from \tilde{Q} is the same way as N is defined from Q . Since $M_0(x)$ and $M_1(x)$ have the same distribution, (3.1) yields

$$\mathbb{E}[N_t(x)] = \mathbb{P}(t \geq \tau_1) + 2\mathbb{E}\left[\tilde{N}_{M_0(x)(t-\tau_1)}(X_0(x))\right], \quad (3.2)$$

with the convention $\tilde{N}_t(x) = 0$ whenever $t < 0$. More generally, if we write $\mathfrak{z}_k \in \mathcal{A}$ for $\mathfrak{z}_k = 0 \dots 0$ repeated k times, then for every positive integer k ,

$$\mathbb{E}[N_t(x)] = g_k(t) + 2^k \mathbb{E}\left[\tilde{N}_{M_{\mathfrak{z}_1}(x) \dots M_{\mathfrak{z}_k}(x)t - F_k}(X_{\mathfrak{z}_k}(x))\right], \quad (3.3)$$

where g_k is a function such that $0 \leq g_k \leq 2^k - 1$ and F_k is a nonnegative random variable defined by

$$F_k = \sum_{i=1}^k \tilde{\tau}_i \prod_{j=i}^k M_{\mathfrak{z}_j}(x),$$

with $(\tilde{\tau}_i)_{i \geq 1}$ a sequence of independent exponential variables with parameter 1.

We now compute the joint distribution of $(M_0(x), X_0(x))$ which will be of great use throughout this work. If f is a nonnegative measurable function, easy calculations yield

$$\begin{aligned} & \mathbb{E}\left[f(M_0(x), X_0(x))\right] \\ &= \int_0^1 du \int_0^1 dv \left(\mathbf{1}_{x < u} f\left(uv, \frac{x}{u}\right) + \mathbf{1}_{x > u} f\left((1-u)v, \frac{x-u}{1-u}\right) \right) \\ &= \int_x^1 \frac{dy}{y} \int_0^{\frac{x}{y}} dm f(m, y) + \int_0^x \frac{dy}{1-y} \int_0^{\frac{1-x}{1-y}} dm f(m, y) \end{aligned} \quad (3.4)$$

$$\begin{aligned} &= \int_0^x dm \int_x^1 \frac{dy}{y} f(m, y) + \int_x^1 dm \int_x^{\frac{x}{m}} \frac{dy}{y} f(m, y) \\ &+ \int_0^{1-x} dm \int_0^x \frac{dy}{1-y} f(m, y) + \int_{1-x}^1 dm \int_{1-\frac{1-x}{m}}^x \frac{dy}{1-y} f(m, y). \end{aligned} \quad (3.5)$$

3.2.3 Depoissonization

The following lemma contains a large deviations argument that will enable us to shift results from the continuous-time model to the discrete-time one.

Lemma 3.1. *For every $\varepsilon > 0$, we have*

$$\mathbb{E}\left[\sup_{x \in [0,1]} |N_{\tau_n}(x) - N_n(x)|^2 \mathbf{1}_{\tau_n \notin [n(1-\varepsilon), n(1+\varepsilon)]}\right] \xrightarrow{n \rightarrow \infty} 0.$$

Proof. Note that for every $x \in [0, 1]$, $t \mapsto N_t(x)$ is non-decreasing and that $N_t(x)$ is at most the number of points fallen so far: $N_t(x) \leq \max\{i \in \mathbb{Z}_+ : \tau_i \leq t\}$. In particular $N_{\tau_n}(x) \leq n$, thus we have

$$\sup_{x \in [0,1]} |N_{\tau_n}(x) - N_n(x)|^2 \mathbf{1}_{\tau_n > n(1+\varepsilon)} \leq n^2 \mathbf{1}_{\tau_n > n(1+\varepsilon)}.$$

A large deviations argument ensures that $n^2 \mathbb{P}(\tau_n > n(1 + \varepsilon))$ tends to 0 as $n \rightarrow \infty$. On the other hand, applying the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} & \mathbb{E} \left[\sup_{x \in [0,1]} |N_{\tau_n}(x) - N_n(x)|^2 \mathbf{1}_{\tau_n < n(1-\varepsilon)} \right] \\ & \leq \sqrt{\mathbb{E}[(\max\{i \in \mathbb{Z}_+ : \tau_i \leq n\})^4]} \sqrt{\mathbb{P}(\tau_n < n(1 - \varepsilon))}. \end{aligned}$$

As $\mathbb{E}[(\max\{i \in \mathbb{Z}_+ : \tau_i \leq n\})^4] = O(n^4)$, large deviations ensure that the quantity in the right-hand side tends to 0 as $n \rightarrow \infty$. Finally, Lemma 3.1 is proved. \square

3.3 Particular cases and fragmentation theory

We give below the definition of a particular case of fragmentation process. For more details, we refer the reader to [Ber06, pages 6-65]. Let ν be a probability measure on $\{(s_1, s_2) : s_1 \geq s_2 > 0 \text{ and } s_1 + s_2 \leq 1\}$. A self-similar fragmentation $(\mathcal{F}_t)_{t \geq 0}$ with dislocation measure ν and index of self-similarity 1 is a Markov process with values in the set $\mathcal{S}^\downarrow = \{(s_1, s_2, \dots) : s_1 \geq s_2 \geq \dots \geq 0 \text{ and } \sum_i s_i \leq 1\}$ describing the evolution of the masses of particles that undergo fragmentation. The process is informally characterized as follows: if at time t we have $\mathcal{F}(t) = (s_1(t), s_2(t), \dots)$, then for every $i \geq 1$, the i -th ‘‘particle’’ of mass $s_i(t)$ lives an exponential time with parameter $s_i(t)$ before splitting into two particles of masses $r_1 s_i(t)$ and $r_2 s_i(t)$, where (r_1, r_2) has been sampled from ν independently of the past and of the other particles. In other words, each particle undergoes a self-similar fragmentation with time rescaled by its mass. In the next section we establish a link between fragmentation theory and the process $N_i(U)$, where U is a r.v. uniformly distributed over $[0, 1]$ and independent of $(Q(t))_{t \geq 0}$. This connection will provide a new proof of a result of [FGCR93] and [CH03]. See also [CLG11] for another recent application of fragmentation theory to a combinatorial problem where the exponent $\frac{\sqrt{17}-3}{2}$ appears.

3.3.1 The uniform case

We consider here the case where the point x is chosen at random uniformly over $[0, 1]$ and independently of $(Q(t))_{t \geq 0}$.

Proposition 3.1. *Let U be a random variable uniformly distributed over $[0, 1]$ and independent of the quadtree $(Q(t))_{t \geq 0}$. Let $u \in \mathcal{A}$ and denote by $u_0 = \emptyset, u_1, \dots, u_k = u$ its ancestors. Then $X_u(U)$ is uniform over $[0, 1]$ and independent of $(M_{u_1}(U), \dots, M_{u_k}(U))$, which is a sequence of independent random variables all having density $2(1 - m) \mathbf{1}_{m \in [0,1]}$.*

Proof. We prove Proposition 3.1 by induction on k . Let $u \in \mathcal{A}$. Denote by $u_0 = \emptyset, u_1, \dots, u_k = u$ its ancestors. Integrating (3.4) for $x \in [0, 1]$, we deduce that for every $v \in \{0, 1\}$, $X_v(U)$ and $M_v(U)$ are independent and distributed according to

$$\mathbf{1}_{u \in [0,1]} du \otimes \mathbf{1}_{m \in [0,1]} 2(1-m) dm. \quad (3.6)$$

Recalling that $Q_{u_1, U}(t) = \{R \in Q(t + \tau_1) : R \subset R_{u_1}(U)\}$, conditionally on $(X_{u_1}(U), M_{u_1}(U))$, the process $\Phi_{R_{u_1}(U)}(Q_{u_1, U})$ has the same distribution as $(\tilde{Q}(M_{u_1}(U)t))_{t \geq 0}$, where \tilde{Q} is an independent copy of Q . Since $X_{u_1}(U)$ is uniform over $[0, 1]$, we deduce by induction on the subquadtrees $Q_{u_1, U}$ that $X_u(U)$ is uniform over $[0, 1]$ and independent of $(M_{u_2}(U), \dots, M_{u_k}(U))$ which is a sequence of independent r.v. all having density $2(1-m)\mathbf{1}_{m \in [0,1]}$. Furthermore it is easy to see that

$$\mathbb{E}\left[(X_{u_i}(U), M_{u_i}(U))_{2 \leq i \leq k} \mid (X_{u_1}(U), M_{u_1}(U))\right] = \mathbb{E}\left[(X_{u_i}(U), M_{u_i}(U))_{2 \leq i \leq k} \mid X_{u_1}(U)\right].$$

Hence by (3.6), $X_u(U)$ is also independent of $M_{u_1}(U)$. \square

Letting $\mathbf{m}(t) = \mathbb{E}[N_t(U)]$, (recall that when $t < 0$, $N_t(x) = 0$ for all $x \in [0, 1]$) equation (3.2) becomes

$$\mathbf{m}(t) = \mathbb{P}(t \geq \tau_1) + 2\mathbb{E}[\mathbf{m}(M(t - \tau_1))], \quad (3.7)$$

where M is independent of τ_1 and has density $2(1-m)\mathbf{1}_{m \in [0,1]}$.

Proposition 3.2. *Let U be uniform over $[0, 1]$ and independent of $(Q(t))_{t \geq 0}$. We have the following convergence*

$$\lim_{t \rightarrow \infty} t^{-\beta^*} \mathbb{E}[N_t(U)] = \frac{\Gamma(2(\beta^* + 1))}{2\Gamma^3(\beta^* + 1)}, \quad \text{where } \beta^* = \frac{\sqrt{17} - 3}{2}.$$

Proof. We consider an auxiliary fragmentation process $(\mathcal{F}_t)_{t \geq 0}$ with index of self-similarity 1 and dislocation probability measure ν given by

$$\int \nu(ds_1, ds_2) f(s_1, s_2) = \mathbb{E}\left[f(M_1(U) \vee M_0(U), M_1(U) \wedge M_0(U))\right].$$

In other words, the dislocation measure is given by the law of the decreasing ordering of $\{M_0(U), M_1(U)\}$. More precisely $(\mathcal{F}_t)_{t \geq 0}$ takes its values in \mathcal{S}^\downarrow and satisfies the following equation in distribution which completely characterizes its law:

$$(\mathcal{F}_t) \stackrel{(d)}{=} \left((\mathbf{1}_{t < \tau}) \dot{+} \left(\mathbf{1}_{t \geq \tau} M_0(U) \cdot \mathcal{F}_{M_0(U)(t-\tau)}^{(0)} \right)_{t \geq 0} \dot{+} \left(\mathbf{1}_{t \geq \tau} M_1(U) \cdot \mathcal{F}_{M_1(U)(t-\tau)}^{(1)} \right)_{t \geq 0} \right)^\downarrow,$$

with $(\mathcal{F}_t^{(0)})_{t \geq 0}$ and $(\mathcal{F}_t^{(1)})_{t \geq 0}$ two independent copies of $(\mathcal{F}_t)_{t \geq 0}$ also independent of $(M_0(U), M_1(U), \tau)$ and τ an independent exponential variable with parameter 1. The symbol $\dot{+}$ means concatenation of sequences and $(\cdot)^\downarrow$ is the decreasing reordering (and erasing of zeros). Then, it is straightforward to see that the expectation of the number $\#\mathcal{F}_t$ of fragments of \mathcal{F}_t minus 1 satisfies the same equation as $\mathbb{E}[N_t(U)]$, namely letting $\mathbf{m}(t) = \mathbb{E}[\#\mathcal{F}_t - 1]$ for $t \geq 0$, and $\mathbf{m}(t) = 0$ for $t < 0$ we have

$$\mathbf{m}(t) = \mathbb{P}(t \geq \tau_1) + 2\mathbb{E}[\mathbf{m}(M(t - \tau_1))], \quad (3.8)$$

where M is independent of τ_1 and has density $2(1-m)\mathbf{1}_{m \in [0,1]}$. By (3.7) and (3.8), the functions \mathbf{m} and \mathbf{m} satisfy the same integral equation,

$$f(t) = 1 - e^{-t} + 2 \int_0^1 dm 2(1-m) \int_0^t ds e^{-s} f(m(t-s)).$$

Differentiating with respect to t , we see that both \mathbf{m} and \mathbf{m} are solutions of the Cauchy problem for the integro-differential equation

$$\begin{cases} \partial_t f(t) = 1 - f(t) + \int_0^1 dm 2(1-m) f(mt), \\ f(0) = 0. \end{cases}$$

Uniqueness of solution of this kind of integro-differential equation is known, see *e.g.* [IL97]. We deduce that for every $t \geq 0$, $\mathbf{m}(t) = \mathbf{m}(t)$. We now focus on $\mathbf{m}(t)$. Following [BG04, Section 3], we let for every $\beta > 0$, $\psi(\beta) = 1 - \int \nu(ds_1, ds_2)(s_1^\beta + s_2^\beta)$. An easy calculation yields:

$$\psi(\beta) = \frac{\beta^2 + 3\beta - 2}{(\beta + 1)(\beta + 2)}.$$

In particular the Malthusian exponent associated to ν , which is characterized by $\psi(\beta) = 0$ (see [Ber06, Section 1.2.2]), is

$$\beta^* = \frac{\sqrt{17} - 3}{2}.$$

Applying [BG04, Theorem 1], we get:

$$\begin{aligned} & \lim_{t \rightarrow \infty} t^{-\beta^*} \mathbb{E}[\#\mathcal{F}_t] \\ &= \frac{\Gamma(1 - \beta^*)}{\beta^*} \frac{4}{2\beta^* + 3} \prod_{k=1}^{\infty} \left(1 - \frac{\beta^*}{k}\right) \left(1 - \frac{\beta^*}{k + \sqrt{17}}\right) \left(1 + \frac{\beta^*}{k + 1}\right) \left(1 + \frac{\beta^*}{k + 2}\right). \end{aligned}$$

Finally, we use the Weierstrass identity for the gamma function: for every complex number $z \in \mathbb{C} \setminus \mathbb{Z}_-$,

$$\Gamma(z + 1) = e^{-\gamma z} \prod_{k=1}^{\infty} \left(1 + \frac{z}{k}\right)^{-1} e^{z/k},$$

where γ is the Euler–Mascheroni constant. We conclude that

$$\lim_{t \rightarrow \infty} t^{-\beta^*} \mathbb{E}[N_t(U)] = \frac{4}{\beta^*(2\beta^* + 3)} \frac{\Gamma(\sqrt{17} + 1)}{\Gamma(\sqrt{17} - \beta^* + 1)} \frac{1}{\Gamma^2(\beta^* + 2)} \frac{1}{1 + \beta^*/2} = \frac{\Gamma(2(\beta^* + 1))}{2\Gamma^3(\beta^* + 1)},$$

which completes the proof of the proposition. \square

Remark 3.1. One can derive the following equality in distribution from (3.1):

$$N_t(U) \stackrel{(d)}{=} \mathbf{1}_{\tau_1 \leq t} + N_{M_0(U)(t-\tau_1)}^{(0)}(X_0(U)) + N_{M_1(U)(t-\tau_1)}^{(1)}(X_1(U)),$$

where $(N_t^{(0)})_{t \geq 0}$ and $(N_t^{(1)})_{t \geq 0}$ are independent copies of the process $(N_t)_{t \geq 0}$. We have already noticed that $X_0(U)$ and $X_1(U)$ are also uniform and independent of $(N_t^{(0)})_{t \geq 0}$, of $(N_t^{(1)})_{t \geq 0}$ and of $(M_0(U), M_1(U))$. If $X_0(U)$ and $X_1(U)$ were independent, then $N_t(U)$ would satisfy the same distributional equation as $(\#\mathcal{F}_t - 1)_{t \geq 0}$. However, this is not the case since we have $X_0(U) = X_1(U)$. This explains why we had to work with expectations.

Corollary 3.1 ([FGCR93], [CH03]). We have

$$\lim_{n \rightarrow \infty} n^{-\beta^*} \mathbb{E}[\mathcal{N}_n(U)] = \frac{\Gamma(2(\beta^* + 1))}{2\Gamma^3(\beta^* + 1)}.$$

Proof. This is a straightforward application of Lemma 3.1 and Proposition 3.2. \square

Remark 3.2. Observe that Chern and Hwang [CH03] obtained a more precise asymptotic behavior of $\mathbb{E}[\mathcal{N}_n(U)]$. They proved that

$$\mathbb{E}[\mathcal{N}_n(U)] = \frac{\Gamma(2(\beta^* + 1))}{2\Gamma^3(\beta^* + 1)} n^{\beta^*} + O(1).$$

3.3.2 Case $x = 0$

As a further example of the connection with fragmentation theory, we derive asymptotics properties for $N_t(0)$. In this case, the sequence of the areas of the rectangles crossed by S_0 is a fragmentation process, enabling us to state a convergence of $N_t(0)$, once rescaled, in \mathbb{L}^2 . A convergence in mean has already been obtained in [FGCR93, Theorem 6] and [FLLS95].

