Combinatorial and algorithmic approaches to random trees and graphs

Jordan Barrett

Department of Mathematics and Statistics
Faculty of Science
McGill University, Montreal

November 2022

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

Contents

Abstract

To help us understand some of the fascinating properties emerging from real-world networks such as social networks and the world wide web, we study graphs that are constructed with simple rules involving some degree of randomness, i.e., random graphs. The field of random graph theory, although much younger than some other fields such as number theory and topology, contains a wealth of deep and interesting results, as well as an an ever growing number of constructive methods for generating random graphs. Despite this huge number of constructive methods, many of the results in random graph theory tell us information about just a few graphs with very simple constructions.

In this thesis, we study constructive models for random graphs that are slight variations of some of the most well known and well understood models. The benefit here is that we may pull from the vast sea of existing results on the simpler models to help us prove results on the more complicated models.

The first process for generating random graphs that we study is a variation of the Erdős-Rényi process [37]. The modification comes in the form of forbidding connections between special vertices that refuse to connect to one another. We show that if the number of special vertices in this construction is above a certain threshold, then with high probability no linear-sized connected components will emerge by the end of the graph's construction. Together with a result from Logan, Molloy and Prałat [53], our result establishes the existence of a phase transition in the behaviour of the component sizes.

The second random graph process that we study is a modification of the configuration model. The configuration model attempts to construct a uniformly random element of $\mathcal{G}_{\mathrm{d}}$ where $\mathcal{G}_{\mathrm{d}}$ is the set of simple graphs with degree sequence d. Although the configuration model does not always succeed in building a simple graph, it is known that, under certain conditions on d, results for the configuration model translate to results for uniform elements of $\mathcal{G}_{\mathrm{d}}$. In a similar fashion, we present a model that attempts to construct uniform elements from the set of simple tree-rooted graphs with degree sequence d. Similar to the configuration

model, we show that results for our model translate to results for uniform elements of simple tree-rooted graphs. We then use this translation property to prove that the sequence of rooted trees emerging from a sequence of uniform simple tree-rooted graphs with a growing number of vertices, after rescaling, converges in distribution to the Brownian continuum random tree.

Lastly, we take an alternative approach to the concept of using simple graphs to help construct more complicated graphs. Specifically, given a weighted graph (g, w) and an arbitrary connected spanning subgraph h of g, we present an algorithm that transforms h into a minimum-weight spanning tree of g via a series of local optimizations. We show that, starting from the complete graph with independent Uniform$[0, 1]$ edge weights, we can with high probability transform a connected spanning subgraph into a minimum-weight spanning tree while only ever changing subgraphs with total weight bounded by $1 + \epsilon$, and that such a transformation is with high probability impossible if we change the bound to $1 - \epsilon$.

Abrégé

Pour nous aider à comprendre certaines des propriétés fascinantes qui émergent dans les réseaux du monde réel, tels que les réseaux sociaux et le web, nous étudions des graphes construits à partir de règles simples impliquant un certain degré d'aléa, c'est-à-dire des graphes aléatoires. Le domaine de la théorie des graphes aléatoires, bien que beaucoup plus jeune que certains autres domaines tels que la théorie des nombres et la topologie, contient une multitude de résultats fondamentaux et fascinants, ainsi qu'un nombre toujours croissant de méthodes de construction pour générer de tels graphes aléatoires. Cependant, malgré ce nombre croissant de méthodes constructives, de nombreux résultats de la théorie des graphes aléatoires se limitent souvent aux quelques graphes ayant des constructions très simples.

Dans cette thèse, nous développons et étudions de nouveaux modèles constructifs pour les graphes aléatoires obtenus après avoir légèrement modifié les modèles les plus connus et les mieux compris. L'avantage ici est que nous pouvons puiser dans la vaste étendue de résultats existants sur ces modèles plus simples afin de nous aider à prouver des résultats sur les modèles plus complexes.

Le premier processus de génération de graphes aléatoires que nous étudions est une variante du processus Erdős-Rényi [37]. La modification se présente sous la forme de connexions interdites entre un ensemble de sommets spécifiques, qui refusent donc de se connecter les uns aux autres. Nous montrons que si le nombre de sommets spécifiques dans cette construction est supérieur à un certain seuil, alors, avec une forte probabilité, aucune composante connexe de taille linéaire n'émergera lors de la construction du graphe. Combiné avec un résultat de Logan, Molloy et Prałat [53], nous établissons l'existence d'une transition de phase concernant la taille des composantes.

Le deuxième graphe aléatoire que nous étudions est une modification du modèle de configuration. Le modèle de configuration tente de construire aléatoirement un élément uniformément choisi dans $\mathcal{G}_{\mathrm{d}}$, où $\mathcal{G}_{\mathrm{d}}$ est l'ensemble des graphes simples dont les degrés sont donnés par d. Bien que le modèle de configuration ne réussisse pas toujours à construire un

tel graphe simple, il est connu que, sous certaines conditions sur d, les résultats du modèle de configuration sont intimement liés aux éléments uniformes de $\mathcal{G}_{\mathrm{d}}$. Partant de cette observation, nous présentons un modèle qui essaye de construire des éléments uniformes à partir de l'ensemble de graphes arborescents simples avec une séquence de degrés d. Semblable au modèle de configuration, nous montrons que les résultats de notre modèle sont liés aux éléments uniformes de graphes arborescents simples. Nous utilisons ensuite cette propriété pour prouver que la séquence d'arbres enracinés émergeant d'une séquence de graphes arborescents simples uniformes avec un nombre croissant de sommets converge en distribution vers l'arbre brownien (sous condition de renormalization adéquate).

Pour finir, nous adoptons une approche alternative au concept d'utilisation de graphes simples pour aider à construire des graphes plus complexes. Plus précisément, étant donné un graphe pondéré (g, w) et un sous-graphe couvrant arbitraire h de g, nous présentons un algorithme qui transforme h en l'arbre couvrant minimal de g via une série d'optimisations locales. Nous montrons qu'à partir du graphe complet pondéré par des variables Uniform$[0, 1]$ indépendantes, nous pouvons avec une forte probabilité transformer un sous-graphe couvrant connexe en un arbre couvrant de poids minimal tout en ne changeant que des sous-graphes avec un poids total borné par $1 + \epsilon$, et qu'une telle transformation est impossible avec une forte probabilité si nous changeons la borne en $1 - \epsilon$.

Acknowledgements

I cannot overstate how much I need to thank my supervisor, Louigi Addario-Berry, for all that he has done for me in the past four and a half years. Louigi has helped me mature as a mathematician, and I could not possibly give a comprehensive list of reasons to thank him. I am grateful for the weekly meetings, the lessons in turning heuristic arguments into rigorous proofs, the thorough edits, the dedication to good writing, and the fun and friendly conversations in-between. All in all, I owe a large part of my success as a student to him, and I was incredibly lucky to have him as my supervisor.

I would like to thank Benoît Corsini for his contribution to my success, as well as my enjoyment, during my time at McGill. Benoît taught me a lot about probabilistic arguments while he and I were working together on [6]. Outside of research, I am grateful to have Benoît as a friend, and I have particularly enjoyed sharing our love of music.

I would like to thank Louigi Addario-Berry, Nicolas Broutin, and Gábor Lugosi for organizing the Bellairs probability and combinatorics workshops, and in particular I would like to thank Louigi for inviting me to said workshops in 2019 and 2022. The Bellairs workshops were possibly my two favourite weeks of the past four years. I met so many incredible mathematicians, attended some insightful lectures, and got the opportunity to work on some intriguing open problems. In particular, [6] spawned from an open problem presented by Gábor.

The pandemic was, to put it lightly, a pretty massive speed bump in the middle of an already bumpy road for me. However, the Graduate Student Association for Mathematics and Statistics (GSAMS) hosted a number of remote events that made the pandemic a bit less awful. In particular, I want to thank Gavin Barill for organizing the weekly social event "quarantini time". Although I was stuck at home, I always looked forward to Friday evenings where I could hop on Zoom and hang out with a great group of friends from the department.

I would like to thank the friends I made at McGill for making my time there as enjoyable as possible. A special thanks to Carlos and Jake for letting me win all of those games of pool.

Lastly, I would like to thank Mom and Dad. You guys have supported my goal to be a mathematician since the day I decided it was what I wanted to do. I am forever grateful, and I love you both.

Contributions

This thesis is a collection of three original works [4–6] either already published or submitted for publication. The papers appear in Chapters 3, 2, and 4 respectively. The contributions are summarized below.

Chapter 2, *Multi-source invasion percolation on the complete graph*, is a reproduction of the article [5], co-authored with Louigi Addario-Berry, available on arXiv and submitted for publication in 2022. For this work, both authors provided equal contributions.

Chapter 3, *Random tree-weighted graphs*, is a reproduction of the article [4], co-authored with Louigi Addario-Berry, and published in *Stochastic Analysis, Random Fields and Integrable Probability* in 2021. For this work, both authors provided equal contributions.

Chapter 4, *Finding minimum spanning trees via local improvements*, is a reproduction of the article [6], co-authored with Louigi Addario-Berry and Benoît Corsini, available on arXiv and submitted for publication in 2022. For this work, all three authors provided equal contributions.

List of Figures

In other words, in order to obtain U_1, U_2, U_3, U_4, and U_5, we sequentially add 8, 9, 3, 2, and 1 to U_0.

Chapter 1

Introduction

Networks are one of the most fundamental structures in nature. Some examples of networks are family trees, city bus routes, social networks, and ecosystems. As our population increases and we progress through the age of technology, understanding the structure of real-world networks becomes more and more important. The irony is that, with an increased population and more advanced technology, some of the most important networks are becoming much larger and more difficult to access. To see this phenomanon, consider the network consisting of websites and hyperlinks between websites. This is a well known network called the *web graph*. In [15], Albert, Jeong, and Barabasi explain that (a) understanding the structure of this network helps us locate information on the world wide web, and (b) it is impossible to reproduce this network in its entirety as its size is expanding at an uncontrollable rate.

To help us understand the structure of real-world networks, we study *random graphs*. A random graph is a collection of vertices and edges that is constructed with some element of randomness (typically involving the edges). The randomness in a graph's construction exists as a proxy for the unpredictable behaviour of whatever entities might make up a network. However, this behaviour is sometimes incredibly nuanced. For example, how could a randomly constructed graph account for a city bus avoiding a certain road because of potholes? Although the structure of real-world networks is multi-variable by nature, we can study networks through the lens of mean-field theory [23]. In brief, mean-field theory takes the perspective that by averaging over all of the factors that contribute to a system (in our case a network), we can use a much simpler structure to model the system. Although we will lose some information about the underlying network, studying a simpler random model can still give us meaningful insight into the network's structure.

In the three papers that make up this thesis, we take existing random graph models with very simple constructions and we use them as the foundations for slightly more complex models.

Before continuing, I will take a moment to go over some important definitions and notation used throughout this thesis.

- A graph $g = (v(g), e(g))$ is a set of vertices $v(g)$ and a set of edges $e(g)$ between elements of $v(g)$. We typically write $v(g)$ and $e(g)$ for the respective vertex set and edge set of g. We denote an edge between vertices u and v by uv or by vu, and we consider edges to be unordered unless otherwise specified.

- A *multigraph* is a graph g in which we allow self-loops and multiple instances of the same edge. In contrast, g is a *simple graph* if there are no self-loops and there is at most one edge between any pair of distinct vertices. Unless otherwise specified, we always assume that graphs are simple.

- For $n \in \mathbb{N}$ we write $[n] := \{1, \ldots, n\}$. The set $[n]$ will frequently be the vertex set of a graph.

- For a graph g, a *subgraph* of g is any graph h with $v(h) \subseteq v(g)$ and $e(h) \subseteq e(g)$. We say that h is an *induced subgraph* of g if $e(h)$ contains every edge in $e(g)$ with both endpoints in $v(h)$. We say that h is a *spanning subgraph* of g if $v(h) = v(g)$. We say that h is a *spanning tree* of g if h is a spanning subgraph of g that is also a tree.

- For a graph g and a vertex $v \in v(g)$, the *degree of v in g* is the number of endpoints of edges incident to v in $e(g)$; note that self-loops add 2 to the degree. We write $d_g(v)$ for the degree of v in g and, given some ordering $(v_1, \ldots, v_n)$ of the vertices, we write $(d_g(v_1), \ldots, d_g(v_n))$ for the *degree sequence* of g with vertex ordering $(v_1, \ldots, v_n)$. A sequence $(d_1, \ldots, d_n)$ is a degree sequence if it the degree sequence of some graph g.

- Given a sequence of events $(A_n, n \geq 1)$, we say that A_n occurs *with high probability* if $\mathbf{P}\{A_n\} \to 1$ as $n \to \infty$.

Dating back to 1959, the most well studied and well understood process for generating random graphs is the Erdős-Rényi process [37]. Let K_n be the complete graph with vertex set $v(K_n) = [n]$ and edge set $e(K_n) = \big\{ij, 1 \leq i < j \leq n\big\}$, and let $\mathbb{X} : e(K_n) \to [0,1]$ be the random weight function with $\mathbb{X}(e) \sim \text{Uniform}[0,1]$, independently for all $e \in e(K_n)$. Next, for all $0 \leq p \leq 1$ let $G(n,p)$ be the graph with vertex set $[n]$ and edge set $\{e \in e(K_n) : \mathbb{X}(e) \leq p\}$. Then $G(n,p)$ is an *Erdős-Rényi graph* on vertex set $[n]$ with parameter p, and

$\big(G(n,p), 0 \leq p \leq 1\big)$ is a *continuous time Erdős-Rényi process.* The Erdős-Rényi graph is arguably the simplest random graph model, and its structure is well-understood. The Erdős-Rényi process is the building block for the model studied in Chapter 2, and the particular Erdős-Rényi graph $G(n, 1)$ (i.e., the complete graph with independent Uniform$[0, 1]$ edge weights) is the building block for the model studied in Chapter 4.

Another well known and well studied random graph process is the configuration model. As motivation for the model, suppose we wish to understand the typical structure of graphs with a fixed degree sequence. More specifically, let $\mathrm{d} = (d_1, \dots, d_n)$ be a degree sequence and let $\mathcal{G}_{\mathrm{d}}$ be the set of simple graphs with vertex set $[n]$ and degree sequence d. What does a "typical" graph in $\mathcal{G}_{\mathrm{d}}$ look like? The configuration model construction is a sort of brute force attempt at answering this question. We start the construction with a set of vertices $[n]$ and no edges. Then, for all $i \in [n]$ endow vertex i with d_i half-edges. Next, recursively choose two distinct half-edges uniformly at random from the set of pairs of remaining half-edges and join them to form an edge. Once all half-edges have been paired, the resulting random graph is an instance of the configuration model with vertex set $[n]$ and degree sequence d; we denote it by $\mathrm{CM}_n(\mathrm{d})$. The configuration model does not always yield an element of $\mathcal{G}_{\mathrm{d}}$, though. In fact, the construction can result in a graph with loops and multi-edges, i.e., a multigraph. However, it is straightforward to show that, conditionally given that $\mathrm{CM}_n(\mathrm{d})$ is simple, then $\mathrm{CM}_n(\mathrm{d})$ is a uniformly random element of $\mathcal{G}_{\mathrm{d}}$. Thus, if the probability that $\mathrm{CM}_n(\mathrm{d})$ is simple is not too small, then we can prove results about the distribution of $\mathcal{G}_{\mathrm{d}}$ by proving corresponding results about $\mathrm{CM}_n(\mathrm{d})$. The configuration model is the building block for the model studied in Chapter 3.

1.1 Competitive networks

Suppose a select few people in a community are running for mayor. Then a network can be formed by considering people as vertices and by creating edges between two vertices if the corresponding two people vote for the same mayoral candidate. Assuming mayoral candidates always vote for themselves, this graph is simply a collection of disjoint cliques, one for each candidate. However, the distribution of the clique sizes is precisely what determines the outcome of the election. We would therefore like to understand how these cliques are formed and how their sizes compare to one another.

In Chapter 2 we study a model for competitive networks that was first introduced by Logan, Molloy, and Prałat in [53]. Consider a modified version of the Erdős-Rényi process, written as $(G(n,k,p), 0 \leq p \leq 1)$, where we do not allow any paths between vertices in $[k]$. More specifically, let $m = \binom{n}{2}$, let $e_1, \ldots, e_m$ be the order in which the edges are added during the Erdős-Rényi process $(G(n,p), 0 \leq p \leq 1)$, let $p_0 = 0$, and write $p_i = \mathbb{X}(e_i)$ for $1 \leq i \leq m$. Then the process $(G(n,k,p), 0 \leq p \leq 1)$ can be defined as follows. Starting from $G(n,k,p_0) = G(n,0)$, which is the empty graph on vertex set $[n]$, inductively for $0 \leq i < \binom{n}{2}$, construct $G(n,k,p_{i+1})$ from $G(n,k,p_i)$ by setting $G(n,k,p_{i+1}) = G(n,k,p_i)$ if $e(G(n,k,p_i)) \cup \{e_{i+1}\}$ contains a path joining two vertices in $[k]$, and otherwise setting $v(G(n,k,p_{i+1})) = v(G(n,k,p_i)) = [n]$ and $e(G(n,k,p_{i+1})) = e(G(n,k,p_i)) \cup \{e_{i+1}\}$. In either case, set $G(n,k,p) = G(n,k,p_i)$ for $p_i \leq p < p_{i+1}$ and set $G(n,k,p) = G(n,k,p_m)$ for $p_m \leq p \leq 1$. We justify calling this a modified Erdős-Rényi process by noting that the processes $(G(n,p), 0 \leq p \leq 1)$ and $(G(n,1,p), 0 \leq p \leq 1)$ are identical. In general, $(G(n,p), 0 \leq p \leq 1)$ and $(G(n,k,p), 0 \leq p \leq 1)$ add the same edges in the same order, except that the latter process refuses to add edges that would create paths between vertices in $[k]$. Informally, we can think of contacts between the n voters happening in a uniformly random order, with each voter making the same decision about who to vote for as the first person they speak to.

The following is a slightly modified but equivalent version of the main result from Chapter 2.

Theorem (Theorem 2.1.1)**.** *Fix positive integers $(k(n), n \geq 1)$, and for $n \geq 1$ let M_n be the size of the largest component of $G(n, k(n), 1)$.*

- *If $k(n)/n^{1/3} \to 0$ then $M_n/n \to 1$ in probability.*
- *If $k(n)/n^{1/3} \to \infty$ then $M_n/n \to 0$ in probability.*

Logan, Molloy and Prałat proved one half of this result: if $k(n)/n^{1/3} \to 0$ then $M_n/n \to 1$ in probability. In our paper we prove the other half, and together the two halves establish the existence of a phase transition in the behaviour of M_n around $k(n) \asymp n^{1/3}$.

The first step in understanding the growth of the largest component in $(G(n,k(n),p), 0 \leq p \leq 1)$ is understanding the growth of the largest component in $(G(n,p), 0 \leq p \leq 1)$. Luckily for us, Erdős and Rényi give a description of this growth in [38], and Béla Bollobás improves on this description in [30]. Together they show that, almost surely,

1. If $pn < 1$ then the size of the largest component of $G(n,p)$ is $O(\log n)$,
2. if $pn = 1$ then the size of the largest component of $G(n,p)$ is $\Theta(n^{2/3})$, and

3. if $pn > 1$ then the size of the largest component of $G(n,p)$ is $\Theta(n)$.

The second step in understanding the growth is comparing the processes $(G(n,p), 0 \leq p \leq 1)$ and $(G(n,k,p), 0 \leq p \leq 1)$. When the two processes are coupled as described before the statement of Theorem 2.1.1, it is straightforward to show that $G(n,k,p)$ is a subgraph of $G(n,p)$. Furthermore, if H is a connected component of $G(n,p)$ and $v(H) \cap [k] = \ell$, then the subgraph of $G(n,k,p)$ induced by $v(H)$ has precisely ℓ components, each containing exactly one vertex in $v(H) \cap [k]$. Informally, the components of $G(n,p)$ get "chopped up" in $G(n,k,p)$, and the amount of chopping a component receives depends on the number of vertices from $[k]$ in said component.

We now state a simplified version of Proposition 2.2.1, a key input to the proof of Theorem 2.1.1.

Proposition (Proposition 2.2.1)**.** *Fix positive integers $(k(n), n \geq 1)$ with $k(n) \in [n]$ and $k(n)/n^{1/3} \to \infty$. Next, let $M_{n,1/n}(k(n))$ be the size of the largest connected component of $G(n,k(n),1/n)$. Then $M_{n,1/n}(k(n))/n^{2/3} \to 0$ in probability.*

Informally, this proposition says that the components of size $\Theta(n^{2/3})$ in $G(n,1/n)$ are chopped into $o(n^{2/3})$-sized pieces in $G(n,k(n),1/n)$. At a high level, Proposition 2.2.1 is a consequence of there being roughly $k(n)n^{-1/3} \gg 1$ vertices from $[k]$ in a component of $G(n,1/n)$ of size $\Theta(n^{2/3})$. However, the proof of Proposition 2.2.1 is very involved and the bulk of Chapter 2 is dedicated to proving it.

To prove Theorem 2.1.1, we combine Proposition 2.2.1 with some existing results by Addario-Berry, Broutin, Goldschmidt and Miermont in [10], by Addario-Berry and Sen in [14], and by Aldous in [18], which together tell us that, for $k(n)/n^{1/3} \to \infty$, with high probability none of the components in $G(n,k(n),1/n)$ are able to grow into a component of size $\Theta(n)$ by the end of the process $(G(n,k(n),p, 0 \leq p \leq 1))$.

1.2 Tree-weighted graphs

Recall that, for a degree sequence $\mathrm{d} = (d_1, \ldots, d_n)$, $\mathcal{G}_{\mathrm{d}}$ is the set of simple graphs with degree sequence d. Suppose that we want to sample from $\mathcal{G}_{\mathrm{d}}$ but we want to ensure that our sampled graph is connected. One strategy is to sample uniformly at random from the set of pairs (g,t) where $g \in \mathcal{G}_{\mathrm{d}}$ and t is a spanning tree of g. By sampling this way, the probability of choosing $g \in \mathcal{G}_{\mathrm{d}}$ is proportional to the number of spanning trees in g, and in particular the probability of choosing a disconnected graph is 0. A natural question then arises: If (G,T)

is sampled in this way, then what is the distribution of T? This is precisely the question we tackle in Chapter 3.

In Chapter 3 we study *tree-rooted graphs*. A tree-rooted graph is a triple (g, t, γ) where g is a connected graph, t is a spanning tree of g, and $\gamma = uv$ is a distinguished oriented edge in $e(g) \setminus e(t)$; we view t as a rooted tree with root u. We say (g, t, γ) is *simple* if g is simple. We call a triple (G, T, Γ) sampled uniformly at random from the set of simple tree-rooted graphs with degree sequence d a *random simple tree-weighted graph with degree sequence* d. Note that if (G, T, Γ) is sampled in this way and (G', T') is sampled as in the previous paragraph, then $G \overset{d}{=} G'$ and $T \overset{d}{=} T'$ (as non-rooted trees) since the number of options for Γ is constant with respect to d. We include the oriented edge because the main result in Chapter 3 relies on a combinatorial construction which is most naturally seen as building tree-rooted graphs.

Given a graph g and a constant $c > 0$, write cg for the measured metric space $(v(g), \mathrm{dist}, \pi)$ whose points are elements of $v(g)$, with $\mathrm{dist}(x, y)$ equalling c times the graph distance between vertices x and y, and with π the uniform measure on $v(g)$. The following is a simplified and slightly informal statement of the main result from Chapter 3.

Theorem (Theorem 3.1.1)**.** *For each $n \geq 1$ let $\mathrm{d}^n = (d^n(i), 1 \leq i \leq n)$ be a degree sequence and let (G_n, T_n, Γ_n) be a random simple tree-weighted graph with degree sequence d^n. Then, under certain convergence conditions on the sequence of degree sequences $(\mathrm{d}^n, n \geq 1)$, there exists a constant σ such that*

$$\frac{\sigma}{n^{1/2}} T_n \xrightarrow{d} \mathcal{T}$$

as $n \to \infty$ with respect to the Gromov-Hausdorff-Prokhorov topology, where $\mathcal{T}$ is the Brownian continuum random tree.

Some insight for the necessary conditions on $(\mathrm{d}^n, n \geq 1)$ is given throughout the remainder of this section.

The Brownian continuum random tree, sometimes referred to as the Aldous' continuum random tree or simply as the continuum random tree, was presented by David Aldous in [16] and is described, in his words, "as the set-representation of some (Platonic) continuum random tree" [16, page 10]. Notably, Aldous shows that if $\mathcal{T}_n$ is chosen uniformly at random from the set of n^{n-2} trees on $[n]$, then by letting 1 be the root of $\mathcal{T}_n$, a $\Theta(n^{-1/2})$ rescaling of $\mathcal{T}_n$ converges in distribution, with respect to the Gromov-Hausdorff-Prokhorov topology, to the continuum random tree. (In fact, Aldous proved this result in a different topology; it was proved for convergence relative to the Gromov-Hausdorff-Prokhorov topology by Jean-François Le Gall in [50].)

The Gromov-Hausdorff-Prokhorov topology is discussed thoroughly by Abraham, Delmas and Hoscheit in [1]. In summary, a sequence $((X_n, d_n, \mu_n), n \geq 1)$ of (deterministic) measured metric spaces converges to a limit $(X_\infty, d_\infty, \mu_\infty)$ in the Gromov-Hausdorff-Prokhorov sense if there exists a metric space (Z, D) and a sequence of maps $(f_n : X_n \to Z, 1 \leq n \leq \infty)$ with $Y_n := \left(f_n(X_n), D\big|_{f_n(X_n)}\right)$ isometric to (X_n, d_n) for all n, such that $\mathrm{dist}_{\mathrm{Haus}}(Y_n, Y_\infty) \to 0$ as $n \to \infty$, and such that the push-forwards $\nu_n := f_n^*(\mu_n)$, defined by $f_n^*(\mu_n)(A) = \mu_n(f_n^{-1}(A))$ for all measurable $A \subseteq Z$, satisfy $\mathrm{dist}_{\mathrm{Prok}}(\nu_n, \nu_\infty) \to 0$ as $n \to \infty$, where $\mathrm{dist}_{\mathrm{Haus}}$ and $\mathrm{dist}_{\mathrm{Prok}}$ are the Hausdorff distance and Prokhorov distance respectively. A sequence of *random* measured metric spaces $(M_n, n \geq 1)$ converges in distribution to some limit M_∞ in the Gromov-Hausdorff-Prokhorov sense if there is a coupling $(M_n', 1 \leq n \leq \infty)$ of $(M_n, 1 \leq n \leq \infty)$ such that $\mathrm{dist}_{\mathrm{GHP}}(M_n'(\omega), M_\infty'(\omega)) \to 0$ for almost every $\omega \in \Omega$, as $n \to \infty$.

To prove Theorem 3.1.1, we apply a result from Broutin and Marckert [33]. For a rooted tree t on vertex set $[n]$, let $c_t(u)$ be the number of children of vertex u in t. Then $\mathrm{c}_t = (c_t(i), 1 \leq i \leq n)$ is called the *child sequence of* t. Broutin and Marckert's result states (roughly) that if $(\mathrm{c}^n, n \geq 1)$ is a sequence of child sequences and, for all $n \geq 1$, T_n is chosen uniformly at random from the set of rooted trees with n vertices and child sequence c^n, then under certain convergence conditions on $(\mathrm{c}^n, n \geq 1)$, there exists a constant σ such that $\sigma n^{-1/2} T_n \xrightarrow{d} \mathcal{T}$ as $n \to \infty$, in the Gromov-Hausdorff-Prokhorov sense, where $\mathcal{T}$ is the Brownian continuum random tree. Thanks to Broutin and Marckert's result, our main result now boils down to (a) showing that the spanning trees in random simple tree-weighted graphs are uniformly random trees conditioned on their child sequences, and (b) determining the conditions under which the child sequences of these spanning trees satisfy the conditions needed to apply the result of Broutin and Marckert; the convergence conditions on $(\mathrm{d}^n, n \geq 1)$ stated in Theorem 3.1.1 exist, in part, to help prove (b).

To study the child sequences of tree-weighted graphs, we first need a way to construct these graphs. Our construction procedure is inspired by two well known processes: Pitman's additive coalescent [64] and the configuration model. Starting with Pitman's additive coalescent, this is a Markov process starting from a set of masses $(x_i, i \geq 1)$ with $\sum_{i \geq 1} x_i = 1$ where we repeatedly merge two masses x_i and x_j into a single mass $x_i + x_j$ at rate $x_i + x_j$. We use a new version of Pitman's additive coalescent as the first part of our construction process, as follows:

1. Given a degree sequence $\mathrm{d} = (d_1, \ldots, d_n)$, start with a graph with vertex set $[n]$ and no edges. Then, for each $i \in [n]$, endow vertex i with d_i half-edges, one of which is distinguished as the root half-edge.

2. Repeatedly until the graph is connected, choose an unpaired root half-edge and an unpaired non-root half-edge, ensuring that the chosen half-edges do not belong to the same tree but otherwise uniformly at random, and pair them to form an edge.

Figure 3.1 gives an example of this pairing process with the degree sequence $(4, 4, 3, 1)$. Assuming that $d_i \geq 1$ for $1 \leq i \leq n$ and that $\sum_{i \in [n]} d_i \geq 2(n-1)$, This process always yields a tree, and each vertex of the tree is endowed with a random number of remaining unpaired half-edges; there is exactly one vertex for which one of the unpaired half-edges is a root half-edge. This brings us to the second part of the construction process, which is a configuration model on the remaining half-edges. Repeatedly, until there are no unpaired half-edges, choose a pair of unpaired half-edges uniformly at random and pair them to create an edge. Together, these two parts yield a graph G with degree sequence d, a spanning tree T given by the first part of the construction, and a distinguished oriented edge $\Gamma = UV$ given by the unpaired root half-edge that becomes paired during the second part of the construction; we consider U to be the vertex containing the root half-edge. We call a triple (G, T, Γ) constructed in this way a *random tree-weighted graph with degree sequence* d, and we show that if G is a simple graph then (G, T, Γ) is a random simple tree-weighted graph with degree sequence d, justifying the similar name.

The majority of Chapter 3 is dedicated to proving two results. Let $(\mathrm{d}^n, n \geq 1)$ be a sequence of degree sequences that satisfies the convergence conditions in Theorem 3.1.1 and, for each $n \geq 1$, let (G_n, T_n, Γ_n) be a random tree-weighted graph with degree sequence d^n. Firstly, we prove that the child sequences $(\mathrm{c}^n, n \geq 1)$ corresponding to $(T_n, n \geq 1)$ satisfy the conditions required for Broutin and Marckert's result with high probability. Secondly, we prove, under our conditions on the degree sequences, that the probability of G_n being a simple graph asymptotically approaches a strictly positive constant. This second result tells us that conditioning on G_n being simple has an asymptotically negligible effect on the law of T_n. Thus, similar to how results on the configuration model can be used to prove results for uniform samples from $\mathcal{G}_{\mathrm{d}^{\mathrm{n}}}$, we can use the results on random tree-weighted graphs to get results for random simple tree-weighted graphs.

1.3 Minimum-weight spanning trees

Given a graph g, a weight function $w : e(g) \to [0, \infty)$, and a subgraph $h \subseteq g$, write $w(h) = \sum_{e \in e(h)} w(e)$ for the *weight of* h. Then a *minimum-weight spanning tree of* (g, w) is a spanning tree $t \subseteq g$ such that $w(t) \leq w(t')$ for all spanning trees $t' \subseteq g$. Minimum-weight spanning trees have been the focus of many works in the fields of graph theory, combinatorics and optimization. Dating back to 1926, Otakar Borůvka presented the first algorithm for finding minimum weight spanning trees [31]. Although this was a ground-breaking result for the field of combinatorial optimization, Borůvka's algorithm is somewhat involved, and simpler algorithms have since been discovered.

The two most famous algorithms for finding minimum weight spanning trees are Kruskal's algorithm [49] and Prim's algorithm [65]. Let (g, w) be a weighted connected graph with n vertices and m edges and let $(e_1, \ldots, e_m)$ be an ordering of $e(g)$ such that, for all $1 \leq i < m$, $w(e_i) \leq w(e_{i+1})$. Then Kruskal's algorithm works as follows.

1. Let $h_0 \subset g$ be the spanning subgraph of g with no edges.

2. For $0 \leq i < m$, construct h_{i+1} from h_i as follows. If $e(h_i) \cup \{e_{i+1}\}$ contains a cycle, then set $h_{i+1} = h_i$. Otherwise, set $v(h_{i+1}) = v(h_i)$ and set $e(h_{i+1}) = e(h_i) \cup \{e_{i+1}\}$.

3. Return h_m.

Likewise, Prim's algorithm (started from a given vertex $v \in v(g)$) works as follows.

1. Let $h_0 \subseteq g$ be the subgraph of g consisting of a single vertex v and no edges.

2. Given h_i for $0 \leq i < n-1$, let e be the lowest-weight edge with one endpoint in $v(h_i)$ and the other endpoint in $v(g) \setminus v(h_i)$. Let u be the endpoint of e in $v(g) \setminus v(h_i)$ and construct h_{i+1} from h_i by setting $v(h_{i+1}) = v(h_i) \cup \{u\}$ and setting $e(h_{i+1}) = e(h_i) \cup \{e\}$.

3. Return h_{n-1}.

In both cases, the returned graph is indeed the minimum weight spanning tree of (g, w). Kruskal's algorithm and Prim's algorithm are excellent algorithms for finding minimum weight spanning trees due to their simplicity. However, one major draw back to both algorithms is that at every step we must compare the weights of a potentially large number of edges. In particular, Kruskal's algorithm requires us to compare the weights of every edge in the graph during the first step. A motivating question for chapter 4 is then: can we find minimum weight spanning trees without requiring global information at every step?

In Chapter 4, we take a different approach to the idea of using simple graphs as building blocks for more complex graphs. Specifically, we take some initial spanning subgraph and morph it into a minimum-weight spanning tree by performing a series of "local" changes to the initial subgraph. Let g be a graph, let $w : e(g) \to [0, \infty)$ be injective, and let h be a connected spanning subgraph of g (we want w to be injective as this ensures that there is only one minimum weight spanning tree). For a set $S \subseteq v(g)$, write $h[S]$ for the graph with vertex set S and edge set $\{uv \in e(h) : u, v \in S\}$. Next, define an *optimization function* $\Phi_g(h, S)$ as follows. If $h[S]$ is not connected, then set $\Phi_g(h, S) = h$. Otherwise, construct $\Phi_g(h, S)$ from h by removing the edge set $e(h[S])$ and replacing it with the edge set of the minimum weight spanning tree of $g[S]$. Now given a connected spanning subgraph h and a sequence of vertex sets $\mathbb{S} = (S_i, 1 \leq i \leq m)$, define a corresponding sequence of spanning subgraphs $(h_0, \ldots, h_m)$ by setting $h_0 = h$ and inductively setting $h_{i+1} = \Phi_g(h_i, S_{i+1})$ for $0 \leq i < m$. If h_m is the minimum weight spanning tree of g, then we call $\mathbb{S}$ an MST *sequence for the pair* $((g, w), h)$. Finally, define the *weight of the sequence* $\mathbb{S}$, written $\mathrm{wt}(\mathbb{S})$, as

$$\mathrm{wt}(\mathbb{S}) = \max_{0 \leq i < m} \left\{ w(h_i[S_{i+1}]) \right\}.$$

We are interested in finding MST sequences with low weight. Without such a constraint, finding MST sequences is trivial: given the pair $((g, w), h)$ we could simply find the minimum-weight spanning tree of (g, w) by choosing the MST sequence $\mathbb{S} = (v(g))$. However, in this case $\mathrm{wt}(\mathbb{S}) = w(h)$ is likely very large, and computing our optimization function in this case requires the same amount of information as Kruskal's algorithm performed on the same initialization (g, w). On the other hand, if we can find an MST sequence $\mathbb{S}$ with $\mathrm{wt}(\mathbb{S})$ much smaller than $w(h)$ then our optimization function will only ever modify low-weight subgraphs and, if $\mathrm{wt}(\mathbb{S})$ is very small, we will be able to obtain the minimum weight spanning tree without requiring "global" information at any step.