Theorem 3.2. *The random variable*

$$\mathfrak{M}_t = \sum_{u \in \mathcal{A}} \text{Leb}(R_u(0))^{\sqrt{2}-1} \mathbf{1}_{R_u(0) \in \mathcal{Q}(t)}, \quad t \geq 0,$$

is a uniformly integrable martingale which converges almost surely to \mathfrak{M}_∞ as $t \rightarrow \infty$. The distribution of \mathfrak{M}_∞ is characterized by

$$\mathbb{E}[\mathfrak{M}_\infty] = 1 \quad \text{and} \quad \mathfrak{M}_\infty \stackrel{(d)}{=} M_0(0)^{\sqrt{2}-1} \mathfrak{M}_\infty^{(0)} + M_1(0)^{\sqrt{2}-1} \mathfrak{M}_\infty^{(1)}, \quad (3.9)$$

where $\mathfrak{M}_\infty^{(0)}$ and $\mathfrak{M}_\infty^{(1)}$ are two independent copies of \mathfrak{M}_∞ also independent of $(M_0(0), M_1(0))$. Furthermore, we have the following convergence in \mathbb{L}^2 :

$$t^{1-\sqrt{2}} N_t(0) \xrightarrow[t \rightarrow \infty]{} \frac{\Gamma(2\sqrt{2})}{\sqrt{2}\Gamma^3(\sqrt{2})} \mathfrak{M}_\infty.$$

Proof. It is easy to check from properties of Poisson measures that the rearrangement in decreasing order of the masses of the rectangles living at time t and intersecting S_0 ,

$$\left(\text{Leb}(R_u(0)) \mathbf{1}_{R_u(0) \in \mathcal{Q}(t)} \right)_{t \geq 0}^\downarrow,$$

is a self-similar fragmentation with index 1 and dislocation probability measure given by the decreasing ordering of $\{M_0(0), M_1(0)\}$. As in the proof of Proposition 3.2, we introduce for every $\beta > 0$, $\Psi(\beta) = 1 - \mathbb{E}[M_0(0)^\beta + M_1(0)^\beta]$, which is easily computed:

$$\Psi(\beta) = \frac{(\beta + 1)^2 - 2}{(\beta + 1)^2}.$$

Thus the Malthusian exponent p^* of this fragmentation satisfying $\Psi(p^*) = 0$ is

$$p^* = \sqrt{2} - 1.$$

The first two points of the theorem follow from classical results of fragmentation theory, see [Ber06, Theorem 1.1]. We refer to [Liu97] for the characterization of the law of \mathfrak{M}_∞ via the distributional equation (3.9) and to [Liu01] for some of its properties. The last point comes from [BG04, Corollary 6] and the Weierstrass identity for the gamma function used in a similar manner as in the proof of Proposition 3.2. \square

Corollary 3.2. *We have the following convergence in \mathbb{L}^2 :*

$$n^{1-\sqrt{2}}N_{\tau_n}(0) \xrightarrow[n \rightarrow \infty]{} \frac{\Gamma(2\sqrt{2})}{\sqrt{2}\Gamma^3(\sqrt{2})}\mathfrak{M}_\infty.$$

Proof. This proposition easily derives from Lemma 3.1 and Theorem 3.2. \square

Remark 3.3. Observe that Corollary 3.2 implies the following convergence in distribution:

$$n^{1-\sqrt{2}}\mathcal{N}_n(0) \xrightarrow[n \rightarrow \infty]{} \frac{\Gamma(2\sqrt{2})}{\sqrt{2}\Gamma^3(\sqrt{2})}\mathfrak{M}_\infty.$$

Remark 3.4. It is worthwhile to notice that the behavior of the cost of the partial match query in the case $x = 0$ is drastically different from its behavior in the case when x is uniform or x is fixed in $(0, 1)$ (see Theorem 3.1 and Proposition 3.2).

3.3.3 An *a priori* uniform bound

This section is devoted to the proof of an *a priori* uniform bound on $s^{-\beta^*}\mathbb{E}[N_s(x)]$ over $(x, s) \in (0, 1) \times (0, \infty)$ that will be useful in many places.

Lemma 3.2. *There exists $C < \infty$ such that*

$$\sup_{x \in (0, 1)} \sup_{s > 0} \mathbb{E} \left[s^{-\beta^*} N_s(x) \right] \leq C. \quad (3.10)$$

Proof. As a first step, we start by proving that there exists $C_1 < \infty$ such that for every $x \in (0, 1)$,

$$\sup_{s > 0} \mathbb{E} \left[s^{-\beta^*} N_s(x) \right] \leq \frac{C_1}{x \wedge (1-x)}. \quad (3.11)$$

Combining (3.2) with the densities computed in (3.4), we deduce that for every $x \in (0, 1)$

$$\begin{aligned} t^{-\beta^*}\mathbb{E}[N_t(x)] &= t^{-\beta^*}\mathbb{P}(t \geq \tau_1) + 2 \left(\int_x^1 \frac{dy}{y} \int_0^{\frac{x}{y}} dm \mathbb{E} [t^{-\beta^*} N_{m(t-\tau_1)}(y)] \right. \\ &\quad \left. + \int_0^x \frac{dy}{1-y} \int_0^{\frac{1-x}{1-y}} dm \mathbb{E} [t^{-\beta^*} N_{m(t-\tau_1)}(y)] \right). \end{aligned} \quad (3.12)$$

By the monotonicity of $t \mapsto N_t(x)$ we have $\mathbb{E} [t^{-\beta^*} N_{m(t-\tau_1)}(y)] \leq \mathbb{E} [t^{-\beta^*} N_t(y)]$. Furthermore, recalling that $\beta^* < 1$, there exists a constant C' such that for every $t > 0$,

$t^{-\beta^*} \mathbb{P}(t \geq \tau_1) \leq C'$. Hence

$$\begin{aligned} t^{-\beta^*} \mathbb{E}[N_t(x)] &\leq C' + 2 \left(\int_x^1 \frac{x dy}{y^2} \mathbb{E} [t^{-\beta^*} N_t(y)] + \int_0^x \frac{(1-x) dy}{(1-y)^2} \mathbb{E} [t^{-\beta^*} N_t(y)] \right) \\ &\leq C' + \frac{2}{x \wedge (1-x)} \int_0^1 dy \mathbb{E} [t^{-\beta^*} N_t(y)] \\ &= C' + \frac{2}{x \wedge (1-x)} \mathbb{E} [t^{-\beta^*} N_t(U)]. \end{aligned}$$

It has been shown in Proposition 3.2 that $\mathbb{E} [t^{-\beta^*} N_t(U)]$ has a finite limit as $t \rightarrow \infty$, and for every $t > 0$, $\mathbb{E} [N_t(U)] \leq t$. Thus the quantity $\mathbb{E} [t^{-\beta^*} N_t(U)]$ is bounded over $(0, \infty)$. The inequality (3.11) follows from these considerations.

Introducing $S(x) = \sup_{s>0} s^{-\beta^*} \mathbb{E}[N_s(x)]$ for every $x \in [0, 1]$, we have just shown that $S(x) \leq C_1(x \wedge (1-x))^{-1}$. Using (3.12), we have for every $x \in (1/2, 1)$:

$$\begin{aligned} S(x) &= \sup_{t>0} \left\{ t^{-\beta^*} \mathbb{P}(t \geq \tau_1) + 2 \left(\int_x^1 \frac{dy}{y} \int_0^{\frac{x}{y}} dm \mathbb{E} [t^{-\beta^*} N_{m(t-\tau_1)}(y)] \right. \right. \\ &\quad \left. \left. + \int_0^x \frac{dy}{1-y} \int_0^{\frac{1-x}{1-y}} dm \mathbb{E} [t^{-\beta^*} N_{m(t-\tau_1)}(y)] \right) \right\} \\ &\leq C' + 2 \sup_{t>0} \left\{ \int_x^1 \frac{dy}{y} \int_0^1 dm \mathbb{E} [t^{-\beta^*} N_t(y)] + \int_0^{1/2} \frac{dy}{1-y} \int_0^1 dm \mathbb{E} [t^{-\beta^*} N_t(y)] \right\} \\ &\quad + 2 \sup_{t>0} \int_{1/2}^x \frac{dy}{1-y} \int_0^{\frac{1-x}{1-y}} dm m^{\beta^*} \mathbb{E} [(mt)^{-\beta^*} N_{mt}(y)] \\ &\leq C' + 8 \sup_{t>0} \int_0^1 dy \mathbb{E} [t^{-\beta^*} N_t(y)] + 2 \int_{1/2}^x \frac{dy}{1-y} \int_0^{\frac{1-x}{1-y}} dm m^{\beta^*} S(y) \\ &\leq C_2 + \frac{2}{\beta^* + 1} (1-x)^{\beta^*+1} \int_{1/2}^x dy \frac{1}{(1-y)^{\beta^*+2}} S(y). \end{aligned} \tag{3.13}$$

Let us show that this implies that for every $x \in (0, 1)$, $S(x) \leq 100C_2$. Arguing by contradiction, suppose that there exists $a \in (1/2, 1)$ such that $S(a) > 100C_2$. Let $S = \sup_{x \in [1/2, a]} S(x)$. By (3.11), S is finite; there exists $b \in [1/2, a]$ such that $S(b) \geq 0.9S$. In particular, $S(b) \geq 0.9 \sup_{x \in [1/2, b]} S(x)$ and $S(b) > 90C_2$. Applying (3.13) at b , we get

$$\begin{aligned} S(b) &\leq 90^{-1} S(b) + \frac{2}{\beta^* + 1} (1-b)^{\beta^*+1} \int_{1/2}^b dy \frac{1}{(1-y)^{\beta^*+2}} 0.9^{-1} S(b) \\ &\leq 90^{-1} S(b) + \frac{2 \cdot 0.9^{-1}}{(\beta^* + 1)^2} S(b), \end{aligned}$$

leading to a contradiction since $(\beta^* + 1)^2 > \frac{2 \cdot 0.9^{-1}}{1-90^{-1}}$. Finally, $S(x) \leq 100C_2$ for every $x \in (0, 1)$. \square

3.4 The convergence at fixed $x \in (0, 1)$

We prove in this section that when $x \in [0, 1]$ is fixed, $t^{-\beta^*} \mathbb{E}[N_t(x)]$ admits a finite limit as $t \rightarrow \infty$. The results of the preceding section do not directly apply since the place $X_0(x)$ of x in the rectangle $R_0(x)$ highly depends on x . Recall notation \mathfrak{z}_k for the word composed of k zeros $0 \dots 0 \in \mathcal{A}$. The guiding idea is that the splittings tend to make $X_{\mathfrak{z}_k}(x)$ uniform and independent of $M_{\mathfrak{z}_k}(x)$.

3.4.1 A key Markov chain

Fix $x \in (0, 1)$. To simplify notation, for every $k \geq 1$, we write X_k for $X_{\mathfrak{z}_k}(x)$ and M_k for $M_{\mathfrak{z}_k}(x)$. We shall focus on the process $(X_k, M_k)_{k \geq 0}$, which is obviously a homogeneous Markov chain starting from $(x, 1)$ whose transition probability is given by (3.4) or (3.5). Let $k \geq 1$. We denote by \mathcal{F}_k the filtration generated by $(X_i, M_i)_{1 \leq i \leq k}$. It is easy to see that the transition probability only depends on X_k , that is

$$\mathbb{E}[(X_{k+i}, M_{k+i})_{i \geq 1} | \mathcal{F}_k] = \mathbb{E}[(X_{k+i}, M_{k+i})_{i \geq 1} | X_k].$$

Proposition 3.1. *Fix $x \in (0, 1)$. There exists a coupling of the chain $(X_k, M_k)_{k \geq 0}$ with a random time $T \in \mathbb{Z}_+$ such that for any $k \geq 0$, conditionally on $\{T \leq k\}$, the r.v. X_k is uniformly distributed over $[0, 1]$, independent of $(M_i)_{1 \leq i \leq k}$ and of T . Furthermore, we have*

$$\mathbb{E}[1.15^T] < +\infty.$$

Proof. For any $k \geq 1$ we consider the event

$$E_k = \{M_k < X_{k-1} \wedge (1 - X_{k-1})\}.$$

Using the explicit densities (3.4) and (3.5), one sees that conditionally on \mathcal{F}_{k-1} and on the event E_k of probability $-(X_{k-1} \wedge (1 - X_{k-1})) \ln(X_{k-1}(1 - X_{k-1}))$, the conditional distribution of X_k is

$$\frac{1}{-\ln(X_{k-1}(1 - X_{k-1}))} \left(\frac{1}{1-y} \mathbf{1}_{y \in (0, X_{k-1})} + \frac{1}{y} \mathbf{1}_{y \in (X_{k-1}, 1)} \right) dy.$$

In particular, conditionally on E_k and \mathcal{F}_{k-1} , the variable X_k is independent of M_k and has a density bounded from below by $-1/\ln(X_{k-1}(1 - X_{k-1}))$. Thus, we can construct simultaneously with $(X_k, M_k)_{k \geq 0}$ a sequence of random variables $(B_k)_{k \geq 0} \in \{0, 1\}^{\mathbb{Z}_+}$ as follows. Suppose that we have constructed $(X_i, M_i, B_i)_{0 \leq i \leq k-1}$. Then independently of \mathcal{F}_{k-1} , toss a Bernoulli variable of parameter $-(X_{k-1} \wedge (1 - X_{k-1})) \ln(X_{k-1}(1 - X_{k-1}))$. If 0 comes out, we consider that we are on the event E_k^c , then put $B_k = 0$ and sample (X_k, M_k) with the conditional distribution on E_k^c and \mathcal{F}_{k-1} . If 1 comes out, we consider that we are on the event E_k and we proceed to the following.

1. First sample M_k from its distribution conditionally on E_k and \mathcal{F}_{k-1} .
2. Then independently of M_k , toss a Bernoulli variable B_k of parameter

$$\frac{1}{-\ln(X_{k-1}(1 - X_{k-1}))}.$$

If $B_k = 1$, sample X_k uniformly from $[0, 1]$ and independently of (M_1, \dots, M_k) . Otherwise, sample X_k with density

$$\frac{1}{-\ln(X_{k-1}(1-X_{k-1})) - 1} \left(\left(\frac{1}{1-y} - 1 \right) \mathbf{1}_{y \in (0, X_{k-1})} + \left(\frac{1}{y} - 1 \right) \mathbf{1}_{y \in (X_{k-1}, 1)} \right) dy,$$

independently of (M_1, \dots, M_k) .

The device provides us with a Markov chain $(X_k, M_k, B_k)_{k \geq 0}$ such that the first two coordinates have the law of the process introduced before Proposition 3.1. We then let

$$T = \inf\{k \geq 0, B_k = 1\}.$$

By definition of T , the random variable X_T is sampled uniformly over $[0, 1]$ and independently of (M_1, \dots, M_T) . We deduce that the process $(X_{T+i}, M_{T+i})_{i \geq 1}$ has the same distribution as the process $(X_{\mathfrak{J}_k}(U), M_{\mathfrak{J}_k}(U))_{k \geq 1}$ defined in Proposition 3.1, hence an easy adaptation of Proposition 3.1 shows that for every positive integer i , X_{T+i} is uniformly distributed over $[0, 1]$ independent of (M_1, \dots, M_{T+i}) and of T . This proves the first part of Proposition 3.1.

For the second part, we need to evaluate the tail of the random time T . We introduce the following variation. Let $(\hat{X}_k)_{k \geq 0}$ be a Markov chain with space state $(0, 1) \cup \{\partial\}$, where ∂ is a cemetery point. Informally, this chain is the chain (X_k) until we reach the time T , then it is killed and sent to the cemetery point. Thanks to the calculation presented at the beginning of the proof, it should be clear that given X_{k-1} and conditionally on $\{T \geq k-1\}$, the probability of the event $\{T = k\}$ is $X_{k-1} \wedge (1 - X_{k-1})$. Thus the transition probability for the chain (\hat{X}_k) is defined as follows: for every $x \in (0, 1)$,

$$p(x, dy) = x \wedge (1-x) \delta_\partial + \left(\frac{1-x}{(1-y)^2} \mathbf{1}_{y \in (0, x)} + \frac{x}{y^2} \mathbf{1}_{y \in (x, 1)} - x \wedge (1-x) \right) dy,$$

and $p(\partial, dy) = \delta_\partial$. By construction of this chain, the stopping time $\hat{T} = \inf\{k \geq 1 : \hat{X}_k = \partial\}$ has the same distribution as T . In order to estimate \hat{T} , we define the following potential function $V : (0, 1) \cup \{\partial\} \rightarrow [1, \infty]$:

$$V(x) = \mathbf{1}_{x=\partial} + \frac{10}{\sqrt{x}} \mathbf{1}_{x \in (0, 1/2)} + \frac{10}{\sqrt{1-x}} \mathbf{1}_{x \in [1/2, 1]}.$$

Then one can show that for every $x \in (0, 1) \cup \{\partial\}$,

$$\int p(x, dy) V(y) \leq 0.85V(x) + \mathbf{1}_{\{\partial\}}(x),$$

so that [MT09, Theorem 15.2.5] may be applied: there exists $\varepsilon > 0$ such that for all $x \in (0, 1)$,

$$\mathbb{E} \left[\sum_{k=0}^{\hat{T}-1} V(\hat{X}_k) 1.15^k \right] \leq \varepsilon^{-1} 1.15^{-1} V(x),$$

from which we deduce that

$$\mathbb{E} \left[1.15^{\hat{T}} \right] < \infty$$

(note that the last quantity is not uniformly bounded for $x \in (0, 1)$). This completes the proof of Proposition 3.1. \square

In the remaining part of this section, x is fixed in $(0, 1)$. Coming back to (3.3) and writing $\overline{M}_k = M_1 M_2 \dots M_k$ for the Lebesgue measure of $R_{3^k}(x)$, we have

$$\begin{aligned} & t^{-\beta^*} \mathbb{E}[N_t(x)] \\ &= t^{-\beta^*} \left(g_k(t) + 2^k \mathbb{E} \left[\tilde{N}_{\overline{M}_k t - F_k}(X_k) \mathbf{1}_{T > k} \right] + 2^k \mathbb{E} \left[\tilde{N}_{\overline{M}_k t - F_k}(X_k) \mathbf{1}_{T \leq k} \right] \right). \end{aligned} \quad (3.14)$$

We shall treat separately the last two terms of (3.14).

3.4.2 Study of $t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\overline{M}_k t - F_k}(X_k) \mathbf{1}_{T > k}]$

We shall see that $t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\overline{M}_k t - F_k}(X_k) \mathbf{1}_{T > k}]$ is arbitrarily small uniformly in t provided that the integer k is chosen large enough. Observe

$$\begin{aligned} & t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\overline{M}_k t - F_k}(X_k) \mathbf{1}_{T > k}] \\ &\leq t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\overline{M}_k t}(X_k) \mathbf{1}_{T > k}] \\ &= 2^k \mathbb{E} \left[\overline{M}_k^{\beta^*} (\overline{M}_k t)^{-\beta^*} \tilde{N}_{\overline{M}_k t}(X_k) \mathbf{1}_{T > k} \right] \\ &= 2^k \mathbb{E} \left[\overline{M}_k^{\beta^*} \mathbf{1}_{T > k} \mathbb{E} \left[(\overline{M}_k t)^{-\beta^*} \tilde{N}_{\overline{M}_k t}(X_k) \mid \sigma(\overline{M}_k, X_k, T) \right] \right]. \end{aligned}$$

Letting ϕ be the map $(s, u) \mapsto \mathbb{E}[s^{-\beta^*} N_s(u)]$, we have:

$$t^{-\beta^*} 2^k \mathbb{E}[\tilde{N}_{\overline{M}_k t - F_k}(X_k) \mathbf{1}_{T > k}] \leq 2^k \mathbb{E} \left[\overline{M}_k^{\beta^*} \mathbf{1}_{T > k} \phi(\overline{M}_k t, X_k) \right].$$

Thanks to (3.10), $\phi \leq C$, so that the quantity in the last display is at most

$$C 2^k \mathbb{E} \left[\overline{M}_k^{\beta^*} \mathbf{1}_{T > k} \right].$$

Hölder's inequality yields for every $p > 1$

$$C 2^k \mathbb{E} \left[\overline{M}_k^{\beta^*} \mathbf{1}_{T > k} \right] \leq C 2^k \mathbb{E} \left[\overline{M}_k^{\beta^* p} \right]^{1/p} \mathbb{E} \left[\mathbf{1}_{T > k} \right]^{1-1/p}.$$

The last term is easily treated, by Markov's inequality we have

$$\mathbb{E} \left[\mathbf{1}_{T > k} \right] \leq 1.15^{-k} \mathbb{E}[1.15^T].$$

Concerning $\mathbb{E}[\overline{M}_k^{\beta^* p}]$ we have

$$\begin{aligned} \mathbb{E} \left[\overline{M}_k^{\beta^* p} \right] &\leq \mathbb{E} \left[M_{3^2}(x)^{\beta^* p} \dots M_{3^k}(x)^{\beta^* p} \right] \\ &= \int_0^1 f^{(x)}(y) dy \mathbb{E} \left[M_{3^1}(y)^{\beta^* p} \dots M_{3^{k-1}}(y)^{\beta^* p} \right], \end{aligned}$$

where $f^{(x)}$ is the density of X_1 under \mathbb{P} . It is easy to see from (3.4) that $f^{(x)}$ is bounded from above by $(x \wedge (1-x))^{-1}$. Hence

$$\mathbb{E} \left[\overline{M}_k^{\beta^* p} \right] \leq \frac{1}{x \wedge (1-x)} \int_0^1 dy \mathbb{E} \left[\overline{M}_{k-1}(y)^{\beta^* p} \right].$$

Recall from Proposition 3.1 that when $x = U$ is uniform over $[0, 1]$ and independent of $(Q(t))_{t \geq 0}$, then $M_{j_1}(U), \dots, M_{j_k}(U)$ are independent and distributed according to $\mathbf{1}_{m \in [0, 1]} 2(1 - m)dm$. In particular

$$\mathbb{E} \left[M_0(U)^{\beta^* p} \right] = \frac{2}{(\beta^* p + 1)(\beta^* p + 2)}$$

and thus

$$\int_0^1 dy \mathbb{E} \left[\overline{M}_{k-1}(y)^{\beta^* p} \right] = \left(\frac{2}{(\beta^* p + 1)(\beta^* p + 2)} \right)^{k-1}.$$

Gathering all these estimates, we obtain

$$\begin{aligned} & t^{-\beta^*} 2^k \mathbb{E} \left[N_{\overline{M}_k t - F_k}(X_k) \mathbf{1}_{T > k} \right] \\ & \leq C 2^k \left(\frac{1}{x \wedge (1 - x)} \right)^{1/p} \left(\frac{2}{(\beta^* p + 1)(\beta^* p + 2)} \right)^{(k-1)/p} \mathbb{E} \left[1.15^T \right]^{1-1/p} 1.15^{-k(1-1/p)} \\ & = K_{p,x} \left(2 \left\{ \frac{2}{(\beta^* p + 1)(\beta^* p + 2)} \right\}^{1/p} 1.15^{1/p-1} \right)^k, \end{aligned}$$

where $K_{p,x}$ is a constant that only depends on p and x but not on k . Now, one can easily prove that for $p > 1$ sufficiently close to 1, the term in the curly brackets in the last display becomes strictly less than 1. Consequently, letting $\varepsilon > 0$ fixed, there exists an integer k sufficiently large such that for every $t > 0$,

$$t^{-\beta^*} 2^k \mathbb{E} \left[N_{\overline{M}_k t - F_k}(X_k) \mathbf{1}_{T > k} \right] \leq \varepsilon. \quad (3.15)$$

3.4.3 Conclusion

Observe that we have for every $t > 0$

$$\begin{aligned} & t^{-\beta^*} 2^k \mathbb{E} \left[\tilde{N}_{\overline{M}_k t - F_k}(X_k) \mathbf{1}_{T \leq k} \right] \\ & = 2^k \mathbb{E} \left[\mathbf{1}_{T \leq k} \mathbb{E} \left[t^{-\beta^*} \tilde{N}_{\overline{M}_k t - F_k}(X_k) \middle| \sigma(\overline{M}_k, F_k, T) \right] \right] \\ & = 2^k \mathbb{E} \left[\mathbf{1}_{T \leq k} (\overline{M}_k - t^{-1} F_k)_+^{\beta^*} \mathbb{E} \left[(\overline{M}_k t - F_k)_+^{-\beta^*} \tilde{N}_{\overline{M}_k t - F_k}(X_k) \middle| \sigma(\overline{M}_k, F_k, T) \right] \right], \end{aligned}$$

where y_+ denotes $y \vee 0$. By Proposition 3.1, on the event $\{T \leq k\}$, the r.v. X_k is uniformly distributed over $[0, 1]$ and independent of M_1, \dots, M_k thus of \overline{M}_k . It is also independent of F_k and T . Hence, letting θ be the map $s \mapsto \mathbb{E}[s_+^{-\beta^*} N_s(U)]$, where U is a random variable uniformly distributed on $(0, 1)$ independent of N , we have:

$$t^{-\beta^*} 2^k \mathbb{E} \left[\tilde{N}_{\overline{M}_k t - F_k}(X_k) \mathbf{1}_{T \leq k} \right] = 2^k \mathbb{E} \left[\mathbf{1}_{T \leq k} (\overline{M}_k - t^{-1} F_k)_+^{\beta^*} \theta(\overline{M}_k t - F_k) \right].$$

Applying Proposition 3.2, $\theta(\overline{M}_k t - F_k)$ a.s. tends to a finite limit as $t \rightarrow \infty$. Hence by dominated convergence $t^{-\beta^*} 2^k \mathbb{E} \left[N_{\overline{M}_k t - F_k}(X_k) \mathbf{1}_{T \leq k} \right]$ has a finite limit as $t \rightarrow \infty$. We deduce from this fact, (3.14) and (3.15) that

$$\limsup_{t \rightarrow \infty} t^{-\beta^*} \mathbb{E} [N_t(x)] - \liminf_{t \rightarrow \infty} t^{-\beta^*} \mathbb{E} [N_t(x)] \leq \varepsilon.$$

Since that inequality holds for every $\varepsilon > 0$, $t^{-\beta^*} \mathbb{E}[N_t(x)]$ has a finite limit as $t \rightarrow \infty$ which we denote by $n_\infty(x)$:

$$n_\infty(x) = \lim_{t \rightarrow \infty} t^{-\beta^*} \mathbb{E}[N_t(x)].$$

3.5 Identifying the limit

In this section, we show that $x \mapsto n_\infty(x)$ is proportional to $x \mapsto (x(1-x))^{\beta^*/2}$ using a fixed point argument for integral equation (see also [CLG11, Section 4.1] for a similar application). The normalizing constant will come from the \mathbb{L}^1 -norm of $x \mapsto (x(1-x))^{\beta^*/2}$ and the constant of Proposition 3.2.

Combining (3.2) with the densities computed in (3.4), we deduce that

$$\begin{aligned} t^{-\beta^*} \mathbb{E}[N_t(x)] &= t^{-\beta^*} \mathbb{P}(t \geq \tau_1) + 2 \left(\int_x^1 \frac{dy}{y} \int_0^{\frac{x}{y}} dmm^{\beta^*} \mathbb{E}[(mt)^{-\beta^*} N_{m(t-\tau_1)}(y)] \right. \\ &\quad \left. + \int_0^x \frac{dy}{1-y} \int_0^{\frac{1-x}{1-y}} dmm^{\beta^*} \mathbb{E}[(mt)^{-\beta^*} N_{m(t-\tau_1)}(y)] \right). \end{aligned}$$

Thanks to Lemma 3.2, we get by dominated convergence

$$n_\infty(x) = \frac{2}{\beta^* + 1} \left(x^{\beta^*+1} \int_x^1 dy \frac{1}{y^{\beta^*+2}} n_\infty(y) + (1-x)^{\beta^*+1} \int_0^x dy \frac{1}{(1-y)^{\beta^*+2}} n_\infty(y) \right).$$

In other words, if we define

$$g_x(y) = \frac{2}{\beta^* + 1} \left(x^{\beta^*+1} \frac{1}{y^{\beta^*+2}} \mathbf{1}_{x < y < 1} + (1-x)^{\beta^*+1} \frac{1}{(1-y)^{\beta^*+2}} \mathbf{1}_{0 < y < x} \right),$$

we have

$$n_\infty(x) = \int_0^1 dy g_x(y) n_\infty(y).$$

Let G be the operator that maps a function $f \in \mathbb{L}^1[0, 1]$ to the function

$$G(f)(x) = \int_0^1 dy g_x(y) f(y).$$

In particular, n_∞ is a fixed point of G . It is easy to check that $x \in (0, 1) \mapsto g_x(\cdot) \in \mathbb{L}^1[0, 1]$ is continuous for the \mathbb{L}^1 -norm. Furthermore, Lemma 3.2 ensures that $|n_\infty(x)| \leq C$ for every $x \in (0, 1)$. As a consequence, $x \mapsto n_\infty(x)$ is continuous over $(0, 1)$. An easy computation shows that for every $y \in (0, 1)$, $\int_0^1 dx g_x(y) = 1$. Let p be another fixed point of G having the same integral as n_∞ . Then

$$\begin{aligned} \int_0^1 dx |n_\infty(x) - p(x)| &= \int_0^1 dx \left| \int_0^1 dy g_x(y) (n_\infty - p)(y) \right| \\ &\leq \int_0^1 dx \int_0^1 dy g_x(y) |n_\infty(y) - p(y)| \\ &= \int_0^1 dy |n_\infty(y) - p(y)|, \end{aligned}$$

which shows that the inequality is in fact an equality. Hence $n_\infty - p$ has a.e. a constant sign. As we know that the integral of $n_\infty - p$ is zero, we deduce that $n_\infty = p$ a.e. Straightforward calculations prove that $p_0 : x \mapsto (x(1-x))^{\beta^*/2}$ is also a fixed point of G of \mathbb{L}^1 -norm, so that

$$n_\infty(x) = \|n_\infty\|_1 \|p_0\|_1^{-1} (x(1-x))^{\beta^*/2} \quad \text{a.e.}$$

Since n_∞ and p_0 are continuous, we can remove the a.e. statement (observe that $n_\infty(0) = n_\infty(1) = 0$ by Theorem 3.2). Plainly,

$$\|p_0\|_1 = \frac{\Gamma^2\left(\frac{\beta^*}{2} + 1\right)}{\Gamma(\beta^* + 2)}.$$

On the other hand, (3.10) and the dominated convergence theorem ensure that $\|n_\infty\|_1 = \lim_{t \rightarrow \infty} t^{-\beta^*} \mathbb{E}[N_t(U)]$, which was computed in Proposition 3.2:

$$\|n_\infty\|_1 = \frac{\Gamma(2(\beta^* + 1))}{2\Gamma^3(\beta^* + 1)}.$$

Proof of Theorem 3.1. To sum up, we have for every $x \in [0, 1]$:

$$t^{-\beta^*} \mathbb{E}[N_t(x)] \xrightarrow{t \rightarrow \infty} \frac{\Gamma(2\beta^* + 2) \Gamma(\beta^* + 2)}{2\Gamma^3(\beta^* + 1) \Gamma^2\left(\frac{\beta^*}{2} + 1\right)} (x(1-x))^{\beta^*/2}.$$

Applying Lemma 3.1, Theorem 3.1 is shown. \square

3.6 Extensions and comments

3.6.1 Various convergences

In this paper, we only proved a convergence in mean of $t^{-\beta^*} N_t(x)$. We may wonder whether this quantity also converges in distribution, in probability, or even almost surely. A more interesting question is the following: does the process $((t^{-\beta^*} N_t(x))_{x \in [0,1]}, t > 0)$ converge in distribution in the Skorokhod sense to a random function $(\mathcal{C}(x))_{x \in [0,1]}$ as $t \rightarrow \infty$? Observe that if it does, then there exists a random point U uniformly distributed over $(0, 1)$ such that $\mathcal{C}(U) = 0$, U corresponding to the point x_1 of the first atom of Π ($N_t(x_1)$ is indeed of order $t^{\sqrt{2}-1}$ by Theorem 3.2).

Conjecture 3.1. *We have the functional limit law $(t^{-\beta^*} N_t(x))_{x \in [0,1]} \rightarrow (\mathcal{C}(x))_{x \in [0,1]}$ as $t \rightarrow \infty$ in $(\mathbb{D}([0,1]), \|\cdot\|_\infty)$, where \mathcal{C} satisfies the distributional fixed point equation*

$$\begin{aligned} (\mathcal{C}(x))_{x \in [0,1]} \stackrel{(d)}{=} & \left(\mathbf{1}_{x < U_0} \left\{ (U_0 U_1)^{\beta^*} \mathcal{C}^{(00)}\left(\frac{x}{U_0}\right) + (U_0(1-U_1))^{\beta^*} \mathcal{C}^{(01)}\left(\frac{x}{U_0}\right) \right\} \right. \\ & + \mathbf{1}_{x > U_0} \left\{ ((1-U_0)U_1)^{\beta^*} \mathcal{C}^{(10)}\left(\frac{x-U_0}{1-U_0}\right) \right. \\ & \left. \left. + ((1-U_0)(1-U_1))^{\beta^*} \mathcal{C}^{(11)}\left(\frac{x-U_0}{1-U_0}\right) \right\} \right)_{x \in [0,1]}, \end{aligned}$$

where $U_0, U_1, \mathcal{C}^{(00)}, \mathcal{C}^{(01)}, \mathcal{C}^{(10)}, \mathcal{C}^{(11)}$ are independent, U_0 and U_1 are uniformly distributed on $[0, 1]$ and $\mathcal{C}^{(00)}, \mathcal{C}^{(01)}, \mathcal{C}^{(10)}, \mathcal{C}^{(11)}$ have all the same distribution as \mathcal{C} .

3.6.2 Multidimensional case

The strategy adopted in Section 3.3.1 may be generalized to higher dimensions. As for the convergence in mean of the number of hyper-rectangles crossed by a *fixed* affine subspace having a direction generated by some vectors of the canonical basis, our approach may also be followed.

3.6.3 Quadtree as a model of random geometry

On top of its numerous applications in theoretical computer science, the model of random quadtree may be considered as a model of random geometry. More precisely one can view, for $t \geq 0$, the set of rectangles $Q(t)$ as a random graph, assigning length 1 to each edge of the rectangles. We denote this graph by $\tilde{Q}(t)$. A natural question would be to understand the metric behavior of $\tilde{Q}(t)$ as $t \rightarrow \infty$. The study of the graph distance L_t in $\tilde{Q}(t)$ between the upper-left and upper-right corners would be a first step in understanding the global geometry of \tilde{Q}_t . Observe that Theorem 3.2 already shows that L_t is less than the order $t^{\sqrt{2}-1}$.

CHAPTER 4

THE COMPONENT SIZES OF A CRITICAL RANDOM GRAPH WITH GIVEN DEGREE SEQUENCE

Consider a critical random multigraph \mathcal{G}_n with n vertices constructed by the configuration model such that its vertex degrees are independent random variables with the same distribution ν (criticality means that the second moment of ν is finite and equals twice its first moment). We specify the scaling limits of the ordered sequence of component sizes of \mathcal{G}_n as n tends to infinity in different cases. When ν has finite third moment, the components sizes rescaled by $n^{-2/3}$ converge to the excursion lengths of a Brownian motion with parabolic drift above past minima, whereas when ν is a power law distribution with exponent $\gamma \in (3, 4)$, the components sizes rescaled by $n^{-(\gamma-2)/(\gamma-1)}$ converge to the excursion lengths of a certain drifted process with independent increments above past minima. We deduce the asymptotic behavior of the component sizes of a critical random simple graph when ν has finite third moment.

Contents

4.1	Introduction	82
4.2	Formulation of the results	84
4.3	The depth-first search	86
4.3.1	An algorithmic construction of \mathcal{G}_n	86
4.3.2	The depth-first walk	87
4.4	Convergence of the walk \bar{s}_n	89
4.5	Study of the cycle half-edges	94
4.5.1	Upper bound of the number of cycle half-edges	94
4.5.2	End of the proof of Theorem 4.3	95

4.6	Study of the random simple graph \mathcal{SG}_n	96
4.6.1	Time arrival of loops and multiple edges	96
4.6.2	End of the proof of Theorem 4.4	97
4.7	The power law distribution setting	98
4.7.1	The Poissonian argument	98
4.7.2	The main results	102
4.7.3	Open questions	106

4.1 Introduction

The classical random graph model $G(n, p)$ has received a lot of attention since its introduction by Erdős and Rényi [ER60], especially because of the existence of a phase transition. In this model, a graph on n labeled vertices is constructed randomly by joining any pair of vertices by an edge with probability p , independently of the other pairs. For large n , the structure of this random graph depends on the value of np : for $p \sim c/n$ with $c < 1$, the largest connected component contains $O(\ln n)$ vertices, whereas when $p \sim c/n$ with $c > 1$, the largest component has $\Theta(n)$ vertices while the second largest component has $O(\ln n)$ vertices. The cases $c < 1$ and $c > 1$ are called subcritical and supercritical respectively. Much attention has been devoted to the critical case $p \sim 1/n$. When p is exactly equal to $1/n$, the largest components of $G(n, p)$ have sizes of order $n^{2/3}$.