Recall that K_n is the complete graph on vertex set $[n]$ and $\mathbb{X} : e(K_n) \to [0, 1]$ is the random weight function with $\mathbb{X}(e) \sim \mathrm{Uniform}[0, 1]$, independently for all $e \in e(K_n)$. The main result of Chapter 4 is a slightly modified version of the following theorem.

Theorem (Theorem 4.1.1)**.** *Fix a sequence $(H_n, n \geq 1)$ of connected graphs with H_n being a spanning subgraph of K_n. Then for any $\epsilon > 0$, as $n \to \infty$,*

1. *with high probability there exists an* MST *sequence $\mathbb{S}$ for $((K_n, \mathbb{X}), H_n)$ with* $\mathrm{wt}(\mathbb{S}) \leq 1 + \epsilon$*, and*

2. *there exists $\delta > 0$ such that with high probability, given any sequence $\mathbb{S} = (S_1, \ldots, S_m)$ for $((K_n, \mathbb{X}), H_n)$ with* $\mathrm{wt}(\mathbb{S}) < 1 - \epsilon$*, the final spanning subgraph $H_{n,m}$ has weight $w(H_{n,m}) \geq \delta n w(T)$ where T is the minimum weight spanning tree of $(K_n, \mathbb{X})$.*

The first statement in this result tells us that we can start from a connected spanning subgraph H_n of K_n and apply a series of optimizations to H_n and eventually obtain the minimum weight spanning tree of $(K_n, \mathbb{X})$. Moreover, assuming H_n was not chosen by a malicious adversary, then with high probability we need only change pieces of H_n with weight at most $1 + \epsilon$.

Proving the second statement in Theorem 4.1.1 is easy since H_n almost surely contains linearly many edges with weight greater than $1 - \epsilon$. Proving the first statement is much harder and is the bulk of Chapter 4. We show the existence of an MST sequence in two steps. In step one, we describe a sequence that can transform a spanning subgraph H containing a minimum-weight spanning tree on k vertices into a subgraph H' containing a minimum-weight spanning tree on $k + 1$ vertices. We call this transformation *the eating algorithm*, and it can be applied iteratively to yield the minimum-weight spanning tree of $(K_n, \mathbb{X})$. However, we require an initial value of k that is quite large in order to ensure that the largest weight of a subgraph considered by the eating algorithm is small with high probability. Thus, in step two we use an easy Ramsey-type argument to show that every graph contains a large subgraph that is either a path, a star, or a complete graph. We then give a detailed description of how to construct a low-weight MST sequence starting from each of these three initial subgraphs. Thus, we have a way to transform a large subgraph of H_n into a minimum weight spanning tree, after which we can apply the eating algorithm iteratively until we output the minimum-weight spanning tree of $(K_n, \mathbb{X})$.

I would like to note that our methods are not entirely constructive, and we do not claim to out-perform Kruskal's algorithm with respect to time complexity.

Chapter 2

Multi-Source Invasion Percolation on the Complete Graph

2.1 Introduction

Fix a locally finite weighted graph $G = (v(G), e(G), \mathrm{w})$ such that $\mathrm{w} : e(G) \to (0, \infty)$ is injective. The *invasion percolation* process on G works as follows.

- Fix a finite starting set $\mathcal{S} \subseteq v(G)$, and let $\mathcal{S}_0 = \mathcal{S}$.
- For $1 \leq i < |v(G)| + 1 - |\mathcal{S}|$, let $e_i \in E$ be the smallest-weight edge from $\mathcal{S}_{i-1}$ to the rest of the graph. That is, $e_i = uv$ minimizes

$$\{\mathrm{w}_e : e = uv, u \in \mathcal{S}_{i-1}, v \notin \mathcal{S}_{i-1}\}.$$

- Let $v_i = v$, and set $\mathcal{S}_i = \mathcal{S}_{i-1} \cup \{v_i\}$.

Write $F(G, \mathcal{S}) = (v(F(G, \mathcal{S})), e(F(G, \mathcal{S})))$ for the subgraph of G with vertex set $\mathcal{S} \cup \{v_i, 0 \leq i < |v(G)| + 1 - |\mathcal{S}|\}$ and edge set $\{e_i, 1 \leq i < |v(G) + 1 - |\mathcal{S}|\}$. Since each edge added by invasion percolation connects to a vertex not incident to any previous edge, the result $F(G, \mathcal{S})$ of the invasion percolation is a forest with $|\mathcal{S}|$ connected components, in which each of the elements of $\mathcal{S}$ lies in a distinct connected component of $F_{\mathcal{S}}$.

Invasion percolation was introduced in [34], and independently (with a slightly different formulation, using vertex rather than edge weights) in [70]. The latter paper, which coined the term "invasion percolation", considered the process on 2- and 3-dimensional lattice rectangles, with the starting set given by the vertices of one boundary side (or boundary face), and with independent random Uniform$[0,1]$ weights.

The behaviour of invasion percolation with random weights is known to be closely linked to that of *critical percolation* on the corresponding graph, and indeed, invasion percolation is one of the simplest examples of self-organized criticality in random systems [66].

Invasion percolation has been extensively studied in the probability and statistical physics communities: on lattices [36, 42, 62, 67, 71], on trees [13, 21, 22, 57, 63], and in the mean-field or general graph setting [2, 19, 42, 55, 61]. However, past work has almost exclusively focussed on invasion percolation run from a single starting vertex.

The purpose of this paper is to study mean-field invasion percolation run from starting sets of variable sizes. We establish a phase transition in the structure of the resulting forest, depending on the size of the starting set. Write $\mathrm{K}_n = ([n], \binom{[n]}{2}, \mathrm{U})$ for the randomly-weighted complete graph, with vertex set $[n] := \{1, \ldots, n\}$, edge set $\binom{[n]}{2} := \{e \subset [n] : |e| = 2\}$, and independent Uniform$[0,1]$ edge weights $\mathrm{U} = \{U(e), e \in \binom{[n]}{2}\}$.

Theorem 2.1.1. *Fix positive integers $(k(n), n \geq 1)$, and for $n \geq 1$ let M_n be the size of the largest connected component of $F(\mathrm{K}_n, [k(n)])$.*

- *If $k(n)/n^{1/3} \to 0$ then $M_n/n \to 1$ in probability.*
- *If $k(n)/n^{1/3} \to \infty$ then $M_n/n \to 0$ in probability.*

Remarks.

$\star$ By the symmetries of the model, the starting set $[k(n)]$ could be replaced by any other set $\mathcal{S}(n)$ of size $k(n)$ and the same result would hold.

$\star$ The first assertion of the theorem, that if $k(n)/n^{1/3} \to 0$ then $M_n/n \to 1$ in probability, was proved in [53]. That work also proved that $M_n/n \to 0$ in probability provided that $k(n)/(n^{1/3}(\log n)^{4/3}(\log\log n)^{1/3}) \to \infty$, and provided more quantitative upper bounds on M_n for such values of $k(n)$. Thus, the main contribution of this work is to pin down the location of the phase transition in the behaviour of M_n to $k(n)$ of order precisely $n^{1/3}$.

$\star$ We conjecture that if $k(n)/n^{1/3} \to c \in \mathbb{R}$ then M_n/n converges in distribution to a non-degenerate limit $M_\infty(c)$. More strongly, we make the following conjecture. Write $L_{n,i}$ for the size of the i'th largest connected component of $F(\mathrm{K}_n, [k(n)])$, with $L_{n,i} = 0$ if $F(\mathrm{K}_n, [k(n)])$ has fewer than i connected components. For each $c \in \mathbb{R}$ there exists a random vector $(L_{\infty,i}(c), i \geq 1)$ taking values in the set $\Delta_\infty^{\downarrow} = \{(\ell_i, i \geq 1) \in (0,1)^{\mathbb{N}} : \sum_{i\geq 1} \ell_i = 1\}$, such that if $k(n)/n^{1/3} \to c$ then $(L_{n,i}/n, i \geq 1) \xrightarrow{\mathrm{d}} (L_{\infty,i}(c))$ in the sense of finite-dimensional distributions.

2.1.1 Overview of the rest of the paper

In Section 2.2, we explain several useful connections between invasion percolation, critical percolation, and minimum spanning trees. We then use these connections to prove Theorem 2.1.1, modulo a key input to the proof. This key input, Proposition 2.2.1, roughly states the following. In the case that $k(n)/n^{1/3} \to \infty$, if we run the multi-source invasion percolation process for $n/2 + O(n^{2/3})$ steps, then the size of the largest tree is with high probability much smaller than the size of the largest component in an Erdős–Rényi random graph process run for the same number of steps (which precisely builds a critical Erdős–Rényi random graph).

The proof of Proposition 2.2.1, which occupies the bulk of the paper, appears in Section 2.3. It makes use of the connections between invasion percolation and critical percolation, and the fact that the components of the critical Erdős–Rényi random graph are with high probability treelike, to reduce the analysis to that of a fragmentation process on large random binary trees.

Finally, Section 2.4 proposes some future research directions suggested by the current work.

2.2 A sketch proof of Theorem 2.1.1.

2.2.1 Invasion percolation, Prim's algorithm, and Kruskal's algorithm

Suppose that $G = (v(G), e(G), \mathrm{w})$ is a finite graph. If $\mathcal{S} = \{v\}$ consists of a single vertex $v \in v(G)$, then invasion percolation is equivalent to *Prim's algorithm* [65] started from v, and $F(G, \mathcal{S})$ is thus the minimum-weight spanning tree (MST) of the weighted graph G. If $\mathcal{S}$ consists of more than one vertex, the invasion percolation process can still be viewed as a form of Prim's algorithm, as follows. Augment G by adding a new vertex ρ and edges from ρ to all elements of $\mathcal{S}$. Fix $0 < \epsilon < \min(\mathrm{w}_e, e \in e(G))$ and augment w by giving the edges $\{\rho x, x \in \mathcal{S}\}$ each a distinct weight less than ϵ. Write $G'_{\mathcal{S}} = (v(G'_{\mathcal{S}}), e(G'_{\mathcal{S}}), \mathrm{w}')$ for the augmented graph. Then invasion percolation on $G'_{\mathcal{S}}$ with starting set $\{\rho\}$ will first add edges $\{\rho x, x \in \mathcal{S}\}$, and will then add the same edges as invasion percolation on G with starting set $\mathcal{S}$, in the same order. It follows that the subgraph of $F(G'_{\mathcal{S}}, \rho)$ obtained by removing ρ and its incident edges is precisely $F(G, \mathcal{S})$.

In the setting of finite graphs, an alternative construction of $F(G, \mathcal{S})$ is given by *Kruskal's* *algorithm* [49], which works as follows. Write $m = |e(G)|$ and list the edges of G in increasing order of weight as $e(1), \ldots, e(m)$. Let $F_0 = F_0^{G,\mathcal{S}} = (v(G), \emptyset)$. Then, for $1 \le i \le m$:

- If $e(i) = u(i)v(i)$ joins distinct connected components of F_{i-1}, and $u(i)$ and $v(i)$ do not both lie in components containing elements of $\mathcal{S}$, then set $F_i = F_{i-1} + e(i) := (v(F_{i-1}), e(F_{i-1}) \cup \{e(i)\})$.

- Otherwise, set $F_i = F_{i-1}$.

The output of Kruskal's algorithm is the forest $F_m = F_m^{G,\mathcal{S}}$. To see that $F_m = F(G, \mathcal{S})$, it suffices to consider running Kruskal's algorithm on the augmented graph $G'_{\mathcal{S}}$ defined above. The result is the MST of $G'_{\mathcal{S}}$, and is therefore equal to $F(G'_{\mathcal{S}}, \{\rho\})$. However, Kruskal's algorithm run on $G'_{\mathcal{S}}$ and $\{\rho\}$ will begin by adding the edges ρx for $x \in \mathcal{S}$, since these edges have lower weight than all other edges in $G'_{\mathcal{S}}$. Once these edges are added, the vertices of $\mathcal{S}$ all lie in a single connected component, so the remaining steps of Kruskal's algorithm run on $G'_{\mathcal{S}}$ and $\{\rho\}$ add the same edges as Kruskal's algorithm run on G and $\mathcal{S}$, in the same order. It follows that F_m can be obtained from $F(G'_{\mathcal{S}}, \{\rho\})$ by removing ρ and its incident edges. We saw using Prim's algorithm that performing this operation to $F(G'_{\mathcal{S}}, \{\rho\})$ yields $F(G, \mathcal{S})$, and so indeed $F_m^{G,\mathcal{S}} = F(G, \mathcal{S})$.

It will be useful that the above construction couples the processes $(F_i^{G,\mathcal{S}}, 0 \le i \le m)$ for different starting sets $\mathcal{S}$: if $\mathcal{S}' \subset \mathcal{S}$ then $F_i^{G,\mathcal{S}}$ is a subgraph of $F_i^{G,\mathcal{S}'}$ for all $0 \le i \le m$. More specifically, suppose that $\mathcal{S} = \mathcal{S}' \cup \{z\}$ for some fixed $z \in v(G) \setminus \mathcal{S}'$. Let $v \in \mathcal{S}'$ be the unique element of $\mathcal{S}'$ in the same component of $F_m^{G,\mathcal{S}'}$ as z, and let $e(j)$ be the largest-weight edge on the path from v to z in $F_m^{G,\mathcal{S}'}$. Then

$$F_i^{G,\mathcal{S}} = \begin{cases} F_i^{G,\mathcal{S}'} & \text{if } i < j \\ F_i^{G,\mathcal{S}'} - e(j) & \text{if } i \ge j\,. \end{cases} \tag{2.1}$$

2.2.2 Kruskal's algorithm and the Erdős–Rényi process

There is a second useful coupling, between $(F_i^{G,\mathcal{S}}, 0 \le i \le m)$ and a graph process which does not forbid cycles but maintains the condition that vertices in the starting set $\mathcal{S}$ are not allowed to join the same connected component. The restricted process, which we call the *Erdős–Rényi process* and denote $(G_i, 0 \le i \le m) = (G_i^{\mathcal{S}}, 0 \le i \le m)$, works as follows. List the edges of G in increasing order of edge weight as $(e(i), 0 \le i \le m)$. For $1 \le i \le m$, if the edge $e(i)$ joins connected components of G_{i-1} containing distinct elements of $\mathcal{S}$ then set

$G_i = G_{i-1} = ([n], e(G_{i-1}))$; otherwise, set $G_i = G_{i-1} + e(i)$. The final graph G_m consists of $|\mathcal{S}|$ connected components, each containing exactly one of the vertices of $\mathcal{S}$.

The orderings of edges in Kruskal's algorithm and in the Erdős–Rényi process are identical. Moreover, if the same starting set $\mathcal{S}$ is used for both processes, then the only edges which are added by the Erdős–Rényi process but not by Kruskal's algorithm join vertices which already lie in the same connected component. It follows that $F_i^{G,\mathcal{S}}$ and $G_i^{\mathcal{S}}$ have the same connected components for all $1 \le i \le m$. (More strongly, for each connected component C of $G_i^{\mathcal{S}}$, the corresponding component of $F_i^{G,\mathcal{S}}$ is the minimum weight spanning tree of C.) In particular, this yields that the size of the largest connected component is the same in $F(G,\mathcal{S})$ and in $G_m^{\mathcal{S}}$.

To justify the name "Erdős–Rényi process", note that if $G = \mathrm{K}_n$ is the randomly-weighted complete graph and $|\mathcal{S}| = 1$, then $(G_i^{\mathcal{S}}, 0 \le i \le m) = (G_i^{\mathcal{S}}, 0 \le i \le \binom{n}{2})$ is precisely the classical Erdős–Rényi random graph process, in which the edges of the complete graph are added one-at-a-time in exchangeable random order.

2.2.3 The critical random graph and the proof of Theorem 2.1.1

We now specialize to the setting of this paper, the randomly-weighted complete graph K_n. It is useful to continuize both the Erdős–Rényi process and Kruskal's algorithm; write $(G(n,\mathcal{S},p), 0 \le p \le 1)$ for the random graph process in which $G(n,\mathcal{S},p)$ has vertex set $[n]$ and edge set

$$\left\{e \in e\big(\mathrm{K}^{\mathcal{S}}_{n,\binom{n}{2}}\big) : U_e \le p\right\},$$

and $(F(n,\mathcal{S},p), 0 \le p \le 1)$ for the process in which $F(n,\mathcal{S},p)$ has vertex set $[n]$ and edge set

$$\left\{e \in e\big(F^{\mathrm{K}_n,\mathcal{S}}_{\binom{n}{2}}\big) : U_e \le p\right\}.$$

The continuous-time processes add the same edges as the discrete processes, and in the same order. More strongly, $G(n,\mathcal{S},U_{e(i)}) = \mathrm{K}^{\mathcal{S}}_{n,i}$ for all $1 \le i \le \binom{n}{2}$, and $(G(n,\mathcal{S},p), 0 \le p \le 1)$ is constant except at times $(U_i, 1 \le i \le \binom{n}{2})$; the corresponding relation holds for the discrete- and continuous-time Kruskal processes.

When $|\mathcal{S}| = 1$ we omit $\mathcal{S}$ from the notation, writing, e.g., $G(n,p)$ rather than $G(n,\mathcal{S},p)$, as in this case the processes do not in fact depend on $\mathcal{S}$. Note that $F(n,1)$ is then the MST of K_n.

The relation (2.1) implies that for any $p \in (0,1)$ and $\mathcal{S} \subset [n]$, the connected components of $F(n,\mathcal{S},p)$ refine those of $F(n,p)$, in that for any component C of $F(n,p)$ the vertex set of

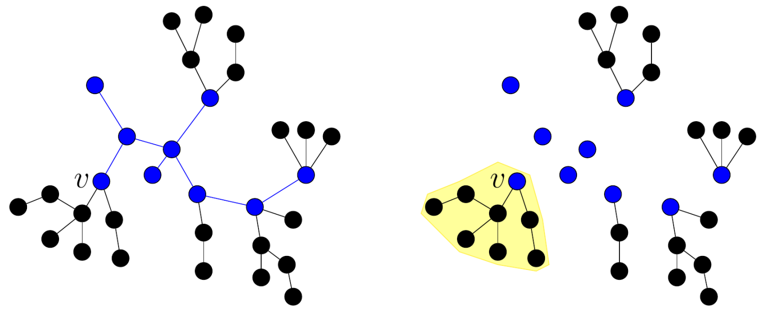

Figure 2.1: Left: An instantiation of $F(n,1)$ with $F^1(n,p_{n,\lambda})$ drawn in blue. Right: the forest obtained from $F(n,1)$ by removing the edges of $F^1(n,p_{n,\lambda})$. On the right the tree $T^n_{v,\lambda}$ is highlighted.

C may be written as a union of the vertex sets of components of $F(n,\mathcal{S},p)$. Since $F(n,\mathcal{S},p)$ and $G(n,\mathcal{S},p)$ have the same components for all $\mathcal{S} \subset [n]$ and $p \in [0,1]$, the same fact holds for $G(n,\mathcal{S},p)$ and $G(n,p)$.

The heart of the proof that $M_n/n \to 0$ in probability when $k(n)/n^{1/3} \to \infty$ consists in establishing that in the *critical window* of the Erdős-Rényi process, when $p = 1/n + O(1/n^{4/3})$, the connected components of $F(n,[k(n)],p)$ all have size $o(n^{2/3})$ with high probability. For $\lambda \in \mathbb{R}$, write

$$p_{n,\lambda} = 1/n + \lambda/n^{4/3}\,.$$

Proposition 2.2.1. *Fix positive integers $(k(n), n \geq 1)$ with $k(n) \in [n]$ and $k(n)/n^{1/3} \to \infty$. Next, fix $\lambda \in \mathbb{R}$, and let $M_{n,\lambda}(k(n))$ be the size of the largest connected component of $F(n,[k(n)],p_{n,\lambda})$. Then $M_{n,\lambda}(k(n))/n^{2/3} \to 0$ in probability.*

The proof of Proposition 2.2.1 appears in Section 2.3. To prove Theorem 2.1.1, we combine this proposition with the following two pre-existing results about the structure of the minimum spanning tree of K_n. For $p \in [0,1]$, write $F^1(n,p)$ for the largest connected component of $F(n,p)$, with ties broken uniformly at random. Fix $\lambda \in \mathbb{R}$, and consider the forest obtained from the minimum spanning tree, $F(n,1)$, by removing the edges of $F^1(n,p_{n,\lambda})$. For each vertex v of $F^1(n,p_{n,\lambda})$, write $T^n_{v,\lambda}$ for the tree of this forest containing v; see Figure 2.1. Let $q^v_{n,\lambda} = |T^n_{v,\lambda}|/n$ be the proportion of vertices of $F(n,1)$ lying in $T^n_{v,\lambda}$.

Proposition 2.2.2 ([10], Lemma 4.11)**.** *Write $\Delta^n_\lambda = \max(q^v_{n,\lambda}, v \in F^1(n,p_{n,\lambda}))$. Then for all $\delta > 0$,*

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \mathbf{P}\left\{\Delta^n_\lambda > \delta\right\} = 0\,.$$

Next, let

$$\mathcal{F}_{n,\lambda} := \sigma\big(U(e)\mathbf{1}_{[U(e)\le p_{n,\lambda}]}, e \in e(K_n)\big) = \sigma\big(U(e)\mathbf{1}_{[e\in e(G(n,p_{n,\lambda}))]}, e \in e(K_n)\big)$$

be the σ-algebra containing all information about the weights of edges in $G(n, p_{n,\lambda})$.

Proposition 2.2.3 ([14], Lemma 6.19)**.** *For every $\lambda \in \mathbb{R}$, conditionally given $\mathcal{F}_{n,\lambda}$, the collection of random variables $(q^v_{n,\lambda}, v \in F^1(n, p_{n,\lambda}))$ is exchangeable.*

The exchangeability in [14, Lemma 6.19] is stated conditionally given $G(n, p_{n,\lambda})$, rather than given $\mathcal{F}_{n,\lambda}$. In other words, in [14] the conditioning is only on the graph structure of $G(n, p_{n,\lambda})$, but not on the weights of its edges. However, an essentially identical proof to that given in [14] establishes the slightly stronger statement above.

We also require a fact about concentration of exchangeable random sums, which is a consequence of a result of Aldous [18].

Proposition 2.2.4 ([18], Theorem 20.7)**.** *For all $\epsilon > 0$ there exists $\delta > 0$ such that the following holds. Let $(q_i, 1 \le i \le m)$ be non-negative real numbers with $\sum_{1\le i\le m} q_i = 1$, and let $(\pi(i), 1 \le i \le m)$ be a uniformly random permutation of $[m]$. If $\max_{1\le i\le m} q_i \le \delta$ then*

$$\mathbf{P}\left\{\max_{1\le i\le m}\left|\sum_{j=1}^{i}\left(q_{\pi(j)} - \frac{1}{m}\right)\right| > \epsilon\right\} < \epsilon\,.$$

See [28, Lemma 7.5] and [27, Lemma 4.9] for quantitative versions of this result. Proposition 2.2.4 is the last fact we need for the proof of our main result.

Proof of Theorem 2.1.1. As noted just after the statement of Theorem 2.1.1, the fact that $M_n/n \to 1$ in probability when $k(n)/n^{1/3} \to 0$ was proved in [53], so we need only handle the other assertion of the theorem. For the remainder of the proof we therefore assume that $k(n)/n^{1/3} \to \infty$.

A result of Łuczak [52, Theorem 3 ii.] implies that if $p = p_n$ satisfies that $p_n = (1+o(1))/n$ and $n^{4/3}(p_n - 1/n) \to \infty$, then the largest component of $G(n, p_n)$ has size $(2+o(1))(n^2p_n - n)$ in probability. Since the components of $G(n, p_n)$ and of $F(n, p_n)$ are identical, recalling that $p_{n,\lambda} = 1/n + \lambda/n^{4/3}$, it follows that if $\lambda = \lambda(n) \to \infty$ with $\lambda(n) = o(n^{1/3})$, then $|F^1(n, p_{n,\lambda(n)})|/(n^{2/3}\lambda(n)) \to 2$ in probability. By a subsubsequence argument, this implies that for all $\delta > 0$,

$$\lim_{\lambda\to\infty}\limsup_{n\to\infty}\mathbf{P}\left\{|F^1(n, p_{n,\lambda})|/(n^{2/3}\lambda(n)) < 2 - \delta\right\} = 0. \tag{2.2}$$

Fix $\epsilon \in (0,1)$. Then fix $\delta \in (0,\epsilon/2)$ small enough that Proposition 2.2.4 holds for this ϵ and δ, then let λ be large enough that for all n sufficiently large,

$$\mathbf{P}\left\{|F^1(n,p_{n,\lambda})|/n^{2/3} \leq 1\right\} < \epsilon \tag{2.3}$$

and

$$\mathbf{P}\left\{\Delta_\lambda^n \geq \delta\right\} < \epsilon\,. \tag{2.4}$$

This is possible by (2.2) and by Proposition 2.2.2. Since λ is fixed, by Proposition 2.2.1, we also have that

$$\mathbf{P}\left\{M_{n,\lambda}(k(n))/n^{2/3} \geq \delta\right\} < \epsilon \tag{2.5}$$

for n sufficiently large.

List the connected components of $F(n,[k(n)],p_{n,\lambda})$ contained in $F^1(n,p_{n,\lambda})$ as $C_{n,\lambda}^1,\ldots,C_{n,\lambda}^K$; here K is a random variable. These components are subtrees of $F^1(n,p_{n,\lambda})$, and their vertex sets partition $v(F^1(n,p_{n,\lambda}))$. Note that, writing $\mathcal{S}_{n,\lambda}=[k(n)]\cap v(F^1(n,p_{n,\lambda}))$, then each of $C_{n,\lambda}^1,\ldots,C_{n,\lambda}^K$ contains exactly one vertex of $\mathcal{S}_{n,\lambda}$.

We now consider the restricted process $(F(n,\mathcal{S}_{n,\lambda},p), 0 \leq p \leq 1)$. Due to the relation (2.1), the only edges added in the Kruskal process $(F(n,p), 0 \leq p \leq 1)$ which are not added in the restricted process $(F(n,\mathcal{S}_{n,\lambda},p), 0 \leq p \leq 1)$ are the edges of $F^1(n,p_{n,\lambda})$ which join distinct components $C_{n,\lambda}^1,\ldots,C_{n,\lambda}^K$, and these edges are already present in $F(n,p_{n,\lambda})$. It follows that for each $1 \leq i \leq K$, the connected component of $F(n,\mathcal{S}_{n,\lambda},1)$ containing $C_{n,\lambda}^i$ is precisely the union of the trees $\{T_{v,\lambda}^n, v \in v(C_{n,\lambda}^i)\}$. On the other hand, since $\mathcal{S}_{n,\lambda} \subset [k(n)]$, the components of $F(n,[k(n)],1)$ partition the components of $F(n,\mathcal{S}_{n,\lambda},1)$, and so

$$\begin{aligned} M_n &= \max(|C| : C \text{ is a component of } F(n,[k(n)],1)) \\ &\leq \max(|C| : C \text{ is a component of } F(n,S_{n,\lambda},1)) = \max_{1\leq i\leq K} \sum_{v \in C_{n,\lambda}^i} |T_{v,\lambda}^n|\,. \end{aligned}$$

Write $m_n = |F^1(n,p_{n,\lambda})|$, then list the vertices of $F^1(n,p_{n,\lambda})$ as $v_1,\ldots,v_{m_n}$ so that for each $1 \leq i \leq K$, the vertices of $C_{n,\lambda}^i$ appear consecutively — as

$$v_{|C_{n,\lambda}^1|+\ldots+|C_{n,\lambda}^{i-1}|+1},\ldots,v_{|C_{n,\lambda}^1|+\ldots+|C_{n,\lambda}^i|}\,,$$

say. Necessarily $\max(|C_{n,\lambda}^i|, 1 \leq i \leq K) \leq M_{n,\lambda}(k(n))$, so on the event that $M_{n,\lambda}(k(n)) \leq$

$\delta n^{2/3}$ and $|F^1(n, p_{n,\lambda})| = m_n \geq n^{2/3}$, we then have

$$\frac{M_n}{n} \leq \max\Big(\sum_{j=i}^{\ell} q_{n,\lambda}^{v_j}, 1 \leq i < \ell \leq m_n, \ell - i < \delta m_n\Big).$$

Therefore, on this event, if $M_n/n \geq 3\epsilon$ then we may find i and ℓ as above so that

$$\sum_{j=i}^{\ell}\Big(q_{n,\lambda}^{v_j} - \frac{1}{m_n}\Big) = \Big(\sum_{j=i}^{\ell} q_{n,\lambda}^{v_j}\Big) - \frac{\ell - i}{m_n} \geq 3\epsilon - \delta > 2\epsilon\,,$$

so either $\sum_{j=1}^{i}(q_{n,\lambda}^{v_j} - 1/m_n) < -\epsilon$ or $\sum_{j=1}^{\ell}(q_{n,\lambda}^{v_j} - 1/m_n) > \epsilon$. It follows that

$$\begin{aligned}
\mathbf{P}\left\{M_n \geq 3\epsilon n\right\} &\leq \mathbf{P}\left\{M_{n,\lambda}(k(n)) > \delta n^{2/3}\right\} + \mathbf{P}\left\{|F^1(n, p_{n,\lambda})| < n^{2/3}\right\} \\
&\quad + \mathbf{P}\left\{\max_{1\leq i\leq m_n}\Big|\sum_{j=1}^{i}\Big(q_{n,\lambda}^{v_j} - \frac{1}{m_n}\Big)\Big| > \epsilon\right\} \\
&< 2\epsilon + \mathbf{P}\left\{\max_{1\leq i\leq m_n}\Big|\sum_{j=1}^{i}\Big(q_{n,\lambda}^{v_j} - \frac{1}{m_n}\Big)\Big| > \epsilon\right\}\,, \qquad (2.6)
\end{aligned}$$

where in the final line we have used (2.3) and (2.5). To bound the third probability we write

$$\begin{aligned}
&\mathbf{P}\left\{\max_{1\leq i\leq m_n}\Big|\sum_{j=1}^{i}\Big(q_{n,\lambda}^{v_j} - \frac{1}{m_n}\Big)\Big| > \epsilon\right\} \\
&= \mathbf{E}\left(\mathbf{P}\left\{\max_{1\leq i\leq m_n}\Big|\sum_{j=1}^{i}\Big(q_{n,\lambda}^{v_j} - \frac{1}{m_n}\Big)\Big| > \epsilon \;\Big|\; \mathcal{F}_{n,\lambda}\right\}\right) \\
&= \mathbf{E}\left(\mathbf{P}\left\{\max_{1\leq i\leq m_n}\Big|\sum_{j=1}^{i}\Big(q_{n,\lambda}^{v_{\pi(j)}} - \frac{1}{m_n}\Big)\Big| > \epsilon \;\Big|\; \mathcal{F}_{n,\lambda}\right\}\right)\,,
\end{aligned}$$

where conditionally given $\mathcal{F}_{n,\lambda}$, π is a uniformly random permutation of m_n independent of the values $(q_{n,\lambda}^{v_i}, 1 \leq i \leq m_n)$. The second equality holds as conditionally given $\mathcal{F}_{n,\lambda}$ the random variables $(q_{n,\lambda}^{v_i}, 1 \leq i \leq m_n)$ are exchangeable, due to Proposition 2.2.3.

Recall that $\Delta_\lambda^n := \max(q_{n,\lambda}^v, v \in F^1(n, p_{n,\lambda}))$; then by Proposition 2.2.4 we have

$$
\begin{aligned}
&\mathbf{P}\left\{\max_{1\le i\le m_n}\Big|\sum_{j=1}^{i}\left(q_{n,\lambda}^{v_{\pi(j)}} - \frac{1}{m_n}\right)\Big| > \epsilon \mid \mathcal{F}_{n,\lambda}\right\} \\
&\le \mathbf{P}\left\{\Delta_\lambda^n \ge \delta \mid \mathcal{F}_{n,\lambda}\right\} + \mathbf{P}\left\{\max_{1\le i\le m_n}\Big|\sum_{j=1}^{i}\left(q_{n,\lambda}^{v_{\pi(j)}} - \frac{1}{m_n}\right)\Big| > \epsilon \mid \mathcal{F}_{n,\lambda}, \Delta_\lambda^n \le \delta\right\} \\
&\le \mathbf{P}\left\{\Delta_\lambda^n \ge \delta \mid \mathcal{F}_{n,\lambda}\right\} + \epsilon\,,
\end{aligned}
$$

so it follows that

$$
\begin{aligned}
\mathbf{P}\left\{\max_{1\le i\le m_n}\Big|\sum_{j=1}^{i}\left(q_{n,\lambda}^{v_j} - \frac{1}{m_n}\right)\Big| > \epsilon\right\} &\le \epsilon + \mathbf{E}\left(\mathbf{P}\left\{\Delta_\lambda^n \ge \delta \mid \mathcal{F}_{n,\lambda}\right\}\right) \\
&= \epsilon + \mathbf{P}\left\{\Delta_\lambda^n \ge \delta\right\} < 2\epsilon\,,
\end{aligned}
$$

the last inequality holding by (2.4). Combining this bound with (2.6), it follows that $\mathbf{P}\{M_n \ge 3\epsilon n\} < 4\epsilon$; since $\epsilon > 0$ was arbitrary, this implies that $M_n/n \to 0$ in probability, as required. □

2.3 Proof of Proposition 2.2.1.

Our proof of Proposition 2.2.1 has three steps. In the first step, we show that it suffices to prove that all components of $F(n, [k(n)], p_{n,\lambda})$ contained in the largest $O(1)$ components of $F(n, p_{n,\lambda})$ have size $o(n^{2/3})$ in probability. This essentially boils down to the application of well-known facts about the structure of the critical random graph. In the second step we analyze the couplings presented above, between Kruskal's algorithm with different starting sets $\mathcal{S}$ and between Kruskal's algorithm and the Erdős-Rényi process. This analysis provides us with a tool for understanding how, distributionally, a given component C of $F(n, p_{n,\lambda})$ is partitioned into pieces in $F(n, [k(n)], p_{n,\lambda})$, depending on the number of elements of $C \cap [k(n)]$. In the third step, which occupies most of the rest of the paper, we use the result of the analysis of the couplings to show that the largest connected components of $F(n, p_{n,\lambda})$ are indeed partitioned into pieces of size $o(n^{2/3})$ in $F(n, [k(n)], p_{n,\lambda})$, with high probability.

2.3.1 Step 1: reducing to the study of large components.

For $p \in [0,1]$, list the components of $F(n,p)$ in decreasing order of size as $(F^i(n,p), i \geq 1)$, with ties broken uniformly at random. (The point of breaking ties this way is so that $v(F^i(n,p))$ is a uniformly random subset of $[n]$ conditional on its size.) Then for $\lambda \in \mathbb{R}$ and $\mathcal{S} \subset [n]$, write $M^i_{n,\lambda}(\mathcal{S})$ for the size of the largest connected component of $F(n,\mathcal{S},p_{n,\lambda})$ contained in $F^i(n,p_{n,\lambda})$. If $\mathcal{S} = [k(n)]$ we write $M^i_{n,\lambda}(k(n))$ instead of $M^i_{n,\lambda}([k(n)])$

In this section, we show how Proposition 2.2.1 is a consequence of the following result.

Proposition 2.3.1. *Fix $\lambda \in \mathbb{R}$ and $i \in \mathbb{N}$. If $k(n)/n^{1/3} \to \infty$ then $M^i_{n,\lambda}(k(n))/n^{2/3} \to 0$ in probability.*

Proof of Proposition 2.2.1. Fix $\epsilon > 0$.

By [17, Corollary 2], there is $j = j(\epsilon) \in \mathbb{N}$ such that for all $n \in \mathbb{N}$,

$$\mathbf{P}\left\{\max(|F^\ell(n,p_{n,\lambda})|, \ell > j) > \epsilon n^{2/3}\right\} < \epsilon\,.$$

Since $M^\ell(n,p_{n,\lambda})(k(n)) \leq |F^\ell(n,p_{n,\lambda})|$, it follows that for this value of j,

$$\begin{aligned}
\mathbf{P}\left\{M_{n,\lambda}(k(n)) \geq \epsilon n^{2/3}\right\} &\leq \mathbf{P}\left\{\max_{1\leq \ell\leq j} M^\ell_{n,\lambda}(k(n)) > \epsilon n^{2/3}\right\} \\
&\quad + \mathbf{P}\left\{\max(|F^\ell(n,p_{n,\lambda})|, \ell > j) > \epsilon n^{2/3}\right\} \\
&\leq \epsilon + \sum_{1\leq \ell\leq j} \mathbf{P}\left\{M^\ell_{n,\lambda}(k(n)) > \epsilon n^{2/3}\right\} \\
&\leq 2\epsilon
\end{aligned}$$

for n sufficiently large, the last bound holding due to Proposition 2.3.1. Since $\epsilon > 0$ was arbitrary, this proves Proposition 2.2.1. $\square$

2.3.2 Step 2: composing the couplings

It is useful to briefly return to the setting of a deterministic connected graph $G = (v(G), e(G), \mathrm{w})$. Fix a starting set $\mathcal{S} \subset v(G)$, and list edges of G in increasing order of weight as $e(1), \ldots, e(m)$. Using the couplings of $F^{G,\mathcal{S}}_i$ and $F^{G,\emptyset}_i$, on the one hand, and of F_i and G_i, on the other hand, allows us to construct $F^{G,\mathcal{S}}_m$ via a *path-and-cycle-breaking* process starting from G. Recall the definition of the augmented graph $G'_{\mathcal{S}}$ from Section 2.2.1, which is formed from G by adding a vertex ρ which is joined to the vertices of $\mathcal{S}$ by edges of very low weight.

Then an edge $e(i)$ is added to G_i but not to $F_i^{G,\emptyset}$ if and only if it lies on a cycle of G_i, which occurs if and only if it is the largest-weight edge on a cycle in G. the edge $e(i)$ is added to $F_i^{G,\emptyset}$ but not $F_i^{G,\mathcal{S}}$ if and only if it is the largest-weight edge on a cycle in $G'_{\mathcal{S}}$, which occurs if and only if there are distinct vertices $u, v \in \mathcal{S}$ such that $e(i)$ lies on a path from u to v in G_i (in which case $e(i)$ is the largest-weight edge on such a path). It follows that we may recover $F_m^{G,\mathcal{S}}$ from G as follows.

- Let $H_0 = G$.
- For $0 \leq i < m$, if either
  - (a) $e(m-i)$ lies on a cycle of H_i, or
  - (b) there exist distinct vertices $u, v \in \mathcal{S}$ such that $e(m-i)$ lies on a path from u to v in H_i,

  then set $H_{i+1} = H_i - e(m-i)$; otherwise set $H_{i+1} = H_i$.

The final graph H_m is precisely $F_m^{G,\mathcal{S}}$. This path-and-cycle-breaking construction of $F_m^{G,\mathcal{S}}$ has the following immediate consequence in the setting of exchangeable edge weights.

Fact 2.3.2 (Path-and-cycle-breaking). *Let $G = (v(G), e(G), \mathrm{w})$ be a connected graph with exchangeable, almost surely distinct edge weights. Fix $\mathcal{S} \subset v(G)$ and an ordering $\mathbf{e} = (e_1, \ldots, e_m)$ of $e(G)$. Generate a subgraph F of G as follows.*

1. *Let $H_0 = G$.*
2. *For $0 \leq i < m$, if e_i lies on a cycle in H_i or e_i lies on a path in H_i between distinct vertices of $\mathcal{S}$, then set $H_{i+1} = H_i - e_i$; otherwise set $H_{i+1} = H_i$.*
3. *Set $F = H_m$.*

If the ordering $\mathbf{e}$ is exchangeable then F is distributed as $F_m^{G,\mathcal{S}}$.

We call the above process *path-and-cycle-breaking on G with starting set $\mathcal{S}$ and edge ordering* $\mathbf{e}$, and refer to F as the outcome of the process. (The edge weights w are not used in the process, but they are used in defining $F_m^{G,\mathcal{S}}$.) Most of our analysis will end up focussing on the case that G is in fact a tree; in this case path-and-cycle-breaking process clearly never breaks cycles, and we simply refer to it as a path-breaking process.

We shall use the path-and-cycle-breaking process to understand how the components of $F(n, [k(n)], p_{n,\lambda})$ partition those of $G(n, p_{n,\lambda})$. Suppose that C is a connected component of

$G(n, p_{n,\lambda})$. Let $N = |v(C)|$ and let $S = |e(C)| - |v(C)| + 1$ be the *surplus* of C. Let C' be obtained from C by relabeling the vertices of C in increasing order as $1, \ldots, N$. Then C' is uniformly distributed over connected graphs with vertex set $[N]$ and surplus S, and its edge weights are exchangeable. In view of these facts, the value of the next proposition should be rather clear.

Proposition 2.3.3. *For all $\epsilon > 0$ and any non-negative integer s, there exists integer $r > 0$ such that the following holds. For $q \geq 1$, let G_q be uniformly distributed over the set of connected graphs with vertex set $[q]$ and surplus s. For $q \geq r$ let $F_q = F_q(r, \mathbf{e})$ be the outcome of the path-and-cycle-breaking process on G_q with starting set $[r]$ and an exchangeable random ordering $\mathbf{e} = (e_1, \ldots, e_m)$ of $e(G_q)$. Then for all q sufficiently large,*

$$\mathbf{E}\left(\max(|C| : C \textit{ is a component of } F_q)\right) \leq \epsilon q\,.$$

This proposition has the following consequence. Fix non-negative integers s and $(r(q), q \geq 1)$ with $r(q) \leq q$ and with $r(q) \to \infty$ as $q \to \infty$. Let G_q be as in Proposition 2.3.3, and let F_q be the outcome of the path-and-cycle-breaking process on G_q with starting set $[r(q)]$ and an exchangeable random ordering $\mathbf{e} = (e_1, \ldots, e_m)$ of $e(G_q)$. Then Proposition 2.3.3 and Markov's inequality together imply that for any $\epsilon > 0$,

$$q^{-1}\mathbf{E}\left(\max(|C| : C \text{ is a component of } F_q)\right) \to 0 \tag{2.7}$$

as $q \to \infty$.

We prove Proposition 2.3.3 in Section 2.3.3, below; before doing so, we use it (or in fact its consequence, (2.7)) to prove Proposition 2.3.1.

Proof of Proposition 2.3.1. Fix $\lambda \in \mathbb{R}$ and $i \in \mathbb{N}$. Write $G^i(n, p_{n,\lambda})$ for the component of $G^i(n, p_{n,\lambda})$ spanned by $F^i(n, p_{n,\lambda})$.

Let $Q = Q(n) = |v(G^i(n, p_{n,\lambda}))| = |v(F^i(n, p_{n,\lambda}))|$, let $R = R(n) = |v(G^i(n, p_{n,\lambda}) \cap [k(n)]|$, and let $S = S(n) = |e(G^i(n, p_{n,\lambda}))| - |v(G^i(n, p_{n,\lambda})| + 1$ be the surplus of $G^i(n, p_{n,\lambda})$.

We will use in the course of the proof that $S(n)$ converges in distribution to an almost surely finite limit, and that $n^{-2/3}|F^i(n, p_{n,\lambda})| = n^{-2/3}|G^i(n, p_{n,\lambda})|$ converges in distribution to an almost surely finite, strictly positive limit; these facts appear in [17, Folk Theorem 1 and Corollary 2].

Conditionally given $Q(n)$, the vertex set $v(G^i(n, p_{n,\lambda}))$ is a uniformly random size-$Q(n)$ subset of $[n]$. The last convergence in distribution referenced in the previous paragraph (and in particular the fact that the limit is almost surely strictly positive) implies that for any $\epsilon > 0$ there exists $\delta > 0$ such that $\mathbf{P}\left\{Q(n) \geq \delta n^{2/3}\right\} > 1 - \epsilon$. Since $k(n)/n^{1/3} \to \infty$, this implies that $k(n)Q(n)/n \to \infty$ in probability. Since $v(F^i(n, p_{n,\lambda}))$ is a uniformly random subset of $[n]$ conditional on its size, it then follows by standard concentration results for sampling without replacement that $R(n) \to \infty$ in probability. Moreover, for any fixed $s \in \mathbb{N}$, by [17, Corollary 2] we have $\liminf_{n\to\infty} \mathbf{P}\{S(n) = s\} > 0$, Since $Q(n)$ and $R(n)$ both tend to infinity in probability, it follows that $Q(n)$ and $R(n)$ still tend to infinity in probability on the event that $S(n) = s$, in the sense that for any $x > 0$,

$$\mathbf{P}\{Q(n) > x, R(n) > x \mid S(n) = s\} \to 1$$

as $n \to \infty$.

Next, recall that

$$\begin{aligned}
&M^i_{n,\lambda}(k(n)) \\
&:= \max(|C| : C \text{ a conn. comp. of } F(n, [k(n)], p_{n,\lambda}) \text{ contained in } F^i(n, p_{n,\lambda}))
\end{aligned}$$

and write $M^i_{n,\lambda} = M^i_{n,\lambda}(k(n))$ for succinctness. Conditionally given $Q(n)$, $R(n)$ and $S(n)$, the random variable $M^i_{n,\lambda}$ has the same distribution as $\max(|C| : C$ is a component of $F_{Q(n)})$, where $F_{Q(n)}$ is the outcome of the path-and-cycle breaking process on $G_{Q(n)}$ with starting set $R(n)$. By the exchangeability of the vertex labels, this distribution is unchanged if rather than $R(n)$ we use the starting set $[|R(n)|] = \{1, \ldots, |R(n)|\}$. It then follows from (2.7) and Markov's inequality that for any $\epsilon > 0$,

$$\begin{aligned}
\mathbf{P}\left\{M^i_{n,\lambda} > \epsilon|F^i(n, p_{n,\lambda})| \mid S(n) = s\right\} &= \mathbf{P}\left\{Q(n)^{-1}M^i_{n,\lambda} > \epsilon \mid S(n) = s\right\} \\
&\to 0\,,
\end{aligned}$$

as $n \to \infty$. Moreover, since $S(n)$ converges in distribution to an almost surely finite limit, it follows that for all $\epsilon > 0$ there is s_0 such that for n sufficiently large, $\mathbf{P}\{S(n) > s_0\} < \epsilon$. Combined with the preceding bound, this yields that for any $\epsilon > 0$,

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbf{P}\left\{M^i_{n,\lambda} > \epsilon|F^i(n, p_{n,\lambda})|\right\} \\
&\leq \limsup_{n\to\infty} \max_{1\leq s\leq s_0} \mathbf{P}\left\{Q(n)^{-1}M^i_{n,\lambda} > \epsilon \mid S(n) = s\right\} + \limsup_{n\to\infty} \mathbf{P}\{S(n) > s_0\} \\
&\leq \epsilon\,.
\end{aligned}$$

It follows that $M^i_{n,\lambda}/|F^i(n,p_{n,\lambda})| \to 0$ in probability. Since $|F^i(n,p_{n,\lambda})|/n^{-2/3}$ converges in distribution to an almost surely finite limit, this implies that $M^i_{n,\lambda}/n^{2/3} \to 0$ in probability, as required. $\square$

2.3.3 Step 3: partitioning a component

The goal of this section is to prove Proposition 2.3.3. We first prove the proposition for the special case $s = 0$, in which case G_q is a uniformly random tree with vertex set q, and the path-and-cycle-breaking process is simply a path-breaking process. We may restate the case $s = 0$ of Proposition 2.3.3 as follows.

Proposition 2.3.4. *For all $\epsilon > 0$, there exists $r > 0$ such that the following holds. For $q \geq 1$, let T_q be uniformly distributed over the set of trees with vertex set $[q]$. Let $F_q = F_q(r, \mathbf{e})$ be the outcome of the path-breaking process on T_q with starting set $[r]$ and an exchangeable random ordering $\mathbf{e} = (e_1, \ldots, e_{q-1})$ of $e(T_q)$. Then for all q sufficiently large,*

$$\mathbf{E}\left(\max(|C| : C \textit{ is a component of } F_q)\right) \leq \epsilon q\,.$$

In Section 2.3.3.1 we prove Proposition 2.3.4, establishing the case $s = 0$ of Proposition 2.3.3. We then use Proposition 2.3.4 to handle the cases when $s \geq 1$, completing the proof of Proposition 2.3.3, in Section 2.3.3.2.

For what follows it is useful to introduce the notation $u \overset{G}{\longleftrightarrow} v$ to mean that there exists a path from u to v in graph G (i.e., u and v are vertices of G lying in the same connected component of G).

2.3.3.1 Proof of Proposition 2.3.4.

For $q \geq 1$, let T_q be uniformly distributed over the set of trees with vertex set $[q]$. Let F_q be the outcome of the path-breaking process on T_q with starting set $[r]$ and an exchangeable random ordering $\mathbf{e} = (e_1, \ldots, e_{q-1})$ of $e(T_q)$. Then, for independent, uniformly random vertices $X, Y \in_u [q]$,

$$\begin{aligned}
\mathbf{P}\left\{X \overset{F_q}{\longleftrightarrow} Y\right\} &= \mathbf{E}\left(\mathbf{P}\left\{X \overset{F_q}{\longleftrightarrow} Y \;\middle|\; F_q\right\}\right) \\
&= \mathbf{E}\left(\sum_{C \text{ is a component of } F_q} \frac{|v(C)|^2}{q^2}\right) \\
&\geq \mathbf{E}\left(\frac{\max(|v(C)| : C \text{ is a component of } F_q)^2}{q^2}\right) \\
&\geq \frac{\mathbf{E}\left(\max(|v(C)| : C \text{ is a component of } F_q)\right)^2}{q^2} . \qquad (2.8)
\end{aligned}$$

We thus analyze the probability that independent samples $X, Y \in_u [q]$ are connected in F_q. For the bulk of the analysis, it is in fact useful to instead consider U, W sampled uniformly *without* replacement, from the set $\{r+1, \ldots, q\}$. Since $\mathbf{P}\{X = Y\} + \mathbf{P}\{X \in [r]\} + \mathbf{P}\{Y \in [r]\} \to 0$ as $q \to \infty$, the error term this adds to the above bound is asymptotically negligible.

For a tree t and a set $S \subseteq v(t)$, write $t\langle S\rangle$ for the smallest subtree of t containing S.

Lemma 2.3.5. *Fix $q \geq r+2$, let U, W be sampled uniformly without replacement from $\{r+1, \ldots, q\}$, let $M_q = |e(T_q\langle\{U, W\} \cup [r]\rangle)|$, and let $(e_{i_1}, \ldots, e_{i_{M_q}})$ be the restriction of the exchangeable random ordering $\mathbf{e}$ to $e(T_q\langle\{U, W\} \cup [r]\rangle)$. Then $U \overset{F_q}{\longleftrightarrow} W$ if and only if $U \overset{F_q'}{\longleftrightarrow} W$ where F_q' is the outcome of the path-breaking process on $T_q\langle\{U, W\} \cup [r]\rangle$ with starting set $[r]$ and edge ordering $(e_{i_1}, \ldots, e_{i_{M_q}})$.*

Proof. For $i \notin \{i_1, \ldots, i_{M_q}\}$, the edge e_i does not lie on a path between any pair of elements of $[r]$, so is not removed by the path-breaking process on T_q with starting set $[r]$ and edge ordering $(e_1, \ldots, e_{q-1})$. It follows (by induction) that the path-breaking process on $T_q\langle\{U, W\} \cup [r]\rangle$ with starting set $[r]$ and edge ordering $(e_{i_1}, \ldots, e_{i_{M_q}})$ removes the same edges, in the same order, as the previously mentioned path-breaking process on T_q. Hence, $F_q' = F_q\langle\{U, W\} \cup [r]\rangle$ so U and W are connected in F_q if and only if they are connected in F_q'. $\square$

For ease of notation, let $T_q' = T_q'(r, \mathbf{e})$ be the tree obtained from $T_q\langle\{U, W\} \cup [r]\rangle$ by relabeling U, W as $r+1, r+2$, relabeling the vertices of $v\big(T_q\langle\{U, W\} \cup [r]\rangle\big) \setminus (\{U, W\} \cup [r])$ in increasing order as $\{r+3, \ldots, M_q\}$, and relabeling the edges $(e_{i_1}, \ldots, e_{i_{M_q}})$ as $\mathbf{e}' = (e_1, \ldots, e_{M_q})$. In a small abuse of notation we continue to denote the relabelings of U, W by

U and W rather than by $r+1$ and $r+2$. Finally, write $F'_q = F'_q(r, \mathbf{e}')$ for the outcome of the path breaking process on T'_q with starting set $[r]$ and edge ordering $\mathbf{e}'$.

For all $m \geq r+2$, let $\mathcal{T}_m = \mathcal{T}_m(r)$ be the set of trees with vertex set $[m]$ and with leaf set a subset of $[r+2]$. Note that since T_q is a uniformly random tree with vertex set $[q]$, by symmetry, $T'_q \in_u \mathcal{T}_{M_q}$, in the sense that for all $m \geq r+2$, conditionally given that $M_q = m$, then $T'_q \in_u \mathcal{T}_m$.

The definitions of this paragraph are illustrated in Figure 2.2. For $T \in \mathcal{T}_m$, let $P_0 = P_0(T) = T\langle\{U, W\}\rangle$, and for all $1 \leq i \leq r$ let $P_i = P_i(T)$ be the path in T connecting i to $T\langle\{U, W\} \cup [i-1]\rangle$. (If $i \in v(T\langle\{U, W\} \cup [i-1]\rangle)$ then P_i consists of a single vertex and no edges; otherwise P_i has edge set $e(T\langle\{U, W\} \cup [i]\rangle) \setminus e(T\langle\{U, W\} \cup [i-1]\rangle)$.) Then, for each $j \in \{1, 2, 3\}$ let $P_{0,j} = P_{0,j}(T)$ be the subpath of $P_0(T)$ with vertex set

$$\left\{ v \in v(P_0(T)) : \left\lfloor \operatorname*{dist}_T(U, W)\frac{j-1}{3} \right\rfloor \leq \operatorname*{dist}_T(U, v) \leq \left\lfloor \operatorname*{dist}_T(U, W)\frac{j}{3} \right\rfloor \right\} ,$$

and let $\mathcal{U}_j = \mathcal{U}_j(T, r, q)$ be the subset of $[r]$ satisfying the following additional properties: for all $i \in \mathcal{U}_j$,

1. $v(P_i) \cap v(P_\ell) = \emptyset$ for all $\ell \in [r] \setminus \{i\}$,
2. $P_i \cap P_0$ is a vertex in $P_{0,j}$, and
3. $1 \leq |e(P_i)| \leq \sqrt{q}$.

Additionally, write $\mathcal{P}_j = \{P_i : i \in \mathcal{U}_j\}$.

For all paths $P \in T'_q$, say P is *targeted* at time t if $e_t \in e(P)$, say P is *targeted for the first time* at time t if P is targeted at time t and not before time t, and say P is *broken* at time t if P is targeted at time t and e_t is removed during the path-breaking process. In the next lemma (and the subsequent Corollary 2.3.7), write $P_i = P_i(T'_q)$ for $0 \leq i \leq r$, and write $P_{0,j} = P_{0,j}(T'_q)$, $\mathcal{U}_j = \mathcal{U}_j(T'_q)$ and $\mathcal{P}_j = \mathcal{P}_j(T'_q)$ for $j \in \{1, 2, 3\}$.

Lemma 2.3.6. *If $U \overset{F'_q}{\longleftrightarrow} W$ and both $|\mathcal{U}_1| \geq 1$ and $|\mathcal{U}_3| \geq 1$, then all but at most one path in $\mathcal{P}_1$ and all but at most one path in $\mathcal{P}_3$ were targeted before $P_{0,2}$ was targeted for the first time.*

Proof. Fix $1 \leq t \leq M_q$. Suppose that $P_i \in \mathcal{P}_1$ and $P_j \in \mathcal{P}_3$ have not been targeted by time t, and that $P_{0,2}$ is targeted for the first time at time t. Let $P_{i,j}$ be the path from i to j in T'_q. Then $P_{i,j} = P_i \cup Q_{0,1} \cup P_{0,2} \cup Q_{0,3} \cup P_j$ where $Q_{0,1} \subseteq P_{0,1}$ and $Q_{0,3} \subseteq P_{0,3}$. Hence, either

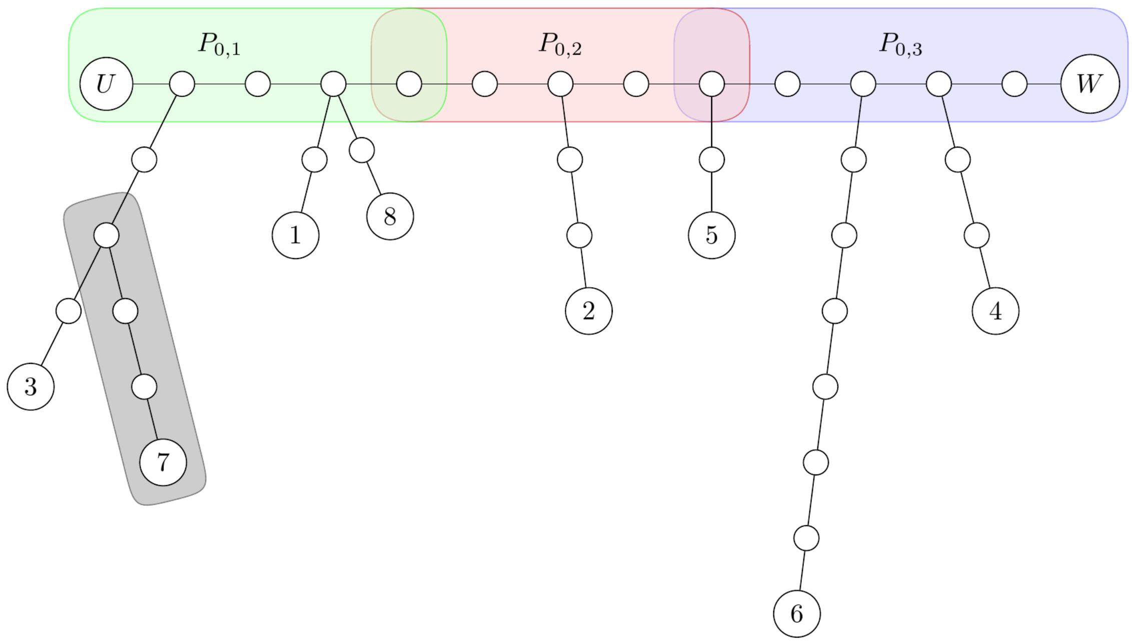

Figure 2.2: An example of $T \in \mathcal{T}_{40}$ with $r = 8$ and $q = 48$. In this example, P_0 connects vertices $U = r + 1$ and $W = r + 2$, the path P_7 is shaded in grey, and all the other paths P_i connect the vertex labeled i to the path P_0. Also, $\mathcal{U}_1(T) = \emptyset$, $\mathcal{U}_2(T) = \{2, 5\}$, and $\mathcal{U}_3(T) = \{4\}$. Note that $6 \notin \mathcal{U}_3$ as $|e(P_6(T))| = 7 > \sqrt{48}$.

$e_t \in P_{0,2}$ is cut, or an edge in $P_{i,j}$ has already been cut. In the latter case, since none of P_i, P_j, or $P_{0,2}$ have been targeted by time t, the edge that was already cut must be an edge of $Q_{0,1} \cup Q_{0,3} \subseteq P_0$. In both cases, an edge in P_0 is cut during the path-breaking process, meaning U and W are not connected in F'_q. $\square$

For the next corollary, it is useful to introduce the shorthand

$$\hat{\mathbf{P}}\{\cdot\} = \mathbf{P}\left\{\cdot \mid T'_q\right\}.$$

Corollary 2.3.7. *Let $E_q(p, u)$ be the event that $|e(P_{0,2})| \geq p$, $|\mathcal{U}_1| \geq u$ and $|\mathcal{U}_3| \geq u$. Then for $u \geq 1$,*

$$\hat{\mathbf{P}}\left\{U \overset{F'_q}{\longleftrightarrow} W \;\middle|\; E_q(p, u)\right\} \leq \prod_{j=0}^{2u-3}\left(1 - \frac{p}{p + \sqrt{q}(2u - j)}\right).$$

Proof. Let $\mathcal{S} = \mathcal{P}_1 \cup \mathcal{P}_3 \cup \{P_{0,2}\}$. For all paths $P \in \mathcal{S}$, let t_P be the first time P is targeted. Then let $s = \left|\{P \in \mathcal{S} : t_P < t_{P_{0,2}}\}\right|$. By Lemma 2.3.6, if U and W are connected in F'_q and both $\mathcal{U}_1$ and $\mathcal{U}_3$ are nonempty, then $s \geq |\mathcal{S}| - 3$.

Let $E_q^{=}(p,u)$ be the event that $|e(P_{0,2})| = p$, $|\mathcal{U}_1| = u$ and $|\mathcal{U}_3| = u$. List the paths in $\mathcal{S}$ in the order they are first targeted as $P^{(1)}, \ldots, P^{(m)}$. By the exchangeability of the ordering $\mathbf{e}' = (e_1, \ldots, e_{M_q})$, this list is a size-biased random ordering of the paths in $\mathcal{S}$: that is, for all $1 \leq j \leq m$ and all $P \in \mathcal{S}$,

$$\hat{\mathbf{P}}\left\{ P^{(j)} = P \;\middle|\; P^{(1)}, \ldots, P^{(j-1)} \right\} = \frac{|P|\mathbf{1}_{[P \notin \{P^{(1)},\ldots,P^{(j-1)}\}]}}{\sum_{Q\in\mathcal{S}} |Q|\mathbf{1}_{[Q \notin \{P^{(1)},\ldots,P^{(j-1)}\}]}}\,.$$

Since all paths in $\mathcal{S}$ aside from $P_{0,2}$ have length at most $\sqrt{q}$, it follows that for all $0 \leq j \leq m$,

$$\hat{\mathbf{P}}\left\{ P^{(j)} = P_{0,2} \;\middle|\; E_q^{=}(p,u), P_{0,2} \notin \{P^{(1)}, \ldots, P^{(j-1)}\} \right\} \geq \frac{p}{p + \sqrt{q}(2u - (j-1))}\,.$$

Thus, by Bayes' formula, and since $m = 2u+1$ when $E_q^{=}(p,u)$ occurs, we have

$$\begin{aligned}
&\hat{\mathbf{P}}\left\{ s \geq |\mathcal{S}| - 3 \;\middle|\; E_q^{=}(p,u) \right\} \\
&= \hat{\mathbf{P}}\left\{ P_{0,2} \notin \{P^{(1)}, \ldots, P^{(m-3)}\} \;\middle|\; E_q^{=}(p,u) \right\} \\
&= \prod_{j=0}^{m-4} \hat{\mathbf{P}}\left\{ P_{0,2} \neq P^{(j+1)} \;\middle|\; E_q^{=}(p,u), P_{0,2} \notin \{P^{(1)}, \ldots, P^{(j)}\} \right\} \\
&\leq \prod_{j=0}^{2u-3} \left(1 - \frac{p}{p + \sqrt{q}(2u - j)} \right)\,.
\end{aligned} \tag{2.9}$$

Since the above product is decreasing in p and in u, and $E_q(p,u)$ is the disjoint union of the events $\{E_q^{=}(p',u'), p' \geq p, u' \geq u\}$, the bound claimed in the corollary follows from (2.9) by the law of total probability. □

In order to make use of Corollary 2.3.7, we now must analyze the typical behaviour of $|e(P_{0,2(T_q')})|$, $|\mathcal{U}_1(T_q')|$ and $|\mathcal{U}_3(T_q')|$. We first gather some auxiliary facts which are crucial for this analysis.

The first facts relate to the asymptotic structure of $T_q'(r)$ for q large. This is described by the so-called *line-breaking construction* of Aldous [16]. Let $(\pi_i, i \geq 0)$ be the ordered sequence of inter-arrival times of a Poisson point sequence on $[0,\infty)$ with intensity measure $\lambda(t) = t$. Such a process may be concretely realized as follows. Let $(E_i, i \geq 1)$ be independent Exp(1) random variables, let $\pi_0 = \sqrt{2}E_1^{1/2}$ and, for each $j \geq 1$, let $\pi_j = \sqrt{2}\,(E_1 + \cdots + E_{j+1})^{1/2} - \sqrt{2}\,(E_1 + \cdots + E_j)^{1/2}$. The atoms of the Poisson process are thus located at the points $\left(\sqrt{2}(E_1 + \ldots + E_i)^{1/2}, i \geq 1\right)$.

Now for $r \geq 0$, construct a binary tree $T_\infty(r)$ with edge lengths and with leaf labels U, W and $1, \ldots, r$, as follows. The tree $T_\infty(0)$ is a line segment $P_{\infty,0}$ of length $|P_{\infty,0}| = \pi_0$, with endpoints labeled U and W. Inductively, for $r \geq 1$ the tree $T_\infty(r)$ is constructed from $T_\infty(r-1)$ by attaching a line segment $P_{\infty,r}$ of length $|P_{\infty,r}| = \pi_r$ to a uniformly chosen point of $T_\infty(r-1)$, and assigning label r to the new leaf at the far end of the line segment.

The next proposition is a consequence of [16, Theorem 8]. In what follows, if G is an unweighted graph then we write $\mathrm{dist}_G(x, y)$ to mean the graph distance between vertices x and y in G (the fewest number of edges in an $x - y$ path). If G is a graph with edge lengths then we write $\mathrm{dist}_G(x, y)$ to mean the length of the shortest $x - y$ path, taking edge lengths into account.

Proposition 2.3.8. *As $q \to \infty$,*

$$\begin{aligned}&(q^{-1/2}\mathrm{dist}_{T'_q(r)}(x, y) : x, y \in \{U, W\} \cup \{1, \ldots, r\})\\ &\xrightarrow{\mathrm{d}} (\mathrm{dist}_{T_\infty(r)}(x, y) : x, y \in \{U, W\} \cup \{1, \ldots, r\}) .\end{aligned}$$

Moreover, for all $r \geq 1$, ignoring its edge lengths, the tree $T_\infty(r)$ is uniformly distributed over the set of binary trees with leaf labels $\{U, W\} \cup [r]$. Finally, for any permutation $\phi : \{U, W, 1, \ldots, r\} \to \{U, W, 1, \ldots, r\}$, the tree $T^\phi_\infty(r)$ obtained from $T_\infty(r)$ by relabeling its leaves according to the permutation ϕ has the same law as $T_\infty(r)$.

Write α_r for the point of $T_\infty(r-1)$ to which $P_{\infty,r}$ is attached. Note that the lengths of the paths $(P_{\infty,i}, 0 \leq i \leq r)$ may be recovered from $T_{\infty,r}$ as $\pi_i = \mathrm{dist}_{T_\infty(r)}(i, \alpha_i)$. Thus, the convergence in Proposition 2.3.8 directly implies that

$$(q^{-1/2}|e(P_0)|, \ldots, q^{-1/2}|e(P_r)|) \xrightarrow{\mathrm{d}} (|P_{\infty,0}|, \ldots, |P_{\infty,r}|) = (\pi_0, \ldots, \pi_r)$$

as $q \to \infty$.

For later use it's handy to describe the reverse construction (recovering the branches from the tree) in a little more generality. Given a binary tree t with edge lengths and with leaves labeled by the elements of $\{U, W\} \cup [r]$, we define a growing sequence of subtrees $t(0), \ldots, t(r)$ where $t(i)$ is the smallest subtree of t containing the leaves in $\{U, W\} \cup \{1, \ldots, i\}$. We then let $P_0(t) = t(0)$, for $i \in [r]$ we let $P_i(t)$ be the path connecting the leaf i to the subtree $t(i-1)$, and let $\alpha_i(t)$ be the point of $t(i-1)$ to which $P_i(t)$ attaches. With these definitions, we have $P_{\infty,i} = P_i(T_{\infty,r})$ for $0 \leq i \leq r$.

The following tail bound for $|e(P_{0,2}(T'_q))|$ is a straightforward consequence of Proposition 2.3.8.

Corollary 2.3.9. *For all $\epsilon > 0$ and positive integers r, for large enough q,*

$$\mathbf{P}\left\{|e(P_{0,2}(T'_q))| < \epsilon\sqrt{q}\right\} \le 5\epsilon^2 .$$

Proof. By Proposition 2.3.8, we have that

$$\begin{aligned}
\mathbf{P}\left\{|e(P_0(T'_q))| < 3\epsilon\sqrt{q}\right\} &= (1+o_q(1))\mathbf{P}\left\{\sqrt{2}(E_1)^{1/2} < 3\epsilon\right\} \\
&= (1+o_q(1))\mathbf{P}\left\{E_1 < \frac{9\epsilon^2}{2}\right\} \\
&= (1+o_q(1))\left(1 - e^{-\frac{9\epsilon^2}{2}}\right) \\
&\le (1+o_q(1))\frac{9\epsilon^2}{2} .
\end{aligned}$$

The result now follows from the fact that $|e(P_{0,2}(T'_q))| \ge \frac{1}{3}|e(P_0(T'_q))| - 1$. □

The next lemma states the tail bound we need for $|\mathcal{U}_1(T'_q)|$ and $|\mathcal{U}_3(T'_q)|$.

Lemma 2.3.10. *For all $\epsilon > 0$ sufficiently small, there exists a positive integer r_0 such that for all $r \ge r_0$, for all q sufficiently large, with $T'_q = T'_q(r, \mathbf{e})$,*

$$\mathbf{P}\left\{|\mathcal{U}_1(T'_q)| < \epsilon\sqrt{r}\right\} < 23\epsilon ,$$

and the same bound holds with $|\mathcal{U}_1(T'_q)|$ replaced by $|\mathcal{U}_3(T'_q)|$.

The proof of Lemma 2.3.10 is somewhat involved, so before proving it we first use it together with the preceding results of the section to prove Proposition 2.3.4.

Proof of Proposition 2.3.4. In this proof, for readability we omit insignificant floors and ceilings. Fix $\epsilon > 0$ small enough that Lemma 2.3.10 applies, let r_0 be as in Lemma 2.3.10, and fix $r \ge r_0$ large enough that $\sum_{j=4}^{\epsilon r^{1/2}} \frac{1}{j} \ge \epsilon^{-1}\log(\epsilon^{-1})$. For the duration of the proof, write $P_j = P_j(T'_q)$ for all $0 \le j \le r$ and $P_{0,i} = P_{0,i}(T'_q)$ and $\mathcal{U}_i = \mathcal{U}_i(T'_q)$ for $i \in \{1, 2, 3\}$.