Molloy and Reed [MR95] showed that a random graph with a given degree sequence exhibits a similar phase transition. More precisely, let $\mathbf{d}^{(n)} = (d_i^{(n)})_{1 \leq i \leq n}$ be a sequence of positive integers such that $\sum_{i=1}^n d_i^{(n)}$ is even. Let $G(n, \mathbf{d}^{(n)})$ be a random graph on n labeled vertices with degree sequence $\mathbf{d}^{(n)}$, uniformly chosen among all possibilities (tacitly assuming that there exists any such graph). Suppose that there exists a probability distribution $(\nu_k)_{k \geq 1}$ such that $\#\{i : d_i^{(n)} = k\}/n \rightarrow \nu_k$ as $n \rightarrow \infty$. Let $\omega(n)$ be the highest degree in the graph. Under some further strong conditions on the sequences $\mathbf{d}^{(n)}$, Molloy and Reed proved that if $Q = \sum_{k=1}^{\infty} k(k-2)\nu_k < 0$ and $\omega(n) \leq n^{1/8-\varepsilon}$ for some $\varepsilon > 0$, then with probability tending to 1, the size of the largest component of $G(n, \mathbf{d}^{(n)})$ is $O(\omega^2(n) \ln n)$, whereas if $Q > 0$ and $\omega(n) \leq n^{1/4-\varepsilon}$ for some $\varepsilon > 0$, then with probability tending to 1, the size of the largest component is $\Theta(n)$, and if furthermore Q is finite, the size of the second largest component is $O(\ln n)$.

More recently, the near-critical behavior of such graphs has been studied. When $Q = 0$, the structure of $G(n, \mathbf{d}^{(n)})$ depends on how fast the quantity

$$\alpha_n = \frac{\sum_{k=1}^{\infty} k(k-2) \#\{i : d_i^{(n)} = k\}}{n} = \frac{\sum_{i=1}^n d_i^{(n)}(d_i^{(n)} - 2)}{n}$$

converges to 0 (see Kang and Seierstad [KS08]). Requiring a fourth moment condition, Janson and Luczak [JL09] proved that if $n^{1/3}\alpha_n \rightarrow \infty$, then the size of the largest component of $G(n, \mathbf{d}^{(n)})$ divided by $n\alpha_n$ converges in probability to $\frac{2\mu}{\beta}$, while the size of the second largest component of $G(n, \mathbf{d}^{(n)})$ divided by $n\alpha_n$ converges in probability

to 0, where $\mu = \sum_{k=1}^{\infty} k\nu_k$ and $\beta = \sum_{k=3}^{\infty} k(k-1)(k-2)\nu_k \in (0, \infty)$. Furthermore, they noticed that their results can also be applied to some other random graph models by conditioning on the vertex degrees, provided that the random graph conditioned on the degree sequence has a uniform distribution over all possibilities. This is the case for $G(n, p)$ with $np \rightarrow 1$ and $n^{1/3}(np-1) \rightarrow \infty$. Note that if $n^{1/3}(np-1) = O(1)$, it is well-known that the largest component and the second largest component have both sizes of the same order $n^{2/3}$, so that their results do not hold.

A major difficulty when dealing with the natural random graph $G(n, \mathbf{d}^{(n)})$ is that, despite its straightforward definition, it cannot be constructed via an easy algorithm. To circumvent that obstacle, it is convenient to work with *multigraphs*, in which multiple edges and loops are allowed, using the explicit procedure provided by the *configuration model*, which was introduced by Bender and Canfield [BC78] and later studied by Bollobás [Bol80] and Wormald [Wor78]. See also Molloy and Reed [MR95, MR98], Kang and Seierstad [KS08], Bertoin and Sidoravicius [BS09], van der Hofstad [vdH10], Hatami and Reed [HM10]. Specifically, take a set of $d_i^{(n)}$ half-edges for the vertex with label i , $i \in \{1, \dots, n\}$, and combine the half-edges into pairs by a uniformly random matching of the set of all half-edges. Observing that every graph $G(n, \mathbf{d}^{(n)})$ may be constructed through the same number, $d_1^{(n)}! \dots d_n^{(n)}!$, of pairing of half-edges, we get that conditional on being a (simple) graph, the multigraph obtained by the configuration model has the same distribution as $G(n, \mathbf{d}^{(n)})$. That is why we shall first deal with multigraphs. We shall then see how to derive results for simple graphs.

The present work is devoted to the study inside the critical window. We suppose that we are given a probability distribution $\nu = (\nu_k)_{k \geq 1}$ with finite second moment such that $\nu_2 < 1$ and $\sum_{k=1}^{\infty} k(k-2)\nu_k = 0$. Let D be a random variable with distribution ν . The multigraph \mathcal{G}_n consisting of n vertices is defined by the configuration model as follows. Let D_1, D_2, \dots, D_n be n independent copies of D . Condition on $\sum_{i=1}^n D_i$ being even. Take a set of D_i half-edges for each vertex, and combine the half-edges into pairs by a uniformly random matching of the set of all half-edges. The random multigraph that this construction leads to is denoted by \mathcal{G}_n . We aim at specifying the asymptotics of the ordered sequence \mathbf{C}_n^ν of component sizes of \mathcal{G}_n in two different settings. First, we shall study the case when ν has finite third moment. We shall prove that $n^{-2/3}\mathbf{C}_n^\nu$ then converges in distribution (with respect to a certain topology that will be detailed below) as $n \rightarrow \infty$ to the ordered sequence of the excursion lengths of a Brownian motion with parabolic drift. This should be viewed as an extension of Aldous' well-known result for the critical behavior of Erdős-Renyi random graphs (see [Ald97]). Next the case when ν is a power law distribution with exponent $\gamma \in (3, 4)$ will be studied. We shall show that $n^{-(\gamma-2)/(\gamma-1)}\mathbf{C}_n^\nu$ converges in distribution as $n \rightarrow \infty$ to the ordered sequence of the excursion lengths of a certain non-trivial drifted process with independent increments.

Similar results have already been obtained for different random graph models. For example, Turova [Tur09] and Bhamidi *et al.* [BvdHvL10a, BvdHvL10b] studied special cases of rank-1 inhomogeneous random graphs constructed as follows. Let F be a distribution function on $[0, \infty)$ and w_1, w_2, \dots, w_n be defined by $w_i = [1 - F]^{-1}(i/n)$. Consider a simple graph on n labeled vertices such that an edge joins the vertices i and j ($i \neq j$) with probability $1 - \exp(-w_i w_j / l_n)$, where $l_n = \sum_{i=1}^n w_i$, different edges being independent. Denoting by W a r.v. with distribution function F , suppose that $\mathbb{E}[W^2] < \infty$. The criticality of the model occurs when $\mathbb{E}[W^2] = \mathbb{E}[W]$. As in the

present work, two different settings have been considered. In the case $\mathbb{E}[W^3] < \infty$, Turova [Tur09] and Bhamidi *et al.* [BvdHvL10a] separately showed that the ordered sequence of component sizes of the inhomogeneous random graph with n vertices once rescaled by $n^{-2/3}$ converges in distribution as $n \rightarrow \infty$ to the ordered sequence of the excursion lengths of a Brownian motion with parabolic drift, thus extending the results of Aldous [Ald97]. As for the power law case, Bhamidi *et al.* [BvdHvL10b] proved that if there exist $\gamma \in (3, 4)$ and $c > 0$ such that $1 - F(x) \sim_{x \rightarrow \infty} cx^{1-\gamma}$, the ordered sequence of component sizes of the inhomogeneous random graph with n vertices once rescaled by $n^{-(\gamma-2)/(\gamma-1)}$ then converges in distribution as $n \rightarrow \infty$ to hitting times of a thinned Lévy process. This should be viewed as a special case of the results obtained by Aldous and Limic in [AL98].

We too shall be interested in random simple graphs. Specifically, let \mathcal{SG}_n be the random simple graph consisting of n vertices such that, conditioned on the degree sequence (D_1, \dots, D_n) , it is uniformly distributed over all graphs with this degree sequence. Observe that \mathcal{SG}_n may also be view as the multigraph \mathcal{G}_n conditioned to be simple. When ν has finite third moment, we shall be able to prove that the ordered sequence \mathbf{SC}_n^ν of component sizes of the graph \mathcal{SG}_n has the same asymptotic behavior as \mathbf{C}_n^ν . We refer to Britton *et al.* [BDML06] for an understanding of the link between inhomogeneous random graphs and \mathcal{SG}_n .

The paper is organized as follows. In Sections 4.2, 4.3, 4.4, 4.5 and 4.6, we deal with the finite third moment case. Apart from Section 4.6, the main techniques developed there are used in Section 4.7, where the power law distribution case is studied. Section 4.6, devoted to \mathbf{SC}_n^ν , is specific to the finite third moment case. The main results will be stated in Section 4.2. In Section 4.3, following the ideas of Aldous [Ald97], we shall observe that the study may be reduced to the understanding of a walk defined via an algorithmic procedure related to the configuration model. Thank to [Ald97], convergence of that walk turns out to be sufficient. It will be obtained in Section 4.4 using standard methodology from stochastic process theory (see, *e.g.*, the CLT for continuous-time martingale). A key technique to obtain martingales is Poissonization. Basically, instead of considering multigraphs with exactly n vertices, we shall deal with multigraphs with Poisson(n) vertices. Our approach also relies on size-biased ordering. Finally, in Section 4.5, we shall be interested in the number of cycles in the multigraph \mathcal{G}_n . To conclude, in Section 4.7, we shall study \mathbf{C}_n^ν when ν is a power law distribution with exponent in $(3, 4)$. We shall follow the same strategy, except we shall apply results of Aldous and Limic [AL98].

4.2 Formulation of the results in the finite third moment setting

In the first sections of the paper, we suppose that ν satisfies:

$$\sum_{k=1}^{\infty} k(k-2)\nu_k = 0, \quad \sum_{k=1}^{\infty} k^3\nu_k < \infty \quad \text{and} \quad \nu_2 < 1. \quad (4.1)$$

The power law case will be studied in Section 4.7.

Let

$$\mu = \sum_{k=1}^{\infty} k\nu_k \quad \text{and} \quad \beta = \sum_{k=3}^{\infty} k(k-1)(k-2)\nu_k.$$

Observe that $\beta > 0$. Define the Brownian motion with parabolic drift

$$W^\nu : t \geq 0 \mapsto \sqrt{\frac{\beta}{\mu}} W(t) - \frac{\beta}{2\mu^2} t^2,$$

where $(W(t), t \geq 0)$ is a standard Brownian motion. The reflected process valued the nonnegative half-line is

$$R^\nu : t \geq 0 \mapsto W^\nu(t) - \min_{0 \leq s \leq t} W^\nu(s).$$

Call excursion interval of R^ν every time interval $\gamma = [l(\gamma), r(\gamma)]$ such that $R^\nu(l(\gamma)) = R^\nu(r(\gamma)) = 0$ and $R^\nu(t) > 0$ on $l(\gamma) < t < r(\gamma)$. The excursion has length $|\gamma| = r(\gamma) - l(\gamma)$. Aldous observed in [Ald97] that we can order excursions by length, that is the set of excursions of R^ν may be written $\{\gamma_j, j \geq 1\}$ so that the lengths $|\gamma_j|$ are decreasing. In the notation of [Ald97], define l_{\searrow}^2 as the set of infinite sequences $x = (x_1, x_2, \dots)$ with $x_1 \geq x_2 \geq \dots \geq 0$ and $\sum_i x_i^2 < \infty$, endowed with the Euclidean metric. Aldous showed in [Ald97, Lemma 25] that $\mathbb{E}[\sum_{j \geq 1} |\gamma_j|^2] < \infty$. In particular $(|\gamma_j|, j \geq 1)$ a.s. belongs to l_{\searrow}^2 . On the other hand, we may regard the finite sequence \mathbf{C}_n^ν as a random element of l_{\searrow}^2 by appending zero entries.

Our main result describes the component sizes of \mathcal{G}_n for large n ; it mirrors that of Aldous [Ald97] for the critical random graph.

Theorem 4.1. *Suppose ν satisfies (4.1). Let \mathbf{C}_n^ν be the ordered sequence of component sizes of \mathcal{G}_n . Then*

$$n^{-2/3} \mathbf{C}_n^\nu \xrightarrow[n \rightarrow \infty]{(d)} (|\gamma_j|, j \geq 1)$$

with respect to the l_{\searrow}^2 topology.

Remark 4.1. Suppose $\nu_2 = 1$, i.e., $D \equiv 2$. Then the components of \mathcal{G}_n are cycles. It is well-known that the distribution of cycle lengths is given by the Ewens's sampling formula $\text{ESF}(1/2)$, and thus the size of the largest component divided by n converges in distribution to a non-degenerate distribution on $[0, 1]$ (see [ABT03, Lemma 5.7]). This is also the case for the k -th largest component, where k is a fixed positive integer. That is why the assumption $\nu_2 < 1$ made in (4.1) is crucial in our setting.

Note that in our setting,

$$\liminf_{n \rightarrow \infty} \mathbb{P}(\mathcal{G}_n \text{ is a simple graph}) > 0 \tag{4.2}$$

(see Bollobás [Bol01], Janson [Jan09]). Recall that $\mathcal{S}\mathcal{G}_n$ is the random simple graph such that, conditioned on the degree sequence (D_1, \dots, D_n) , it is uniformly distributed over all graphs with this degree sequence. But it is also the multigraph \mathcal{G}_n conditioned on being simple. That is why authors usually first focus on \mathcal{G}_n to then deduce results for $\mathcal{S}\mathcal{G}_n$ using Equation (4.2) (see for instance Pittel [Pit08], Janson [Jan08], Janson and Luczak [JL09]). In the finite third moment setting, we shall be able to set up this strategy; we shall prove an analogous result of Theorem 4.1:

Theorem 4.2. *Suppose ν satisfies (4.1). Let \mathbf{SC}_n^ν be the ordered sequence of component sizes of \mathcal{SG}_n . Then*

$$n^{-2/3} \mathbf{SC}_n^\nu \xrightarrow[n \rightarrow \infty]{(d)} (|\gamma_j|, j \geq 1)$$

with respect to the l_{\searrow}^2 topology.

Remark 4.2. Consider the case when ν is the Poisson distribution with parameter 1 (observe though that $\mathbb{P}(D = 0) > 0$, so strictly speaking, it is out of our setting, but our result still holds as vertices with degree 0 play no role). Then, for large integers n , \mathcal{SG}_n is an approximation of the Erdős–Rényi random graph $G(n, 1/n)$. Now, in that case, $\mu = \beta = 1$, so the process W^ν is the Brownian motion with drift $-t$ at time t , which describes the asymptotic component sizes of $G(n, 1/n)$ (see [Ald97]).

4.3 The depth-first search

4.3.1 An algorithmic construction of \mathcal{G}_n

We start by describing a convenient algorithm to construct the multigraph \mathcal{G}_n . Suppose that $\sum_{i=1}^n D_i$ is even. We partition the set of half-edges into three subsets: the set \mathcal{S} of sleeping half-edges, the set \mathcal{A} of active half-edges and the set \mathcal{D} of dead half-edges. $\mathcal{S} \cup \mathcal{A}$ is the set of living half-edges. Initially, all the half-edges are sleeping.

Pick a sleeping half-edge uniformly at random and let v_1 denote the vertex it is attached to. Declare all the half-edges attached to v_1 active. While $\mathcal{A} \neq \emptyset$, proceed as follows.

- Let i be the largest integer k such that there exists an active half-edge attached to v_k .
- Consider an active half-edge l attached to v_i .
- Kill l , *i.e.*, remove it from \mathcal{A} and place it into \mathcal{D} .
- Choose uniformly at random a living half-edge r and pair l to it.
- If r is sleeping, let v_{j+1} denote the vertex it is attached to, where j is the number of vertices which were found before the discovery of the vertex attached to r . Declare all the half-edges attached to v_{j+1} except r active.
- Kill r .

Iterate until $\mathcal{A} = \emptyset$. At that step, the first component has been totally explored. If $\mathcal{S} \neq \emptyset$, proceed similarly with the remaining living vertices until all the half-edges have been killed; the multigraph \mathcal{G}_n is then constructed, and its vertices have been ordered via a depth-first search. See Figure 4.1 below for a simple illustration.

Note also that, by construction, the order in which the components appear in the depth-first search is size-biased order.

4.3.2 The depth-first walk

We now explain how the information on the component sizes may be encoded in a walk constructed via the depth-first search, which, as we shall see, is related to the process W^ν . We shall also need the notion of *cycle half-edge*.

Definition 4.1. *A half-edge l is called a cycle half-edge if there exists a half-edge r such that:*

- *l was killed before r ,*
- *l was paired to r ,*
- *r was active when l was paired to it.*

Observe that there exists a bijection between the set of cycle half-edges and the set of cycles, loops and multiple edges in \mathcal{G}_n .