Recall from Corollary 2.3.7 that $E_q(p, u)$ is the event that $|e(P_{0,q})| \ge p$, $|\mathcal{U}_1| \ge u$ and

$|\mathcal{U}_3| \geq u$. Taking $p = \epsilon q^{1/2}$ and $u = \epsilon r^{1/2}/2$, by that corollary and Bayes' formula we have

$$\begin{aligned}
\hat{\mathbf{P}}\left\{U \xleftrightarrow{F_q'} W\right\} &\leq \hat{\mathbf{P}}\left\{U \xleftrightarrow{F_q'} W \,\middle|\, E_q(p,u)\right\} + \hat{\mathbf{P}}\left\{E_q(p,u)^c\right\} \\
&\leq \prod_{j=0}^{2u-3}\left(1 - \frac{p}{p + q^{1/2}(2u-j)}\right) + \hat{\mathbf{P}}\left\{E_q(p,u)^c\right\} \\
&= \prod_{j=0}^{\epsilon r^{1/2}-3}\left(1 - \frac{\epsilon}{\epsilon(1+r^{1/2}) - j}\right) + \hat{\mathbf{P}}\left\{E_q(p,u)^c\right\} \\
&\leq \exp\left(-\epsilon \sum_{j=0}^{\epsilon r^{1/2}-3} \frac{1}{\epsilon(1+r^{1/2}) - j}\right) + \hat{\mathbf{P}}\left\{E_q(p,u)^c\right\} \\
&\leq \exp\left(-\epsilon \sum_{j=4}^{\epsilon r^{1/2}} \frac{1}{j}\right) + \hat{\mathbf{P}}\left\{E_q(p,u)^c\right\} \\
&\leq \epsilon + \hat{\mathbf{P}}\left\{E_q(p,u)^c\right\} ,
\end{aligned}$$

the last bound holding since we chose r large enough that $\sum_{j=4}^{\epsilon r^{1/2}} \frac{1}{j} \geq \epsilon^{-1}\log(\epsilon^{-1})$. Taking expectations, the tower law yields that

$$\mathbf{P}\left\{U \xleftrightarrow{F_q'} W\right\} \leq \epsilon + \mathbf{P}\left\{E_q(p,u)^c\right\}.$$

By Corollary 2.3.9 we have $\mathbf{P}\left\{|e(P_{0,2})| \leq \epsilon q^{1/2}\right\} \leq 5\epsilon^2$ and by Lemma 2.3.10 we have $\mathbf{P}\left\{\min(|\mathcal{U}_1|, |\mathcal{U}_3|) \leq \epsilon r^{1/2}/2\right\} \leq 46\epsilon$, so $\mathbf{P}\left\{E_q(p,u)^c\right\} \leq 5\epsilon^2 + 46\epsilon$. Also, by Lemma 2.3.5 we have $\mathbf{P}\left\{U \xleftrightarrow{F_q'} W\right\} = \mathbf{P}\left\{U \xleftrightarrow{F_q} W\right\}$, so we obtain the bound

$$\mathbf{P}\left\{U \xleftrightarrow{F_q} W\right\} \leq 47\epsilon + 5\epsilon^2.$$

To conclude, let X, Y be independent uniform samples from $[q]$. The conditional distribution of (X, Y) given that $X \neq Y$ and that $X \notin [r], Y \notin [r]$ is precisely that of (U, V), so

$$\begin{aligned}
\mathbf{P}\left\{X \xleftrightarrow{F_q} Y\right\} &\leq \mathbf{P}\left\{U \xleftrightarrow{F_q} W\right\} + \mathbf{P}\left\{X = Y\right\} + \mathbf{P}\left\{X \in [r]\right\} + \mathbf{P}\left\{Y \in [r]\right\} \\
&\leq 47\epsilon + 5\epsilon^2 + \frac{2r+1}{q}.
\end{aligned}$$

Finally, by (2.8), we have

$$\mathbf{P}\left\{X \overset{F_q}{\longleftrightarrow} Y\right\} \geq \frac{\mathbf{E}\left(\max\left(|C| : C \text{ is a component of } F_q\right)\right)^2}{q^2},$$

and so

$$\begin{aligned}\mathbf{E}\left(\max\left(|C| : C \text{ is a component of } F_q\right)\right) &\leq q\mathbf{P}\left\{X \overset{F_q'}{\longleftrightarrow} X\right\}^{1/2} \\ &\leq q\left(47\epsilon + 5\epsilon^2 + \frac{2r+1}{q}\right)^{1/2}.\end{aligned}$$

The result follows since $\epsilon > 0$ can be taken arbitrarily small, and since we can make $(2r+1)/q$ as small as we like by taking q large. □

The remainder of the section is devoted to proving Lemma 2.3.10. We use Proposition 2.3.8 to allow us to control the large-q behaviour of the probabilities in question by instead studying the limiting tree $T_\infty(r)$. Given a binary tree t with edge lengths and with leaves labeled by $\{U, W\} \cup [r]$, recall that $\alpha_t(i)$ is the attachment point of the line segment $P_i(t)$ to $t(i-1)$, and that $|P_i(t)| = \mathrm{dist}_t(i, \alpha(i))$. Let $\mathcal{U}_1(t)$ be the set of leaves $i \in [r]$ satisfying the following properties.

1. $P_i(t) \cap P_j(t) = \emptyset$ for all $j \in [r] \setminus \{i\}$,
2. $\alpha_i(t) \in P_0(t)$ and $\mathrm{dist}_t(\alpha_i(t), U) \leq \mathrm{dist}_t(U, W)/3$, and
3. $\mathrm{dist}_t(i, \alpha_i(t)) \leq 1$.

Define $\mathcal{U}_3(t)$ in the same way, but with the second condition replaced by the condition that $\mathrm{dist}_t(\alpha_i(t), W) \leq \mathrm{dist}_t(U, W)/3$.

The convergence in Proposition 2.3.8 implies that for any $r \geq 1$, as $q \to \infty$ we have $(|\mathcal{U}_1(T_q', r, q)|, |\mathcal{U}_3(T_q', r, q)| \xrightarrow{\mathrm{d}} (|\mathcal{U}_1(T_\infty(r))|, |\mathcal{U}_3(T_\infty(r))|)$. This allows us to prove the proposition by proving lower tail bounds for $|\mathcal{U}_1(T_\infty(r)|$ and $|\mathcal{U}_3(T_\infty(r)|$. To establish such bounds, we will use the second moment method, applied conditionally given $|\pi_0|$. The application of the method is greatly simplified by the following exchangeability result, which allows us to focus our attention on the final two paths $P_{\infty,r-1}$ and $P_{\infty,r}$

Lemma 2.3.11. *Write $\mathcal{U}_1 = \mathcal{U}_1(T_\infty(r))$. Then for all $i \in [r]$,*

$$\mathbf{P}\left\{P_{\infty,i} \in \mathcal{U}_1 \mid \pi_0\right\} = \mathbf{P}\left\{P_{\infty,r} \in \mathcal{U}_1 \mid \pi_0\right\},$$

and for all $i, j \in [r]$ with $i \neq j$,

$$\mathbf{P}\left\{ P_{\infty,i} \in \mathcal{U}_1, P_{\infty,j} \in \mathcal{U}_1 \mid \pi_0 \right\} = \mathbf{P}\left\{ P_{\infty,r-1} \in \mathcal{U}_1, P_{\infty,r} \in \mathcal{U}_1 \mid \pi_0 \right\} .$$

Moreover, the same identities hold with $\mathcal{U}_1$ replaced by $\mathcal{U}_3 = \mathcal{U}_3(T_\infty(r))$.

Proof. We work on the probability-one event that $|P_{\infty,i}| > 0$ for all $i \in [r]$. (Note that $P_{\infty,i} = P_i(T_\infty(r))$. Fix any permutation ϕ of the leaf labels $\{1, \ldots, r\}$, and let $T_\infty^\phi(r)$ be the tree obtained from $T_\infty(r)$ by permuting the labels $\{1, \ldots, r\}$ according to ϕ; note that labels U and W remain fixed. Any such permutation induces an automorphism of $T_\infty(r)$ and $T_\infty^\phi(r)$ as binary leaf-labeled trees with edge lengths.

We claim that $\mathcal{U}_1(T_\infty^\phi(r)) = \{\phi(i) : i \in \mathcal{U}_1(T_\infty(r))\}$. To see this, fix $i \in [r]$. If $i \in \mathcal{U}_1(T_\infty(r))$, then $P_{\infty,i} \cap P_{\infty,j} = \emptyset$ for all $j \in [r] \setminus i$, so the only point of intersection of $P_{\infty,i}$ with the rest of $T_\infty(r)$ lies on the path $P_{\infty,0} = P_0(T_\infty(r))$ from U to W. Writing $P_{\infty,i}^\phi$ for the image of $P_{\infty,i}$ in $T_\infty^\phi(r)$ under the automorphism induced by ϕ, the only point of intersection of $P_{\infty,i}^\phi$ with the rest of $T_\infty^\phi(r)$ must then lie on the path from U to W in $T_\infty^\phi(r)$, since the labels U and W are unchanged by ϕ. Thus, $P_{\infty,i}^\phi = P_{\phi(i)}(T_\infty^\phi(r))$, and so $P_{\phi(i)}(T_\infty^\phi(r)) \cap P_j(T_\infty^\phi(r)) = \emptyset$ for all $j \in [r] \setminus \{\phi(i)\}$. Since the lengths of $P_{\infty,i}$ and $P_{\infty,i}^\phi$, and their attachment points to the $U - W$ path, are the same in $T_\infty(r)$ and $T_\infty^\phi(r)$, it follows that $\phi(i) \in \mathcal{U}_1(T_\infty^\phi(r))$. A corresponding argument using ϕ^{-1} shows that if $i \in \mathcal{U}_1(T_\infty^\phi(r))$ then $\phi^{-1}(i) \in \mathcal{U}_1(T_\infty(r)$, which establishes the claim.

By Proposition 2.3.8, for any permutation $\phi : [r] \to [r]$, the trees $T_\infty^\phi(r)$ and $T_\infty(r)$ have the same law. Since $\mathcal{U}_1(T_\infty^\phi(r)) = \{\phi(i) : i \in \mathcal{U}_1(T_\infty(r))\}$, by taking ϕ to be a uniformly random permutation of $[r]$ it then follows that for all $i \in [r]$, and $0 \leq s \leq r$,

$$\mathbf{P}\left\{ i \in \mathcal{U}_1(T_\infty(r)) \mid |\mathcal{U}_1(T_\infty(r))| = s \right\} = \frac{s}{r}$$

and hence

$$\mathbf{P}\left\{ i \in \mathcal{U}_1(T_\infty(r)) \mid |\mathcal{U}_1(T_\infty(r))| = s \right\} = \mathbf{P}\left\{ r \in \mathcal{U}_1(T_\infty(r)) \mid |\mathcal{U}_1(T_\infty(r))| = s \right\} .$$

Similarly, for all $1 \leq i < j \leq r$,

$$\mathbf{P}\left\{ i, j \in \mathcal{U}_1(T_\infty(r)) \mid |\mathcal{U}_1(T_\infty(r))| = s \right\} = \frac{s(s-1)}{r(r-1)},$$

and hence

$$
\begin{aligned}
&\mathbf{P}\left\{i,j \in \mathcal{U}_1(T_\infty(r)) \mid |\mathcal{U}_1(T_\infty(r))| = s\right\} \\
&= \mathbf{P}\left\{r-1, r \in \mathcal{U}_1(T_\infty(r)) \mid |\mathcal{U}_1(T_\infty(r))| = s\right\} .
\end{aligned}
$$

The lemma now follows by averaging over $s = |\mathcal{U}_1(T_\infty(r))|$. □

Proof of Lemma 2.3.10. As mentioned earlier, the convergence in distribution from Proposition 2.3.8 implies that for any $r \geq 1$, as $q \to \infty$ we have $(\mathcal{U}_1(T'_q, r, q), \mathcal{U}_3(T'_q, r, q) \xrightarrow{\mathrm{d}} (\mathcal{U}_1(T_\infty(r)), \mathcal{U}_3(T_\infty(r)))$. To prove the lemma it thus suffices to show that for all $\epsilon > 0$ there exists r_0 such that for all $r \geq r_0$,

$$
\mathbf{P}\left\{|\mathcal{U}_1(T_\infty(r))| < \epsilon\sqrt{r}\right\} < 22\epsilon. \tag{2.10}
$$

(The same bound then holds for $\mathcal{U}_3(T_\infty(r))$ by symmetry.) The remainder of the proof is thus devoted to establishing (2.10). In what follows we write $\mathcal{U}_1 = \mathcal{U}_1(T_\infty(r))$.

By Lemma 2.3.11, we have

$$
\mathbf{E}\left\{\mathcal{U}_1 \mid \pi_0\right\} = r\mathbf{P}\left\{r \in \mathcal{U}_1 \mid \pi_0\right\} .
$$

On the probability-one event that $\alpha(1), \ldots, \alpha(r)$ are all distinct, $r \in \mathcal{U}_1$ if and only if $\pi_r \leq 1$, $\alpha(r) \in P_{\infty,0}$, and $\mathrm{dist}_{T_\infty(r)}(\alpha(r), U) \leq \pi_0/3$. Since $\alpha(r)$ is uniformly distributed over $T_\infty(r-1)$, which is the union of the paths $P_{\infty,0}, \ldots, P_{\infty,r-1}$, for $r > 1$ we thus have

$$
\begin{aligned}
\mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\} &= \mathbf{P}\left\{r \in \mathcal{U}_1 \mid \pi_0\right\} \\
&= \mathbf{E}\left\{\mathbf{P}\left\{r \in \mathcal{U}_1 \mid \pi_0, \ldots, \pi_r\right\} \mid \pi_0\right\} \\
&= \mathbf{E}\left\{\left.\frac{\pi_0/3}{|\pi_0| + \ldots + |\pi_{r-1}|}\mathbf{1}_{[|\pi_r| \leq 1]} \,\right|\, \pi_0\right\} \\
&= \mathbf{E}\left\{\left.\frac{1}{3}\frac{E_1^{1/2}}{(E_1 + \ldots + E_r)^{1/2}}\mathbf{1}_{[\pi_r \leq 1]} \,\right|\, E_1\right\} .
\end{aligned}
$$

For the first and last identities above, we used that conditioning on π_0 and on E_1 is equivalent, since $\pi_0 = \sqrt{2}E_1^{1/2}$.

Likewise, still on the event that $\alpha(1), \ldots, \alpha(r)$ are all distinct, provided that $r \geq 2$, the point $r-1$ belongs to $\mathcal{U}_1$ if and only if $\pi_{r-1} \leq 1$, $\alpha(r-1) \in P_{\infty,0}$, and $\mathrm{dist}_{T_\infty(r)}(\alpha(r-1), U) \leq$

$\pi_0/3$. A similar derivation then shows that for $r \geq 2$,

$$\begin{aligned}
&\mathbf{P}\left\{r-1 \in \mathcal{U}_1, r \in \mathcal{U}_1 \;\middle|\; E_1\right\} \\
&= \mathbf{E}\left\{ \frac{1}{9} \frac{E_1^{1/2}}{(E_1+\ldots+E_{r-1})^{1/2}} \frac{E_1^{1/2}}{(E_1+\ldots+E_r)^{1/2}} \mathbf{1}_{[\pi_{r-1},\pi_r \leq 1]} \;\middle|\; E_1 \right\}.
\end{aligned}$$

We let

$$R = \frac{1}{3} \frac{E_1^{1/2}}{(E_1+\cdots+E_r)^{1/2}}$$

and

$$R' = \frac{1}{9} \frac{E_1}{(E_1+\cdots+E_{r-1})^{1/2}(E_1+\cdots+E_r)^{1/2}},$$

so that the above identities may be written more succinctly as

$$\mathbf{P}\left\{r \in \mathcal{U}_1 \;\middle|\; E_1\right\} = \mathbf{E}\left\{ R\mathbf{1}_{[\pi_r \leq 1]} \;\middle|\; E_1\right\} \tag{2.11}$$

and

$$\mathbf{P}\left\{r-1 \in \mathcal{U}_1, r \in \mathcal{U}_1 \;\middle|\; E_1\right\} = \mathbf{E}\left\{ R'\mathbf{1}_{[\pi_{r-1},\pi_r \leq 1]} \;\middle|\; E_1\right\} \tag{2.12}$$

Now fix $\epsilon \in (0,1/2)$ small and let $A(\epsilon, r)$ be the event that

$$(1-\epsilon)r \leq E_2+\ldots+E_{r-1} \leq E_2+\ldots+E_r \leq (1+\epsilon)r$$

and that $E_r \leq r^{1/2}$ and $E_{r+1} \leq r^{1/2}$. On $A(\epsilon, r)$, using the bound $(a+x)^{1/2} - a^{1/2} \leq x/(2a^{1/2})$ for all $a, x > 0$, we have that

$$\begin{aligned}
\pi_r &= \sqrt{2}\left((E_1+\ldots+E_{r+1})^{1/2} - (E_1+\ldots+E_r)^{1/2}\right) \\
&\leq \sqrt{2}\left((E_2+\ldots+E_{r+1})^{1/2} - (E_2+\ldots+E_r)^{1/2}\right) \\
&\leq \sqrt{2}\frac{E_{r+1}}{2((1-\epsilon)r)^{1/2}} < 1,
\end{aligned}$$

and likewise $\pi_{r-1} < 1$ on $A(\epsilon, r)$.

For r large enough, $\mathbf{P}\left\{E_r \geq r^{1/2}\right\} = e^{-r^{1/2}} \leq 1/r^2$ and $\mathbf{P}\left\{E_{r+1} \geq r^{1/2}\right\} \leq 1/r^2$, and by Chebyshev's inequality,

$$\mathbf{P}\left\{E_2+\ldots+E_{r-1} \leq (1-\epsilon)r\right\} \leq \frac{2}{(\epsilon r)^2}$$

and

$$\mathbf{P}\left\{E_2 + \ldots + E_r \geq (1+\epsilon)r\right\} \leq \frac{2}{(\epsilon r)^2},$$

and hence

$$\mathbf{P}\left\{A(\epsilon, r)^c\right\} \leq \frac{2}{r^2} + \frac{4}{(\epsilon r)^2} \leq \frac{6}{(\epsilon r)^2}\,.$$

Moreover, $A(\epsilon, r)$ is independent of E_1, so $\mathbf{P}\left\{A(\epsilon, r) \mid E_1\right\} = \mathbf{P}\left\{A(\epsilon, r)\right\}$.

With the preceding bounds at hand, we have enough information to control (2.11) and (2.12). Writing $A = A(\epsilon, r)$, we have

$$\begin{aligned}
\mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\} &= \mathbf{E}\left\{ R\mathbf{1}_{[\pi_r \leq 1]} \,\middle|\, E_1\right\} \\
&\geq \mathbf{E}\left\{ R\mathbf{1}_{[\pi_r \leq 1]}\mathbf{1}_{[A]} \,\middle|\, E_1\right\} \\
&= \mathbf{E}\left\{ R\mathbf{1}_{[A]} \,\middle|\, E_1\right\} \\
&= \mathbf{E}\left\{R \mid A, E_1\right\}\mathbf{P}\left\{A \mid E_1\right\} \\
&\geq \frac{1}{3}\frac{E_1^{1/2}}{(E_1 + (1+\epsilon)r)^{1/2}}\left(1 - \frac{6}{(\epsilon r)^2}\right)\,, \qquad (2.13)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\} &\leq \mathbf{E}\left\{ R\mathbf{1}_{[\pi_r \leq 1]} \,\middle|\, A, E_1\right\} + \mathbf{P}\left\{A^c \mid E_1\right\} \\
&\leq \mathbf{E}\left\{R \mid A, E_1\right\} + \frac{6}{(\epsilon r)^2} \\
&\leq \frac{1}{3}\frac{E_1^{1/2}}{((1-\epsilon)r)^{1/2}} + \frac{6}{(\epsilon r)^2}\,. \qquad (2.14)
\end{aligned}$$

We similarly have

$$\begin{aligned}
\mathbf{P}\left\{r-1 \in \mathcal{U}_1, r \in \mathcal{U}_1 \mid E_1\right\} &= \mathbf{E}\left\{ R'\mathbf{1}_{[\pi_{r-1}, \pi_r \leq 1]} \,\middle|\, E_1\right\} \\
&\leq \mathbf{E}\left\{ R'\mathbf{1}_{[\pi_{r-1}, \pi_r \leq 1]} \,\middle|\, E_1, A\right\} + \mathbf{P}\left\{A^c \mid E_1\right\} \\
&\leq \mathbf{E}\left\{R' \mid A, E_1\right\} + \frac{6}{(\epsilon r)^2} \\
&\leq \frac{1}{9}\frac{E_1}{(1-\epsilon)r} + \frac{6}{(\epsilon r)^2}\,. \qquad (2.15)
\end{aligned}$$

We use the above bounds in the following formulas, which are immediate consequences of

Lemma 2.3.11 (using the fact that conditioning on π_0 and on E_1 is equivalent):

$$\mathbf{E}\left\{|\mathcal{U}_1| \mid E_1\right\} = r\mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\} \tag{2.16}$$

and

$$\begin{aligned}
&\mathbf{Var}\left\{|\mathcal{U}_1|| E_1\right\} \\
&= r(r-1)\left(\mathbf{P}\left\{r-1, r \in \mathcal{U}_1 \mid E_1\right\} - \mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\}^2\right) \\
&+ r\left(\mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\} - \mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\}^2\right) .
\end{aligned} \tag{2.17}$$

We now temporarily work on the event that $E_1 \in (\epsilon, \epsilon^{-1})$. On this event, there is $r_0 = r_0(\epsilon) > 0$ such that for all $r \geq r_0$,

$$\begin{aligned}
\frac{1}{3}\frac{E_1^{1/2}}{(E_1 + (1+\epsilon)r)^{1/2}}\left(1 - \frac{6}{(\epsilon r)^2}\right) &\geq \frac{1}{3}\frac{E_1^{1/2}}{((1+2\epsilon)r)^{1/2}} , \\
\frac{1}{3}\frac{E_1^{1/2}}{((1-\epsilon)r)^{1/2}} + \frac{6}{(\epsilon r)^2} &\leq \frac{1}{3}\frac{E_1^{1/2}}{((1-2\epsilon)r)^{1/2}} , \text{ and} \\
\frac{1}{9}\frac{E_1}{(1-\epsilon)r} + \frac{6}{(\epsilon r)^2} &\leq \frac{1}{9}\frac{E_1}{(1-2\epsilon)r} .
\end{aligned}$$

Using the first of these bounds together with (2.13) in (2.16) gives that on the event $\{E_1 \in (\epsilon, \epsilon^{-1})\}$,

$$\mathbf{E}\left\{|\mathcal{U}_1| \mid E_1\right\} \geq \frac{r^{1/2}}{3}\frac{E_1^{1/2}}{(1+2\epsilon)^{1/2}} , \tag{2.18}$$

and using all three bounds together with (2.13),(2.14) and (2.15) in (2.17), we obtain that on the event $\{E_1 \in (\epsilon, \epsilon^{-1})\}$,

$$\begin{aligned}
&\mathbf{Var}\left\{|\mathcal{U}_1|| E_1\right\} \\
&\leq r(r-1)\left(\frac{1}{9}\frac{E_1}{(1-2\epsilon)r} - \frac{1}{9}\frac{E_1}{(1+2\epsilon)r}\right) + r\left(\frac{1}{3}\frac{E_1^{1/2}}{((1-2\epsilon)r)^{1/2}}\right) \\
&= (r-1)\frac{E_1}{9}\frac{4\epsilon}{1-4\epsilon^2} + r^{1/2}\frac{E_1^{1/2}}{3}\frac{1}{(1-2\epsilon)^{1/2}} \\
&< 5\epsilon\mathbf{E}\left\{|\mathcal{U}_1| \mid E_1\right\}^2 ,
\end{aligned}$$

the final bound holding by (2.18) for ϵ sufficiently small ($\epsilon < 1/8$ is enough), and still provided that and $r \geq r_0$.

The lower bound in (2.18) is at least $(rE_1)^{1/2}/4$ provided ϵ is small enough, so it now follows by the conditional Chebyshev inequality that

$$\begin{aligned}
&\mathbf{P}\left\{ |\mathcal{U}_1| \le \frac{(rE_1)^{1/2}}{8}, E_1 \in (\epsilon, \epsilon^{-1}) \;\middle|\; E_1 \right\} \\
&\le \frac{\mathbf{Var}\left\{|\mathcal{U}_1|| \; E_1\right\} \mathbf{1}_{[E_1 \in (\epsilon,\epsilon^{-1})]}}{(\mathbf{E}\left\{ |\mathcal{U}_1| \mid E_1 \right\}/2)^2} \\
&\le 20\epsilon \mathbf{1}_{[E_1 \in (\epsilon,\epsilon^{-1})]} \,.
\end{aligned}$$

For ϵ small enough, if $E_1 \ge \epsilon$ then $E_1^{1/2}/8 > 2\epsilon$, so the preceding bound implies that, unconditionally,

$$\begin{aligned}
&\mathbf{P}\left\{|\mathcal{U}_1| \le \epsilon r^{1/2}\right\} \\
&= \mathbf{E}\left(\mathbf{P}\left\{ |\mathcal{U}_1| \le \epsilon r^{1/2} \;\middle|\; E_1 \right\}\right) \\
&\le \mathbf{E}\left(\mathbf{P}\left\{ |\mathcal{U}_1| \le \frac{(rE_1)^{1/2}}{8} \;\middle|\; E_1 \right\} \mathbf{1}_{[E_1 \in \epsilon, \epsilon^{-1}]}\right) + \mathbf{P}\left\{E_1 \notin (\epsilon, \epsilon^{-1})\right\} \\
&\le 22\epsilon \,,
\end{aligned}$$

the last bound holding since $\mathbf{P}\left\{E_1 \notin \epsilon, \epsilon^{-1}\right\} < 2\epsilon$ for ϵ small. This establishes (2.10) and completes the proof. $\square$

2.3.3.2 Proof of Proposition 2.3.3

Having already proved Proposition 2.3.4, to complete the proof of Proposition 2.3.3 it remains to handle the cases when $s > 0$. So fix $\epsilon > 0$ and integers $s > 0$ and $r > 0$, and let G_q and $F_q = F_q(r, \mathbf{e})$ be as in the statement of Proposition 2.3.3. Like in the case $s = 0$, it suffices to prove that if r is sufficiently large as a function of ϵ and s then for all q sufficiently large, if X and Y are independent, uniformly random elements of $[q]$, independent of G_q and of the ordering $\mathbf{e}$, then

$$\mathbf{P}\left\{ X \xleftrightarrow{F_q} Y \right\} < \epsilon \,.$$

To accomplish this, we decompose G_q into a collection of trees to which we can apply the result from the $s = 0$ case, Proposition 2.3.4. We next turn to defining the necessary decomposition. The definitions of the next four paragraphs are illustrated in Figure 2.3.

Let $\mathrm{core}(G_q)$ be the maximum induced subgraph of G_q with minimum degree 2; equivalently, this is the subgraph of G_q induced by the set of vertices which lie on cycles of G_q.

For $v \in [q]$ let $c(v)$ be the (unique) closest vertex of $\mathrm{core}(G_q)$ to v in G_q. In particular, if $v \in v(\mathrm{core}(G_q))$ then $c(v) = v$.

If $s \geq 2$ then the *kernel* of G_q, denoted $K(G_q)$, is the multigraph obtained from $\mathrm{core}(G_q)$ by contracting each path whose endpoints have degree at least three in $\mathrm{core}(G_q)$ and whose internal vertices have degree two in $\mathrm{core}(G_q)$ into a single edge. For each vertex v of G_q, we define its "attachment location on $K(G_q)$", denoted $\kappa(v)$, as follows. For each edge e of $K(G_q)$, if $c(v)$ is an internal vertex of the path which was contracted to make e, then set $\kappa(v) = e$. Otherwise, if $c(v)$ is a vertex w of $K(G_q)$ then set $\kappa(v) = w$.

If $s = 1$ then $\mathrm{core}(G_q)$ is a cycle. It is still useful for us to define the kernel in this case, but the definition is slightly different (and slightly non-standard). To define it, we first augment the core by adding all vertices of the path from q to $c(q)$; we write $\mathrm{core}^+(G_q)$ for the subgraph of G_q induced by this path together with $\mathrm{core}(G_q)$. We then define the kernel $K(G_q)$ to be the multigraph obtained from G_q by contracting each maximal path or cycle of $\mathrm{core}^+(G_q)$ whose endpoints lie in $\{q \cup c(q)\}$ to form a single edge. If $q \neq c(q)$ then this creates a "lollipop" consisting of a loop edge at $c(q)$ and a single edge from $c(q)$ to q; if $q = c(q)$ then the result is simply a loop edge at $c(q)$.

Provided that $s \geq 1$, so that the kernel is defined, for $a \in v(K(G_q)) \cup e(K(G_q))$ we now set $V_q(a) = \{v \in [q] : \kappa(v) = a\}$. Then the set

$$\mathbf{V}_q = \{V_q(a), a \in v(K(G_q)) \cup e(K(G_q))\} \tag{2.19}$$

is a partition of $v(G_q) = [q]$. For each $a \in v(K(G_q)) \cup e(K(G_q))$, we let $T_q(a)$ be the subgraph of G_q spanned by $V_q(a)$. Also, for $e = xy \in e(K(G_q))$, we write $Z(e, x)$ (resp. $Z(e, y)$) for the unique vertex of $T_q(e)$ incident to x (resp. to y).

By the definition of the core, $T_q(a)$ is necessarily a tree. By the symmetry of the model, conditionally given the partition $\mathbf{V}_q$ in (2.19), the trees $(T_q(a), a \in v(K(G_q)) \cup e(K(G_q)))$ are independent and each is a uniformly random tree on its vertex set. Moreover, also by symmetry, for each $e \in e(K(G_q))$, conditionally given both $\mathbf{V}_q$ and the tree $T_q(e)$, the vertices $Z(e, u)$ and $Z(e, v)$ are independent uniformly random elements of $V_q(e)$.

The next proposition describes the asymptotic structure of the partition of mass in (2.19). For each positive integer k, let

$$\Delta_k = \{(x_1, \ldots, x_k) \in (0, 1)^k : x_1 + \ldots + x_k = 1\}$$

denote the $(k-1)$-dimensional simplex. Then for $(\alpha_1, \ldots, \alpha_k) \in \Delta_k$, the Dirichlet$(\alpha_1, \ldots, \alpha_k)$

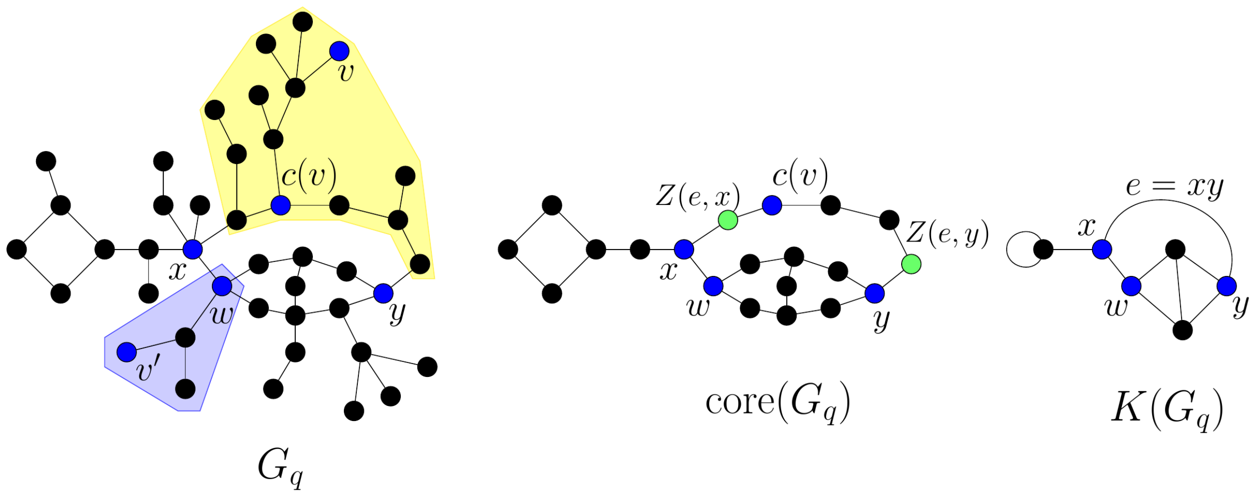

Figure 2.3: Left: An instantiation of graph G_q; here $q = 43$ and $s = 3$. Center: the graph core(G_q). Right: the kernel $K(G_q)$. In the graph G_q, the vertex v has $\kappa(v) = e = xy$ since $c(v)$ lies on the path of core(G_q) which is contracted to form e. The vertex v' has $\kappa(v') = w$ since $c(v) = w$ is a vertex of $K(G_q)$. The trees $T_q(e)$ and $T_q(w)$ are highlighted in yellow and in blue, respectively. In the center, the vertices $Z(e,x)$ and $Z(e,y)$ are green.

distribution on Δ_k has density

$$\frac{\Gamma(\alpha_1 + \ldots + \alpha_k)}{\Gamma(\alpha_1) \cdot \ldots \cdot \Gamma(\alpha_k)} \prod_{j=1}^{k} x_j^{\alpha_j - 1}$$

with respect to $(k-1)$-dimensional Lebesgue measure on Δ_k.

Proposition 2.3.12 ([8] Theorem 22, [9] Theorem 6 (c))**.** *Fix $s \geq 1$ and let G_q be uniformly distributed over the set of connected graphs with vertex set $[q]$ and surplus s. Then as $q \to \infty$, the vector*

$$(q^{-1} V_q(e), e \in e(K(G_q)))$$

converges in distribution to a Dirichlet$(1/2, \ldots, 1/2)$ random vector of length $k = 2s - 1 + \mathbf{1}_{[s=1]}$.

In the vector in Proposition 2.3.12 we may take the edges of $K(G_q)$ to be ordered lexicographically, say, but the precise ordering rule does not play an important role in this paper.

Recall that X and Y are independent, uniformly random elements of $[q]$, independent of

G_q and of the ordering $\mathbf{e}$. Then

$$\begin{aligned}
\mathbf{P}\left\{X \xleftrightarrow{F_q} Y\right\} &= \mathbf{E}\left(\mathbf{P}\left\{E_q \mid F_q\right\}\right) \\
&\geq \mathbf{E}\left(\frac{\max(|C| : C \text{ is a component of } F_q)^2}{q^2}\right) \\
&\geq \frac{\left(\mathbf{E}\left(\max(|C| : C \text{ is a component of } F_q)\right)\right)^2}{q^2},
\end{aligned}$$

so to accomplish our goal it suffices to show that $\mathbf{P}\left\{X \xleftrightarrow{F_q} Y\right\} \leq \epsilon^2$ if r is large enough. Since $\epsilon > 0$ was arbitrary, we may as well just show that $\mathbf{P}\left\{X \xleftrightarrow{F_q} Y\right\} < \epsilon$ for r large.

Let A be the event that $\kappa(X) \in e(K(G_q))$ and $\kappa(Y) \in e(K(G_q))$, let $B = A \cap \{\kappa(X) = \kappa(Y)\}$ and let $C = A \cap \{\kappa(X) \neq \kappa(Y)\}$. By Proposition 2.3.12, $q^{-1}\sum_{e \in e(K(G_q))} V_q(e) \to 1$ in probability, which implies that $\mathbf{P}\{A\} \to 1$ as $q \to \infty$. By the same proposition, the limits

$$\lim_{q\to\infty} \mathbf{P}\{B\} = p = 1 - \lim_{q\to\infty} \mathbf{P}\{C\}$$

both exist, and the value p lies strictly between 0 and 1.

Now let $\delta > 0$ be small enough that for q large,

$$\mathbf{P}\left\{\min(|V_q(e)|, e \in e(K(G_q))) < \delta q\right\} < \min(p, 1-p)\epsilon/7\,;$$

such a value δ exists by Proposition 2.3.12. Then by Bayes' formula, for q sufficiently large,

$$\mathbf{P}\left\{\min(|V_q(e)|, e \in e(K(G_q))) < \delta q \mid B\right\} < \epsilon/6 \tag{2.20}$$

and

$$\mathbf{P}\left\{\min(|V_q(e)|, e \in e(K(G_q))) < \delta q \mid C\right\} < \epsilon/6. \tag{2.21}$$

If $\kappa(X) = e = uv$, then any path from X to Y in F_q must either lie within $T_q(e)$ or else must pass through one of $Z(e,u)$ or $Z(e,v)$. It follows that

$$\begin{aligned}
&\mathbf{P}\left\{X \stackrel{F_q}{\longleftrightarrow} Y \mid \mathbf{V}_q, B\right\} \\
&\leq \mathbf{P}\left\{X \stackrel{T_q(e)\cap F_q}{\longleftrightarrow} Y \mid \mathbf{V}_q, B\right\} + \mathbf{P}\left\{X \stackrel{T_q(e)\cap F_q}{\longleftrightarrow} Z(e,u) \mid \mathbf{V}_q, B\right\} \\
&+ \mathbf{P}\left\{X \stackrel{T_q(e)\cap F_q}{\longleftrightarrow} Z(e,v) \mid \mathbf{V}_q, B\right\} \\
&= 3\mathbf{P}\left\{X \stackrel{T_q(e)\cap F_q}{\longleftrightarrow} Y \mid \mathbf{V}_q, B\right\},
\end{aligned}$$

where for the final equality we have used that where we have used that conditionally given B and $\mathbf{V}_q$, the vertices $Z(e,u)$ and $Z(e,v)$ and Y are all uniformly random elements of $V_q(e)$ independent of X, and where we write $T_q(e) \cap F_q$ to mean the subgraph of $T_q(e)$ with edge set $e(T_q(e)) \cap e(F_q)$.

Now note that the path-and-cycle-breaking process on F_q, when restricted to $T_q(e)$, removes a superset of the edges that would be removed by running the path-breaking process on $T_q(e)$ with the induced edge ordering. (This holds since removing an edge e' of $T_q(e)$ may separate a pair of elements of $[r]$ one or both of which lie outside of $V_q(e)$; in this case, the edge e' is removed in the path-and-cycle-breaking process. However, e' may not be removed in the path-breaking process, if e' does not separate a pair of elements of $[r] \cap V_q(e)$.) In other words, writing $F_q(e)'$ for the forest obtained by running the path-breaking process on $T_q(e)$ with starting set $[r] \cap V_q(e)$ and edge ordering given by the restriction of $\mathbf{e}$ to $e(T_q(e))$, then $T_q(e) \cap F_q$ is a sub-forest of $F_q(e)'$. It follows that, writing $e = \kappa(X)$, which is also equal to $\kappa(Y)$ when B occurs, we have

$$\begin{aligned}
\mathbf{P}\left\{X \stackrel{F_q}{\longleftrightarrow} Y \mid \mathbf{V}_q, B\right\} &\leq 3\mathbf{P}\left\{X \stackrel{T_q(e)\cap F_q}{\longleftrightarrow} Y \mid \mathbf{V}_q, B\right\} \\
&\leq 3\mathbf{P}\left\{X \stackrel{F_q(e)'}{\longleftrightarrow} Y \mid \mathbf{V}_q, B\right\}. \qquad (2.22)
\end{aligned}$$

Now note that if G is a fixed graph with vertex set $[q]$ whose largest connected component has c vertices, and X and Y are independent uniformly random elements of $[q]$, then $\mathbf{P}\left\{X \stackrel{G}{\longleftrightarrow} Y\right\} \leq \frac{c}{q}$. If G is instead random, then this bound and the tower law give that

$$\mathbf{P}\left\{X \stackrel{G}{\longleftrightarrow} Y\right\} \leq q^{-1}\mathbf{E}\left(\max(|C| : C \text{ is a component of } F_q)\right).$$

It thus follows from Proposition 2.3.4 there is q_0 such that if $|V_q(e)| \geq q_0$ then the conditional probability on the right of (2.22) is less than $\epsilon/12$, so we have

$$\mathbf{P}\left\{X \stackrel{F_q}{\longleftrightarrow} Y \mid \mathbf{V}_q, B\right\} < 3(\epsilon/12)\mathbf{1}_{[V_q(\kappa(X))\geq q_0]} + 3\mathbf{1}_{[V_q(\kappa(X))<q_0]},$$

which together with (2.20) yields that for q large enough (and in particular large enough that $\delta q > q_0$),

$$\begin{aligned}\mathbf{P}\left\{X \stackrel{F_q}{\longleftrightarrow} Y \mid B\right\} &= \mathbf{E}\left(\mathbf{P}\left\{X \stackrel{F_q}{\longleftrightarrow} Y \mid \mathbf{V}_q, B\right\} \mid B\right) \\ &(\epsilon/4)\mathbf{P}\left\{V_q(\kappa(X)) \geq q_0 \mid B\right\} + 3\mathbf{P}\left\{V_q(\kappa(X)) < q_0 \mid B\right\} \\ &< 3\epsilon/4\,.\end{aligned}$$

A nearly identical proof, but using (2.21) in place of (2.20), shows that $\mathbf{P}\left\{X \stackrel{F_q}{\longleftrightarrow} Y \mid C\right\}$ $< 3\epsilon/4$ for all q sufficiently large. (In fact, in this case we could obtain a slightly better bound, since when C occurs, in order for X and Y to be connected in F_q there must be a path from X to $Z(e,u)$ or $Z(e,v)$ in $T_q(e) \cap F_q$; the term $X \stackrel{T_q(e)\cap F_q}{\longleftrightarrow} Y$ does not appear.) Since if A occurs then either B or C must occur, it follows that

$$\begin{aligned}\mathbf{P}\left\{X \stackrel{F_q}{\longleftrightarrow} Y\right\} &\leq \mathbf{P}\left\{X \stackrel{F_q}{\longleftrightarrow} Y, B\right\} + \mathbf{P}\left\{X \stackrel{F_q}{\longleftrightarrow} Y, C\right\} + \mathbf{P}\left\{A^c\right\} \\ &\leq \frac{3\epsilon}{4} + \mathbf{P}\left\{A\right\} < \epsilon\end{aligned}$$

the last two inequalities holding for all q sufficiently large. This completes the proof of Proposition 2.3.3 in the case $s > 0$.

2.4 Conclusion

In addition to the conjectures raised directly after the statement of Theorem 2.1.1, there are numerous avenues for future research suggested by the current work.

First, we expect that a version of the dichotomy established in Theorem 2.1.1 should hold for other high-dimensional random graphs, at least those with sufficient symmetry. For example, we expect that the same theorem should hold if K_n is replaced by a uniformly random d-regular graph (for $d \geq 3$), or by the nearest-neighbour hypercube $\{0,1\}^N$ with $2^N \asymp n$. A version of the theorem may well also hold in high-dimensional lattice tori (i.e. with K_n replaced by $(\mathbb{Z}/m\mathbb{Z})^d$, where $m^d \asymp n$, with d fixed and large). However, in Euclidean settings there is less symmetry; nearby sources are in more direct competition than far-off sources, and it is not clear to us how substantially this will affect the behaviour of the multi-source invasion process.

The behaviour in low-dimensional settings is of course also interesting. It's possible that enough is known about two-dimensional critical percolation (at least on the triangular lattice [41]) to be able to make some progress on the structure of multi-source invasion percolation.

Our results suggest the following behaviour for multi-source invasion percolation on large conditioned critical Bienaymé trees[1] with finite variance offspring distribution. For such trees, invasion percolation from boundedly many sources (i.e. with $k(n) = k$ fixed) will result in all components having macroscopic sizes which are random to first order; on the other hand, invasion percolation from unboundedly many sources (i.e. with $k(n) \to \infty$) will with high probability result in all components of sublinear size. This can likely be proved in detail using weak convergence arguments similar to those used to study the "Markov chainsaw" in [11]. In both cases, it would be would be of interest to understand the distribution of component sizes; in the case of unboundedly many sources, the precise behaviour of the size of the largest connected component is unclear to us, and may depend more sensitively on the offspring distribution, at least if $k(n) \to \infty$ sufficiently quickly.

It is less clear to us what should happen for conditioned critical Bienaymé trees with infinite variance (e.g. stable trees). In this setting, the presence of hubs – nodes with very large degree - could play an important role in the dynamics of the invasion process.

For other models of random trees and networks (e.g. preferential attachment networks, inhomogeneous random graphs, or networks with community structure, or any sort of directed models), the subject is wide open.

[1]We follow the terminological suggestion of [7], using the term "Bienaymé trees" rather than "Galton-Watson trees" for the family trees of branching processes.

Chapter 3

Random Tree-Weighted Graphs

3.1 Introduction

By a *rooted tree* we mean a labeled tree $t = (\mathrm{v}(t), \mathrm{e}(t))$, with a distinguished root node denoted $r(t)$. A *tree-rooted graph* is a pair (g, t, γ) where $g = (\mathrm{v}(g), \mathrm{e}(g))$ is a labeled graph, $t = (\mathrm{v}(t), \mathrm{e}(t))$ is a spanning tree of g, and $\gamma = uv$ is a distinguished oriented edge with $\{u, v\} \in \mathrm{e}(g) \setminus \mathrm{e}(t)$. We view t as a rooted tree by setting $r(t) = u$.

Throughout this work, we allow our graphs to have multiple edges and loops; in tree-rooted graphs, the root edge is allowed to be a loop. We say (g, t, γ) is *simple* if g is simple, i.e., if g contains no multiple edges or loops.

For a node u of a rooted tree t, we write $c_t(u)$ for the number of children of u in t. Given a rooted tree t with $\mathrm{v}(t) = [n] := \{1, 2, \ldots, n\}$, the *child sequence* of t is the sequence $\mathrm{c}_t = (c_t(i), 1 \leq i \leq n)$. Similarly, given a tree-rooted graph (g, t, γ) with vertex set $\mathrm{v}(g) = [n]$, the *degree sequence* of (g, t, γ) is the sequence $(d_g(i), 1 \leq i \leq n)$, where $d_g(i)$ is the number of endpoints of edges incident to i in g; here loops are counted twice.

For any sequence $\mathrm{d} = (d(1), \ldots, d(n))$ of non-negative integers, we define the *degree distribution* $p_{\mathrm{d}} = (p_{\mathrm{d}}(k), k \geq 1)$ of d by letting $p_{\mathrm{d}}(k) = \#\{i \in [n] : d(i) = k\}/n$.

The following theorem contains our main result, which is an invariance principle for the spanning trees in random tree-rooted graphs with a fixed degree sequence. To state it, two further pieces of notation are needed. Given a finite graph $g = (v, e)$ and a constant $c > 0$, we write cg for the measured metric space (v, dist, π) whose points are the elements of v, with $\mathrm{dist}(x, y) := c \cdot \mathrm{dist}_g(x, y)$, where $\mathrm{dist}_g(x, y)$ denotes graph distance in g, and with π the uniform probability measure on v. Also, for a sequence $p = (p(k), k \geq 1)$ of real numbers, we write $\mu_1(p) := \sum_{k \geq 1} k p(k)$ and $\mu_2(p) := \sum_{k \geq 1} k^2 p(k)$.

Theorem 3.1.1. *For each $n \geq 1$ let $\mathrm{d}^n = (d^n(i), 1 \leq i \leq n)$ be a degree sequence with $\min_{1\leq i\leq n} d^n(i) \geq 1$, with $\sum_{i\in[n]} d^n(i) \geq 2n$ and with $\sum_{i\in[n]} d^n(i)$ even. Let p^n be the degree distribution of d^n. Suppose that there exists a probability distribution $p = (p(k), k \geq 1)$ such that (a) $p^n \to p$ pointwise and $p(2) < 1$, and (b) $\mu_2(p^n) \to \mu_2(p) \in (0, \infty)$. Then there exists $\sigma = \sigma(p) \in (0, \infty)$ such that the following holds.*

For $n \geq 1$ let (G_n, T_n, Γ_n) be chosen uniformly at random among all simple tree-rooted graphs with vertex set $[n]$ and degree sequence d^n. Then

$$\frac{\sigma}{n^{1/2}} T_n \xrightarrow{\mathrm{d}} \mathcal{T}$$

as $n \to \infty$ with respect to the Gromov-Hausdorff-Prokhorov topology, where $\mathcal{T}$ is the Brownian continuum random tree.

We refer the reader to [1] for a good discussion of the Gromov-Hausdorff-Prokhorov topology aimed at probabilists. The technical insight underlying the proof of Theorem 3.1.1 is the fact that Pitman's additive coalescent [64] can be modified to yield a simple construction procedure for random tree-weighted graphs with a given degree sequence. We anticipate that this procedure has further interesting features to be explored.

3.1.1 Related work

The enumerative combinatorics of tree-rooted maps was developed in the 1960's and 1970's [60, 69]. The area has seen renewed attention over the last decade or so [25, 26, 32]. Random tree-rooted maps can be interpreted as samples from a Fortuin-Kastelyn model at zero temperature, and are an active object of study in the planar probability community (see, e.g.,[24, 45–48, 51, 58]).

There has also been some work on the typical number of spanning trees in uniformly random graphs [43, 44, 56] with given degree sequences. (In such models, the underlying graph is sampled uniformly at random from some set of allowed graphs; in our model, it is the tree-weighted graph which is uniformly random, which means the underlying measure on graphs is biased in favour of graphs with a greater number of spanning trees.)

Except in the setting of graphs on surfaces, we have not found any previous work on tree-weighted graphs, random or otherwise.

3.1.2 Overview of the proof

We begin with a small number of facts and definitions that are required for the overview. We say a sequence $\mathrm{c} = (c(i), 1 \leq i \leq n)$ of non-negative integers is a child sequence if it is the child sequence of some tree. Note that c is a child sequence if and only if $\sum_{1\leq i\leq n} c(i) = n-1$, in which case

$$\#\{\text{rooted trees } t : \mathrm{c}_t = \mathrm{c}\} = \binom{n-1}{c(1), \ldots, c(n)} = \frac{1}{n}\frac{n!}{\prod_{i=1}^n c(i)!}\,; \tag{3.1}$$

see [59], Section 3.3.

Given any sequence $\mathrm{c} = (c(i), 1 \leq i \leq n)$ of non-negative integers, for $k \geq 1$ we write $Q_\mathrm{c}(k) = \#\{i \in [n] : c(i) = k\}$. We call $Q_\mathrm{c} = (Q_\mathrm{c}(k), k \geq 0)$ the *child statistics vector* of c. For a tree t with child sequence c_t, we will sometimes write $Q_t = Q_{\mathrm{c}_t}$ for succinctness.

Given a graph g, for an edge $e \in \mathrm{e}(g)$ we write $m_g(e)$ for the multiplicity of edge e in g. Given a degree sequence $\mathrm{d} = (d(1), \ldots, d(n))$, the classical *configuration model* [68, Chapter 7] produces a random graph G such that for any fixed graph g with degree sequence d,

$$\mathbf{P}\left\{G = g\right\} \propto \frac{1}{\prod_{i=1}^n 2^{m_g(ii)} \prod_{e\in\mathrm{e}(g)} m_g(e)!}\,. \tag{3.2}$$

In Section 3.2, we define a sampling procedure, inspired by the configuration model and by Pitman's additive coalescent [64], which produces a random tree-weighted graph (G, T, Γ) with the property that for any fixed tree-weighted graph (g, t, γ) with degree sequence d,

$$\mathbf{P}\left\{(G, T, \Gamma) = (g, t, \gamma)\right\} \propto \frac{2^{\mathbf{1}_{[\gamma \text{ is a loop}]}} \cdot m_{g-t}(\gamma)}{\prod_{i=1}^n 2^{m_{g-t}(ii)} \cdot \prod_{e\in\mathrm{e}(g)} m_{g-t}(e)!}\,, \tag{3.3}$$

where $g - t$ is the graph with the same vertex set as g and with edge multiplicities given by

$$m_{g-t}(e) = \begin{cases} m_g(e) & \text{if } e \notin \mathrm{e}(t) \\ m_g(e) - 1 & \text{if } e \in \mathrm{e}(t)\,. \end{cases}$$

We call a random tree-weighted graph (G, T, Γ) with distribution given by (3.3) a *random tree-weighted graph with degree sequence* d. Note that in this case, conditionally given that G is simple, (G, T, Γ) is uniformly distributed over simple tree-rooted graphs with degree sequence d.

The sampling procedure we use has enough exchangeability that, conditional on its child sequence, the resulting spanning tree T is uniformly distributed; that is, for any fixed child sequence $\mathrm{c} = (c(i), 1 \le i \le n)$, and any tree t with $\mathrm{c}_t = \mathrm{c}$,

$$\mathbf{P}\{T = t \mid \mathrm{c}_T = \mathrm{c}\} = \binom{n-1}{c(1), \ldots, c(n)}^{-1}.$$

Now let $(\mathrm{d}^n, n \ge 1)$ be a sequence of degree sequences satisfying the conditions of Theorem 3.1.1; for each n let $(G(\mathrm{d}^n), T(\mathrm{d}^n), \Gamma(\mathrm{d}^n))$ be a random tree-weighted graph with degree sequence d^n. We prove (see Proposition 3.3.1) that there is a probability distribution $q = (q(i), i \ge 0)$ with $\mu_2(q) < \infty$ such that the child statistics vector $Q_{T(\mathrm{d}^n)}$ satisfies that $n^{-1}Q_{T(\mathrm{d}^n)}(a) \to q(a)$ in probability for all $a \ge 0$, and moreover that $\mu_2(n^{-1}Q_{T(\mathrm{d}^n)}) \to \mu_2(q)$ in probability. It then follows from a result of Broutin and Marckert [33] that

$$\frac{\sigma}{n^{1/2}} T(\mathrm{d}^n) \xrightarrow{\mathrm{d}} \mathcal{T} \tag{3.4}$$

in the Gromov-Hausdorff-Prokhorov sense, where $\sigma^2 = \mu_2(q) - 1$ and $\mathcal{T}$ is the Brownian continuum random tree.

This is not quite the convergence claimed in Theorem 3.1.1, because $(G(\mathrm{d}^n), T(\mathrm{d}^n), \Gamma(\mathrm{d}^n))$ is a random tree-weighted graph with degree sequence d^n, whereas Theorem 3.1.1 concerns random *simple* tree-weighted graphs. To obtain Theorem 3.1.1 from (3.4), we show that there is $\alpha \in (0, 1]$ such that as $n \to \infty$,

$$\mathbf{P}\{G(\mathrm{d}^n) \text{ is simple } \mid T(\mathrm{d}^n)\} \to \alpha \tag{3.5}$$

in probability. The value of (3.5), informally, is that it implies that conditioning $G(\mathrm{d}^n)$ to be simple has an asymptotically negligible effect on the law of $T(\mathrm{d}^n)$. Since the law of $(G(\mathrm{d}^n), T(\mathrm{d}^n), \Gamma(\mathrm{d}^n))$, conditional on the simplicity of $G(\mathrm{d}^n)$, is uniform over simple tree-weighted graphs with degree sequence d^n, we can reach our conclusion in a straightforward manner.

We finish the overview with a brief discussion of how we prove (3.5). Our procedure for constructing random tree-weighted graphs with a given degree sequence first constructs the tree $T(\mathrm{d}^n)$, then randomly pairs the remaining half-edges as in the standard configuration model. Viewing $T(\mathrm{d}^n)$ as fixed, this leads us to the following more general question. Let $G = (V, E)$ be a random graph with a given degree sequence generated according to the configuration model, and let $T = (V, E')$ be a fixed, simple graph with the same vertex set. What is the probability that the union of G and T forms a simple graph (i.e. that

G is a simple graph and that E and E' are disjoint)? We provide a partial answer to this question by proving a fairly general Poisson approximation theorem for the number of loops and multiple edges in the superposition of a fixed graph and a random graph drawn from the configuration model (see Theorem 3.4.1). In order to apply Theorem 3.4.1, we need that the joint degree statistics in $G(\mathrm{d}^n)$ and in $T(\mathrm{d}^n)$) are sufficiently well-behaved; proving this is the task of Section 3.3. We then state and prove Theorem 3.4.1 in Section 3.4, and finally put all the pieces together to prove Theorem 3.1.1 in Section 3.5.

3.2 Pitman's additive coalescent with a fixed degree sequence

3.2.1 The sampling process

Let $\mathrm{d} = (d(1), \dots, d(n))$ be a degree sequence, which is to say that $d(1), \dots, d(n)$ are non-negative integers. To be well-defined, the next process requires that $\sum_{1 \le i \le n} d(i) \ge 2n - 1$ and that $d(i) \ge 1$ for all $1 \le i \le n$.

Pitman's additive coalescent. The process has $n - 1$ steps, and at the start of step k consists of a rooted forest $F_k(\mathrm{d}) = \{T_1^k(\mathrm{d}), \dots, T_{n+1-k}^k(\mathrm{d})\}$ with $n + 1 - k$ trees. At the start of step 1, these trees are isolated vertices with labels $1, \dots, n$. Vertex i has $d(i)$ half-edges $(i1, i2, \dots, id(i))$ attached to it, and $id(i)$ is distinguished as the *root half-edge*.
Step k:

Choose a uniformly random pair (r_k, s_k), where r_k is a root half-edge which is not paired in $F_k(\mathrm{d})$ and s_k is a non-root half-edge which is not paired in $F_k(\mathrm{d})$ and additionally belongs to a different tree of $F_k(\mathrm{d})$ from r_k.

Pair the half-edges r_k and s_k to create an edge e_k connecting their endpoints; this merges two trees of $F_k(\mathrm{d})$. The root of the new tree is the same as the root of the tree of $F_k(\mathrm{d})$ containing s_k. In the new tree, the vertex incident to r_k is the child of the vertex incident to s_k.

Define $F_{k+1}(\mathrm{d})$ to be the forest consisting of the new tree thus created, together with the remaining $n - k - 1$ unaltered trees of $F_k(\mathrm{d})$.

An example is shown in Figure 3.1. Write $T(\mathrm{d}) = T_1^n(\mathrm{d})$ for the single tree in the random forest $F_n(\mathrm{d})$. Attached to the tree $T(\mathrm{d})$ there is a single pendant (unpaired) root half-edge

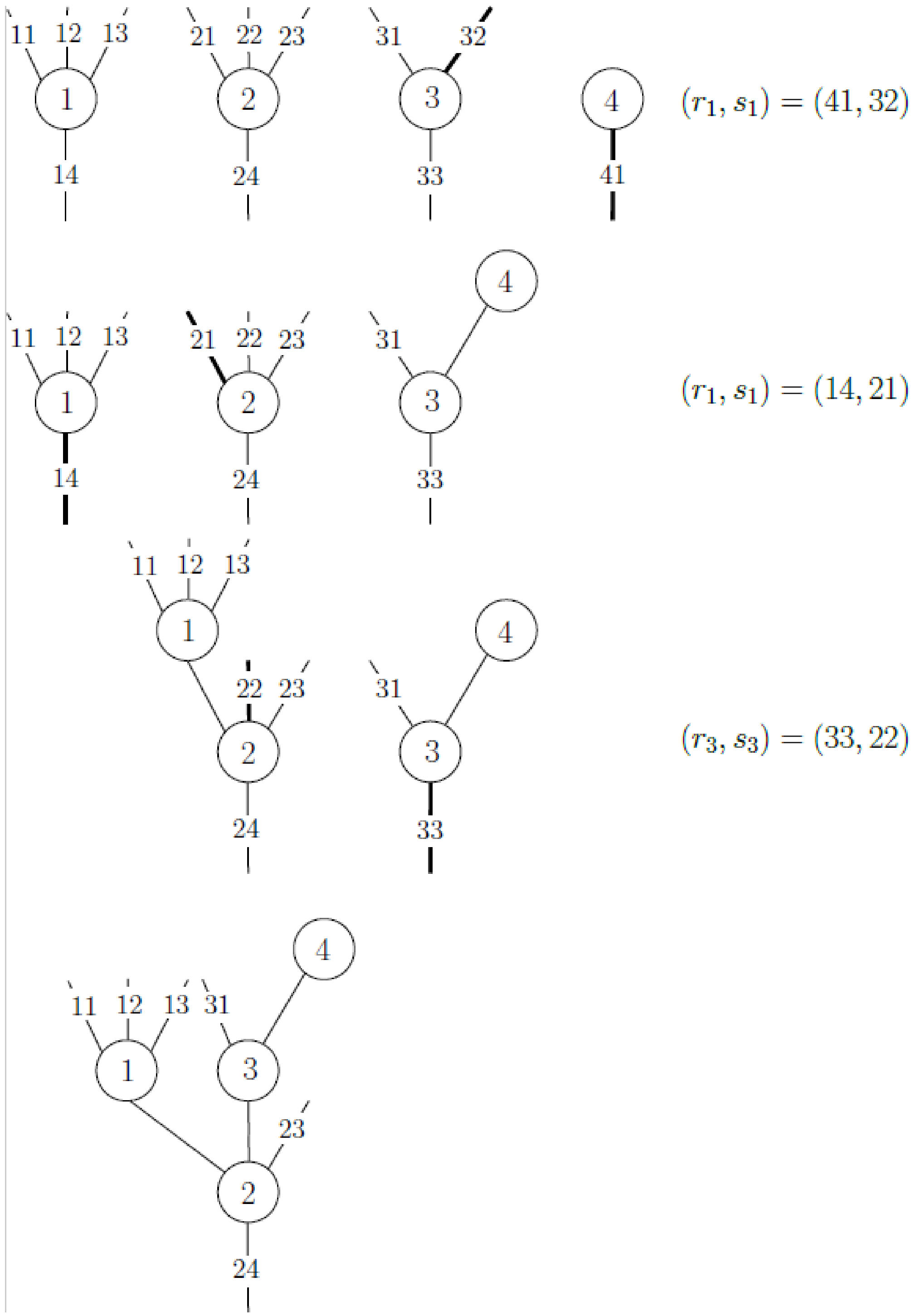

Figure 3.1: An example of an execution path of Pitman's additive coalescent. The forests F_1, F_2, F_3 and F_4 are displayed in successive rows.

which is incident to the root of $T(\mathrm{d})$, and if $\sum_{i=1}^n d(i) > 2n-1$ then there are also other pendant half-edges. By ignoring pendant half-edges, we may view $T(\mathrm{d})$ as a random rooted tree with vertex set $[n]$.

It will be useful to additionally define two edge labellings of $T(\mathrm{d})$, denoted K and H. We define $\mathrm{K}(e)$ to be the step at which edge e was added; so $\mathrm{K}(e_k) := k$. We define $\mathrm{H}(e)$ to be the non-root half-edge used in creating e; so $\mathrm{H}(e_k) = s_k$. Note that K is a bijection between $\mathrm{e}(T(\mathrm{d}))$ and $[n-1]$. Also, if $i \in [n]$ has $c_{T(\mathrm{d})}(i) = c$, then H assigns c distinct half-edges from the set $\{i1, \ldots, i(d(i)-1)\}$ to the edges between i and its children in $T(\mathrm{d})$.

We use the phrase "execution path" to mean a sequence of pairs $(r_1, s_1), \ldots, (r_{n-1}, s_{n-1})$ which may concievably appear as the ordered sequence of pairs of half-edges added during the course of Pitman's coalescent.

The next proposition fully describes the joint distribution of $T(\mathrm{d})$, K, and H. In its proof, and in what follows, for a rooted tree t and a node $u \in \mathrm{v}(t) \setminus \{r(t)\}$ we write $\mathrm{par}(u)$ for the parent of u in t. Also, we use the falling factorial notation $(k)_\ell := k(k-1)\cdot \ldots \cdot (k-\ell+1) = k!/(k-\ell)!$.

Proposition 3.2.1. *Let $\mathrm{d} = (d(1), \ldots, d(n))$ be a degree sequence with $d(i) \geq 1$ for all $i \in [n]$ and with $\sum_{i=1}^n d(i) \geq 2(n-1)$, and write $m = \frac{1}{2}\sum_{i=1}^n d(i)$. (Note: we allow that $\sum_{i=1}^n d(i)$ is odd.) Then the following properties all hold.*

1. *For any fixed rooted tree t with vertex set $[n]$,*

$$\mathbf{P}\{T(\mathrm{d}) = t\} = \frac{1}{(2m-n)_{n-1}} \prod_{i=1}^{n} (d(i)-1)_{c_t(i)} \,.$$

2. *Fix any set $\mathcal{H} \subset \bigcup_{i=1}^n \{i1, \ldots, i(d(i)-1)\}$ with $|\mathcal{H}| = n-1$. Conditionally given that $\{s_1, \ldots, s_{n-1}\} = \mathcal{H}$, the triple $(T(\mathrm{d}), \mathrm{K}, \mathrm{H})$ is uniformly distributed over the $((n-1)!)^2$ triples which are consistent with the event $\{s_1, \ldots, s_{n-1}\} = \mathcal{H}$.*

3. *The sequence $(s_1, \ldots, s_{n-1})$ of non-root half-edges, added by Pitman's coalescent, is uniformly distributed over the set of sequences of $(n-1)$ distinct elements of $\bigcup_{1\leq i\leq n}\{i1, \ldots, i(d(i)-1)\}$. Consequently, $\{s_1, \ldots, s_{n-1}\}$ is a uniformly random size-$(n-1)$ subset of $\bigcup_{1\leq i\leq n}\{i1, \ldots, i(d(i)-1)\}$.*

4. *Finally, conditionally given that $T(\mathrm{d}) = t$ and given the set $\{s_1, \ldots, s_{n-1}\}$ of non-root half-edges added by Pitman's coalescent, the ordering $(e_1, \ldots, e_{n-1})$ of $\mathrm{e}(t)$ is uniformly distributed over the $(n-1)!$ possible orderings of $\mathrm{e}(t)$.*

Proof. At step i of the process, there are $n+1-i$ components and $2m-n+1-i$ unpaired non-root half-edges. We may specify the pair (r_i, s_i) by first revealing the non-root half-edge s_i, then revealing r_i. Whatever the choice of s_i, there are $n-i$ possibilities for r_i, so the number of distinct choices for the pair (r_i, s_i) is $(2m-n+1-i)(n-i)$. Thus, the total number of possible execution paths for the process is

$$\prod_{i=1}^{n-1}(2m-n+1-i)(n-i) = (n-1)!(2m-n)_{n-1}. \tag{3.6}$$

The execution path followed by the process is uniquely determined by the tree $T(\mathrm{d})$ and the functions $\mathrm{K} : \mathrm{e}(T(\mathrm{d})) \to [n-1]$ and $\mathrm{H} : \mathrm{e}(T(\mathrm{d})) \to \bigcup_{i=1}^{n}\{i1, \ldots, i(\mathrm{d}_i - 1)\}$. To see this, fix any $k \in [n-1]$. Then the edge e_k created at step k of Pitman's coalescent may be recovered as $e_k = \mathrm{K}^{-1}(k)$; and, if $e_k = uv$ with $v = \mathrm{par}(u)$ then the half-edges paired to create e_k are the root half-edge $vd(v)$ incident to v and the half-edge $\mathrm{H}^{-1}(e_k)$.

Now, fix any tree t with degree sequence d, any bijection $\mathrm{k} : \mathrm{e}(t) \to [n-1]$, and any function $\mathrm{h} : \mathrm{e}(t) \to \mathbb{N}$ which, for all $i \in [n]$, assigns $c_t(i)$ distinct values from the set $\{1, \ldots, (d(i)-1)\}$ to the edges between i and its children in t. Together with (3.6), the observation of the preceding paragraph implies that

$$\mathbf{P}\{T(\mathrm{d}) = t, \mathrm{K} = \mathrm{k}, \mathrm{H} = \mathrm{h}\} = \frac{1}{(n-1)!(2m-n)_{n-1}}.$$

Having fixed the tree t, the number of possible values for K is $(n-1)!$ and the number of possible values for H is $\prod_{i\in[n]}(d(i)-1)_{c_t(i)}$. It follows that

$$\mathbf{P}\{T(\mathrm{d}) = t\} = \frac{(n-1)! \cdot \prod_{i\in[n]}(d(i)-1)_{c_t(i)}}{(n-1)!(2m-n)_{n-1}} = \frac{\prod_{i\in[n]}(d(i)-1)_{c_t(i)}}{(2m-n)_{n-1}},$$

which proves the first claim of the proposition.

Next, fix $\mathcal{H}$ as in the second assertion of the proposition, and any ordering of $\mathcal{H}$ as $(h_1, \ldots, h_{n-1})$. Then the number of execution paths which yield that $s_k = h_k$ for $k \in [n-1]$ is precisely $(n-1)!$. To see this, note that if $s_j = h_j$ for $1 \le j \le k$ then, whatever the choices of the root half-edges $(r_j, 1 \le j \le k)$, the forest F_k^n has $n+1-k$ component trees so there are $n-k$ unpaired root half-edges in components different from that of s_k; any such root half-edge may be chosen as r_k. Since there are also $(n-1)!$ possible orderings of $\mathcal{H}$, the second assertion of the proposition follows.

To prove the third statement, fix a set $\mathcal{H}$ and an ordering $(h_1, \ldots, h_{n-1})$ of its elements, as in the previous paragraph. For each $1 \le k < n-1$, given that $s_j = h_j$ for $1 \le j < k$, whatever the choices of $(r_j, 1 \le j < k)$ may be, there are $n-k$ ways to choose r_k in a distinct tree from h_k. It follows that there are $\prod_{k=1}^{n-2}(n-k) = (n-1)!$ execution paths with the property that $s_k = h_k$ for each $1 \le k \le n-1$. Since this number does not depend on $\mathcal{H}$, it follows that each size-$(n-1)$ subset of $\bigcup_{1\le i\le n}\{i1,\ldots,i(d(i)-1)\}$ is equally likely.

Finally, fix both the tree t and an unordered set $\mathcal{H}$ of non-root half-edges with $|\mathcal{H} \cap \{i1,\ldots,i(d(i)-1)\}| = c_t(i)$ for all $i \in [n]$. We consider the number of execution paths which yield $T(\mathrm{d}) = t$ and $\{s_1,\ldots,s_{n-1}\} = \mathcal{H}$. The number of choices of an ordering function $\mathrm{k} : \mathrm{e}(t) \to [n-1]$ consistent with these constraints is still $(n-1)!$. Moreover, whatever the choice of k, under the further constraint $\mathrm{K} = \mathrm{k}$, the number of possibilities for H is $\prod_{i\in[n]} c_t(i)!$. To see this, note that for each $i \in [n]$, the constraints precisely imply that $\mathcal{H} \cap \{i1,\ldots,i(d(i)-1)\} = \{s_1,\ldots,s_{n-1}\} \cap \{i1,\ldots,i(d(i)-1)\}$, and H is fixed once we additionally specify which of these $c_t(i)$ half-edges is matched to which child of i, for each $i \in [n]$. It follows that the number of execution paths which yield that $T(\mathrm{d}) = t$, that $\mathrm{K} = \mathrm{k}$ and that $\{s_1,\ldots,s_n\} = \mathcal{H}$ is

$$\prod_{i=1}^{n} c_t(i)!\,.$$

As this quantity doesn't depend on the choice of the ordering function k, the final assertion of the proposition follows. ☐

We state a corollary of the above proposition, for later use.

Corollary 3.2.2. *The tree $T(\mathrm{d})$ is a uniformly random rooted tree with child sequence c_T.*

The corollary follows since the formula for $\mathbf{P}\{T(\mathrm{d}) = t\}$ from Proposition 3.2.1 only depends on t through c_t.

We now assume that $\sum_{i=1}^n d(i) \ge 2n$ and that $\sum_{i=1}^n d(i)$ is even, and define a random tree-rooted graph $(G, T, \Gamma) = (G(\mathrm{d}), T(\mathrm{d}), \Gamma(\mathrm{d}))$ as follows: First, let $T = T(\mathrm{d})$ be the random tree built by Pitman's coalescent, and let $\Gamma^+ = \Gamma^+(\mathrm{d})$ be its root half-edge. We refer to T as the *spanning tree-elect* of a to-be-constructed tree-rooted graph. Next, choose a uniformly random matching of the $2m - 2(n-1)$ pendant half-edges attached to T, and pair the half-edges according to this matching to create $G = G(\mathrm{d})$. Then let Γ be the edge containing Γ^+, oriented so that Γ^+ is at the head; for later use, let $\Gamma^- = \Gamma^-(\mathrm{d})$ be the other half-edge of Γ. We call $(G(\mathrm{d}), T(\mathrm{d}), \Gamma(\mathrm{d}))$, or any other graph with the same distribution, a *random tree-rooted graph with degree sequence* d. The tree T has now taken office.