Let us now define the walk associated to the depth-first search which will encode all the information that we need to study the component sizes. Write $(\hat{D}_i, i \in \{1, 2, \dots, n\})$ the sequence of the degrees of the vertices of \mathcal{G}_n ordered by their appearances in the depth-first search: for every $i \in \{1, \dots, n\}$,

$$\hat{D}_i = \text{degree of } v_i.$$

Define the depth-first walk $(W_n(i), 0 \leq i \leq n)$ by letting for every $i \in \{0, \dots, n\}$,

$$W_n(i) = \sum_{j=1}^i \left(\hat{D}_j - 2 - 2\#\{\text{cycle half-edges attached to } v_j\} \right). \quad (4.3)$$

Order the components $\mathcal{C}_1, \mathcal{C}_2, \dots$ according to the depth-first search. Let

$$\begin{aligned} \zeta(k) &= \sum_{j=1}^k |\mathcal{C}_j|, \\ \zeta^{-1}(i) &= \min\{k : \zeta(k) \geq i\}, \end{aligned}$$

so that $\zeta^{-1}(i)$ is the index of the component containing v_i . It is easily seen that

$$W_n(\zeta(k)) = -2k \quad \text{and} \quad W_n(i) \geq -2k - 1 \quad \text{for all } \zeta(k) \leq i < \zeta(k+1). \quad (4.4)$$

It follows that we can recover component sizes and indices from the walk via

$$\begin{aligned} \zeta(k) &= \min\{i : W_n(i) = -2k\}, \\ |\mathcal{C}_j| &= \zeta(j) - \zeta(j-1), \\ \zeta^{-1}(i) &= 1 - \left\lceil \min_{j < i} \frac{W_n(j)}{2} \right\rceil. \end{aligned}$$

Our main result relates the walk to the process W^ν :

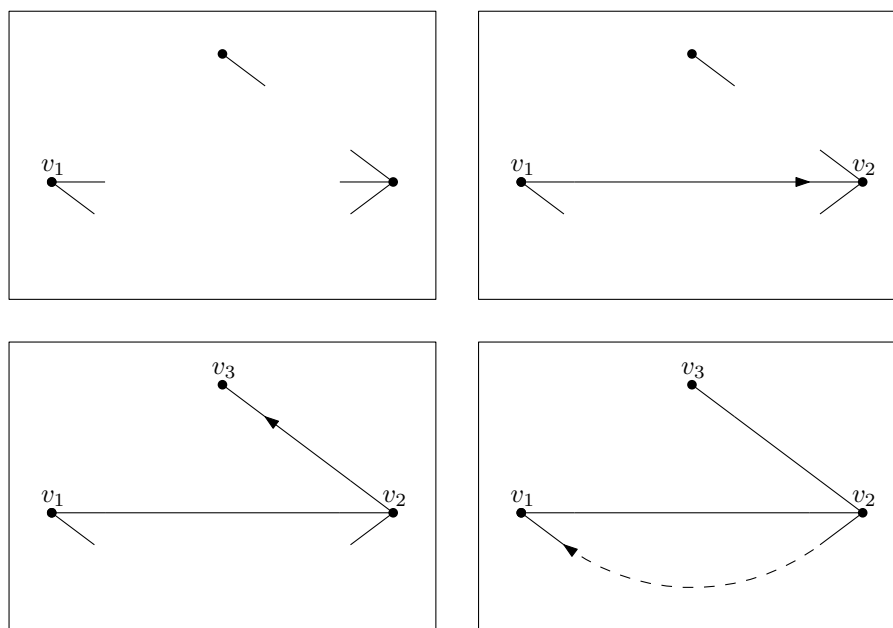


Figure 4.1: A realization of the algorithm constructing \mathcal{G}_3 . The dashed oriented edge of the last picture contains a cycle half-edge at its origin; v_2 has a cycle half-edge. The sequence of the steps of the walk W_3 is $(0, -1, -2)$.

Theorem 4.3. *Suppose ν satisfies (4.1). Rescale the depth-first walk W_n by defining for every $t \in [0, n^{1/3}]$*

$$\bar{W}_n(t) = n^{-1/3} W_n(\lfloor tn^{2/3} \rfloor).$$

Then

$$\bar{W}_n \xrightarrow[n \rightarrow \infty]{(d)} W^\nu.$$

To see how Theorem 4.1 follows from Theorem 4.3, we refer to Section 3.4 of the remarkable paper [Ald97] of Aldous¹. Intuitively, the result should be clear from property (4.4) of depth-first walk. Component sizes are indeed encoded as lengths of path segments above past even minima; these converge to lengths of excursions of W^ν above past minima, which are just lengths of excursions of the reflected process $(W^\nu(t) - \min_{0 \leq s \leq t} W^\nu(s), t \geq 0)$ above 0. Similarly, Theorem 4.2 is proven as soon as the following result is shown:

Theorem 4.4. *Suppose ν satisfies (4.1). The law of the rescaled walk \bar{W}_n conditioned on the event $\{\mathcal{G}_n \text{ is simple}\}$ converges in distribution to W^ν as $n \rightarrow \infty$.*

The next two sections are devoted to the proof of Theorem 4.3. In Section 4.4, we shall be interested in the depth-first walk $(\sum_{j=1}^i (\hat{D}_j - 2), 0 \leq i \leq n)$. It is easier to study the latter than the walk W_n since it ignores cycle half-edges and its law only

¹Recall that components appeared in size-biased order in the depth-first walk.

depends on the sequence $(\hat{D}_j, 1 \leq j \leq n)$, which has the law of the size-biased ordering of n independent copies of D . Let

$$\bar{s}_n : t \in [0, n^{1/3}] \mapsto n^{-1/3} \sum_{1 \leq j \leq tn^{2/3}} (\hat{D}_j - 2).$$

We shall show that the walk \bar{s}_n converges in distribution to W^ν as $n \rightarrow \infty$. In Section 4.5, we shall see that the difference between the two rescaled depth-first walks \bar{W}_n and \bar{s}_n is so small that in the limit, these processes have the same behavior. The combination of the two remarks yields Theorem 4.3. As for Theorem 4.4, it will be proved in Section 4.6.

4.4 Convergence of the walk \bar{s}_n

As mentioned above, in this section, we forget the contribution of the cycle half-edges to the depth-first walk W_n (we shall see in Section 4.5 that there are indeed few cycle half-edges up to time $tn^{2/3}$ for every fixed $t > 0$) and we only focus on the simpler walk $(\sum_{j=1}^i (\hat{D}_j - 2), 0 \leq i \leq n)$.

It is easily seen that the configuration model defining \mathcal{G}_n induces a degree-biased ordering of its vertices: conditionally on the degree sequence (D_1, \dots, D_n) , the sequence $(\hat{D}_1, \dots, \hat{D}_n)$ has the law of a size-biased reordering of (D_1, \dots, D_n) . Conditionally on $(D_1, \dots, D_n) = (d_1, \dots, d_n)$, a convenient way to degree-biased order the vertices of \mathcal{G}_n is to assign an exponential clock with parameter d_i to the vertex i , $i \in \{1, 2, \dots, n\}$, and to order the vertices according to the times the clocks they are attached to ring. Furthermore, since we want to discover vertices regularly as the walk $(\sum_{j=1}^i (\hat{D}_j - 2), 0 \leq i \leq n)$ does, time should elapse quicker and quicker, so we need to speed it up.

Specifically, let

$$\mathcal{L} : t \geq 0 \mapsto \sum_{k \in \mathbb{N}^*} e^{-kt} \nu_k$$

be the Laplace transform of ν . Let $\phi = 1 - \mathcal{L}$. Write ψ for the inverse of ϕ . Following [PPY92], consider a Poisson point process Π_n on $\mathbb{N}^* \times (0, n)$ with intensity π_n , where

$$\pi_n(k, ds) = k e^{-k\psi(s/n)} \psi'(s/n) \nu_k ds.$$

The k -components of the atoms of Π_n should be viewed as degrees whereas the s -components should be seen as time. Observe that since ψ is a convex increasing function, time is indeed speeded up.

Let $(\hat{D}_1, \hat{D}_2, \dots)$ be the sequence of the k -components of the atoms ordered according to the s -components increasing. Conditionally on Π_n having exactly n atoms, $(\hat{D}_1, \dots, \hat{D}_n)$ has the same law as the random vector $(\hat{D}_1, \dots, \hat{D}_n)$. Nonetheless, since the number of atoms of Π_n is a Poisson variable with parameter n and, as we shall soon see, we are only interested in what happens up to time $O(n^{2/3})$, we shall study the process Π_n without the latter conditioning. We thus get a Markovian process.

Observe that for every $s \in (0, n)$, $\sum_{k \in \mathbb{N}^*} k e^{-k\psi(s/n)} \psi'(s/n) \nu_k = 1$. Hence the following result, which shows that the atoms of Π_n are discovered regularly with respect to time.

Lemma 4.1. *The point process*

$$\{s \in (0, n) : \text{there exists } k \in \mathbb{N}^* \text{ such that } (k, s) \in \Pi_n\}$$

is Poisson point process on $(0, n)$ with intensity dt .

It should now be natural to introduce the process $(S_n(t))_{t \geq 0}$ defined as the sum of the k -components of the atoms of Π_n minus 2 with s -components less than or equal to t :

$$S_n(t) = \sum_{(k,s) \in \Pi_n} (k-2) \mathbf{1}_{s \leq t}.$$

In other words, letting $N_n(t) = \#\{s \in (0, t] : \text{there exists } k \in \mathbb{N}^* \text{ such that } (k, s) \in \Pi_n\}$,

$$S_n(t) = \sum_{j=1}^{N_n(t)} (\hat{\mathcal{D}}_j - 2).$$

We can now state the key result of the present work:

Proposition 4.1. *Rescale S_n by defining $\bar{S}_n : t \geq 0 \mapsto n^{-1/3} S_n(tn^{2/3})$. Then*

$$\bar{S}_n \xrightarrow[n \rightarrow \infty]{(d)} W^\nu.$$

Proof. We follow the ideas of Aldous [Ald97]. Let

$$A_n : t \mapsto \int \pi_n(k, ds) (k-2) \mathbf{1}_{s \leq t}$$

be the continuous bounded variation process such that

$$M_n(t) = S_n(t) - A_n(t), \quad t \geq 0,$$

is a martingale. Observe that A_n is deterministic. Rescaling as usual to define \bar{A}_n and \bar{M}_n , we shall see in Lemma 4.2 below that for every $t_0 > 0$,

$$\lim_{n \rightarrow \infty} \sup_{t \leq t_0} \left| \bar{A}_n(t) + \frac{\beta}{2\mu^2} t^2 \right| = 0$$

and in Lemma 4.3 below that

$$\bar{M}_n \xrightarrow[n \rightarrow \infty]{(d)} \sqrt{\frac{\beta}{\mu}} B,$$

where B denotes a standard Brownian motion, which will imply Proposition 4.1. \square

Lemma 4.2. *For every $t_0 > 0$,*

$$\lim_{n \rightarrow \infty} \sup_{t \leq t_0} \left| \bar{A}_n(t) + \frac{\beta}{2\mu^2} t^2 \right| = 0.$$

Proof. By definition,

$$\begin{aligned} A_n(t) &= \int_0^t \sum_{k \in \mathbb{N}^*} (k^2 - 2k) e^{-k\psi(s/n)} \psi'(s/n) \nu_k ds \\ &= \int_0^t (a_n(s) - 2) ds, \end{aligned}$$

where

$$a_n(s) = \frac{\mathcal{L}''(\psi(s/n))}{-\mathcal{L}'(\psi(s/n))}.$$

Since $\mathbb{E}[D^2] = 2\mathbb{E}[D]$, $a_n(s)$ tends to 2 as $n \rightarrow \infty$. Moreover, it is easily seen by approximating $\psi(s/n)$ by $\frac{s}{\mu n}$ that $a_n(s) - 2$ is approximatively $-\frac{\beta}{\mu^2} \frac{s}{n}$. Let us be more precise.

Let $s \in [0, t_0 n^{2/3}]$. Observe that

$$\mathcal{L}''(x) = \mathbb{E}[D^2] - \mathbb{E}[D^3]x + o(x) \quad \text{and} \quad -\mathcal{L}'(x) = \mathbb{E}[D] - \mathbb{E}[D^2]x + o(x). \quad (4.5)$$

Because $\psi(x) = \frac{x}{\mu} + o(x)$, we deduce that there exists a function $\varepsilon(\cdot)$ tending to 0 at 0 such that:

$$a_n(s) - 2 = -\frac{\beta}{\mu^2} \frac{s}{n} + \frac{s}{n} \varepsilon\left(\frac{s}{n}\right).$$

We deduce that for every $t \in [0, t_0 n^{2/3}]$,

$$\left| A_n(t) + \frac{\beta}{\mu^2} \frac{t^2}{2n} \right| \leq \frac{1}{n} \int_0^t s \left| \varepsilon\left(\frac{s}{n}\right) \right| ds \leq \frac{1}{n} \int_0^{t_0 n^{2/3}} s \left| \varepsilon\left(\frac{s}{n}\right) \right| ds.$$

As a result, for every $\eta > 0$, there exists an integer $n_0(\eta)$ such that for every integer $n \geq n_0(\eta)$,

$$\sup_{t \leq t_0 n^{2/3}} \left| A_n(t) + \frac{\beta}{2\mu^2} \frac{t^2}{n} \right| \leq \frac{1}{n} \int_0^{t_0 n^{2/3}} s \eta ds = \frac{t_0^2}{2} \eta n^{1/3},$$

which proves Lemma 4.2. \square

Lemma 4.3. $\bar{M}_n \xrightarrow[n \rightarrow \infty]{(d)} \sqrt{\frac{\beta}{\mu}} B$, where B denotes a standard Brownian motion.

Proof. We want to apply the functional CLT for continuous-time martingales (see [EK86, Theorem 7.1.4(b)]). Since M_n is a purely discontinuous martingale, $[M_n]_t = \sum_{s \leq t} \Delta M_n(s)^2$, so that its predictable projection

$$\langle M_n \rangle : t \mapsto \int \pi_n(k, ds) (k - 2)^2 \mathbf{1}_{s \leq t}$$

is the continuous, increasing process such that $M_n^2 - \langle M_n \rangle$ is a martingale. Observe that $\langle M_n \rangle$ is deterministic. Defining $\overline{\langle M_n \rangle} : t \mapsto n^{-2/3} \langle M_n \rangle (tn^{2/3})$, all we have to prove is that for every $t_0 > 0$,

$$\overline{\langle M_n \rangle}(t_0) \xrightarrow[n \rightarrow \infty]{} \frac{\beta}{\mu} t_0 \quad (4.6)$$

and

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\sup_{t \leq t_0} |\overline{M}_n(t) - \overline{M}_n(t-)|^2 \right] = 0. \quad (4.7)$$

Let us establish (4.6). Elementary calculations yield:

$$\langle M_n \rangle(t) = \int_0^t b_n(s) ds,$$

where

$$b_n(s) = \frac{\mathcal{L}^{(3)} + 4\mathcal{L}'' + 4\mathcal{L}'}{\mathcal{L}'} \circ \psi \left(\frac{s}{n} \right).$$

It is then easily seen that there exists a function $\varepsilon(\cdot)$ tending to 0 at 0 such that:

$$b_n(s) = \frac{\beta}{\mu} + \varepsilon \left(\frac{s}{n} \right).$$

We deduce that

$$\left| \langle M_n \rangle(t_0 n^{2/3}) - \frac{\beta}{\mu} t_0 n^{2/3} \right| \leq \int_0^{t_0 n^{2/3}} \left| \varepsilon \left(\frac{s}{n} \right) \right| ds.$$

Hence, for every $\eta > 0$, there exists an integer $n_1(\eta)$ such that for every integer $n \geq n_1(\eta)$,

$$\left| \langle M_n \rangle(t_0 n^{2/3}) - \frac{\beta}{\mu} t_0 n^{2/3} \right| \leq \eta t_0 n^{2/3},$$

which proves (4.6).

We next turn our attention to (4.7). Note that $M_n(t) - M_n(t-) = S_n(t) - S_n(t-)$, so

$$\begin{aligned} \sup_{t \leq t_0 n^{2/3}} |M_n(t) - M_n(t-)|^2 &= \sup \{ (k-2)^2 : (k, s) \in \Pi_n \text{ and } s \leq t_0 n^{2/3} \} \\ &\leq \sup \{ k^2 : (k, s) \in \Pi_n \text{ and } s \leq t_0 n^{2/3} \}. \end{aligned}$$

Let S_n denote $\sup \{ k : (k, s) \in \Pi_n \text{ and } s \leq t_0 n^{2/3} \}$ (we drop the dependency on t_0 in the notation). We have:

$$\mathbb{E} [S_n^2] = \sum_{k=1}^{\lfloor n^{1/3} \rfloor - 1} \mathbb{P} (S_n \geq \sqrt{k}) + \sum_{k \geq n^{1/3}} \mathbb{P} (S_n \geq \sqrt{k}) \leq n^{1/3} + \sum_{k \geq n^{1/3}} \mathbb{P} (S_n \geq \sqrt{k}).$$

Now, for every $m \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{P}(S_n \geq m) &= 1 - \mathbb{P}(S_n < m) \\
&= 1 - \mathbb{P}(\Pi_n(\{m, m+1, \dots\} \times [0, t_0 n^{2/3}]) = 0) \\
&= 1 - \exp(-\pi_n(\{m, m+1, \dots\} \times [0, t_0 n^{2/3}])) \\
&\leq \pi_n(\{m, m+1, \dots\} \times [0, t_0 n^{2/3}]) \\
&= \sum_{l \geq m} \nu_l \int_0^{t_0 n^{2/3}} ds l e^{-l\psi(s/n)} \psi'(s/n) \\
&= n \sum_{l \geq m} \nu_l \left(1 - e^{-l\psi(t_0 n^{-1/3})}\right) \\
&\leq n\psi(t_0 n^{-1/3}) \sum_{l \geq m} l \nu_l.
\end{aligned}$$

As a result,

$$\begin{aligned}
\mathbb{E}[S_n^2] &\leq n^{1/3} + \sum_{k \geq n^{1/3}} n\psi(t_0 n^{-1/3}) \sum_{l \geq \sqrt{k}} l \nu_l \\
&= n^{1/3} + n\psi(t_0 n^{-1/3}) \sum_{l \geq n^{1/6}} l \nu_l \sum_{k=\lfloor n^{1/3} \rfloor}^{l^2} 1.
\end{aligned}$$

We deduce that for every integer n ,

$$n^{-2/3} \mathbb{E} \left[\sup_{t \leq t_0 n^{2/3}} |M_n(t) - M_n(t-)|^2 \right] \leq n^{-1/3} + n^{1/3} \psi(t_0 n^{-1/3}) \sum_{l \geq n^{1/6}} l^3 \nu_l.$$

Now, $n^{1/3} \psi(t_0 n^{-1/3})$ tends to $\frac{t_0}{\mu}$ and since $\mathbb{E}[D^3]$ is finite, $\sum_{l \geq n^{1/6}} l^3 \nu_l$ tends to 0. Equation (4.7) is therefore proved. \square

We now give a key consequence of Proposition 4.1 concerning the depth-first walk \bar{s}_n .

Corollary 4.1. *The rescaled depth-first walk \bar{s}_n converges in distribution to W^ν as $n \rightarrow \infty$.*

Proof. Let T_n^j be the first time when j atoms of Π_n have been visited:

$$T_n^j = \inf \{t \geq 0 : N_n(t) \geq j\}.$$

Applying Lemma 4.1 and Proposition 4.1, the process

$$\bar{S}_n \left(n^{-2/3} T_n^{\lfloor tn^{2/3} \rfloor} \right), \quad t \geq 0,$$

converges in distribution to W^ν as $n \rightarrow \infty$. Now, for every $t \geq 0$,

$$\bar{S}_n \left(n^{-2/3} T_n^{\lfloor tn^{2/3} \rfloor} \right) = n^{-1/3} \sum_{1 \leq j \leq \lfloor tn^{2/3} \rfloor} \left(\hat{D}_j - 2 \right).$$

A depoissonization completes the proof. \square

4.5 Study of the cycle half-edges

In this section, we turn our attention to the cycle half-edges. In Section 4.5.1, we shall prove that there are few cycle half-edges in \mathcal{G}_n (see Lemma 4.4 below). We shall then show in Section 4.5.2 how to derive Theorem 4.3 from Corollary 4.1 and Lemma 4.4.