The next proposition describes the distribution of $(G(\mathrm{d}), T(\mathrm{d}), \Gamma(\mathrm{d}))$. For a tree-rooted graph (g, t, γ),

Proposition 3.2.3. *Let* $\mathrm{d} = (d(1), \ldots, d(n))$ *be a degree sequence with* $d(i) \geq 1$ *for all* $i \in [n]$*, and write* $m = \frac{1}{2}\sum_{i=1}^{n} d(i)$*. Fix a tree-rooted graph* (g, t, γ) *where* g *is a graph with degree sequence* d*. Then*

$$\mathbf{P}\left\{(G(\mathrm{d}), T(\mathrm{d}), \Gamma(\mathrm{d})) = (g, t, \gamma)\right\} \propto \frac{2^{\mathbf{1}_{[\gamma \text{ is a loop}]}} \cdot m_{g-t}(\gamma)}{\prod_{i=1}^{n} 2^{m_{g-t}(ii)} \cdot \prod_{e \in \mathrm{e}(g)} m_{g-t}(e)!}\,.$$

Proof. Proposition 3.2.1 gives us a formula for $\mathbf{P}\left\{T(\mathrm{d}) = t\right\}$. We next focus on computing

$$\mathbf{P}\left\{G(\mathrm{d}) = g \mid T(\mathrm{d}) = t\right\}.$$

Write r for the root of t, and $\gamma = qr$ for the oriented root edge of g. Given that $T(\mathrm{d}) = t$, each $i \in [n]$ with $i \neq r(t)$ has $d'(i) := d(i) - c_t(i) - 1$ pendant half-edges attached to it, and r has $d'(r) := d(r) - c_t(r)$ half-edges attached to it. Conditionally given that $T(\mathrm{d}) = t$, the graph $G(\mathrm{d}) - T(\mathrm{d})$ is distributed as $CM(\mathrm{d}')$, a random graph with degree sequence $\mathrm{d}' = (d'(1), \ldots, d'(n))$ sampled according to the configuration model, so with distribution as in (3.2)), and more Writing $m' := m - (n-1) = \frac{1}{2}\sum_{i=1}^{n} d'(i)$ and $g' = (\mathrm{v}(g), \mathrm{e}(g) \setminus \mathrm{e}(t))$, it follows that

$$\begin{aligned}
\mathbf{P}\left\{G(\mathrm{d}) = g \mid T(\mathrm{d}) = t\right\} &= \mathbf{P}\left\{CM(\mathrm{d}') = g'\right\} \\
&= \frac{2^{m'}(m')!}{(2m')!} \frac{\prod_{i=1}^{n} d'(i)!}{\prod_{i=1}^{n} 2^{m_{g'}(ii)} \cdot \prod_{e \in \mathrm{e}(g')} m_{g'}(e)!} \\
&= \frac{2^{m'}(m')!}{(2m')!} \frac{\prod_{i=1}^{n} d'(i)!}{\prod_{i=1}^{n} 2^{m_{g-t}(ii)} \cdot \prod_{e \in \mathrm{e}(g)} m_{g-t}(e)!}\,.
\end{aligned}$$

For the second equality we have used the exact expression for the distribution of $CM(\mathrm{d}')$, which can be found in, e.g., [68], equation (7.2.6). For the last equality, we use that $m_{g'}(ii) = m_{g-t}(ii)$ since t is a tree so contains no loops, and that $m_{g'}(e) = m_{g-t}(e)$ by definition when $e \in \mathrm{e}(g')$.

Given that $T(\mathrm{d}) = t$ and that $G(\mathrm{d}) = g$, in order to have $(G(\mathrm{d}), T(\mathrm{d}), \Gamma(\mathrm{d})) = (g, t, \gamma)$ it is necessary and sufficient that $\Gamma(\mathrm{d}) = \gamma$. This occurs precisely if γ^+, the half-edge of γ incident to r, was matched with some half-edge incident to q. Since the matching of half-edges in $G(\mathrm{d}) - T(\mathrm{d})$ is chosen uniformly at random, by symmetry the conditional probability that this occurred is $m_{g-t}(\gamma)/d'(r)$ if γ is not a loop, and is $2m_{g-t}(\gamma)/d'(r)$ if γ is a loop. We may unify these two formulas by writing

$$\mathbf{P}\left\{\Gamma(\mathrm{d})=\gamma \mid T(\mathrm{d})=t, G(\mathrm{d})=g\right\}=\frac{2^{\mathbf{1}_{[\gamma \text{ is a loop}]}} m_{g-t}(\gamma)}{d'(r)}.$$

Combined with the formula for $\mathbf{P}\left\{T(\mathrm{d})=t\right\}$ from Proposition 3.2.1, this gives

$$\begin{aligned}
&\mathbf{P}\left\{(G(\mathrm{d}), T(\mathrm{d}), \Gamma(\mathrm{d}))=(g,t,\gamma)\right\}\\
&=\frac{1}{(2m-n)_{n-1}}\prod_{i=1}^{n}(d(i)-1)_{c_t(i)}\\
&\quad\cdot\frac{2^{m'}(m')!}{(2m')!}\frac{\prod_{i=1}^{n} d'(i)!}{\prod_{i=1}^{n} 2^{m_{g-t}(ii)}\cdot\prod_{e\in \mathrm{e}(g)} m_{g-t}(e)!}\\
&\quad\cdot\frac{2^{\mathbf{1}_{[\gamma \text{ is a loop}]}} m_{g-t}(\gamma)}{d'(r)}\\
&=\prod_{i=1}^{n}(d(i)-1)!\cdot\frac{2^{m-(n-1)}(m-(n-1))!}{2m'(2m-n)!}\cdot\frac{2^{\mathbf{1}_{[\gamma \text{ is a loop}]}} m_{g-t}(\gamma)}{\prod_{i=1}^{n} 2^{m_{g-t}(ii)}\cdot\prod_{e\in \mathrm{e}(g)} m_{g-t}(e)!}.
\end{aligned}$$

In the second equality we have used that $(2m-n)_{n-1}(2m')! = 2m'(2m-n)!$, that $(d(i)-1)_{c_t(i)}d'(i)! = (d(i)-1)!$ for $i \neq r$, and that $(d(r)-1)_{c_t(r)}d'(r)! = d'(r)(d(r)-1)!$. The first two terms on the final line do not depend on the triple (g,t,γ), so the result follows. □

3.3 Concentration of degrees

Throughout this section, let $(\mathrm{d}^n, n \geq 1)$ be a sequence of degree sequences satisfying the conditions of Theorem 3.1.1, and also let p^n and p be as in Theorem 3.1.1. Next, for $n \geq 1$ let $T(\mathrm{d}^n)$ be the tree built by Pitman's additive coalescent applied to the degree sequence $\mathrm{d}^n = (d^n(i), 1 \leq i \leq n)$. Let $\mathrm{c}^n = (c^n(i), 1 \leq i \leq n)$ be the child sequence of $T(\mathrm{d}^n)$, and recall that $Q_{\mathrm{c}^n} = (Q_{\mathrm{c}^n}(a), a \geq 0)$ is the child statistics vector of c^n. Also, for $0 \leq a < b$, let $P^n_{b,a} = \#\{1 \leq i \leq n : d^n(i) = b, c^n(i) = a\}$. Finally, let $\rho := 1/(\mu_1(p) - 1)$. Note that since $\sum_{i\in[n]} d^n(i) \geq 2n$, necessarily $\mu_1(p^n) \geq 2$; since $p^n \to p$ pointwise and $\mu_2(p^n) \to \mu_2(p)$, it follows that $\mu_1(p^n) \to \mu_1(p)$, so $\mu_1(p) \geq 2$ and hence $\rho \in (0,1]$.

Proposition 3.3.1. *For $a \geq 0$ let*

$$q(a) := \sum_{b=a+1}^{\infty} p(b)\cdot \mathbf{P}\left\{\mathrm{Bin}(b-1,\rho)=a\right\}.$$

Then $\mu_2(q) < \infty$ *and* $\mu_2(n^{-1}Q_{c^n}) \to \mu_2(q)$ *in probability as* $n \to \infty$. *Moreover, for all* $0 \le a < b$, $n^{-1}P^n_{b,a} \xrightarrow{\text{prob}} p(b) \cdot \mathbf{P}\{\mathrm{Bin}(b-1,\rho) = a\}$, *and* $n^{-1}Q_{c^n}(a) \xrightarrow{\text{prob}} q(a)$, *in both cases as* $n \to \infty$.

Let $(G(\mathrm{d}^n), T(\mathrm{d}^n), \Gamma(\mathrm{d}^n))$ be a random tree-weighted graph with degree sequence d^n. Using Proposition 3.3.1, together with existing results from the literature, it is fairly straightforward to establish that $(\sigma n^{-1/2})T(\mathrm{d}^n) \xrightarrow{\mathrm{d}} \mathcal{T}$, with $\sigma = \mu_2(q) - 1 \in (0,\infty)$, where $\mathcal{T}$ is the Brownian continuum random tree. However, in order to show that such convergence holds for the corresponding random *simple* tree-weighted graphs, we additionally need the next proposition, which establishes that the number of pairs of tree-adjacent vertices in $T(\mathrm{d}^n)$ with given fixed degrees is well-concentrated around its expected values. This will be used in order to show that the probability of $G(\mathrm{d}^n)$ being simple given $T(\mathrm{d}^n)$ asymptotically behaves like a constant.

Write $G_-(\mathrm{d}^n) = G(\mathrm{d}^n) - T(\mathrm{d}^n)$ and let $\mathrm{d}^n_- = (d^n_-(i), 1 \le i \le n)$ be the degree sequence of $G_-(\mathrm{d}^n)$. For integers $k, \ell \ge 0$, let

$$\alpha(k,\ell) = \sum_{a_1,a_2\ge 0} a_2 p(\ell+a_2+1)\mathbf{P}\{\mathrm{Bin}(\ell+a_2,\rho) = a_2\} \cdot p(k+a_1+1)\mathbf{P}\{\mathrm{Bin}(k+a_1,\rho) = a_1\}\ . \tag{3.7}$$

Proposition 3.3.2. *For integers* $k, \ell \ge 0$ *let*

$$A^n(k,\ell) = \left|\left\{uv \in \mathrm{e}(T(\mathrm{d}^n)) : \mathrm{d}^n_-(u) = k, \mathrm{d}^n_-(v) = \ell\right\}\right| .$$

Then for all $k, \ell \ge 0$,

$$\frac{1}{n}A^n(k,\ell) \xrightarrow{\text{prob}} \alpha(k,\ell)$$

as $n \to \infty$, *and also*

$$\frac{1}{n}\sum_{k,\ell\ge 0} k\ell A^n(k,\ell) \xrightarrow{\text{prob}} \sum_{k,\ell\ge 0} k\ell\alpha(k,\ell).$$

The proofs of Propositions 3.3.1 and 3.3.2 appear in Appendix 3.A.

To conclude the section, we observe that $\alpha(k,\ell)$ defines a probability distribution on pairs of non-negative integers. Indeed,

$$\begin{aligned}\sum_{k\ge 0}\sum_{a_1\ge 0} p(k+a_1+1)\mathbf{P}\{\mathrm{Bin}(k+a_1,\rho) = a_1\} &= \sum_{m\ge 0}\sum_{a=0}^{m} p(m+1)\mathbf{P}\{\mathrm{Bin}(m,\rho) = a\} \\ &= \sum_{m\ge 0} p(m+1) = 1 - p(0) = 1\,,\end{aligned}$$

and

$$\begin{aligned}
&\sum_{\ell\geq 0}\sum_{a_2\geq 0} a_2 p(\ell+a_2+1)\mathbf{P}\left\{\mathrm{Bin}(\ell+a_2,\rho)=a_2\right\}\\
&=\sum_{m\geq 0}\sum_{a=0}^{m} ap(m+1)\mathbf{P}\left\{\mathrm{Bin}(m,\rho)=a\right\}\\
&=\sum_{m\geq 0} p(m+1)\cdot m\rho=(\mu_1(p)-(1-p(0)))\rho=(\mu_1(p)-1)\rho=1,
\end{aligned}$$

so by factorizing $\sum_{k,\ell\geq 0}\alpha(k,\ell)$ we obtain

$$\sum_{k,\ell\geq 0}\alpha(k,\ell)=\left(\sum_{m\geq 0}p(m+1)\right)\cdot\left(\sum_{m\geq 0}\sum_{m\geq 0}p(m+1)\cdot m\rho\right)=1\,;$$

the fact that $\sum_{k,\ell\geq 0}\alpha(k,\ell)=1$ will be used in the proof of Proposition 3.3.2. A similar computation shows that

$$\sum_{k,\ell\geq 0}k\ell\cdot\alpha(k,\ell)\leq\left(\sum_{m\geq 0}mp(m+1)\right)\cdot\left(\sum_{m\geq 0}m^2p(m+1)\rho\right)=\mu_2(p)-2\mu_1(p)+1<\infty\,, \quad (3.8)$$

a fact we will use in bounding the probability of simplicity of $G(\mathrm{d}^n)$.

3.4 Poisson approximation for graph superpositions.

In this section we state a Poisson approximation theorem for the number of loops and multiple edges in the superposition of a fixed simple graph and a random graph with a fixed degree sequence; this in particular allows us to control the probability that such a superposition yields a simple graph.

Let H be a simple graph with vertex set $v(H)=[n]$. Fix a degree sequence $\mathrm{d}=(d(1),\ldots,d(n))$ whose sum of degrees is even, and let G be a random graph with degree sequence d sampled according to the configuration model. For vertices $u,v\in[n]$ and $i\in[d(u)], j\in[d(v)]$, let $\mathbf{1}_{[ui,vj]}$ be the indicator of the event that half-edge ui is matched with half-edge vj in G. Now write

$$\begin{aligned}
\mathcal{L} = \mathcal{L}(G) &= \{(ui, uj) : u \in [n], i, j \in [d(u)], i < j\} \\
\mathcal{M} = \mathcal{M}(G, H) &= \{((ui_1, vj_1), (ui_2, vj_2)) : u, v \in [n], uv \notin e(H), \\
&\qquad i_1, i_2 \in [d(u)], j_1, j_2 \in [d(v)], u < v, i_1 < i_2, j_1 \neq j_2\}, \text{ and} \\
\mathcal{N} = \mathcal{N}(G, H) &= \{(ui, vj) : uv \in e(H), i \in [d(u)], j \in [d(v)]\},
\end{aligned}$$

and let

$$\begin{aligned}
L = L(G) &= \sum_{(ui,uj)\in\mathcal{L}} \mathbf{1}_{[ui,uj]}\,, \\
M = M(G, H) &= \sum_{((ui_1,vj_1),(ui_2vj_2))\in\mathcal{M}} \mathbf{1}_{[(ui_1,vj_1)]}\mathbf{1}_{[(ui_2vj_2)]}\,, \text{ and} \\
N = N(G, H) & \sum_{(ui,vj)\in\mathcal{N}} \mathbf{1}_{[ui,vj]}\,.
\end{aligned}$$

Note that the graph with edge set $e(G) \cup e(H)$ is simple precisely if $L + M + N = 0$.

Theorem 3.4.1. *Fix a sequence of simple graphs $(h_n, n \geq 1)$ with $v(h_n) = [n]$ for all $n \geq 1$ and $\max_{v\in[n]}\{\deg_{h_n}(v)\} = o(n)$. For each $n \geq 1$ let $\mathrm{d}^n = (d^n(v), 1 \leq v \leq n)$ be a degree sequence and let p^n be the degree distribution of d^n. Suppose that there exists a probability distribution $p = (p(k), k \geq 0)$ with $\mu_2(p) \in [0, \infty)$ and $p(0) < 1$ such that the following holds.*

First, $\mathrm{p}^n \to p$ pointwise and $\mu_2(\mathrm{p}^n) \to \mu_2(p)$. Second, there are non-negative numbers $(\alpha(a, b), a, b \geq 0)$ such that for any $a, b \geq 0$

$$\alpha^n(a, b) := \frac{1}{n}\left|\{uv \in e(h_n) : d^n(u) = a, d^n(v) = b\}\right| \to \alpha(a, b),$$

and

$$\sum_{k,\ell\geq 0} kl\alpha^n(k, \ell) \to \sum_{k,\ell\geq 0} kl\alpha(k, \ell) < \infty \tag{3.9}$$

For $n \geq 1$ let G_n be distributed according to the configuration model on graphs with vertex set $[n]$ and degree sequence d^n. Then with $L_n = L(G_n)$, $M_n = M(G_n, h_n)$ and $N_n = N(G_n, h_n)$, we have

$$\|\mathrm{Dist}(L_n, M_n, N_n) - \mathrm{Poi}(\nu/2) \otimes \mathrm{Poi}(\nu^2/4) \otimes \mathrm{Poi}(\eta)\|_{\mathrm{TV}} \to 0$$

as $n \to \infty$, where $\nu = (\mu_2(p)/\mu_1(p)) - 1$ and $\eta = \frac{1}{\mu_1(p)} \sum_{i,j\geq 1} ij\alpha(i, j)$.

In the statement of Theorem 3.4.1 we have introduced the notation $\deg_{h_n}(v)$ for the degree of vertex v in h_n, and the notation $\|\mu - \nu\|_{\mathrm{TV}}$ for the total variation distance between probability measures. The proof of Theorem 3.4.1 appears in Appendix 3.B. This theorem has the following consequence for random tree-weighted graphs, which we will use in the next section.

Corollary 3.4.2. *Let* $(\mathrm{d}^n, n \geq 1)$ *and* $(p^n, n \geq 1)$ *be as in Theorem 3.1.1, and for* $n \geq 1$ *let* $(G(\mathrm{d}^n), T(\mathrm{d}^n), \Gamma(\mathrm{d}^n))$ *be a random tree-weighted graph with degree sequence* d. *Then*

$$\mathbf{P}\left\{G(\mathrm{d}^n) \text{ simple } \mid T(\mathrm{d}^n)\right\} \xrightarrow{\text{prob}} \exp(-\nu/2 - \nu^2/4 - \eta)\,,$$

as $n \to \infty$.

This corollary follows straightforwardly from Theorem 3.4.1 when $\mu_2(p^n) > 2$, in which case $G_-(\mathrm{d}^n) = G(\mathrm{d}^n) - T(\mathrm{d}^n)$ has a linear number of edges. However, when $\mu_2(p^n) = 2$, and the graph $G_-(\mathrm{d}^n)$ has a sub-linear number of edges, a separate argument is needed. The proof of Corollary 3.4.2 also appears in Appendix 3.B.

3.5 Proof of Theorem 3.1.1

Let $(\mathrm{d}^n, n \geq 1)$ be a sequence of degree sequences satisfying the conditions of Theorem 3.1.1. For $n \geq 1$ let $T(\mathrm{d}^n)$ be the tree built by Pitman's additive coalescent applied to degree sequence d^n, and let c^n be the child sequence of $T(\mathrm{d}^n)$. By Proposition 3.2.1 (1), conditionally given c^n, the tree $T(\mathrm{d}^n)$ is uniformly distributed over the set of trees with child sequence c^n.

By Proposition 3.3.1, the child statistics vectors $(Q_{\mathrm{c}^n}, n \geq 1)$ satisfy that, as $n \to \infty$, for all $a \geq 0$,

$$n^{-1} Q_{\mathrm{c}^n}(a) \xrightarrow{\text{prob}} q(a), \tag{3.10}$$

and moreover that $\mu_2(n^{-1}Q_{\mathrm{c}^n}(a)) \to \mu_2(q)$. Here $q = (q(a), a \geq 0)$ is as in Proposition 3.3.1, and in particular satisfies $\mu_2(q) < \infty$. We will also need that $\mu_2(q) > 1$, and we now justify this.

The convergence (3.10) and the fact that $\mu_2(n^{-1}Q_{\mathrm{c}^n}(a)) \to \mu_2(q)$ together imply that $\mu_1(n^{-1}Q_{\mathrm{c}^n}(a)) \to \mu_1(q)$. But $\mu_1(n^{-1}Q_{\mathrm{c}^n}(a)) = (n-1)/n$ since Q_{c^n} is a child sequence, so necessarily $\mu_1(q) = 1$. By the definition of q, if $\rho = 1$ then $q(1) = p(2)$, and $p(2) < 1$ by assumption. If $\rho > 1$ then

$$q(0) := \sum_{b=1}^{\infty} p(b) \cdot \mathbf{P}\left\{\mathrm{Bin}(b-1,\rho) = 0\right\} > 0,$$

so again $q(1) \le (1 - q(0)) < 1$. Thus, we always have $q(1) < 1$, which together with the fact that $\mu_1(q) = 1$ implies that $\mu_2(q) > 1$.

Writing $\sigma = \mu_2(q) - 1 \in (0,\infty)$, it then follows by Theorem 1 of [33] that

$$\overline{T}(\mathrm{d}^n) := \frac{\sigma}{n^{1/2}} T(\mathrm{d}^n) \xrightarrow{\mathrm{d}} \mathcal{T},$$

in the Gromov-Hausdorff-Prokhorov sense.[1]

We aim to prove the same statement with $\overline{T}(\mathrm{d}^n)$ replaced by $\overline{T}_n := (\sigma/n^{1/2})T_n$, where (G_n, T_n, Γ_n) is is a uniformly random simple tree-rooted graph with degree sequence d^n. To accomplish this, we use that the law of (G_n, T_n, Γ_n) is precisely the conditional law of $(G(\mathrm{d}^n), T(\mathrm{d}^n), \Gamma(\mathrm{d}^n))$ given that $G(\mathrm{d}^n)$ is a simple graph.

Writing $\mathbb{K}$ for Gromov-Hausdorff-Prokhorov space as in [1], for any bounded continuous function $f : \mathbb{K} \to \mathbb{R}$ we have

$$\begin{aligned}
\mathbf{E}\left(f(\overline{T}(\mathrm{d}^n)) \cdot \mathbf{1}_{[G(\mathrm{d}^n)\ \mathrm{simple}]}\right) &= \mathbf{E}\Big(\mathbf{E}\left(f(\overline{T}(\mathrm{d}^n)) \cdot \mathbf{1}_{[G(\mathrm{d}^n)\ \mathrm{simple}]} \mid T(\mathrm{d}^n)\right)\Big) \\
&= \mathbf{E}\left(f(\overline{T}(\mathrm{d}^n)) \cdot \mathbf{P}\left\{G(\mathrm{d}^n)\ \mathrm{simple} \mid T(\mathrm{d}^n)\right\}\right).
\end{aligned}$$

Since $\mathbf{E}f(\overline{T}(\mathrm{d}^n)) \to \mathbf{E}f(\mathcal{T})$, and $\mathbf{P}\left\{G(\mathrm{d}^n)\ \mathrm{simple} \mid T(\mathrm{d}^n)\right\} \xrightarrow{\mathrm{prob}} \exp(-\nu/2 - \nu^2/4 - \eta)$ by Corollary 3.4.2, it follows that

$$\mathbf{E}\left(f(\overline{T}(\mathrm{d}^n)) \cdot \mathbf{1}_{[G(\mathrm{d}^n)\ \mathrm{simple}]}\right) \to \exp(-\nu/2 - \nu^2/4 - \eta)\mathbf{E}\left(f(\mathcal{T})\right).$$

Furthermore,

$$\mathbf{P}\left\{G(\mathrm{d}^n)\ \mathrm{simple}\right\} = \mathbf{E}\left(\mathbf{P}\left\{G(\mathrm{d}^n)\ \mathrm{simple} \mid T(\mathrm{d}^n)\right\}\right) \to \exp(-\nu/2 - \nu^2/4 - \eta),$$

[1]Theorem 1 of [33] is stated for plane trees with a fixed degree sequence, rather than labelled trees with a fixed degree sequence. However, as noted by Broutin and Marckert [33, page 295], a straightforward combinatorial argument shows that the same result holds for labeled trees. Also, as stated, the theorem only yields convergence in the Gromov-Hausdorff sense; but the proof proceeds by establishing convergence distributional of coding functions. As explained in [1, Section 3], such proofs immediately yield the stronger Gromov-Hausdorff-Prokhorov convergence.

and therefore

$$\mathbf{E}\left(f(\overline{T}(\mathrm{d}^n)) \;\middle|\; G(\mathrm{d}^n) \text{ simple}\right) = \frac{\mathbf{E}\left(f(\overline{T}(\mathrm{d}^n)) \cdot \mathbf{1}_{[G(\mathrm{d}^n) \text{ simple}]}\right)}{\mathbf{P}\left\{G(\mathrm{d}^n) \text{ simple}\right\}} \to \mathbf{E}\left(f(\mathcal{T})\right) .$$

Since

$$\mathbf{E}\left(f(\overline{T}_n)\right) = \mathbf{E}\left(f(\overline{T}(\mathrm{d}^n) \;\middle|\; G(\mathrm{d}^n) \text{ simple}\right) ,$$

the fact that $\overline{T}_n \xrightarrow{\mathrm{d}} \mathcal{T}$ now follows by the Portmanteau theorem. □

3.A Proofs of Propositions 3.3.1 and 3.3.2

Before beginning the proofs in earnest, we state and prove a simple bound on the asymptotic behaviour of maximum degrees and sums of small sets of degrees, for sequences of degree sequences as in Theorems 3.1.1 and 3.4.1, which will be used multiple times below.

Fact 3.A.1. *For each $n \geq 1$ let $\mathrm{d}^n = (d^n(v), 1 \leq v \leq n)$ be a degree sequence and let p^n be the degree distribution of d^n. Suppose that there exists a probability distribution $p = (p(k), k \geq 0)$ such that $p^n \to p$ pointwise and $\mu_2(p^n) \to \mu_2(p) \in [0, \infty)$. Then $\max_{1 \leq i \leq n} d^n(i) = o(n^{1/2})$. Also, for any sets $(A_n, n \geq 1)$ with $A_n \subset [n]$ and $|A_n| = o(n)$, it holds that $\sum_{i \in A_n} d^n(i) = o(n)$.*

Proof. If $p^n \to p$ pointwise and $\mu_2(p^n) \to \mu_2(p) \in [0, \infty)$, then for all $\epsilon > 0$ there is M such that

$$\liminf_{n \to \infty} \sum_{k=1}^{M} k^2 p^n(k) \geq \mu_2(p) - \epsilon,$$

so $\sup_{M \geq 1} \liminf_{n \to \infty} \sum_{k=1}^{M} k^2 p^n(k) \geq \mu_2(p)$. If additionally there is $\delta > 0$ such that $\max_{1 \leq i \leq n} d^n(i) \geq \delta n^{1/2}$ for infinitely many n, then

$$\limsup_{n \to \infty} \mu_2(p^n) \geq \delta^2 + \sup_{M \geq 1} \liminf_{n \to \infty} \sum_{k=1}^{M} k^2 p^n(k) > \mu_2(p),$$

so $\mu_2(p^n) \not\to \mu_2(p)$.

Similarly, for sets $(A_n, n \geq 1)$ as in the statement, since $|A_n| = o(n)$, for any $M \in \mathbb{N}$ we have $\sum_{i \in A_n} (d^n(i))^2 \mathbf{1}_{[d^n(i) \leq M]} = o(n)$, so for any $\epsilon > 0$ there is $M \in \mathbb{N}$ such that

$$\liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (d^n(i))^2 \mathbf{1}_{[d^n(i) \leq M]} \mathbf{1}_{[i \notin A_n]} \geq \mu_2(p) - \epsilon.$$

This implies that $\liminf_{n\to\infty} n^{-1}\sum_{i=1}^{n}(d^n(i))^2\mathbf{1}_{[i\notin A_n]} \geq \mu_2(p)$. If also there is $\delta > 0$ such that $\sum_{i\in A_n}(d^n(i))^2 > \delta n$ for infinitely many n, then

$$\limsup_{n\to\infty} \mu_2(p^n) = \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(d^n(i))^2 \geq \mu_2(p) + \delta\,,$$

so $\mu_2(p^n) \not\to \mu_2(p)$. □

Note that the conditions on the degree sequences in both Theorem 3.1.1 and Theorem 3.4.1 allow Fact 3.A.1 to be applied.

To prove Proposition 3.3.1, we will make use of the following lemma, which uses the second moment method to control how subsampling affects degree distributions. The proof of the proposition immediately follows that of the lemma.

Lemma 3.A.2. *For any integer $b \geq 1$ there exists n_0 such that for all $n \geq n_0$ the following holds. Let $\mathrm{d} = (d(1), \ldots, d(n))$ be a degree sequence with $d(i) \geq 1$ for all $i \in [n]$ and with $\sum_{i=1}^{n} d(i) \geq 2n-1$, set $S = \bigcup_{i=1}^{n}\{i1, \ldots, i(d(i)-1)\}$ and write $s = |S|$. Let U be a uniformly random subset of S with $|\mathrm{U}| = n-1$, and for $1 \leq i \leq n$ write $U_i = \#\{1 \leq j < d(i) : (i,j) \in \mathrm{U}\}$. For $0 \leq a < b$, write $P_{b,a} = \#\{1 \leq i \leq n : (d(i), U_i) = (b,a)\}$. Then for all $\epsilon > 0$,*

$$\mathbf{P}\left\{|P_{b,a} - \mathbf{E}P_{b,a}| > \epsilon \mathbf{E}P_{b,a}\right\} < \frac{1}{\epsilon^2}\left(\frac{1}{\mathbf{E}P_{b,a}} + \frac{2b^2}{s}\right).$$

Proof. We fix $0 \leq a \leq b$ and compute the first and second moments of $P_{b+1,a}$; this makes the calculations slightly easier to read than they would be for $P_{b,a}$.

Fix indices k and ℓ with $k \neq \ell$ and $d(k) = d(\ell) = b+1$. Since U is a uniformly random subset of S, by symmetry we have

$$\mathbf{P}\left\{|U_k| = a\right\} = \mathbf{P}\left\{|U_\ell| = a\right\} = \binom{b}{a} \cdot \binom{s-b}{n-1-a}\binom{s}{n-1}^{-1},$$

and

$$\mathbf{P}\left\{|U_k| = |U_\ell| = a\right\} = \binom{b}{a}^2 \cdot \binom{s-2b}{n-1-2a}\binom{s}{n-1}^{-1},$$

so writing $n_{b+1} = \#\{1 \leq i \leq n : d(i) = b+1\}$, we have

$$\mathbf{E}P_{b+1,a} = n_{b+1}\binom{b}{a} \cdot \binom{s-b}{n-1-a}\binom{s}{n-1}^{-1} \tag{3.11}$$

and

$$
\begin{aligned}
&\mathbf{Var}\left\{P_{b+1,a}\right\} \\
&= n_{b+1}(n_{b+1}-1)\binom{b}{a}^2 \left(\binom{s-2b}{n-1-2a}\binom{s}{n-1}^{-1} - \binom{s-b}{n-1-a}^2 \binom{s}{n-1}^{-2} \right) \\
&+ n_{b+1} \left(\binom{b}{a}\binom{s-b}{n-1-1}\binom{s}{n-1}^{-1} - \binom{b}{a}^2 \binom{s-b}{n-1-1}^2 \binom{s}{n-1}^{-2} \right),
\end{aligned}
$$

where the final line accounts for the diagonal terms. Bounding the final line from above by $\mathbf{E}P_{b+1,a}$ and cancelling terms in the parenthetical expression in the middle line gives

$$
\begin{aligned}
&\mathbf{Var}\left\{P_{b+1,a}\right\} - \mathbf{E}P_{b+1,a} \\
&\le n_{b+1}(n_{b+1}-1)\binom{b}{a}^2 \left(\frac{(n-1)_{2a}(s-(n-1))_{2(b-a)}}{(s)_{2b}} - \frac{(n-1)_a^2(s-(n-1))_{b-a}^2}{(s)_b^2} \right).
\end{aligned}
$$

The ratio of the first and the second term in the final parentheses is

$$
\frac{(n-1)_{2a}}{(n-1)_a^2}\frac{(s-(n-1))_{2(b-a)}}{(s-(n-1))_{b-a}^2}\frac{(s)_b^2}{(s)_{2b}} \le \frac{(s)_b^2}{(s)_{2b}} \le \left(1+\frac{b}{s-2b}\right)^b \le 1+\frac{2b^2}{s},
$$

the last bound holding for b fixed and s large. This gives

$$
\begin{aligned}
\mathbf{Var}\left\{P_{b+1,a}\right\} &\le \mathbf{E}P_{b+1,a} + n_{b+1}(n_{b+1}-1)\binom{b}{a}^2 \frac{2b^2}{s}\binom{s-b}{n-1-a}^2\binom{s}{n-1}^{-2} \\
&< \mathbf{E}P_{b+1,a} + \frac{2b^2}{s}(\mathbf{E}P_{b+1,a})^2,
\end{aligned}
$$

and the lemma follows by Chebyshev's inequality. □

Proof of Proposition 3.3.1. We first bound $\mu_2(q)$ by writing

$$
\begin{aligned}
\mu_2(q) &= \sum_{a\ge 0} a^2 q(a) \\
&= \sum_{a\ge 0} a^2 \sum_{b>a} p(b)\cdot \mathbf{P}\left\{\mathrm{Bin}(b-1,\rho)=a\right\} \\
&\le \sum_{b>0} b^2 p(b) \cdot \sum_{0\le a<b} \mathbf{P}\left\{\mathrm{Bin}(b-1,\rho)=a\right\} \\
&= \mu_2(p)^2 < \infty .
\end{aligned}
$$

Next, since $p^n \to p$ pointwise and $\mu_2(p^n) \to \mu_2(p) < \infty$, for any $\epsilon > 0$ there is k such that $\sum_{d \geq k} d^2 p^n(d) < \epsilon$ and $\sum_{d \geq k} d^2 p(d) < \epsilon$. If node i has a children in $T(\mathrm{d}^n)$ then $d^n(i) \geq a+1$, so it follows that

$$\begin{aligned}
\sum_{a=k}^{\infty} a^2 \frac{Q_{\mathrm{c}^n}(a)}{n} &= \sum_{a \geq k} \sum_{b > a} a^2 \frac{\#\{i \leq n : c^n(i) = a, d^n(i) = b\}}{n} \\
&\leq \sum_{b > k} b^2 \sum_{a < b} \frac{\#\{i \leq n : c^n(i) = a, d^n(i) = b\}}{n} \\
&= \sum_{b > k} b^2 p^n(b) < \epsilon\,.
\end{aligned}$$

To complete the proof it thus suffices to show that $n^{-1} P^n_{b,a} \to p(b) \cdot \mathbf{P}\left\{\mathrm{Bin}(b-1, \rho) = a\right\}$ in probability for all $0 \leq a < b$ and that $n^{-1} Q_{\mathrm{c}^n} \to q$ pointwise in probability; the fact that $\mu_2(n^{-1} Q_{\mathrm{c}^n}) \to \mu_2(q)$ in probability then immediately follows.

By the third statement of Proposition 3.2.1, the set of non-root half-edges in $T(\mathrm{d}^n)$ is a uniformly random size-$(n-1)$ subset of the set $S^n := \bigcup_{1 \leq i \leq n} \{i1, \ldots, i(d^n(i)-1)\}$. We will apply Lemma 3.A.2 to control the numbers of nodes with a given number of children in $T(\mathrm{d}^n)$. To make the coming applications of that lemma transparent, we write $s^n := |S^n| = \sum_{1 \leq i \leq n} (d^n(i) - 1)$.

We handle the cases $\mu_1(p) = 2$ and $\mu_1(p) > 2$ separately. If $\mu_1(p) = 2$ then $|S^n| = \sum_{1 \leq i \leq n} (d^n(i) - 1) = (1 + o(1))n$ as $n \to \infty$. Note that in this case $\rho(p) = 1/(\mu_1(p) - 1) = 1$ so $q(a) = p(a+1)$ for all $a \geq 0$. For any $a \geq 0$, by (3.11) we then have

$$\mathbf{E} P^n_{a+1,a} = (1 - o(1)) n p^n(a+1) \binom{a}{a} \binom{|S^n| - 1 - a}{n - 1 - a} \binom{|S^n|}{n-1}^{-1} = (1 - o(1)) n p^n(a+1).$$

If $p(a+1) > 0$ then $np^n(a+1) = \Theta(n)$, so

$$\frac{1}{\mathbf{E}\left(P^n_{a+1,a}\right)} + \frac{2(a+1)^2}{|S^n|} = o(1),$$

and hence by Lemma 3.A.2,

$$\frac{P^n_{a+1,a}}{n} \xrightarrow{\text{prob}} p(a+1) = q(a).$$

If $p(a+1) = 0$ then $p^n(a+1) = o(1)$, so $\mathbf{E}\left(P^n_{a+1,a}\right)/n \to 0$ and thus $P^n_{a+1,a}/n \xrightarrow{\text{prob}} 0 = q(a)$ by Markov's inequality. Since this holds for all $a \geq 0$, and $\sum_{a \geq 0} P^n_{a+1,a}/n \leq 1 = \sum_{a \geq 0} q(a)$, it follows that $\sum_{a\geq 0} P^n_{a+1,a}/n \to 1$ in probability. This implies that $\sum_{b>a+1} P^n_{b,a}/n \to 0$ in probability, so

$$\frac{Q_{c^n}(a)}{n} = \frac{1}{n}\sum_{b>a} P^n_{b,a} = \frac{P^n_{a+1,a}}{n} + \sum_{b>a+1} \frac{P^n_{b,a}}{n} \xrightarrow{\text{prob}} q(a)\,,$$

and that for all $b > a+1$, $P^n_{b,a}/n \to 0 = p(b)\cdot\mathbf{P}\left\{\text{Bin}(b-1,\rho) = a\right\}$ in probability, as required.