4.5.1 Upper bound of the number of cycle half-edges

In this section, we prove the following result:

Lemma 4.4. *Let $t > 0$ and $M > 0$. Introduce the event*

$$E_n(t, M) = \left\{ \max_{i \leq t} \left\{ \bar{s}_n(i) - \min_{k \leq i} \bar{s}_n(k) \right\} \leq M \right\}.$$

Then we have

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[\# \{ \text{cycle half-edges attached to } v_i, i \leq tn^{2/3}, \text{ in } \mathcal{G}_n \} \mathbf{1}_{E_n(t, M)} \right] < \infty.$$

Proof. We first study the number of active half-edges, given they contribute to the appearance of cycle half-edges. It is easily seen that, during the depth-first search, when a component is explored, the number of active half-edges when k vertices have been partially or totally explored is less than or equal to the maximal height of the current excursion of the depth-first walk $(\sum_{j=1}^i (\hat{D}_j - 2), 0 \leq i \leq n)$ above its past minimum plus 1, up to time k . (Note that this would not be true if we had considered the breadth-first search.) Consequently, under the event $E_n(t, M)$, during the first $\lfloor tn^{2/3} \rfloor$ steps, $\#\mathcal{A}$ is always less than or equal to $Mn^{1/3} + 1$.

For every deterministic sequence (x_1, \dots, x_n) of positive integers such that $\sum_{i=1}^n x_i$ is even, conditionally on the event

$$\mathcal{E}_{x_1, \dots, x_n} = \{(\hat{D}_1, \dots, \hat{D}_n) = (x_1, \dots, x_n)\},$$

one has:

$$\begin{aligned} & \mathbb{E} \left[\# \{ \text{cycle half-edges attached to } v_i, i \leq tn^{2/3} \} \mathbf{1}_{E_n(t, M)} \mid \mathcal{E}_{x_1, \dots, x_n} \right] \\ &= \mathbb{E} \left[\sum_{i=1}^{\lfloor tn^{2/3} \rfloor} \sum_{k=1}^{\hat{D}_i} \mathbf{1}_{\{\text{the } k\text{-th half-edge of } v_i \text{ is a cycle half-edge}\}} \mathbf{1}_{E_n(t, M)} \mid \mathcal{E}_{x_1, \dots, x_n} \right] \\ &\leq \sum_{i=1}^{\lfloor tn^{2/3} \rfloor} \sum_{k=1}^{x_i} \mathbb{P}(\text{the } k\text{-th half-edge of } v_i \text{ is a cycle half-edge} \mid \mathcal{E}_{x_1, \dots, x_n} \text{ and } E_n(t, M)) \\ &\leq \sum_{i=1}^{\lfloor tn^{2/3} \rfloor} x_i \frac{Mn^{1/3} + 1}{\sum_{m=1}^n x_m - \sum_{m=1}^{\lfloor tn^{2/3} \rfloor} x_m} \\ &\leq \sum_{i=1}^{\lfloor tn^{2/3} \rfloor} x_i \frac{Mn^{1/3} + 1}{n - tn^{2/3}}. \end{aligned}$$

Consequently,

$$\begin{aligned} & \mathbb{E} \left[\# \left\{ \text{cycle half-edges attached to } v_i, i \leq n^{2/3}, \text{ in } \mathcal{G}_n \right\} \mathbf{1}_{E_n(t,M)} \right] \\ & \leq \frac{Mn^{1/3} + 1}{n - tn^{2/3}} \mathbb{E} \left[\sum_{i=1}^{tn^{2/3}} \hat{D}_i \right] \\ & \leq \frac{Mn^{1/3} + 1}{n - tn^{2/3}} tn^{2/3} \mathbb{E} \left[\hat{D}_1 \right]. \end{aligned}$$

Note that $\mathbb{E}[\hat{D}_1] \leq \sum_{k=1}^{\infty} k \frac{k\nu_k}{\mu}$. Hence

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[\# \left\{ \text{cycle half-edges attached to } v_i, i \leq tn^{2/3} \right\} \mathbf{1}_{E_n(t,M)} \right] \leq 2Mt,$$

which completes the proof of Lemma 4.4. \square

Remark 4.3. We can prove that in fact, for every $t > 0$,

$$\limsup_{n \rightarrow \infty} \mathbb{E} \left[\# \left\{ \text{cycle half-edges attached to } v_i, i \leq tn^{2/3}, \text{ in } \mathcal{G}_n \right\} \right] < \infty.$$

Remark 4.4. We stress that a consequence of [BS09, Theorem 1] is that the expected number of cycle half-edges in \mathcal{G}_n is $o(n)$.

4.5.2 End of the proof of Theorem 4.3

In this section, we prove Theorem 4.3. We keep the notation of Section 4.5.1. Let $t > 0$. Let f be a bounded, Lipschitz function defined on $(\mathcal{C}[0, t], \mathbb{R})$. Applying Corollary 4.1 and the Portmanteau theorem, it suffices to prove that $\mathbb{E}[f(\overline{W}_n)] - \mathbb{E}[f(\overline{s}_n)]$ tends to 0 as $n \rightarrow \infty$.

There exists $K > 0$ such that for every $w, w' \in (\mathcal{C}[0, t], \mathbb{R})$, $|f(w)| \leq K$ and $|f(w) - f(w')| \leq K\|w - w'\|$. Let $M > 0$. One has:

$$\begin{aligned} & \left| \mathbb{E} [f(\overline{W}_n)] - \mathbb{E} [f(\overline{s}_n)] \right| \\ & = \mathbb{E} \left[|f(\overline{W}_n) - f(\overline{s}_n)| \mathbf{1}_{E_n(t,M)} \right] + \mathbb{E} \left[|f(\overline{W}_n) - f(\overline{s}_n)| (1 - \mathbf{1}_{E_n(t,M)}) \right] \\ & \leq \mathbb{E} \left[K \|\overline{W}_n - \overline{s}_n\| \mathbf{1}_{E_n(t,M)} \right] + \mathbb{E} \left[2K (1 - \mathbf{1}_{E_n(t,M)}) \right] \\ & \leq 2Kn^{-1/3} \mathbb{E} \left[\# \left\{ \text{cycle half-edges attached to } v_i, i \leq tn^{2/3} \right\} \mathbf{1}_{E_n(t,M)} \right] \\ & \quad + 2K \mathbb{P} \left(\max_{i \leq t} \left\{ \overline{s}_n(i) - \min_{k \leq i} \overline{s}_n(k) \right\} \geq M \right). \end{aligned}$$

Lemma 4.4 ensures that

$$\lim_{n \rightarrow \infty} n^{-1/3} \mathbb{E} \left[\# \left\{ \text{cycle half-edges attached to } v_i, i = 1 \leq tn^{2/3} \right\} \mathbf{1}_{E_n(t,M)} \right] = 0.$$

Moreover, applying Corollary 4.1 and the Portmanteau theorem,

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(\max_{i \leq t} \left\{ \overline{s}_n(i) - \min_{k \leq i} \overline{s}_n(k) \right\} \geq M \right) \leq \mathbb{P} \left(\max_{s \leq t} \left\{ W^\nu(s) - \min_{u \leq s} W^\nu(u) \right\} \geq M \right).$$

Therefore, for every $M > 0$,

$$\limsup_{n \rightarrow \infty} |\mathbb{E}[f(\overline{W}_n)] - \mathbb{E}[f(\overline{s}_n)]| \leq 2K \mathbb{P} \left(\max_{s \leq t} \left\{ W^\nu(s) - \min_{u \leq s} W^\nu(u) \right\} \geq M \right).$$

Now, the continuity of W^ν implies that

$$\lim_{M \rightarrow \infty} \mathbb{P} \left(\max_{s \leq t} \left\{ W^\nu(s) - \min_{u \leq s} W^\nu(u) \right\} \geq M \right) = 0.$$

Hence

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(\overline{W}_n)] - \mathbb{E}[f(\overline{s}_n)] = 0.$$

Theorem 4.3 is therefore proved.

4.6 Study of the random simple graph \mathcal{SG}_n

The setting of this section is the same as before: ν is supposed to satisfy (4.1). We intend to show Theorem 4.4. The proof is divided into two steps. First (see Lemma 4.5 below) we shall prove that, with probability tending to 1, the possible loops and multiple edges in \mathcal{G}_n arrive only after the first $\lfloor n^{3/4} \rfloor$ vertices have been explored during the depth-first search. We shall then deduce that the walk \overline{W}_n conditioned on the event $\{\mathcal{G}_n \text{ is simple}\}$ has the same asymptotic behavior as the walk \overline{W}_n (see Section 4.6.2).

4.6.1 Time arrival of loops and multiple edges

In this section, we prove the following result:

Lemma 4.5. *Let $T(n)$ be the minimal index of a vertex of \mathcal{G}_n having a loop or a multiple edge:*

$$T(n) = \inf\{i \in \{1, \dots, n\} : v_i \text{ has a loop or a multiple edge}\}.$$

Then,

$$\lim_{t \rightarrow \infty} \mathbb{P}(T(n) > n^{3/4}) = 1.$$

Observe that $T(n) = \infty$ if and only if \mathcal{G}_n is simple.

Proof. Obviously, it suffices to show that

$$\lim_{n \rightarrow \infty} \mathbb{E}[\#\{\text{loops or multiple edges attached to } v_i, i \leq n^{3/4}, \text{ in } \mathcal{G}_n\}] = 0.$$

Let us establish that assertion. We proceed the same way as in the proof of Lemma 4.4. For every deterministic sequence (x_1, \dots, x_n) of positive integers such that $\sum_{i=1}^n x_i$ is even, conditionally on the event

$$\mathcal{E}_{x_1, \dots, x_n} = \{(\hat{D}_1, \dots, \hat{D}_n) = (x_1, \dots, x_n)\},$$

one has:

$$\begin{aligned} & \mathbb{E}[\#\{\text{loops or multiple edges attached to } v_i, i \leq n^{3/4}\} \mid \mathcal{E}_{x_1, \dots, x_n}] \\ & \leq \sum_{i=1}^{n^{3/4}} \sum_{k=1}^{x_i} \mathbb{P}(\text{the } k\text{-th half-edge of } v_i \text{ creates a loop or a multiple edge} \mid \mathcal{E}_{x_1, \dots, x_n}). \end{aligned}$$

Now, the k -th half-edge of a vertex with degree x_i ($i \leq n^{3/4}$) creates a loop with probability at most $\frac{x_i - k}{n - n^{3/4}}$. It creates a multiple edge with probability at most $\frac{k-1}{n - n^{3/4}}$. Consequently

$$\begin{aligned} & \mathbb{E} \left[\# \{ \text{loops or multiple edges attached to } v_i, i \leq n^{3/4} \} \mid \hat{D}_1 = x_1, \dots, \hat{D}_n = x_n \right] \\ & \leq \sum_{i=1}^{n^{3/4}} \sum_{k=1}^{x_i} \frac{x_i}{n - n^{3/4}} \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E} \left[\# \{ \text{loops or multiple edges attached to } v_i, i \leq n^{3/4}, \text{ in } \mathcal{G}_n \} \right] \\ & \leq \frac{1}{n - n^{3/4}} \mathbb{E} \left[\sum_{i=1}^{n^{3/4}} \hat{D}_i^2 \right] \\ & \leq \frac{n^{3/4}}{n - n^{3/4}} \mathbb{E} \left[\hat{D}_1^2 \right]. \end{aligned}$$

Since ν has finite third moment, this completes the proof. \square

4.6.2 End of the proof of Theorem 4.4

Let $t > 0$. Let f be a bounded, continuous function defined on $(\mathcal{C}[0, t], \mathbb{R})$. It suffices to prove that:

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[f(\overline{W}_n) \mid T(n) = \infty \right] = \mathbb{E} \left[f(\overline{W}^\nu) \right],$$

where $f(\overline{W}_n)$ denotes the image of the restriction of the walk \overline{W}_n to $[0, t]$ by f . Let us show that result. Observe that the event

$$\{ \text{neither loop nor multiple edge is attached to } v_i, i > n^{3/4} \}$$

is asymptotically independent of the r.v. $f(\overline{W}_n)$:

$$\begin{aligned} & \mathbb{E} \left[f(\overline{W}_n) \mathbf{1}_{\text{neither loop nor multiple edge is attached to } v_i, i > n^{3/4}} \right] \\ & \underset{n \rightarrow \infty}{\sim} \mathbb{E} \left[f(\overline{W}_n) \right] \mathbb{P} \left(\text{neither loop nor multiple edge is attached to } v_i, i > n^{3/4} \right). \end{aligned}$$

Deciding whether or not a loop or a multiple edge is created after step $n^{3/4}$ does indeed not depend on the first $tn^{2/3}$ steps². Now, by Theorem 4.3,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[f(\overline{W}_n) \right] = \mathbb{E} \left[f(\overline{W}^\nu) \right],$$

and according to Lemma 4.5 and Equation (4.2)

$$\mathbb{P} \left(\text{neither loop nor multiple edge is attached to } v_i, i > n^{3/4} \right) \underset{n \rightarrow \infty}{\sim} \mathbb{P} (T(n) = \infty).$$

²To make this argument rigorous, consider the Poissonian model introduced in Section 4.4; independence is then straightforward and the fact that the two models are asymptotically close has already been seen.

Consequently

$$\mathbb{E} [f(\overline{W}_n) \mathbf{1}_{\text{neither loop nor multiple edge is attached to } v_i, i > n^{3/4}}] \underset{n \rightarrow \infty}{\sim} \mathbb{E} [f(\overline{W}^\nu)] \mathbb{P}(T(n) = \infty).$$

Recalling Lemma 4.5 and Equation (4.2), this proves that

$$\mathbb{E} [f(\overline{W}_n) \mathbf{1}_{T(n)=\infty}] \underset{n \rightarrow \infty}{\sim} \mathbb{E} [f(\overline{W}^\nu)] \mathbb{P}(T(n) = \infty),$$

completing the proof of Theorem 4.4.

4.7 The power law distribution setting

In this section, we do not suppose the finiteness of the moment of order 3 for distribution ν , and rather we replace assumption (4.1) by the following

$$\sum_{k=1}^{\infty} k(k-2)\nu_k = 0 \quad \text{and} \quad \nu_k \underset{k \rightarrow \infty}{\sim} ck^{-\gamma}, \quad (4.8)$$

where $c > 0$ and $\gamma \in (3, 4)$. This implies that (4.5) has to be replaced by

$$\mathcal{L}''(x) = 2\mu - \frac{c \Gamma(4-\gamma)}{\gamma-3} x^{\gamma-3} + o_{x \rightarrow 0}(x^{\gamma-3}). \quad (4.9)$$

We are interested in the component sizes of the multigraph constructed the same way as before. To have a good idea of what the order of the component sizes should be, we adopt the same strategy, using Poissonian calculus (see Section 4.7.1). We shall then show in Section 4.7.2 how to deduce the asymptotic behavior of the component sizes of \mathcal{G}_n in our new situation. In Section 4.7.3 we shall state some open problems.

4.7.1 The Poissonian argument

Taking the same notation as in Section 4.4, we consider here the process $(S_n(t))_{t \geq 0}$ defined by:

$$S_n(t) = \sum_{(k,s) \in \Pi_n} (k-2) \mathbf{1}_{s \leq t}.$$

Recall that Π_n is a Poisson point process on $\mathbb{N}^* \times (0, n)$ with intensity π_n , where $\pi_n(k, ds) = ke^{-k\psi(s/n)}\psi'(s/n)\nu_k ds$. We intend to prove the following result:

Theorem 4.5. *Rescale S_n by defining $\overline{S}_n : t \geq 0 \mapsto n^{-1/(\gamma-1)} S_n(tn^{(\gamma-2)/(\gamma-1)})$. Then*

$$\overline{S}_n \underset{n \rightarrow \infty}{\xrightarrow{(d)}} X^\nu + A^\nu,$$

where

$$A^\nu : t \mapsto -\frac{c \Gamma(4-\gamma)}{(\gamma-3)(\gamma-2)\mu^{\gamma-2}} t^{\gamma-2}$$

and X^ν is the unique process with independent increments such that for every $t \geq 0$ and $u \in \mathbb{R}$,

$$\mathbb{E}[\exp(iuX_t^\nu)] = \exp\left(\int_0^t ds \int_0^\infty dx (e^{iux} - 1 - iux) \frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xs/\mu}\right).$$

Proof. As before, let

$$A_n : t \mapsto \int \pi_n(k, ds)(k-2)\mathbf{1}_{s \leq t}$$

be the deterministic continuous bounded variation function such that

$$M_n(t) = S_n(t) - A_n(t), \quad t \geq 0,$$

is a martingale. Rescaling as in Theorem 4.5 to define \bar{A}_n and \bar{M}_n , we can easily see as in Lemma 4.2 (recall (4.9)) that for every $t > 0$,

$$\limsup_{n \rightarrow \infty} \sup_{s \leq t} |\bar{A}_n(s) - A^\nu(s)| = 0.$$

To complete the proof of Theorem 4.5, it thus suffices to show that

$$\bar{M}_n \xrightarrow[n \rightarrow \infty]{(d)} X^\nu.$$

This will be done in Lemmas 4.6 and 4.7 by applying general results on convergence of processes with independent increments borrowed from [JS03]. \square

The rest of this section is organized as follows. We shall first study in Lemma 4.6 the martingale $\bar{M}_n^{(1)}$ related to the small jumps of \bar{M}_n . Then, in Lemma 4.7, we shall be interested in the martingale $\bar{M}_n^{(2)}$ which counts the big jumps. The fact that $\bar{M}_n = \bar{M}_n^{(1)} + \bar{M}_n^{(2)}$ converges to X^ν , which is the sum of the limits of $\bar{M}_n^{(1)}$ and $\bar{M}_n^{(2)}$, stems from the independence of $\bar{M}_n^{(1)}$ and $\bar{M}_n^{(2)}$ (since they never jump simultaneously). To ease notation, let

$$a = \frac{1}{\gamma - 1}.$$

Lemma 4.6. *The martingale $\bar{M}_n^{(1)}$ defined for every $t \geq 0$ by*

$$\bar{M}_n^{(1)}(t) = \sum_{(k,s) \in \Pi_n} \mathbf{1}_{k < n^a} (k-2)n^{-a} \mathbf{1}_{s \leq tn^{1-a}} - \int \pi_n(k, ds) \mathbf{1}_{k < n^a} (k-2)n^{-a} \mathbf{1}_{s \leq tn^{1-a}}$$

converges in distribution as $n \rightarrow \infty$ to a process $(X_t^{(1)})_{t \geq 0}$ with independent increments characterized by: for every $t \geq 0$ and $u \in \mathbb{R}$,

$$\mathbb{E} \left[\exp \left(iu X_t^{(1)} \right) \right] = \exp \left(\int_0^t ds \int_0^1 dx (e^{iux} - 1 - iux) \frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xs/\mu} \right).$$

Proof. First observe that the process $X^{(1)}$ may be defined as the limit for the metric induced by the norm

$$\|Y\| = \mathbb{E} \left[\sup \{ Y_s^2 : 0 \leq s \leq t \} \right]^{1/2}$$

of the Cauchy family

$$t \mapsto \sum_{s \leq t} \mathbf{1}_{\Delta_s > \varepsilon} \Delta_s - \int_0^t ds \int_\varepsilon^1 dx x \frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xs/\mu}$$

as ε tends to 0, where Δ is a Poisson point process with intensity $\mathbf{1}_{x \in (0,1)} \nu(ds, dx)$ with

$$\nu(ds, dx) = \frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xs/\mu} ds dx.$$

To prove Lemma 4.6, we rely on [JS03, Theorem VII.3.7]. Dealing with small jumps of the martingale \overline{M}_n indeed enables us to work with “square-integrable” processes (note that $\int_0^t \int_{\mathbb{R}} x^2 \mathbf{1}_{x \in (0,1)} \nu(ds, dx) < \infty$).