We now assume $\mu_1(p) > 2$, so that $\rho(p) = 1/(\mu_1(p)-1) < 1$. Since $p = (p_k, k \geq 1)$ is supported on the positive integers,

$$\sum_{a\geq 0} q(a) = \sum_{a\geq 0}\sum_{b=a+1}^{\infty} p(b)\mathbf{P}\left\{\text{Bin}(b-1,\rho) = a\right\} = \sum_{b\geq 1} p(b) = 1\,.$$

Recalling that $Q_{c^n}(a) = \sum_{b>a} P^n_{b,a}$, to show that $n^{-1}Q_{c^n}(a) \to q(a)$ in probability, it therefore suffices to prove that $P^n_{b+1,a}/n \to p(b+1)\cdot\mathbf{P}\left\{\text{Bin}(b,\rho) = a\right\}$ for all $0 \leq a \leq b$, and we now turn to this.

Since $\mu_1(p^n) \to \mu_1(p)$, it follows that $|\sum_{i=1}^n d^n(i) - \mu_1(p)n| = n|\mu_1(p^n) - \mu_1(p)| = o(n)$ as $n \to \infty$, so $s^n = (1+o(1))n(\mu_1(p)-1)$. Thus, for any $b \geq 1$ and $0 \leq a \leq b$ we have

$$\begin{aligned}
\binom{b}{a}\cdot\binom{s^n - b}{n-1-a}\binom{s^n}{n-1}^{-1} &= \binom{b}{a}\frac{(n-1)_a(s^n-(n-1))_{b-a}}{(s^n)_b} \\
&= (1-o(1))\binom{b}{a}\frac{n^a((\mu_1(p)-2)n)^{b-a}}{((\mu_1(p)-1)n)^b} \\
&= (1-o(1))\binom{b}{a}\frac{(\mu_1(p)-2)^{b-a}}{(\mu_1(p)-1)^b} \\
&= (1-o(1))\mathbf{P}\left\{\text{Bin}(b,\rho) = a\right\}.
\end{aligned}$$

Using (3.11) we thus have

$$\mathbf{E}P^n_{b+1,a} = (1-o(1))np^n(b+1)\mathbf{P}\left\{\text{Bin}(b,\rho) = a\right\} = (1-o(1))np(b+1)\mathbf{P}\left\{\text{Bin}(b,\rho) = a\right\}\,,$$

so again, applying Lemma 3.A.2 in the case that $p(b+1) > 0$, and applying Markov's inequality in the case that $p(b+1) = 0$, we obtain that, as $n \to \infty$,

$$\frac{P^n_{b+1,a}}{n} \xrightarrow{\text{prob}} p(b+1)\mathbf{P}\left\{\text{Bin}(b,\rho) = a\right\}\,,$$

as required. $\square$

We now turn to controlling the joint degrees of pairs of tree-adjacent vertices in tree-weighted graphs. Given a degree sequence $\mathrm{d} = (d(1), \ldots, d(n))$ and a tree t with $\mathrm{v}(t) = [n]$, for integers b_1, b_2, a_1, a_2 let

$$\begin{aligned}
&R_{b_1,b_2,a_1,a_2}(t, \mathrm{d}) \\
&= \#\{u \in \mathrm{v}(t) \setminus \{r(t)\} : d(u) = b_1, d(\mathrm{par}(u)) = b_2, c_t(u) = a_1, c_t(\mathrm{par}(u)) = a_2\} \\
&= \sum_{u \in \mathrm{v}(t) \setminus \{r(t)\}} \mathbf{1}_{[d(u)=b_1, c_t(u)=a_1]} \cdot \mathbf{1}_{[d(\mathrm{par}(u))=b_2, c_t(\mathrm{par}(u))=a_2]} \,.
\end{aligned}$$

If (g, t, γ) is a tree-rooted graph and g has degree sequence d, then $R_{b_1,b_2,a_1,a_2}(t, \mathrm{d})$ counts the number of edges xy of t with $y = \mathrm{par}(x)$ such that $c_t(x) = a_1$, $c_t(y) = a_2$ and $d_g(x) = b_1$, $d_g(y) = b_2$.

Proposition 3.A.3. *Under the assumptions of Theorem 3.1.1, for any integers* $0 \le a_1 < b_1$ *and* $0 \le a_2 < b_2$*, as* $n \to \infty$*,*

$$\frac{R_{b_1,b_2,a_1,a_2}(T(\mathrm{d}^n), \mathrm{d}^n)}{n} \to a_2 p(b_2) \mathbf{P}\left\{\mathrm{Bin}(b_2 - 1, \rho) = a_2\right\} \cdot p(b_1) \mathbf{P}\left\{\mathrm{Bin}(b_1 - 1, \rho) = a_1\right\}$$

in probability, where $\rho = 1/(\mu_1(p) - 1)$*.*

We introduce two pieces of notation before beginning the proof. For a half-edge h we write $v(h)$ for the vertex incident to h. Also, for $r \in \mathbb{R}$ we write $r_+ := \max(r, 0)$.

Proof. First, if $a_2 = 0$ then the right-hand side is zero, and also $R_{b_1,b_2,a_1,a_2}(T(\mathrm{d}^n), \mathrm{d}^n) = 0$, since if $v = \mathrm{par}(u) \in T(\mathrm{d}^n)$ then $c_{T(\mathrm{d}^n)}(v) \ge 1$. The result thus holds trivially when $a_2 = 0$, and we assume hereafter that $a_2 \ge 1$. For the remainder of the proof we write $R_{b_1,b_2,a_1,a_2} = R_{b_1,b_2,a_1,a_2}(T(\mathrm{d}^n), \mathrm{d}^n)$ for succinctness.

Let $\mathcal{H}$ be a fixed, size-$(n-1)$ subset of $S^n := \bigcup_{1 \le i \le n} \{i1, \ldots, i(d^n(i) - 1)\}$. Write $\mathcal{S}^n = \{s_1, \ldots, s_{n-1}\}$ for the (unordered) set of non-root half-edges of $T(\mathrm{d}^n)$. We now show that for any half edge $h \in \mathcal{H}$ and any root half-edge r with $v(r) \ne v(h)$, for all $1 \le i \le n-1$,

$$\mathbf{P}\left\{(r_i, s_i) = (r, h) \mid \mathcal{S}^n = \mathcal{H}\right\} = \mathbf{P}\left\{(r_1, s_1) = (r, h) \mid \mathcal{S}^n = \mathcal{H}\right\} . \tag{3.12}$$

To see this, note that by the second assertion of Proposition 3.2.1, the number of execution paths with $\mathcal{S}^n = \mathcal{H}$ is $((n-1)!)^2$. We claim that for any $i \in [n-1]$, the number of execution paths with $\mathcal{S}^n = \mathcal{H}$ which additionally satisfy that $(r_i, s_i) = (r, h)$ is $((n-2)!)^2$. As this number does not depend on $i \in [n-1]$, the displayed identity follows from this claim.

To prove the claim, simply note that there are $(n-2)!$ possible orderings of $\mathcal{H}$ consistent with the constraint that $s_i = h$. Having fixed such an ordering $(h_1, \ldots, h_{n-1})$, for each $j \in [n-1]$ with $j \neq i$, if $s_k = h_k$ for $1 \leq k < j$ then, excluding r_i there are $n - j - \mathbf{1}_{[j<i]}$ unpaired root half-edges in components different from that of s_j, and any such root half-edge may be chosen as r_j. Thus the number of execution paths with $\mathcal{S}^n = \mathcal{H}$ and such that $(r_i, s_i) = (r, h)$ is $(n-2)! \cdot \prod_{j\in[n-1]\setminus\{i\}}(n - j - \mathbf{1}_{[j<i]}) = ((n-2)!)^2$.

Now fix a second non-root half-edge $h' \neq h$ and a second root half-edge $r' \neq r$ not incident to the same vertex as h'. Then a similar argument to the one leading to (3.12) shows that that for any $1 \leq i < j \leq n$,

$$\begin{aligned} &\mathbf{P}\left\{(r_i, s_i) = (r, h), (r_j, s_j) = (r', s') \mid \mathcal{S}^n = \mathcal{H}\right\} \\ &= \mathbf{P}\left\{(r_1, s_1) = (r, h), (r_2, s_2) = (r', s') \mid \mathcal{S}^n = \mathcal{H}\right\}. \end{aligned} \tag{3.13}$$

In the current case, the number of execution paths leading to the events in both the left- and right-hand probabilities is $((n-3)!)^2$.

We will next use the above identities in order to perform first and second moment computations. For any set $H \subset S^n$, for $0 \leq a < b$ let $V^n_{b,a}(H) = \{i \in [n] : d^n(i) = b, |H \cap \{i1, \ldots, i(d^n(i) - 1)| = a\}$. Note that $V^n_{b,a}(\mathcal{S}^n)$ is simply the set of nodes with degree b in $G(\mathrm{d}^n)$ and with a children in $T(\mathrm{d}^n)$; so $P^n_{b,a} = |V^n_{b,a}(\mathcal{S}^n)|$.

Fix a non-root node $u \in T(\mathrm{d}^n)$, and let $m \in [n-1]$ be such that $e_m = \{\mathrm{par}(u), u\}$. Then $v(r_m) = u$ and $v(s_m) = \mathrm{par}(u)$, so $u \in R_{b_1,b_2,a_1,a_2}$ if and only if $v(r_m) \in V^n_{b_1,a_1}(\mathcal{S}^n)$ and $v(s_m) \in V^n_{b_2,a_2}(\mathcal{S}^n)$. By (3.12), it follows that

$$\begin{aligned} &\mathbf{E}\left(R_{b_1,b_2,a_1,a_2} \mid \mathcal{S}^n = \mathcal{H}\right) \\ &= (n-1)\mathbf{P}\left\{v(r_1) \in V^n_{b_1,a_1}(\mathcal{S}^n), v(s_1) \in V^n_{b_2,a_2}(\mathcal{S}^n) \mid \mathcal{S}^n = \mathcal{H}\right\}. \end{aligned} \tag{3.14}$$

Likewise, by (3.13) it follows that

$$\begin{aligned} &\mathbf{E}\left(\binom{R_{b_1,b_2,a_1,a_2}}{2} \mid \mathcal{S}^n = \mathcal{H}\right) \\ &= \binom{n-1}{2}\mathbf{P}\left\{v(r_1), v(r_2) \in V^n_{b_1,a_1}(\mathcal{S}^n), v(s_1), v(s_2) \in V^n_{b_2,a_2}(\mathcal{S}^n) \mid \mathcal{S}^n = \mathcal{H}\right\}. \end{aligned} \tag{3.15}$$

We develop the latter two identities in turn.

For integers $0 \leq a < b$, the number of non-root half-edges $h \in \mathcal{S}^n$ with $v(h) \in V^n_{b,a}(\mathcal{S}^n)$ is $a \cdot |V^n_{b,a}(\mathcal{S}^n)|$, and the number of root half-edges h with $v(h) \in V^n_{b,a}(\mathcal{S}^n)$ is just $|V^n_{b,a}(\mathcal{S}^n)|$.

Conditionally given that $\mathcal{S}^n = \mathcal{H}$, the half-edge s_1 is a uniformly random element of $\mathcal{H}$, so

$$\mathbf{P}\left\{v(s_1) \in V^n_{b_2,a_2}(\mathcal{S}^n) \mid \mathcal{S}^n = \mathcal{H}\right\} = \frac{a_2|V^n_{b_2,a_2}(\mathcal{H})|}{|\mathcal{H}|} = \frac{a_2|V^n_{b_2,a_2}(\mathcal{H})|}{n-1}\,.$$

Having chosen s_1, if $v(s_1) = v$ then $v(r_1)$ is a uniformly random element of $[n] \setminus \{v\}$, so

$$\mathbf{P}\left\{v(r_1) \in V^n_{b_1,a_1}(\mathcal{S}^n) \mid \mathcal{S}^n = \mathcal{H}, v(s_1) \in V^n_{b_2,a_2}(\mathcal{S}^n)\right\} = \frac{(|V^n_{b_1,a_1}(\mathcal{H})| - \mathbf{1}_{[(b_1,a_1)=(b_2,a_2)]})_+}{n-1}.$$

Using these two identities in (3.14), it follows that

$$(n-1)\mathbf{E}\left(R_{b_1,b_2,a_1,a_2} \mid \mathcal{S}^n = \mathcal{H}\right) = a_2|V^n_{b_2,a_2}(\mathcal{H})|(|V^n_{b_1,a_1}(\mathcal{H})| - \mathbf{1}_{[(b_1,a_1)=(b_2,a_2)]})_+\,,$$

so since $|V^n_{b,a}(\mathcal{S}^n)| = P^n_{b,a}$ for all $0 \le a < b$, by Proposition 3.3.1 we have

$$\begin{aligned}\mathbf{E}\left(\frac{R_{b_1,b_2,a_1,a_2}}{n} \mid \mathcal{S}^n\right) &= \frac{1}{n(n-1)}a_2 P^n_{b_2,a_2}(P^n_{b_1,a_1} - \mathbf{1}_{[(b_1,a_1)=(b_2,a_2)]}) \\ &\xrightarrow{\text{prob}} a_2 p(b_2)\mathbf{P}\left\{\mathrm{Bin}(b_2-1,\rho) = a_2\right\} \cdot p(b_1)\mathbf{P}\left\{\mathrm{Bin}(b_1-1,\rho) = a_1\right\}\,.\end{aligned} \tag{3.16}$$

For the second moment calculation, we need to additionally compute

$$\mathbf{P}\left\{v(r_2) \in V^n_{b_1,a_1}(\mathcal{S}^n), v(s_2) \in V^n_{b_2,a_2}(\mathcal{S}^n) \mid \mathcal{S}^n = \mathcal{H}, v(r_1) \in V^n_{b_1,a_1}(\mathcal{S}^n), v(s_1) \in V^n_{b_2,a_2}(\mathcal{S}^n)\right\}\,. \tag{3.17}$$

Under the conditioning in (3.17), the number of non-root half-edges $h \in \mathcal{S}^n \setminus \{s_1\}$ with $v(h) \in V^n_{b_2,a_2}(\mathcal{H})$ is $(a_2 \cdot |V^n_{b_2,a_2}(\mathcal{H})| - 1)_+$, so

$$\begin{aligned}&\mathbf{P}\left\{v(s_2) \in V^n_{b_2,a_2}(\mathcal{S}^n) \mid \mathcal{S}^n = \mathcal{H}, v(r_1) \in V^n_{b_1,a_1}(\mathcal{S}^n), v(s_1) \in V^n_{b_2,a_2}(\mathcal{S}^n)\right\} \\ &= \frac{(a_2 \cdot |V^n_{b_2,a_2}(\mathcal{H})| - 1)_+}{n-2}\,.\end{aligned}$$

Now suppose that $\mathcal{S}^n = \mathcal{H}, v(r_1) \in V^n_{b_1,a_1}(\mathcal{S}^n), v(s_1) \in V^n_{b_2,a_2}(\mathcal{S}^n)$, and that $v(s_2) \in V^n_{b_2,a_2}(\mathcal{H})$, and consider the number of possible values for r_2. We claim that the number of unpaired root half-edges h with $v(h) \in V^n_{b_1,a_1}(\mathcal{S}^n)$ such that $v(h)$ is in a component different from $v(s_2)$ is

$$(|V^n_{b_1,a_1}(\mathcal{H})| - 1 - \mathbf{1}_{[(b_1,a_1)=(b_2,a_2)]})_+.$$

To see this, note that if $(b_1, a_1) = (b_2, a_2)$ and either $v(s_2) = v(s_1)$ or $v(s_2) = v(r_1)$, then we are precisely constrained constrained to choose h so that $v(h) \in V^n_{b_1,a_1}(\mathcal{H}) \setminus \{v(r_1), v(s_1)\}$. On the other hand, if $(b_1, a_1) = (b_2, a_2)$ and $v(s_2) \notin \{v(r_1), v(s_1)\}$ then we are constrained to choose h so that $v(h) \in V^n_{b_1,a_1}(\mathcal{H}) \setminus \{v(r_1), v(s_2)\}$. Both cases agree with the above formula. When $(b_1, a_1) = (b_2, a_2)$, the claim is straightforward, since in that case we are only constrained to choose h so that $v(h) \in V^n_{b_1,a_1}(\mathcal{H}) \setminus \{v(r_1)\}$. It follows that

$$
\begin{aligned}
&\mathbf{P}\left\{v(r_2) \in V^n_{b_1,a_1}(\mathcal{S}^n) \mid \mathcal{S}^n = \mathcal{H}, v(s_2) \in V^n_{b_2,a_2}(\mathcal{S}^n), v(r_1) \in V^n_{b_1,a_1}(\mathcal{S}^n), v(s_1) \in V^n_{b_2,a_2}(\mathcal{S}^n)\right\} \\
&= \frac{(|V^n_{b_1,a_1}(\mathcal{H})| - 1 - \mathbf{1}_{[(b_1,a_1) \neq (b_2,a_2)]})_+}{n-2}.
\end{aligned}
$$

Combining the above identities with (3.15) yields that

$$
\begin{aligned}
&2(n-1)(n-2)\mathbf{E}\left(\binom{R_{b_1,b_2,a_1,a_2}}{2} \mid \mathcal{S}^n = \mathcal{H}\right) \\
&= a_2|V^n_{b_2,a_2}(\mathcal{H})|\big(a_2|V^n_{b_2,a_2}(\mathcal{H})| - 1\big)_+ \\
&\cdot \big(|V^n_{b_1,a_1}(\mathcal{H})| - \mathbf{1}_{[(b_1,a_1) \neq (b_2,a_2)]}\big)_+ \big(|V^n_{b_1,a_1}(\mathcal{H})| - 1 - \mathbf{1}_{[(b_1,a_1) \neq (b_2,a_2)]}\big)_+,
\end{aligned}
$$

so since $a_2 \geq 1$, Proposition 3.3.1 implies that

$$
\begin{aligned}
&\mathbf{E}\left(\frac{R_{b_1,b_2,a_1,a_2}(R_{b_1,b_2,a_1,a_2} - 1)}{n^2} \mid \mathcal{S}^n\right) \\
&= \frac{a_2 P^n_{b_2,a_2}(a_2 P^n_{b_2,a_2} - 1)_+}{n(n-1)} \cdot \frac{(P^n_{b_1,a_1} - \mathbf{1}_{[(b_1,a_1) \neq (b_2,a_2)]})_+ (P^n_{b_1,a_1} - 1 - \mathbf{1}_{[(b_1,a_1) \neq (b_2,a_2)]})_+}{n(n-2)} \\
&\xrightarrow{\text{prob}} \big(a_2 p(b_2)\mathbf{P}\left\{\mathrm{Bin}(b_2 - 1, \rho) = a_2\right\} \cdot p(b_1)\mathbf{P}\left\{\mathrm{Bin}(b_1 - 1, \rho) = a_1\right\}\big)^2.
\end{aligned}
$$

Also, (3.16) implies that $\mathbf{E}\left(n^{-2}R_{b_1,b_2,a_1,a_2} \mid \mathcal{S}^n\right) \xrightarrow{\text{prob}} 0$, which with the preceding asymptotic implies that

$$
\mathbf{E}\left(\frac{R^2_{b_1,b_2,a_1,a_2}}{n^2} \mid \mathcal{S}^n\right) \xrightarrow{\text{prob}} \big(a_2 p(b_2)\mathbf{P}\left\{\mathrm{Bin}(b_2 - 1, \rho) = a_2\right\} \cdot p(b_1)\mathbf{P}\left\{\mathrm{Bin}(b_1 - 1, \rho) = a_1\right\}\big)^2.
$$

Combining this with (3.16) gives that

$$
\mathbf{E}\left(\left(\frac{R_{b_1,b_2,a_1,a_2}}{n}\right)^2 \mid \mathcal{S}^n\right) - \left(\mathbf{E}\left(\frac{R_{b_1,b_2,a_1,a_2}}{n} \mid \mathcal{S}^n\right)\right)^2 \xrightarrow{\text{prob}} 0\,;
$$

the conditional Chebyshev's inequality then gives that for all $\epsilon > 0$,

$$\mathbf{P}\left\{\left|\frac{R_{b_1,b_2,a_1,a_2}}{n} - \mathbf{E}\left(\frac{R_{b_1,b_2,a_1,a_2}}{n} \mid \mathcal{S}^n\right)\right| > \epsilon \mid \mathcal{S}^n\right\} \xrightarrow{\text{prob}} 0.$$

Taking expectations on the left of the previous inequality to remove the conditioning, and again using (3.16), this time to replace the term $\mathbf{E}\left(\frac{R_{b_1,b_2,a_1,a_2}}{n} \mid \mathcal{S}^n\right)$ in the probability by the constant $C := a_2 p(b_2)\mathbf{P}\left\{\mathrm{Bin}(b_2-1,\rho) = a_2\right\} \cdot p(b_1)\mathbf{P}\left\{\mathrm{Bin}(b_1-1,\rho) = a_1\right\}$, we obtain that

$$\mathbf{P}\left\{\left|\frac{R_{b_1,b_2,a_1,a_2}}{n} - C\right| > \epsilon\right\} \to 0,$$

as required. □

Proof of Proposition 3.3.2. We may reexpress $A^n(k,\ell)$ as

$$\begin{aligned} A^n(k,\ell) = \sum_{a_1,a_2\geq 0} \sum_{u\in \mathrm{v}(T(\mathrm{d}^n))} & \mathbf{1}_{[r(T(\mathrm{d}^n))\notin\{u,\mathrm{par}(u)\}]} \\ & \cdot \mathbf{1}_{[\mathrm{d}^n(u)=k+a_1+1,\mathrm{c}^n(u)=a_1]} \\ & \cdot \mathbf{1}_{[\mathrm{d}^n(\mathrm{par}(u))=\ell+a_2+1,\mathrm{c}^n(\mathrm{par}(u))=a_2]} \\ + \sum_{a_1,a_2\geq 0} \sum_{u\in \mathrm{v}(T(\mathrm{d}^n))} & \mathbf{1}_{[\mathrm{par}(u)=r(T(\mathrm{d}^n))]} \\ & \cdot \mathbf{1}_{[\mathrm{d}^n(u)=k+a_1+1,\mathrm{c}^n(u)=a_1]} \\ & \cdot \mathbf{1}_{[\mathrm{d}^n(\mathrm{par}(u))=\ell+a_2,\mathrm{c}^n(\mathrm{par}(u))=a_2]} \end{aligned}$$

For fixed $a_1, a_2 \geq 0$, if we replace $\mathbf{1}_{[r(T(\mathrm{d}^n))\notin\{u,\mathrm{par}(u)\}]}$ by $\mathbf{1}_{[u\neq r(T(\mathrm{d}^n))]}$ in the first double sum, then the inner sum is simply $R_{k+a_1+1,\ell+a_2+1,a_1,a_2}$. It follows that

$$\begin{aligned}
A^n(k,\ell) = &\sum_{a_1,a_2\geq 0} R_{k+a_1+1,\ell+a_2+1,a_1,a_2} \\
&+ \sum_{a_1,a_2\geq 0} \sum_{u\in \mathrm{v}(T(\mathrm{d}^n))} \mathbf{1}_{[\mathrm{par}(u)=r(T(\mathrm{d}^n))]} \\
&\qquad\qquad \cdot \mathbf{1}_{[\mathrm{d}^n(u)=k+a_1+1,\mathrm{c}^n(u)=a_1]} \\
&\qquad\qquad \cdot \mathbf{1}_{[\mathrm{d}^n(\mathrm{par}(u))=\ell+a_2,\mathrm{c}^n(\mathrm{par}(u))=a_2]} \\
&- \sum_{a_1,a_2\geq 0} \sum_{u\in \mathrm{v}(T(\mathrm{d}^n))} \mathbf{1}_{[\mathrm{par}(u)=r(T(\mathrm{d}^n))]} \\
&\qquad\qquad \cdot \mathbf{1}_{[\mathrm{d}^n(u)=k+a_1+1,\mathrm{c}^n(u)=a_1]} \\
&\qquad\qquad \cdot \mathbf{1}_{[\mathrm{d}^n(\mathrm{par}(u))=\ell+a_2+1,\mathrm{c}^n(\mathrm{par}(u))=a_2]}\,.
\end{aligned}$$

But each of the last two double sums is bounded by $c^n(r(T(\mathrm{d}^n)))$, since they both count each child of the root at most once. Under the assumptions of Theorem 3.1.1, by Fact 3.A.1 we have $c^n(r(T(\mathrm{d}^n))) \leq \max_{1\leq i\leq n} d^n(i) = o(n^{1/2})$, so the preceding identity gives

$$\left| A^n(k,\ell) - \sum_{a_1,a_2\geq 0} R_{k+a_1+1,\ell+a_2+1,a_1,a_2} \right| = o(n^{1/2}).$$

Since

$$\begin{aligned}
&n^{-1}R_{k+a_1+1,\ell+a_2+1,a_1,a_2} \\
&\xrightarrow{\text{prob}} a_2 p(\ell+a_2+1)\mathbf{P}\left\{\mathrm{Bin}(\ell+a_2,\rho)=a_2\right\}\cdot p(k+a_1+1)\mathbf{P}\left\{\mathrm{Bin}(k+a_1,\rho)=a_1\right\}
\end{aligned}$$

by Proposition 3.A.3, and summing the right-hand side of the last expression over $a_1,a_2\geq 0$ gives $\alpha(k,l)$, it follows that for any $\epsilon>0$,

$$\mathbf{P}\left\{A^n(k,l)/n \geq \alpha(k,l)-\epsilon\right\} \to 1.$$

But also $n^{-1}\sum_{k,l\geq 0} A^n(k,l) = |\mathrm{e}(T(\mathrm{d}^n))| = (n-1)/n \to 1$; so since $\sum_{k,l\geq 0}\alpha(k,l)=1$, we must in fact have that

$$\frac{A^n(k,l)}{n} \xrightarrow{\text{prob}} \alpha(k,l)$$

for all $k,l\geq 0$, as required.

It remains to show that $n^{-1}\sum_{k,l\geq 0} klA^n(k,l) \xrightarrow{\text{prob}} \sum_{k,l\geq 0} kl\alpha(k,l)$. For this we will exploit the exchangeability of Pitman's additive coalescent. Recall the notation $v(h)$ for the vertex incident to half-edge h. Note that for any $M \in \mathbb{N}$ we have

$$\begin{aligned}
\sum_{k,l\geq 0} klA^n(k,l) - \sum_{0\leq k,l\leq M} klA^n(k,l) &= \sum_{uv\in e(T^n)} d_-^n(u)d_-^n(v)\mathbf{1}_{[\max(d_-^n(u),d_-^n(v))>M]} \\
&\leq \sum_{uv\in e(T^n)} d^n(u)d^n(v)\mathbf{1}_{[\max(d^n(u),d^n(v))>M]} \\
&= \sum_{i=1}^{n-1} d^n(v(r_i))d^n(v(s_i))\mathbf{1}_{[\max(d^n(v(r_i)),d^n(v(s_i)))>M]}.
\end{aligned}$$

Now, by Proposition 3.2.1 (3) and the identity (3.12), for any $1 \leq i \leq n-1$ we have

$$\begin{aligned}
&\mathbf{E}\left(d^n(v(r_i))d^n(v(s_i))\mathbf{1}_{[\max(d^n(v(r_i)),d^n(v(s_i)))>M]}\right) \\
&= \mathbf{E}\left(d^n(v(r_1))d^n(v(s_1))\mathbf{1}_{[\max(d^n(v(r_1)),d^n(v(s_1)))>M]}\right),
\end{aligned}$$

and by the definition of Pitman's additive coalescent we have

$$\begin{aligned}
&\mathbf{E}\left(d^n(v(r_1))d^n(v(s_1))\mathbf{1}_{[\max(d^n(v(r_1)),d^n(v(s_1)))>M]}\right) \\
&= \sum_{u\in[n]}\sum_{v\in[n]} d^n(u)d^n(v)\mathbf{1}_{[\max(d^n(u),d^n(v)))>M]}\mathbf{P}\left\{v(s_1)=u, v(r_1)=v\right\} \\
&= \sum_{u\in[n]}\sum_{v\in[n]} d^n(u)d^n(v)\mathbf{1}_{[\max(d^n(u),d^n(v))>M]}\cdot\frac{d^n(u)}{n\mu_1(p^n)}\cdot\frac{1}{n-1},
\end{aligned}$$

where we have used that $\sum_{i\in[n]} d^n(i) = n\mu_1(p^n)$. Next,

$$\begin{aligned}
&\sum_{u\in[n]}\sum_{v\in[n]}(d^n(u))^2 d^n(v)\mathbf{1}_{[\max(d^n(u),d^n(v))>M]} \\
&\leq \left(\left(\sum_{u\in[n]:d^n(u)>M}(d^n(u))^2\right)\cdot\sum_{v\in[n]} d^n(v) + \left(\sum_{v\in[n]:d^n(v)>M} d^n(v)\right)\cdot\sum_{u\in[n]}(d^n(u))^2\right), \\
&= n\mu_1(p^n)\cdot\sum_{u\in[n]:d^n(u)>M}(d^n(u))^2 \;+\; n\mu_2(p^n)\cdot\sum_{v\in[n]:d^n(v)>M} d^n(v).
\end{aligned}$$

Since $p^n \to p$, $\mu_1(p^n) \to \mu_1(p)$ and $\mu_2(p_n) \to \mu_2(p)$, for any $\delta > 0$ we may choose $M = M(\delta)$ sufficiently large so that $\sum_{u\in[n]:d^n(u)>M}(d^n(u))^2 < \delta n$ and $\sum_{v\in[n]:d^n(v)>M} d^n(v) < \delta n$, for all

$n \geq 1$. For such M, the previous bound and the two identities which precede it yield that

$$\mathbf{E}\left(\sum_{k,l\geq 0} klA^n(k,l) - \sum_{0\leq k,l\leq M} klA^n(k,l)\right) \leq \frac{1}{\mu_1(p^n)}(\delta n\mu_1(p^n) + \delta n\mu_2(p^n)).$$

By Markov's inequality, it follows that for all $\epsilon > 0$ there is $M \in \mathbb{N}$ such that for all $n \in \mathbb{N}$,

$$\mathbf{P}\left\{\sum_{k,l\geq 0} klA^n(k,l) - \sum_{0\leq k,l\leq M} klA^n(k,l) > \epsilon n\right\} < \epsilon.$$

Finally, since $n^{-1}A^n(k,l) \xrightarrow{\text{prob}} \alpha(k,l)$, it follows that for all $M \in \mathbb{N}$ we have

$$\frac{1}{n}\sum_{0\leq k,l\leq M} klA^n(k,l) \xrightarrow{\text{prob}} \sum_{0\leq k,l\leq M} kl\alpha(k,l)\,,$$

so the preceding probability bound implies that

$$\frac{1}{n}\sum_{k,l\geq 0} klA^n(k,l) \xrightarrow{\text{prob}} \lim_{M\to\infty}\sum_{0\leq k,l\leq M} kl\alpha(k,l) = \sum_{k,l\geq 0} kl\alpha(k,l)\,,$$

as required. $\square$

3.B Proof of Theorem 3.4.1

Let H be a simple graph with vertex set $v(H) = [n]$, and let G be a random graph with degree sequence $\mathrm{d} = (d(1), \ldots, d(n))$ sampled according to the configuration model. Recall the definitions of $\mathcal{L}(G), \mathcal{M}(G,H), \mathcal{N}(G,H)$ and $L(G), M(G,H), N(G,H)$ from Section 3.4. The first subsection will provide a quantitative approximation result for mixed moments of L, M and N. In the second subsection, we will use this approximation to prove Theorem 3.4.1.

3.B.1 Deterministic bounds on loops and multi-edges

Our arguments in this section are based on and fairly closely parallel those from [68, Chapter 7]. We recall the falling factorial notation $(x)_\ell = x(x-1)\ldots(x-\ell+1)$. In what follows, it is convenient to define $(x)_\ell = 1$ if $\ell = 0$, and $(x)_\ell = 0$ if $\ell < 0$.