Taking the same notation as in [JS03], we first have to compute the characteristics (B^n, C^n, ν^n) of $\overline{M}_n^{(1)}$, which are defined via the equation: for every $t \geq 0$ and $u \in \mathbb{R}$,

$$\mathbb{E} \left[\exp \left(iu \overline{M}_n^{(1)}(t) \right) \right] = \exp \left(iu B^n(t) - \frac{1}{2} u^2 C^n(t) + \int_0^t \int_{-n^{-a}}^1 (e^{iux} - 1 - iux) \nu^n(ds, dx) \right).$$

The exponential formula for Poisson point processes yields

$$\mathbb{E} \left[\exp \left(iu \overline{M}_n^{(1)}(t) \right) \right] = \exp \left\{ n \sum_{k < n^a} \nu_k \left(1 - e^{-k\psi(tn^{-a})} \right) \left(e^{iu(k-2)n^{-a}} - 1 - iu(k-2)n^{-a} \right) \right\}.$$

Consequently, $B^n = C^n = 0$ and

$$\nu^n(ds, dx) = ds \sum_{k < n^a} \delta_{(k-2)n^{-a}}(dx) n^{1-a} k \nu_k \psi'(sn^{-a}) e^{-k\psi(sn^{-a})}.$$

According to [JS03, Theorem VII.3.7], Lemma 4.6 will be proved as soon as we have shown that for every $t \geq 0$:

$$\int_0^t \int_{-n^{-a}}^1 x^2 \nu^n(ds, dx) \xrightarrow{n \rightarrow \infty} \int_0^t \int_0^1 x^2 \nu(ds, dx), \quad (4.10)$$

and

$$\text{for every } g \in C_2(\mathbb{R}_+), \quad \int_0^t \int_{-n^{-a}}^1 g(x) \nu^n(ds, dx) \xrightarrow{n \rightarrow \infty} \int_0^t \int_0^1 g(x) \nu(ds, dx), \quad (4.11)$$

where $C_2(\mathbb{R}_+)$ is the set of all continuous bounded functions $\mathbb{R}_+ \rightarrow \mathbb{R}$ which are 0 on a neighborhood 0 and have a limit at infinity.

Let us establish (4.10). Elementary calculations yield:

$$\int_0^t \int_{-n^{-a}}^1 x^2 \nu^n(ds, dx) = n^{1-2a} \sum_{k < n^a} (k-2)^2 \nu_k \left(1 - e^{-k\psi(tn^{-a})} \right).$$

A difficulty stems from the lack of good estimates for ν_k when k is small. That is why we write

$$\begin{aligned} \int_0^t \int_{-n^{-a}}^1 x^2 \nu^n(ds, dx) &= n^{1-2a} \sum_{k \in \mathbb{N}^*} (k-2)^2 \nu_k \left(1 - e^{-k\psi(tn^{-a})} \right) \\ &\quad - n^{1-2a} \sum_{k \geq n^a} (k-2)^2 \nu_k \left(1 - e^{-k\psi(tn^{-a})} \right). \end{aligned}$$

It is easy to see that the first term in the difference tends to $\frac{c\Gamma(4-\gamma)}{(\gamma-3)\mu^{\gamma-3}}t^{\gamma-3}$. As for the second, recalling that $\nu_k \sim ck^{-\gamma}$,

$$n^{1-2a} \sum_{k \geq n^a} (k-2)^2 \nu_k \left(1 - e^{-k\psi(tn^{-a})}\right) \underset{n \rightarrow \infty}{\sim} n^{1-2a} \int_{n^a}^{\infty} dx x^2 c x^{-\gamma} \left(1 - e^{-x\psi(tn^{-a})}\right).$$

A change of variables and an application of the dominated convergence theorem (recall that $\psi(x) = \frac{x}{\mu} + o(x)$) yield

$$n^{1-2a} \sum_{k \geq n^a} (k-2)^2 \nu_k \left(1 - e^{-k\psi(tn^{-a})}\right) \xrightarrow{n \rightarrow \infty} \int_1^{\infty} dx c \frac{1 - e^{-xt/\mu}}{x^{\gamma-2}}.$$

Noticing that

$$\frac{c \Gamma(4-\gamma)}{(\gamma-3)\mu^{\gamma-3}} t^{\gamma-3} = \int_0^{\infty} dx c \frac{1 - e^{-xt/\mu}}{x^{\gamma-2}},$$

we finally get

$$\int_0^t \int_{-n^{-a}}^1 x^2 \nu^n(ds, dx) \xrightarrow{n \rightarrow \infty} \int_0^1 dx c \frac{1 - e^{-xt/\mu}}{x^{\gamma-2}},$$

which proves (4.10).

We now turn our attention to (4.11). Let $\varepsilon \in (0, 1)$ and $g : [\varepsilon, 1] \rightarrow \mathbb{R}$ be a continuous function. Then

$$\int_0^t \int_{-n^{-a}}^1 g(x) \nu^n(ds, dx) = n \sum_{\varepsilon n^a < k < n^a} g\left(\frac{k-2}{n^a}\right) \nu_k \left(1 - e^{-k\psi(tn^{-a})}\right).$$

Proceeding as before, we obtain

$$\int_0^t \int_{-n^{-a}}^1 g(x) \nu^n(ds, dx) \xrightarrow{n \rightarrow \infty} \int_{\varepsilon}^1 dx g(x) c \frac{1 - e^{-xt/\mu}}{x^{\gamma}},$$

completing the proof of Lemma 4.6. \square

In order to finish of the proof Theorem 4.5, we now show the convergence of the martingale related to the big jumps.

Lemma 4.7. *The martingale $\overline{M}_n^{(2)}$ defined for every $t \geq 0$ by*

$$\overline{M}_n^{(2)}(t) = \sum_{(k,s) \in \Pi_n} \mathbf{1}_{k \geq n^a} (k-2) n^{-a} \mathbf{1}_{s \leq tn^{1-a}} - \int \pi_n(k, ds) \mathbf{1}_{k \geq n^a} (k-2) n^{-a} \mathbf{1}_{s \leq tn^{1-a}}$$

converges in distribution as $n \rightarrow \infty$ to a process $(X_t^{(2)})_{t \geq 0}$ with independent increments characterized by: for every $s, t \geq 0$, $u \in \mathbb{R}$,

$$\mathbb{E} \left[\exp \left(iu X_t^{(2)} \right) \right] = \exp \left(\int_0^t ds \int_1^{\infty} dx (e^{iux} - 1 - iux) \frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xs/\mu} \right).$$

Proof. The existence of $X^{(2)}$ is easily obtained as the sum of

$$B^\nu : t \mapsto - \int_0^t ds \int_1^\infty dx x \frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xs/\mu}$$

and the partial sum of the jumps of a Poisson point process with intensity $\mathbf{1}_{x \geq 1} \nu(ds, dx)$ (recall that $\nu(ds, dx) = \frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xs/\mu} ds dx$). Let us see how Lemma 4.7 derives from [JS03, Theorem VII.3.4].

As before, we first have to compute the characteristics (B^n, C^n, ν^n) of $\overline{M}_n^{(2)}$, which are now defined via the equation: for every $s, t \geq 0, u \in \mathbb{R}$,

$$\mathbb{E} \left[\exp \left(iu \overline{M}_n^{(2)}(t) \right) \right] = \exp \left(iu B^n(t) - \frac{1}{2} u^2 C^n(t) + \int_0^t \int_{1-2n^{-a}}^\infty (e^{iux} - 1) \nu^n(ds, dx) \right).$$

The exponential formula for Poisson point processes yields

$$\begin{aligned} \mathbb{E} \left[\exp \left(iu \overline{M}_n^{(2)}(t) \right) \right] &= \exp \left\{ -iun^{1-a} \sum_{k \geq n^a} (k-2) \nu_k \left(1 - e^{-k\psi(tn^{-a})} \right) \right. \\ &\quad \left. + n \sum_{k \geq n^a} \nu_k \left(1 - e^{-k\psi(tn^{-a})} \right) \left(e^{iu(k-2)n^{-a}} - 1 \right) \right\}. \end{aligned}$$

Consequently, $C^n = 0$,

$$B^n(t) = -n^{1-a} \sum_{k \geq n^a} (k-2) \nu_k \left(1 - e^{-k\psi(tn^{-a})} \right)$$

and

$$\nu^n(ds, dx) = ds \sum_{k \geq n^a} \delta_{(k-2)n^{-a}}(dx) k \nu_k n^{1-a} \psi'(sn^{-a}) e^{-k\psi(sn^{-a})}.$$

According to [JS03, Theorem VII.3.4], Lemma 4.7 will be proved as soon as we have shown that for every $t \geq 0$:

$$\sup_{s \leq t} |B^n(t) - B_t^\nu| \xrightarrow[n \rightarrow \infty]{} 0, \quad (4.12)$$

and

$$\text{for every } g \in C_2(\mathbb{R}_+), \quad \int_0^t \int_{1-2n^{-a}}^\infty g(x) \nu^n(ds, dx) \xrightarrow[n \rightarrow \infty]{} \int_0^t \int_1^\infty g(x) \nu(ds, dx). \quad (4.13)$$

Equation (4.13) can be shown exactly the same way as (4.11), and to prove (4.12), it suffices to compare the series to the corresponding integrals as we did above. \square

4.7.2 The main results

Repeating what we did in Section 4.5, we deduce from Theorem 4.5 the following key result. As before, the walk defined via (4.3) is denoted by W_n .

Corollary 4.2. *Rescale the depth-first walk W_n by defining for every $t \in [0, n^{1/(\gamma-1)}]$*

$$\bar{W}_n(t) = n^{-1/(\gamma-1)} W_n(\lfloor tn^{(\gamma-2)/(\gamma-1)} \rfloor).$$

Then

$$\bar{W}_n \xrightarrow[n \rightarrow \infty]{(d)} X^\nu + A^\nu.$$

We now give an analogous result of Theorem 4.1 in the present setting. Let R^ν be the reflected process defined by

$$R^\nu : t \geq 0 \mapsto X_t^\nu + A_t^\nu - \inf_{0 \leq s \leq t} \{X_s^\nu + A_s^\nu\}.$$

We define excursion intervals and excursion lengths of R^ν as in Section 4.2.

Theorem 4.6. *Suppose ν satisfies (4.8). Then a.s. the set of excursions of R^ν may be written $\{\gamma_j, j \geq 1\}$ so that the lengths $|\gamma_j|$ are decreasing and*

$$\sum_{j \geq 1} |\gamma_j|^2 < \infty,$$

and letting \mathbf{C}_n^ν be the ordered sequence of component sizes of \mathcal{G}_n ,

$$n^{-(\gamma-2)/(\gamma-1)} \mathbf{C}_n^\nu \xrightarrow[n \rightarrow \infty]{(d)} (|\gamma_j|, j \geq 1)$$

with respect to the l_{\searrow}^2 topology.

We borrow the technique for showing how Theorem 4.6 can be deduced from Corollary 4.2 from the deep paper of Aldous and Limic [AL98]. Indeed, observe that the component sizes of the multigraph \mathcal{G}_n , in the order of appearance in depth-first walk, are size-biased ordered. Applying [AL98, Proposition 17] (see also [Ald97, Proposition 15 and Lemma 25]), Theorem 4.6 thus derives from Corollary 4.2 and the following lemma (we refer to [AL98, Proposition 14]).

Lemma 4.8. *The following four assertions hold.*

1. $X_t^\nu + A_t^\nu \xrightarrow{p} -\infty$ as $t \rightarrow \infty$.
2. $\sup\{|\gamma| : \gamma \text{ is an excursion of } R^\nu \text{ s.t. } l(\gamma) \geq t\} \xrightarrow{p} 0$ as $t \rightarrow \infty$.
3. The set $\{t : R_t^\nu = 0\}$ contains no isolated points a.s.
4. For every $t > 0$, $\mathbb{P}(R_t^\nu = 0) = 0$.

Proof. Proof of 1. By Lemma 4.6,

$$\mathbb{E} \left[\left(X_t^{(1)} \right)^2 \right] = \frac{c}{\mu} \int_0^t ds \int_0^1 dx \frac{1}{x^{\gamma-3}} e^{-xs/\mu} \leq ct \int_0^{1/t} dx \frac{1}{x^{\gamma-3}} + c \int_{1/t}^\infty dx \frac{1}{x^{\gamma-2}},$$

so that

$$\mathbb{E} \left[\left(X_t^{(1)} \right)^2 \right] \leq \frac{c}{(\gamma-3)(4-\gamma)} t^{\gamma-3}.$$

Applying Markov's inequality, we deduce that

$$t^{-(\gamma-3)}X_t^{(1)} \xrightarrow[t \rightarrow \infty]{p} 0. \quad (4.14)$$

Letting $\eta = (\gamma-3)/2$, this implies that $t^{-(1+\eta)}X_t^{(1)} \xrightarrow{p} 0$ as $t \rightarrow \infty$. Then notice that $X_t^{(2)}$ is less than $\sum_{s \leq t} \Delta_s$, where Δ is a Poisson point process with intensity $\mathbf{1}_{x \geq 1} \nu(ds, dx)$ (recall that $\nu(ds, dx) = \frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xs/\mu} ds dx$). Now $\mathbb{E}[\sum_{s \leq t} \Delta_s] = \frac{c}{\mu} \int_0^t ds \int_1^\infty dx \frac{1}{x^{\gamma-2}} e^{-xs/\mu} \leq \frac{c}{\mu(\gamma-3)} t$. Consequently, by Markov's inequality, $t^{-(1+\eta)} \sum_{s \leq t} \Delta_s \xrightarrow{p} 0$ as $t \rightarrow \infty$. Since $t^{-(1+\eta)} A_t^\nu \rightarrow -\infty$ as $t \rightarrow \infty$, property 1 is proved.

Proof of 2. Restate 2 as follows: for every $\varepsilon > 0$,

$$\text{number of (excursion of } R^\nu \text{ with length } > 2\varepsilon) < \infty \text{ a.s.}$$

Fix $\varepsilon > 0$ and define events $C_n = \{\sup_{s \in [(n-1)\varepsilon, n\varepsilon]} (X_{(n+1)\varepsilon}^\nu + A_{(n+1)\varepsilon}^\nu - X_s^\nu - A_s^\nu) > 0\}$. It is easily seen that it suffices to show that $\mathbb{P}(C_n \text{ infinitely often}) = 0$. By (4.14), it is enough to prove that

$$\sum_{n \geq 1 + s_0/\varepsilon} \mathbb{P}(C_n \cap C^{s_0}) < \infty, \quad \text{for every large } s_0, \quad (4.15)$$

where $C^{s_0} = \{\sup_{t \geq s_0} t^{-(\gamma-3)} |X_t^{(1)}| \leq \delta\}$ for some positive (small) constant $\delta > 0$ to be chosen later. Now

$$C_n \subset \left\{ \begin{aligned} &\sup_{s \in [(n-1)\varepsilon, n\varepsilon]} \left(X_{(n+1)\varepsilon}^{(2)} - X_s^{(2)} \right) \geq \frac{c \Gamma(4-\gamma)}{(\gamma-3)(\gamma-2)\mu^{\gamma-2}} \varepsilon^{\gamma-2} \left((n+1)^{\gamma-2} - n^{\gamma-2} \right) \\ &- \sup_{s \in [(n-1)\varepsilon, n\varepsilon]} \left(X_{(n+1)\varepsilon}^{(1)} - X_s^{(1)} \right) \right\}.$$

For every n larger than $1 + s_0/\varepsilon$, on C^{s_0} , we have:

$$\sup_{s \in [(n-1)\varepsilon, n\varepsilon]} \left(X_{(n+1)\varepsilon}^{(1)} - X_s^{(1)} \right) \leq 2\delta \varepsilon^{\gamma-3} (n+1)^{\gamma-3} \leq 2\delta \varepsilon^{\gamma-3} 2^{\gamma-3} n^{\gamma-3}.$$

Consequently, for every n larger than $1 + s_0/\varepsilon$,

$$C_n \cap C^{s_0} \subset \left\{ \sup_{s \in [(n-1)\varepsilon, n\varepsilon]} \left(X_{(n+1)\varepsilon}^{(2)} - X_s^{(2)} \right) \geq \left(\frac{c \Gamma(4-\gamma)}{(\gamma-3)\mu^{\gamma-2}} \varepsilon^{\gamma-2} - \delta \varepsilon^{\gamma-3} 2^{\gamma-2} \right) n^{\gamma-3} \right\}.$$

Taking $\delta = \varepsilon \frac{c \Gamma(4-\gamma)}{(\gamma-3)\mu^{\gamma-2} 2^{\gamma-1}}$, and denoting $\frac{c \Gamma(4-\gamma)}{2(\gamma-3)\mu^{\gamma-2}} \varepsilon^{\gamma-2}$ by ρ , we thus have for every n large enough:

$$C_n \cap C^{s_0} \subset \left\{ \sup_{s \in [(n-1)\varepsilon, n\varepsilon]} \left(X_{(n+1)\varepsilon}^{(2)} - X_s^{(2)} \right) \geq \rho n^{\gamma-3} \right\}.$$

Now, considering a Poisson point process Δ with intensity $\mathbf{1}_{x \geq 1} \nu(ds, dx)$, where

$$\nu(ds, dx) = \frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xs/\mu} ds dx,$$

observe that

$$\begin{aligned}
& \mathbb{P} \left(\sup_{s \in [(n-1)\varepsilon, n\varepsilon]} \left(X_{(n+1)\varepsilon}^{(2)} - X_s^{(2)} \right) \geq \rho n^{\gamma-3} \right) \\
& \leq \mathbb{P} \left(\sum_{s \in [(n-1)\varepsilon, (n+1)\varepsilon]} \Delta_s \geq \rho n^{\gamma-3} \right) \\
& \leq \rho^{-1} n^{-\gamma+3} \mathbb{E} \left[\sum_{s \in [(n-1)\varepsilon, (n+1)\varepsilon]} \Delta_s \right] \\
& = \rho^{-1} n^{-\gamma+3} \frac{c}{\mu} \int_{(n-1)\varepsilon}^{(n+1)\varepsilon} ds \int_1^\infty dx x \frac{1}{x^{\gamma-1}} e^{-xs/\mu}.
\end{aligned}$$

We deduce that for every n larger than $2 + s_0/\varepsilon$,

$$\mathbb{P}(C_n \cap C^{s_0}) \leq \frac{2\varepsilon c}{\rho\mu} n^{-\gamma+3} \int_1^\infty dx x^{2-\gamma} e^{-nx\varepsilon/(2\mu)} \leq \frac{4c}{\rho} n^{-\gamma+2} e^{-n\varepsilon/(2\mu)},$$

which proves (4.15) and completes the proof of assertion 2.

Proof of 3. To show property 3, we first consider the case $t = 0$. We aim at showing that $\inf\{s > 0 : X_s^\nu + A_s^\nu < 0\} = 0$ a.s. Since for every $s \in [0, \infty)$ and $x \in (0, \infty)$, $\frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xs/\mu} \leq \frac{c}{\mu} \frac{1}{x^{\gamma-1}}$, we can couple the process X^ν and construct a stable process L with index $\gamma - 2$ with no negative jumps such that

$$\forall s \geq 0 \quad \forall u \in \mathbb{R}, \quad \mathbb{E}[\exp(iuL_s)] = \exp\left(s \int_0^\infty dx (e^{iux} - 1 - iux) \frac{c}{\mu} \frac{1}{x^{\gamma-1}}\right)$$

satisfying

$$\forall s \geq 0, \quad X_s^\nu \leq L_s + \frac{c}{\mu} \int_0^s dr \int_0^\infty dx \frac{1}{x^{\gamma-2}} (1 - e^{-xr/\mu}), \quad (4.16)$$

i.e.,

$$\forall s \geq 0, \quad X_s^\nu \leq L_s + \frac{c \Gamma(4 - \gamma)}{(\gamma - 3)(\gamma - 2)\mu^{\gamma-2}} s^{\gamma-2}.$$

Consequently $X^\nu + A^\nu \leq L$. Since $\inf\{s > 0 : L_s < 0\} = 0$ a.s., with probability 1, 0 is not an isolated point of the set $\{t : R_t^\nu = 0\}$. This implies assertion 3 by routine arguments (see [AL98] for details).

Proof of 4. Here again, we shall use a coupling argument. We need an inequality opposite to (4.16). Specifically, we have to bound the increments of $X^\nu + A^\nu$ from below. We first focus on $X^{(1)}$. Let $t \in (0, \infty)$. Arguing as before, since for every $s \in [0, t]$ and $x \in (0, 1)$, $\frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xs/\mu} \geq \frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xt/\mu}$, we can construct a Lévy process $(Q_s^{(1)})_{s \in [0, t]}$ such that

$$\forall (u, s) \in \mathbb{R} \times [0, t], \quad \mathbb{E}[\exp(iuQ_s^{(1)})] = \exp\left(s \int_0^1 dx (e^{iux} - 1 - iux) \frac{c}{\mu} \frac{1}{x^{\gamma-1}} e^{-xt/\mu}\right)$$

satisfying

$$\forall s \in [0, t], \quad X_t^{(1)} - X_s^{(1)} \geq Q_t^{(1)} - Q_s^{(1)} + \frac{c}{\mu} \int_s^t dr \int_0^1 dx \frac{1}{x^{\gamma-2}} (e^{-xt/\mu} - e^{-xr/\mu}).$$

Since for every $a, b \in (0, \infty)$ such that $a < b$, $e^{-a} - e^{-b} \leq b - a$, we have for every $s \in [0, t]$

$$X_t^{(1)} - X_s^{(1)} \geq Q_t^{(1)} - Q_s^{(1)} - \frac{c}{2(4-\gamma)\mu^2}(t-s)^2.$$

As a result, for every $s \in [0, t]$,

$$X_t^\nu - X_s^\nu \geq Q_t^{(1)} - Q_s^{(1)} - \frac{c}{2(4-\gamma)\mu^2}(t-s)^2 + B_t^\nu - B_s^\nu.$$

We easily deduce that there exists $C > 0$ (only depending on t) such that for every $s \in [0, t]$,

$$X_t^\nu + A_t^\nu - (X_s^\nu + A_s^\nu) \geq Q_t^{(1)} - Q_s^{(1)} - C(t-s).$$

Consequently

$$\sup \{X_t^\nu + A_t^\nu - (X_s^\nu + A_s^\nu) : s \in [0, t]\} \geq \sup \{Q_t^{(1)} - Ct - (Q_s^{(1)} - Cs) : s \in [0, t]\}.$$

Now, applying [Ber96, Theorem VII. 2 and page 158] to the Lévy process $(Q_s^{(1)} - Cs)_{s \in [0, t]}$, we have

$$\mathbb{P} \left(Q_t^{(1)} - Ct = \inf \{Q_s^{(1)} - Cs : s \in [0, t]\} \right) = 0.$$

We deduce that

$$\mathbb{P} (X_t^\nu + A_t^\nu = \inf \{X_s^\nu + A_s^\nu : s \in [0, t]\}) = 0,$$

which is assertion 4. □

4.7.3 Open questions

The argument used to prove of Lemma 4.5 does not work in our present setting. Observe though that (4.2) still holds here. That is why we believe that the following result is true:

Conjecture 4.1. *Suppose ν satisfies (4.8). Let \mathbf{SC}_n^ν be the ordered sequence of component sizes of \mathcal{SG}_n and $(|\gamma_j|, j \geq 1)$ be the ordered sequence of the excursion lengths of R^ν . Then*

$$n^{-(\gamma-2)/(\gamma-1)} \mathbf{SC}_n^\nu \xrightarrow[n \rightarrow \infty]{(d)} (|\gamma_j|, j \geq 1)$$

with respect to the l_{\searrow}^2 topology.

As before, Conjecture 4.1 would be proven as soon as the following result is shown:

Conjecture 4.2. *Suppose ν satisfies (4.8). The law of the rescaled walk \overline{W}_n conditioned on the event $\{\mathcal{G}_n \text{ is simple}\}$ converges in distribution to $X^\nu + A^\nu$ as $n \rightarrow \infty$.*

BIBLIOGRAPHIE

- [AS64] M. ABRAMOWITZ et I. A. STEGUN : *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 de *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.
- [ABR09] L. ADDARIO-BERRY et B. REED : Minima in branching random walks. *Ann. Probab.*, 37(3):1044–1079, 2009.
- [Ald97] D. J. ALDOUS : Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25(2):812–854, 1997.
- [AL98] D. J. ALDOUS et V. LIMIC : The entrance boundary of the multiplicative coalescent. *Electron. J. Probab.*, 3:No. 3, 59 pp. (electronic), 1998.
- [ABT03] R. ARRATIA, A. D. BARBOUR et S. TAVARÉ : *Logarithmic combinatorial structures: a probabilistic approach*. EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich, 2003.
- [BC78] E. A. BENDER et E. R. CANFIELD : The asymptotic number of labeled graphs with given degree sequences. *J. Combinatorial Theory Ser. A*, 24(3):296–307, 1978.
- [Ber96] J. BERTOIN : *Lévy processes*, volume 121 de *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.
- [Ber01] J. BERTOIN : Homogeneous fragmentation processes. *Probab. Theory Related Fields*, 121(3):301–318, 2001.
- [Ber02] J. BERTOIN : Self-similar fragmentations. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(3):319–340, 2002.
- [Ber06] J. BERTOIN : *Random fragmentation and coagulation processes*, volume 102 de *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.
- [Ber08] J. BERTOIN : Asymptotic regimes for the occupancy scheme of multiplicative cascades. *Stochastic Process. Appl.*, 118(9):1586–1605, 2008.

- [BG04] J. BERTOIN et A. V. GNEDIN : Asymptotic laws for nonconservative self-similar fragmentations. *Electron. J. Probab.*, 9:no. 19, 575–593 (electronic), 2004.
- [BS09] J. BERTOIN et V. SIDORAVICIUS : The structure of typical clusters in large sparse random configurations. *J. Stat. Phys.*, 135(1):87–105, 2009.
- [BvdHvL10a] S. BHAMIDI, R. van der HOFSTAD et J. S.H. van LEEUWAARDEN : Scaling limits for critical inhomogeneous random graphs with finite third moments. *Electron. J. Probab.*, 15:no. 54, 1682–1703, 2010.
- [BvdHvL10b] S. BHAMIDI, R. van der HOFSTAD et J. S.H. van LEEUWAARDEN : Novel scaling limits for critical inhomogeneous random graphs. arXiv:0909.1472v2, 2010.
- [Big92] J. D. BIGGINS : Uniform convergence of martingales in the branching random walk. *Ann. Probab.*, 20(1):137–151, 1992.
- [Bol80] B. BOLLOBÁS : A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4):311–316, 1980.
- [Bol84] B. BOLLOBÁS : The evolution of random graphs. *Trans. Amer. Math. Soc.*, 286(1):257–274, 1984.
- [Bol01] B. BOLLOBÁS : *Random graphs*, volume 73 de *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second édition, 2001.
- [BG97] S. BOUCHERON et D. GARDY : An urn model from learning theory. *Random Structures Algorithms*, 10(1-2):43–67, 1997.
- [BDML06] T. BRITTON, M. DEIJFEN et A. MARTIN-LÖF : Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.*, 124(6):1377–1397, 2006.
- [BF93] J. BUNGE et M. FITZPATRICK : Estimating the number of species : a review. *J. Am. Stat. Assoc.*, 88:364–373, 1993.
- [CH03] H.-H. CHERN et H.-K. HWANG : Partial match queries in random quadtrees. *SIAM J. Comput.*, 32(4):904–915 (electronic), 2003.
- [CJ11] N. CURIEN et A. JOSEPH : Partial match queries in two-dimensional quadtrees: a probabilistic approach. *Adv. in Appl. Probab.*, 43(1):178–194, 2011.
- [CLG11] N. CURIEN et J.-F. LE GALL : Random recursive triangulations of the disk via fragmentation theory. *To appear in Ann. Probab.*, 2011.
- [Dev86] L. DEVROYE : A note on the height of binary search trees. *J. Assoc. Comput. Mach.*, 33(3):489–498, 1986.

-
- [Dev87] L. DEVROYE : Branching processes in the analysis of the heights of trees. *Acta Inform.*, 24(3):277–298, 1987.
- [Dev99] L. DEVROYE : Universal limit laws for depths in random trees. *SIAM J. Comput.*, 28(2):409–432, 1999.
- [ER60] P. ERDŐS et A. RÉNYI : On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.
- [EK86] S. N. ETHIER et T. G. KURTZ : *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.
- [FB74] R. A. FINKEL et J. L. BENTLEY : Quad trees a data structure for retrieval on composite keys. *Acta Informatica*, 4(1):1–9, mars 1974.
- [FGCR93] P. FLAJOLET, G. GONNET, C. PUECH et J.M. ROBSON : Analytic variations on quadtrees. *Algorithmica*, 10(6):473–500, 1993.
- [FLLS95] P. FLAJOLET, G. LABELLE, L. LAFOREST et B. SALVY : Hypergeometrics and the cost structure of quadtrees. *Random Structures Algorithms*, 7(2):117–144, 1995.
- [FS09] P. FLAJOLET et R. SEDGEWICK : *Analytic combinatorics*. Cambridge University Press, Cambridge, 2009.
- [Gar02] D. GARDY : Occupancy urn models in the analysis of algorithms. *J. Statist. Plann. Inference*, 101(1-2):95–105, 2002.
- [GHP07] A. V. GNEDIN, B. HANSEN et J. PITMAN : Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probab. Surv.*, 4:146–171, 2007.
- [HMPW08] B. HAAS, G. MIERMONT, J. PITMAN et M. WINKEL : Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *Ann. Probab.*, 36(5):1790–1837, 2008.
- [HM10] H. HATAMI et M. MOLLOY : The scaling window for a random graph with a given degree sequence. *ACM-SIAM Symposium on Discrete Algorithms*, 2010.
- [HG00] F. HE et K. J. GASTON : Estimating species abundance from occurrence. *The American Naturalist*, 156(5):553–559, 2000.
- [Hol86] L. HOLST : On birthday, collectors’, occupancy and other classical urn problems. *Internat. Statist. Rev.*, 54(1):15–27, 1986.
- [HS09] Y. HU et Z. SHI : Minimal position and critical martingale convergence in branching random walks, and directed polymers on disordered trees. *Ann. Probab.*, 37(2):742–789, 2009.

- [IL97] A. ISERLES et Y. LIU : Integro-differential equations and generalized hypergeometric functions. *J. Math. Anal. Appl.*, 208(2):404–424, 1997.
- [JS03] J. JACOD et A. N. SHIRYAEV : *Limit theorems for stochastic processes*, volume 288 de *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second édition, 2003.
- [Jan08] S. JANSON : The largest component in a subcritical random graph with a power law degree distribution. *Ann. Appl. Probab.*, 18(4):1651–1668, 2008.
- [Jan09] S. JANSON : The probability that a random multigraph is simple. *Combin. Probab. Comput.*, 18(1-2):205–225, 2009.
- [JKŁP93] S. JANSON, D. E. KNUTH, T. ŁUCZAK et B. PITTEL : The birth of the giant component. *Random Structures Algorithms*, 4(3):231–358, 1993.
- [JL09] S. JANSON et M. LUCZAK : A new approach to the giant component problem. *Random Structures Algorithms*, 34(2):197–216, 2009.
- [JK77] N. L. JOHNSON et S. KOTZ : *Urn models and their application*. John Wiley & Sons, New York-London-Sydney, 1977. An approach to modern discrete probability theory, Wiley Series in Probability and Mathematical Statistics.
- [Jos11a] A. JOSEPH : A phase transition for the heights of a fragmentation tree. *To appear in Random Structures Algorithms*, 2011.
- [Jos11b] A. JOSEPH : The component sizes of a critical random graph with given degree sequence. arXiv:1012.2352v2, 2011.
- [KS08] M. KANG et T. G. SEIERSTAD : The critical phase for random graphs with a given degree sequence. *Combin. Probab. Comput.*, 17:67–86, 2008.
- [KSC78] V. F. KOLCHIN, B. A. SEVASTYANOV et V. P. CHISTYAKOV : *Random allocations*. V. H. Winston & Sons, Washington, D.C., 1978. Translated from the Russian, Translation edited by A. V. Balakrishnan, Scripta Series in Mathematics.
- [Liu97] Q. LIU : Sur une équation fonctionnelle et ses applications: une extension du théorème de Kesten-Stigum concernant des processus de branchement. *Adv. in Appl. Probab.*, 29(2):353–373, 1997.
- [Liu01] Q. LIU : Asymptotic properties and absolute continuity of laws stable by random weighted mean. *Stochastic Process. Appl.*, 95(1):83–107, 2001.
- [Łuc90] T. ŁUCZAK : Component behavior near the critical point of the random graph process. *Random Structures Algorithms*, 1(3):287–310, 1990.

-
- [McD95] MCDIARMID : Minimal positions in a branching random walk. *Ann. Appl. Probab.*, 5:128–139, 1995.
- [MT09] S. MEYN et R. L. TWEEDIE : *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second édition, 2009. With a prologue by Peter W. Glynn.
- [MR95] M. MOLLOY et B. REED : A critical point for random graphs with a given degree sequence. *Random Structures Algorithms*, 6(2-3):161–179, 1995.
- [MR98] M. MOLLOY et B. REED : The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.*, 7(3):295–305, 1998.
- [PPY92] M. PERMAN, J. PITMAN et M. YOR : Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Relat. Fields*, 92(1):21–39, 1992.
- [Pit84] B. PITTEL : On growing random binary trees. *J. Math. Anal. Appl.*, 103(2):461–480, 1984.
- [Pit08] B. PITTEL : On the largest component of a random graph with a subpower-law degree sequence in a subcritical phase. *Ann. Appl. Probab.*, 18(4):1636–1650, 2008.
- [Res08] S. I. RESNICK : *Extreme values, regular variation and point processes*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2008. Reprint of the 1987 original.
- [Rob79] J. M. ROBSON : The height of binary search trees. *Austral. Comput. J.*, 11(4):151–153, 1979.
- [Rös92] U. RÖSLER : A fixed point theorem for distributions. *Stochastic Process. Appl.*, 42(2):195–214, 1992.
- [Tur09] T. S. TUROVA : Diffusion approximation for the components in critical inhomogeneous random graphs of rank 1. arXiv:0907.0897v2, 2009.
- [vdH10] R. van der HOFSTAD : Critical behavior in inhomogeneous random graphs. arXiv:0902.0216v2, 2010.
- [Wor78] N. C. WORMALD : *Some Problems in the Enumeration of Labelled Graphs*. Thèse de doctorat, Newcastle University, 1978.