Proposition 3.B.1. *Write $m = \frac{1}{2}\sum_{i=1}^n d(i)$, and write $d_{\max} = \max\{d(1), \ldots, d(n)\}$. For* *any positive integers $q, r, s \in \mathbb{N}$,*

$$\left| \mathbf{E}\left((L)_q (M)_r (N)_s\right) - \frac{(|\mathcal{L}|)_q (|\mathcal{M}|)_r (|\mathcal{N}|)_s}{\prod_{i=0}^{q+2r+s-1} 2m-1-2i} \right| \le C(S_1 + S_2)$$

where C is a constant depending only on q, r and s, S_1 is defined by the following identity,

$$\begin{aligned} S_1 \prod_{i=0}^{q+2r+s-1} (2m-1-2i) =& (|\mathcal{L}|)_{q-2} (|\mathcal{M}|)_r (|\mathcal{N}|)_s \sum_{1 \le u \le n} d(u)^3 \\ &+ (|\mathcal{L}|)_{q-1} (|\mathcal{M}|)_{r-1} (|\mathcal{N}|)_s \sum_{1 \le u \ne v \le n} d(u)^3 d(v)^2 \\ &+ (|\mathcal{L}|)_{q-1} (|\mathcal{M}|)_r (|\mathcal{N}|)_{s-1} \sum_{uv \in e(H)} d(u)^2 d(v) \\ &+ (|\mathcal{L}|)_q (|\mathcal{M}|)_{r-2} (|\mathcal{N}|)_s \sum_{\substack{1 \le u \le n \\ u \notin \{v_1, v_2\}}} d(u)^3 d(v_1)^2 d(v_2)^2 \\ &+ (|\mathcal{L}|)_q (|\mathcal{M}|)_{r-1} (|\mathcal{N}|)_{s-1} \sum_{\substack{1 \le u, v_1, v_2 \le n \\ u, v_1, v_2 \, distinct \\ uv_2 \in e(H)}} d(u)^2 d(v_1)^2 d(v_2) \\ &+ (|\mathcal{L}|)_q (|\mathcal{M}|)_r (|\mathcal{N}|)_{s-2} \sum_{\substack{uv_1, uv_2 \in e(H) \\ v_1 \ne v_2}} d(u) d(v_1) d(v_2), \end{aligned}$$

and S_2 is defined by

$$S_2 = (|\mathcal{L}|)_q (|\mathcal{N}|)_s \sum_{k=1}^{r-1} (|\mathcal{M}|)_{r-k} \sum_{\ell=0}^{k} d_{\max}^{2\ell} \prod_{i=0}^{q+2r+s-1-(2k-\ell)} \frac{1}{2m-1-2i}.$$

Proof. Throughout the proof, write

$$(x, y, z) = \big((x_1, \ldots, x_q), (y_1, \ldots, y_r), (z_1, \ldots, z_s)\big) \in \mathcal{L}^q \times \mathcal{M}^r \times \mathcal{N}^s$$

to denote a generic element of $\mathcal{L}^q \times \mathcal{M}^r \times \mathcal{N}^s$. For $(x, y, z) \in \mathcal{L}^q \times \mathcal{M}^r \times \mathcal{N}^s$, write

$$\mathbf{1}_{[x]} = \prod_{i=1}^{q} \mathbf{1}_{[x_i]}, \qquad \mathbf{1}_{[y]} = \prod_{i=1}^{r} \mathbf{1}_{[y_i]}, \qquad \mathbf{1}_{[z]} = \prod_{i=1}^{s} \mathbf{1}_{[z_i]}.$$

We say (x, y, z) is *non-repeating* if $x_1, x_2, \ldots, x_q$ are pairwise distinct, $y_1, y_2, \ldots, y_r$ are pairwise distinct, and $z_1, z_2, \ldots, z_s$ are pairwise distinct. In what follows, write

$$\sum\nolimits^{\star} := \sum_{\substack{(x,y,z)\in\mathcal{L}^q\times\mathcal{M}^r\times\mathcal{N}^s \\ (x,y,z) \text{ is non-repeating}}},$$

and, for $S \subset \mathcal{L}^q \times \mathcal{M}^r \times \mathcal{N}^s$, write

$$\sum_S{}^{\star} := \sum_{\substack{(x,y,z)\in S \\ (x,y,z) \text{ is non-repeating}}}.$$

Note that $(L)_q(M)_r(N)_s = \sum^{\star} \mathbf{1}_{[x]}\mathbf{1}_{[y]}\mathbf{1}_{[z]}$, so

$$\mathbf{E}\left((L)_q(M)_r(N)_s\right) = \sum\nolimits^{\star} \mathbf{P}\left\{\mathbf{1}_{[x]}\mathbf{1}_{[y]}\mathbf{1}_{[z]} = 1\right\}. \tag{3.18}$$

We say (x, y, z) is *non-conflicting* if the $2q + 4r + 2s$ half-edges appearing in x, y and z are pairwise distinct, and otherwise we say (x, y, z) is conflicting. Since half-edges in G are paired uniformly at random, for non-conflicting (x, y, z) we have

$$\mathbf{P}\left\{\mathbf{1}_{[x]}\mathbf{1}_{[y]}\mathbf{1}_{[z]} = 1\right\} = \prod_{i=0}^{q+2r+s-1} \frac{1}{2m-1-2i}. \tag{3.19}$$

Now, for a given (x, y, z), let $er(x, y, z)$ be defined as follows:

$$er(x,y,z) := \mathbf{P}\left\{\mathbf{1}_{[x]}\mathbf{1}_{[y]}\mathbf{1}_{[z]} = 1\right\} - \prod_{i=0}^{q+2r+s-1} \frac{1}{2m-1-2i}.$$

By (3.19), if (x, y, z) is non-conflicting then $er(x, y, z) = 0$, so

$$\begin{aligned}
\mathbf{E}\left((L)_q(M)_r(N)_s\right) &= \sum\nolimits^{\star} \prod_{i=0}^{q+2r+s-1} \frac{1}{2m-1-2i} + \sum\nolimits^{\star} er(x,y,z) \\
&= \frac{(|\mathcal{L}|)_q(|\mathcal{M}|)_r(|\mathcal{N}|)_s}{\prod_{i=0}^{q+2r+s-1} 2m-1-2i} + \sum_{(x,y,z) \text{ is conflicting}}^{\star} er(x,y,z).
\end{aligned}$$

By the triangle inequality this implies

$$\left|\mathbf{E}\left((L)_q(M)_r(N)_s\right) - \frac{(|\mathcal{L}|)_q(|\mathcal{M}|)_r(|\mathcal{N}|)_s}{\prod_{i=0}^{q+2r+s-1} 2m-1-2i}\right| \le \sum_{(x,y,z) \text{ is conflicting}}^{\star} |er(x,y,z)|. \tag{3.20}$$

To bound the error terms, we must make a distinction between two types of conflicts. This distinction is most easily understood by way of an example. On the one hand, suppose $x_1 = (ui, uj_1)$ and $x_2 = (ui, uj_2)$ for $u \in V(G)$ and distinct $i, j_1, j_2 \in [d(u)]$. Then $\mathbf{1}_{[x_1]}\mathbf{1}_{[x_2]} = 1$ is the event that the half edge ui is joined to uj_1 and uj_2 simultaneously, and $\mathbf{P}\left\{\mathbf{1}_{[x_1]}\mathbf{1}_{[x_2]} = 1\right\} = 0$. On the other hand, suppose $y_1 = (ui_1, vj_1), (ui_2, vj_2)$ and $y_2 = (ui_1, vj_1), (ui_3, vj_3)$ for distinct $u, v \in V(G)$, distinct $i_1, i_2, i_3 \in [d(u)]$, and distinct $j_1, j_2, j_3 \in [d(v)]$. Then $\mathbf{1}_{[y_1]}\mathbf{1}_{[y_2]} = 1$ is the event that u and v are connected by a triple edge, and $\mathbf{P}\left\{\mathbf{1}_{[y_1]}\mathbf{1}_{[y_2]} = 1\right\} > 0$.

We say a conflicting triple (x, y, z) is a *bad conflict* if $\mathbf{P}\left\{\mathbf{1}_{[x]}\mathbf{1}_{[y]}\mathbf{1}_{[z]} = 1\right\} = 0$, and otherwise we say (x, y, z) is a *good conflict*. In the above examples, the first is a bad conflict and the second is a good conflict. Let $\mathcal{B} = \mathcal{B}(q, r, s) \subseteq \mathcal{L}^q \times \mathcal{M}^r \times \mathcal{N}^s$ and $\mathcal{G} = \mathcal{G}(q, r, s) \subseteq \mathcal{L}^q \times \mathcal{M}^r \times \mathcal{N}^s$ be the collections of bad conflicts and good conflicts respectively. The rest of this proof is dedicated to bounding $|\mathcal{B}|, |\mathcal{G}|$, and $er(x, y, z)$ for $(x, y, z) \in \mathcal{B} \cup \mathcal{G}$.

(Bounding $|\mathcal{B}|$): If (x, y, z) is a bad conflict then one of the following must hold (in reading the below descriptions, it may be useful to consult Figure 3.2):

1. There exists $1 \leq a < b \leq q$, $u \in V(G)$ and distinct $i_1, i_2, i_3 \in [d(u)]$ such that $x_a = (ui_1, ui_2)$ and $x_b = (ui_1, ui_3)$. Write $\mathcal{B}_{xx}$ for the set of $(x, y, z) \in \mathcal{B}$ that contain a pair (x_a, x_b) of this form.

2. There exists $1 \leq a \leq q$, $1 \leq b \leq r$, distinct $u, v \in V(G)$, distinct $i_1, i_2, i_3 \in [d(u)]$, and distinct $j_1, j_2 \in [d(v)]$ such that $x_a = (ui_1, ui_3)$ and $y_b = (ui_1, vj_1), (ui_2, vj_2)$. Write $\mathcal{B}_{xy}$ for the set of $(x, y, z) \in \mathcal{B}$ that contain a pair (x_a, y_b) of this form.

3. There exists $1 \leq a \leq q$, $1 \leq b \leq s$, $uv \in e(H)$, distinct $i_1, i_2 \in [d(u)]$, and $j \in [d(v)]$ such that $x_a = (ui_1, ui_2)$ and $z_b = (ui_1, vj)$. Write $\mathcal{B}_{xz}$ for the set of $(x, y, z) \in \mathcal{B}$ that contain a pair (x_a, z_b) of this form.

4. There exists $1 \leq a < b \leq r$, distinct $u, v_1, v_2 \in V(G)$, distinct $i_1, i_2, i_3 \in [d(u)]$, distinct $j_1, j_2 \in [d(v_1)]$, and distinct $k_1, k_2 \in [d(v_2)]$ such that $y_a = (ui_1, v_1j_1), (ui_2, v_1j_2)$ and $y_b = (ui_1, v_2k_1), (ui_3, v_2k_2)$. Write $\mathcal{B}_{yy}$ for the set of $(x, y, z) \in \mathcal{B}$ that contain a pair (y_a, y_b) of this form.

5. There exists $1 \leq a \leq r$, $1 \leq b \leq s$ and distinct $u, v_1, v_2 \in V(G)$ such that $uv_2 \in e(H)$, distinct $i_1, i_2 \in [d(u)]$, distinct $j_1, j_2 \in [d(v_1)]$, and $k \in [d(v_2)]$ such that $y_a =$

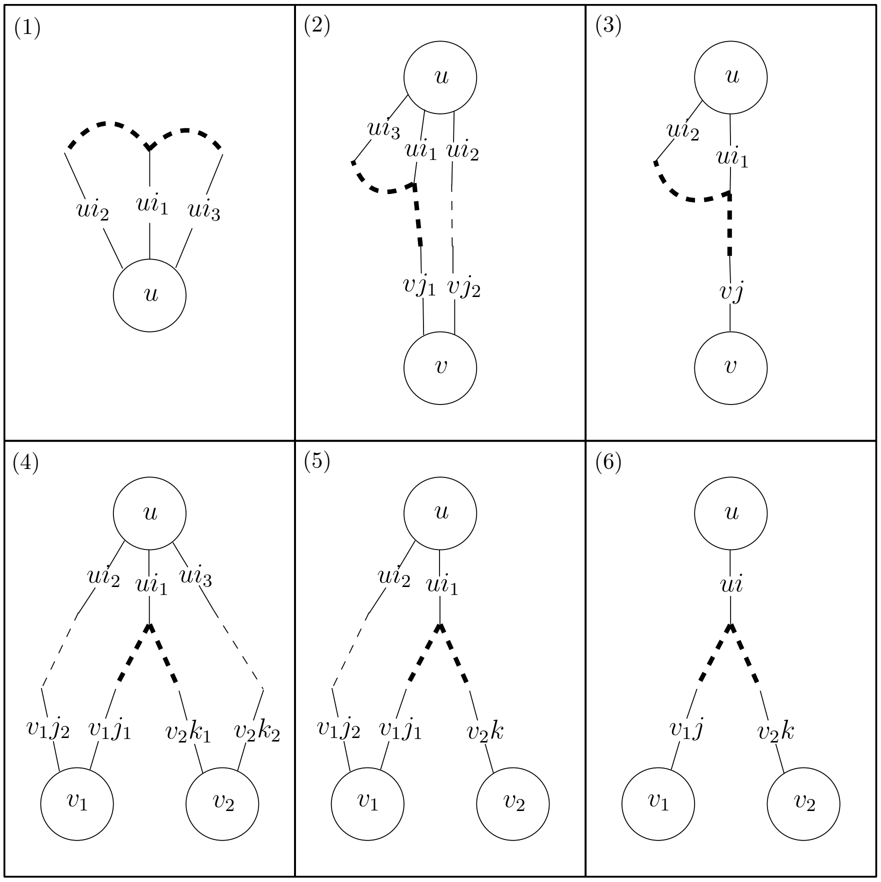

Figure 3.2: An example of each of the six types of "bad" conflicts, depicted in the same order as they are described in the text. In the drawing, dashed lines represent the connections between half-edges required by the event. The bold dashed lines show the locations at which two half-edges must pair with a single half-edge, rendering the corresponding event impossible.

$(ui_1, v_1j_1), (ui_2, v_1j_2)$ and $z_b = (ui_1, v_2k)$. Write $\mathcal{B}_{yz}$ for the set of $(x, y, z) \in \mathcal{B}$ that contain a pair (y_a, z_b) of this form.

6. There exists $1 \leq a < b \leq s$, distinct $uv_1, uv_2 \in e(H)$, $i \in [d(u)]$, $j \in [d(v_1)]$, and $k \in [d(v_2)]$ such that $z_a = (ui, v_1j)$ and $z_b = (ui, v_2k)$. Write $\mathcal{B}_{zz}$ for the set of $(x, y, z) \in \mathcal{B}$ that contain a pair (z_a, z_b) of this form.

We next turn to bounding the sizes of each set, starting with $\mathcal{B}_{xx}$. The number of choices for a and b is $q(q-1)/2$. Having chosen these, for each possible choice of $u \in V(G)$, there are less than $d(u)^3$ choices for i_1, i_2 and i_3. Then, having chosen these, we must choose one of i_1, i_2 and i_3 to be repeated in x_a and x_b. Lastly, we must choose the remaining $q-2$ entries for x, r entries for y, and s entries for z. Hence,

$$\begin{aligned}|\mathcal{B}_{xx}| &\leq (|\mathcal{L}|)_{q-2}(|\mathcal{M}|)_r(|\mathcal{N}|)_s q(q-1)/2 \sum_{1\leq u\leq n} 3d(u)^3\\ &= C_1(|\mathcal{L}|)_{q-2}(|\mathcal{M}|)_r(|\mathcal{N}|)_s \sum_{1\leq u\leq n} d(u)^3,\end{aligned}$$

where $C_1 = C_1(q) = 3q(q-1)/2$. Note that $\mathcal{B}_{xx}$ is empty if $q \leq 1$, so $|\mathcal{B}_{xx}| = 0$. In this case the right hand side is also zero by our convention that $(k)_\ell = 0$ for $\ell < 0$. Therefore, the bound also holds for $q \leq 1$. The subsequent bounds can likewise be seen to hold when the right hand side is zero, though we do not explicitly verify this in every case.

When building an element of $\mathcal{B}_{xy}$, the number of ways to choose a and b is qr. Having chosen these, for each pair $u, v \in V(G)$, there are less than $d(u)^3d(v)^2$ choices for $i_1, i_2, i_3 \in [d(u)]$ and $j_1, j_2 \in [d(v)]$. Then, there are a constant number of ways to arrange the half-edges in x_a and y_b. Lastly, we must choose the remaining entries for x, y and z. Hence,

$$|\mathcal{B}_{xy}| \leq C_2(|\mathcal{L}|)_{q-1}(|\mathcal{M}|)_{r-1}(|\mathcal{N}|)_s \sum_{1\leq u\neq v\leq n} d(u)^3d(v)^2,$$

where C_2 depends only on q and r.

For an element of $\mathcal{B}_{xz}$, for each $uv \in e(H)$, there are less than $d(u)^2d(v)$ ways to choose $i_1, i_2 \in [d(u)]$ and $j \in [d(v)]$. Hence,

$$|\mathcal{B}_{xz}| \leq C_3(|\mathcal{L}|)_{q-1}(|\mathcal{M}|)_r(|\mathcal{N}|)_{s-1} \sum_{uv\in e(H)} d(u)^2d(v).$$

For $\mathcal{B}_{yy}$, for each $u, v_1, v_2 \in V(G)$, there are less than $d(u)^3d(v_1)^2d(v_2)^2$ ways to choose $i_1, i_2, i_3 \in [d(u)]$, $j_1, j_2 \in [d(v_1)]$, and $k_1, k_2 \in [d(v_2)]$. Hence,

$$|\mathcal{B}_{yy}| \leq C_4(|\mathcal{L}|)_q(|\mathcal{M}|)_{r-2}(|\mathcal{N}|)_s \sum_{\substack{1\leq u,v_1,v_2\leq n\\ u,v_1,v_2 \text{ distinct}}} d(u)^3d(v_1)^2d(v_2)^2.$$

For $\mathcal{B}_{yz}$, for each $u, v_1, v_2 \in V(G)$ such that $uv_2 \in e(H)$, there are less than $d(u)^2d(v_1)^2d(v_2)$ ways to choose $i_1, i_2 \in [d(u)]$, $j_1, j_2 \in [d(v_1)]$, and $k \in [d(v_2)]$. Hence,

$$|\mathcal{B}_{yz}| \leq C_5(|\mathcal{L}|)_q(|\mathcal{M}|)_{r-1}(|\mathcal{N}|)_{s-1} \sum_{\substack{1\leq u,v_1,v_2\leq n\\ u,v_1,v_2 \text{ distinct}\\ uv_2\in e(H)}} d(u)^2d(v_1)^2d(v_2).$$

Lastly, for $\mathcal{B}_{zz}$, for each $uv_1, uv_2 \in e(H)$, there are less than $d(u)d(v_1)d(v_2)$ ways to choose $i \in [d(u)]$, $j \in [d(v_1)]$ and $k \in [d(v_2)]$. Hence,

$$|\mathcal{B}_{zz}| \leq C_6(|\mathcal{L}|)_q(|\mathcal{M}|)_r(|\mathcal{N}|)_{s-2} \sum_{\substack{uv_1,uv_2 \in e(H) \\ v_1 \neq v_2}} d(u)d(v_1)d(v_2).$$

Note that the values of C_1, C_2, C_3, C_4, C_5 and C_6 depend only on q, r and s.

(Bounding $|er(x,y,z)|$ for $(x,y,z) \in \mathcal{B}$): If $(x,y,z) \in \mathcal{B}$ then $\mathbf{P}\left\{\mathbf{1}_{[x]}\mathbf{1}_{[y]}\mathbf{1}_{[z]} = 1\right\} = 0$, meaning

$$|er(x,y,z)| = \prod_{i=0}^{q+2r+s-1} \frac{1}{2m-1-2i} \tag{3.21}$$

(Bounding $|\mathcal{G}|$): Suppose $(x,y,z) \in \mathcal{G}$. Then (x,y,z) is conflicting, meaning a half-edge appears more than once in $x \cup y \cup z$. However, since $\mathbf{P}\left\{\mathbf{1}_{[x]}\mathbf{1}_{[y]}\mathbf{1}_{[z]} = 1\right\} > 0$ for $(x,y,z) \in \mathcal{G}$, it cannot be the case where $\mathbf{1}_{[x]}\mathbf{1}_{[y]}\mathbf{1}_{[z]}$ contains the event that a half-edge is paired to two different half-edges simultaneously. Hence, there must be a half-edge pair that appears more than once in $x \cup y \cup z$. Furthermore, this half-edge pair must appear more than once in y, since the edges in y and z are disjoint. It follows that any $(x,y,z) \in \mathcal{G}$ can be constructed in the following way:

1. Choose x and z arbitrarily. The number of choices here is $(|\mathcal{L}|)_q(|\mathcal{N}|)_s$.

2. Choose a set of indices $1 \leq a_1 < a_2 < \cdots < a_k \leq r$ and arbitrarily choose the elements $y_a \in y$ such that $a \neq a_i$ for all $1 \leq i \leq k$. The number of choices here is $(|\mathcal{M}|)_{r-k}$ times a constant in terms of r.

3. Choose $1 \leq b_1 \leq \cdots \leq b_\ell \leq r$ such that $\{b_1, \ldots, b_\ell\} \subseteq \{a_1, \ldots, a_k\}$. For each $1 \leq i \leq \ell$, build y_{b_i} by choosing a half-edge pair already in y, then choosing the other half-edge pair arbitrarily. The number of choices for each i is less than $d_{\max}^2$ times a constant in terms of r.

4. For each $a \in \{a_1, \ldots, a_k\} \setminus \{b_1 \ldots, b_\ell\}$, build y_a by choosing two half-edge pairs already in y. The number of choices here is a constant in terms of r.

Since every element of $\mathcal{G}$ can be constructed in this way, we get

$$|\mathcal{G}| \leq C(r)(|\mathcal{L}|)_q(|\mathcal{N}|)_s \sum_{k=1}^{r-1}(|\mathcal{M}|)_{r-k} \sum_{\ell=0}^{k} d_{\max}^{2\ell}.$$

(Bounding $er(x,y,z)$ for $(x,y,z) \in \mathcal{G}$): Let $(x,y,z) \in \mathcal{G}$ and suppose we can construct (x,y,z) as above with a particular k and ℓ. Then there are $2k-\ell$ half-edge pairs that are redundant when calculating $\mathbf{P}\left\{\mathbf{1}_{[x]}\mathbf{1}_{[y]}\mathbf{1}_{[z]}=1\right\}$. Hence,

$$\mathbf{P}\left\{\mathbf{1}_{[x]}\mathbf{1}_{[y]}\mathbf{1}_{[z]}=1\right\} = \prod_{i=0}^{q+2r+s-1-(2k-\ell)} \frac{1}{2m-1-2i}.$$

Therefore, for such $(x,y,z) \in \mathcal{G}$,

$$\begin{aligned} |er(x,y,z)| &= \left| \prod_{i=0}^{q+2r+s-1} \frac{1}{2m-1-2i} - \prod_{i=0}^{q+2r+s-1-(2k-\ell)} \frac{1}{2m-1-2i} \right| \\ &\leq \prod_{i=0}^{q+2r+s-1-(2k-\ell)} \frac{1}{2m-1-2i}\,. \end{aligned} \tag{3.22}$$

Finally, by (3.20), we know that

$$\left| \mathbf{E}\left((L)_q(M)_r(N)_s\right) - \frac{(|\mathcal{L}|)_q(|\mathcal{M}|)_r(|\mathcal{N}|)_s}{\prod_{i=0}^{q+2r+s-1} 2m-1-2i} \right| \leq \sum_{(x,y,z)\in\mathcal{B}}^{\star} |er(x,y,z)| + \sum_{(x,y,z)\in\mathcal{G}}^{\star} |er(x,y,z)|\,,$$

from which the result now follows by using the bounds on $|\mathcal{B}_{xx}|, |\mathcal{B}_{xy}|, |\mathcal{B}_{xz}|, |\mathcal{B}_{yy}|, |\mathcal{B}_{yz}|, |\mathcal{B}_{zz}|$ and on $|\mathcal{G}|$, together with (3.21) and (3.22) □

3.B.2 The probability of simplicity for a random superposition of graphs

Before proving Theorem 3.4.1, it will be useful to show some auxiliary bounds.

Lemma 3.B.2. *Under the assumptions of Theorem 3.4.1, we have*

$$\sum_{v\in[n]} d^n(v) = O(n), \tag{3.23}$$

$$\sum_{v\in[n]} \left(d^n(v)\right)^2 = O(n), \tag{3.24}$$

$$\sum_{uv\in e(h_n)} d^n(u)d^n(v) = O(n), \tag{3.25}$$

and

$$\sup_{u\in[n]} \sum_{v:uv\in e(h_n)} d^n(v) = o(n), \tag{3.26}$$

Proof. Equations (3.23) and (3.24) follow from the fact that $\mu_1(p^n) \to \mu_1(p) < \infty$ and $\mu_2(p^n) \to \mu_2(p) < \infty$. Indeed, we have

$$\sum_{v\in[n]} d^n(v) = n\sum_{k\geq 1} kp^n(k) = n\mu_1(p^n) = O(n),$$

and

$$\sum_{v\in[n]} (d^n(v))^2 = n\sum_{k\geq 1} k^2p^n(k) = n\mu_2(p^n) = O(n).$$

Equation (3.25) follows from the convergence of $\sum_{i,j\geq 1} ij\alpha^n(i,j)$. Notice that, by the definition of α^n,

$$\sum_{uv\in e(h_n)} d^n(u)d^n(v) = \sum_{i,j\geq 1}\left(\sum_{uv\in e(h_n):d^n(u)=i,d^n(v)=j} ij\right) = \sum_{i,j\geq 1} ij(n\alpha^n(i,j)).$$

Hence,

$$\frac{1}{n}\sum_{uv\in e(h_n)} d^n(u)d^n(v) \to \sum_{i,j\geq 1} ij\alpha(i,j) < \infty,$$

implying that

$$\sum_{uv\in e(h_n)} d^n(u)d^n(v) = O(n).$$

We will prove the fourth bound by contradiction. To this end, suppose (3.26) fails. Then we can find $c > 0$ and a sequence of vertices $(u_n, n \geq 1)$ with $u_n \in v(h_n)$ such that for all n sufficiently large,

$$\sum_{v:u_nv\in e(h_n)} d^n(v) \geq cn. \tag{3.27}$$

Write $\deg(u_n) = \deg_{h_n}(u_n)$. List the neighbours of u_n in h_n as $N_{h_n}(u_n) = \{v_1^n, \ldots, v_{\deg(u_n)}^n\}$ so that $d^n(v_i^n) \geq d^n(v_{i+1}^n)$ for all $1 \leq i < \deg(u_n)$.

Next, fix $D \in \mathbb{N}$ and let $k = k(n) = \max\{i : d^n(v_i^n) \geq D\}$. Then

$$\sum_{i=k+1}^{\deg(u_n)} d^n(v_i^n) \leq (\deg(u_n) - k)(D-1) = o(n),$$

the last bound holding since $\deg(u_n) = o(n)$ by assumption. Thus,

$$\begin{aligned}
\sum_{v:u_n v \in e(h_n)} d^n(v)^2 &= \sum_{i=1}^{\deg(u_n)} d^n(v_i^n)^2 \\
&\geq \sum_{i=1}^{k} d^n(v_i^n)^2 \\
&\geq D \sum_{i=1}^{k} d^n(v_i^n) \\
&= D\left(\sum_{v:u_n v \in e(h_n)} d^n(v) - o(n) \right) \\
&\geq D(c - o(1))n,
\end{aligned}$$

the last bound holding by (3.27). Since $D \in \mathbb{N}$ was arbitrary, it follows that

$$\sum_{v \in [n]} d^n(v)^2 \geq \sum_{v:u_n v \in e(h_n)} d^n(v)^2 = \omega(n),$$

contradicting (3.24). $\square$

The next lemma is the last ingredient needed, and also assumes $p(0) + p(1) < 1$, i.e. an asymptotically non-zero proportion of the degrees in G_n are 2 or greater. We will show later that Theorem 3.4.1 is straightforward when $p(0) + p(1) = 1$.

Lemma 3.B.3. *Under the assumptions of Theorem 3.4.1, suppose additionally that $p(0) + p(1) < 1$. Then*

$$|\mathcal{L}(G_n)| = \Theta(n), \tag{3.28}$$

and

$$|\mathcal{M}(G_n, h_n)| = \Theta(n^2). \tag{3.29}$$

Proof. Let $\mathcal{L}_n = \mathcal{L}(G_n), \mathcal{M}_n = \mathcal{M}(G_n, h_n)$, and $\mathcal{N}_n = \mathcal{N}(G_n, h_n)$. By their definitions, we have

$$\begin{aligned}
|\mathcal{L}_n| &= \sum_{v \in [n]} \frac{d^n(v)\,(d^n(v) - 1)}{2}, \text{ and} \\
|\mathcal{M}_n| &= \sum_{\{u<v:uv \notin e(h_n)\}} \frac{d^n(u)\,(d^n(u) - 1)}{2} d^n(v)\,(d^n(v) - 1)\,.
\end{aligned}$$

For the upper bounds, by Lemma 3.B.2 we have

$$|\mathcal{L}_n| = \sum_{v\in[n]} \frac{d^n(v)\left(d^n(v)-1\right)}{2} \leq \sum_{v\in[n]} d^n(v)^2 = O(n), \text{ and}$$

$$|\mathcal{M}_n| = \sum_{\{u<v:uv\notin e(h_n)\}} \frac{d^n(u)\left(d^n(u)-1\right)}{2} d^n(v)\left(d^n(v)-1\right) \leq \left(\sum_{v\in[n]} d^n(v)^2\right)^2 = O(n^2).$$

For the lower bounds, first notice that

$$\begin{aligned}
|\mathcal{L}_n| &= \sum_{v\in[n]} \frac{d^n(v)\left(d^n(v)-1\right)}{2} \\
&= \sum_{v\in[n]:d^n(v)>1} \frac{d^n(v)\left(d^n(v)-1\right)}{2} \\
&\geq |\{v\in[n]: d^n(v)>1\}| \\
&= n(1-p^n(0)-p^n(1))
\end{aligned}$$

Since $p^n(0)+p^n(1)\to p(0)+p(1)<1$ by assumption, this implies $|\{v\in[n]: d^n(v)>1\}| = \Theta(n)$, and so

$$|\mathcal{L}_n| \geq |\{v\in[n]: d^n(v)>1\}| = \Theta(n).$$

Similarly, we have

$$|\mathcal{M}_n| \geq \Big|\{(u,v): u<v, uv\notin e(h_n) \text{ and } d^n(u), d^n(v)>1\}\Big|.$$

From the conditions of Theorem 3.4.1 we know that $\max_{v\in[n]}\{\deg_{h_n}(v)\} = o(n)$, which implies that $|e(h_n)| = o(n^2)$. Hence, we have

$$\begin{aligned}
&\Big|\{(u,v): u<v, uv\notin e(h_n) \text{ and } d^n(u), d^n(v)>1\}\Big| \\
=&\Big|\{(u,v): u<v, d^n(u), d^n(v)>1\}\Big| - o(n^2),
\end{aligned}$$

and writing $k=k(n)=|\{v\in[n]: d^n(v)>1\}|$, we have

$$\Big|\{(u,v): u<v, d^n(u), d^n(v)>1\}\Big| = \binom{k}{2} = \Theta(k^2) = \Theta(n^2),$$

and hence, $|\mathcal{M}_n| = \Theta(n^2)$. □

Proof of Theorem 3.4.1. Let $m = m(n) = \frac{1}{2}\sum_{v \in [n]} d^n(v)$, and let q, r and s be positive integers. Also, in what follows write $d^n_{\max} = \max_{1 \le i \le n} d^n(i)$; by Fact 3.A.1 we know that $d^n_{\max} = o(n^{1/2})$.

Assume for the time being that $p(0) + p(1) < 1$ and that $\eta > 0$. Notice that when $\eta > 0$ and $\mu_1(p) > 0$ we have

$$\frac{|\mathcal{N}_n|}{n} = \frac{1}{n} \sum_{uv \in e(h_n)} d^n(u) d^n(v) = \sum_{i,j \ge 1} ij\alpha^n(i,j) \to \sum_{i,j \ge 1} ij\alpha(i,j) = \mu_1(p)\eta \in (0, \infty);$$

we have $\mu_1(p) > 0$ since $p(0) < 1$, and $\mu_1(p) < \infty$ since $\mu_2(p) < \infty$. It follows that $|\mathcal{N}_n| = \Theta(n)$.

We first claim that

$$\mathbf{E}\left((L_n)_q (M_n)_r (N_n)_s\right) = \frac{(|\mathcal{L}_n|)_q (|\mathcal{M}_n|)_r (|\mathcal{N}_n|)_s}{\prod_{i=0}^{q+2r+s-1} 2m - 1 - 2i}(1 + o(1)).$$

From Proposition 3.B.1 we know that

$$\left| \mathbf{E}\left((L_n)_q (M_n)_r (N_n)_s\right) - \frac{(|\mathcal{L}_n|)_q (|\mathcal{M}_n|)_r (|\mathcal{N}_n|)_s}{\prod_{i=0}^{q+2r+s-1} 2m - 1 - 2i} \right| \le C(S_1 + S_2),$$

where S_1 is defined by the relationship

$$
\begin{aligned}
&S_1 \cdot \prod_{i=0}^{q+2r+s-1} (2m-1-2i) \\
&= (|\mathcal{L}_n|)_{q-2}(|\mathcal{M}_n|)_r(|\mathcal{N}_n|)_s \sum_{v\in[n]} (d^n(v))^3 \\
&\quad + (|\mathcal{L}_n|)_{q-1}(|\mathcal{M}_n|)_{r-1}(|\mathcal{N}_n|)_s \sum_{1\le u\neq v\le n} (d^n(u))^3(d^n(v))^2 \\
&\quad + (|\mathcal{L}_n|)_{q-1}(|\mathcal{M}_n|)_r(|\mathcal{N}_n|)_{s-1} \sum_{uv\in e(h_n)} (d^n(u))^2 d^n(v) \\
&\quad + (|\mathcal{L}_n|)_q(|\mathcal{M}_n|)_{r-2}(|\mathcal{N}_n|)_s \sum_{\substack{1\le u,v_1,v_2\le n\\ u,v_1,v_2 \text{ distinct}}} (d^n(u))^3(d^n(v_1))^2(d^n(v_2))^2 \\
&\quad + (|\mathcal{L}_n|)_q(|\mathcal{M}_n|)_{r-1}(|\mathcal{N}_n|)_{s-1} \sum_{\substack{1\le u,v_1,v_2\le n\\ u,v_1,v_2 \text{ distinct}\\ uv_2\in e(h_n)}} (d^n(u))^2(d^n(v_1))^2 d^n(v_2) \\
&\quad + (|\mathcal{L}_n|)_q(|\mathcal{M}_n|)_r(|\mathcal{N}_n|)_{s-2} \sum_{\substack{uv_1,uv_2\in e(h_n)\\ v_1\neq v_2}} d^n(u)d^n(v_1)d^n(v_2),
\end{aligned}
$$

and S_2 is defined by

$$
S_2 = (|\mathcal{L}_n|)_q(|\mathcal{N}_n|)_s \sum_{k=1}^{r-1} (|\mathcal{M}_n|)_{r-k} \sum_{\ell=0}^{k} (d^n_{\max})^{2\ell} \prod_{i=0}^{q+2r+s-1-(2k-\ell)} \frac{1}{2m-1-2i};
$$

recall that we set $(k)_\ell = 0$ if $\ell < 0$. We now show that

$$
S_1 = o\left(\frac{(|\mathcal{L}_n|)_q(|\mathcal{M}_n|)_r(|\mathcal{N}_n|)_s}{\prod_{i=0}^{q+2r+s-1}(2m-1-2i)}\right) \text{ and } S_2 = o\left(\frac{(|\mathcal{L}_n|)_q(|\mathcal{M}_n|)_r(|\mathcal{N}_n|)_s}{\prod_{i=0}^{q+2r+s-1}(2m-1-2i)}\right). \tag{3.30}
$$

We will start by bounding S_1. Since $|\mathcal{L}_n|, |\mathcal{M}_n|, |\mathcal{N}_n| \to \infty$ as $n \to \infty$, to prove the first bound in (3.30) it suffices to establish the following bounds:

$$\begin{aligned}
\sum_{v\in[n]}(d^n(v))^3 &= o(|\mathcal{L}_n|^2),\\
\sum_{1\le u\ne v\le n}(d^n(u))^3(d^n(v))^2 &= o(|\mathcal{L}_n||\mathcal{M}_n|),\\
\sum_{uv\in e(h_n)}(d^n(u))^2 d^n(v) &= o(|\mathcal{L}_n||\mathcal{N}_n|),\\
\sum_{\substack{1\le u,v_1,v_2\le n\\ u,v_1,v_2 \text{ distinct}}}(d^n(u))^3(d^n(v_1))^2(d^n(v_2))^2 &= o(|\mathcal{M}_n|^2),\\
\sum_{\substack{1\le u,v_1,v_2\le n\\ u,v_1,v_2 \text{ distinct}\\ uv_2\in e(h_n)}}(d^n(u))^2(d^n(v_1))^2 d^n(v_2) &= o(|\mathcal{M}_n||\mathcal{N}_n|), \text{ and}\\
\sum_{\substack{uv_1,uv_2\in e(h_n)\\ v_1\ne v_2}} d^n(u)d^n(v_1)d^n(v_2) &= o(|\mathcal{N}_n|^2).
\end{aligned} \tag{3.31}$$

Using Lemmas 3.B.2 and 3.B.3, together with the fact that $d^n_{\max} = o(n^{1/2})$, we get the following results:

$$\sum_{v\in[n]}(d^n(v))^3 \le d^n_{\max}\sum_{v\in[n]}(d^n(v))^2 = o(n^{1/2})\cdot O(n) = o(n^2) = o(|\mathcal{L}_n|^2),$$

$$\begin{aligned}
\sum_{1\le u\ne v\le n}(d^n(u))^3(d^n(v))^2 &\le d^n_{\max}\sum_{1\le u\ne v\le n}(d^n(u))^2(d^n(v))^2\\
&\le d^n_{\max}\left(\sum_{v\in[n]}(d^n(v))^2\right)^2 = o(n^3) = o(|\mathcal{L}_n||\mathcal{M}_n|),
\end{aligned}$$

$$\sum_{uv\in e(h_n)}(d^n(u))^2 d^n(v) \le d^n_{\max}\sum_{uv\in e(h_n)} d^n(u)d^n(v) = d^n_{\max}|\mathcal{N}_n| = o(|\mathcal{L}_n||\mathcal{N}_n|),$$

$$\sum_{\substack{1\le u,v_1,v_2\le n\\ u,v_1,v_2 \text{ distinct}}} (d^n(u))^3(d^n(v_1))^2(d^n(v_2))^2 \le d^n_{\max} \sum_{\substack{1\le u,v_1,v_2\le n\\ u,v_1,v_2 \text{ distinct}}} (d^n(u))^2(d^n(v_1))^2(d^n(v_2))^2$$
$$\le d^n_{\max}\left(\sum_{v\in[n]}^{n}(d^n(v))^2\right)^3 = o(n^4) = o(|\mathcal{M}_n|^2),$$

$$\sum_{\substack{1\le u,v_1,v_2\le n\\ u,v_1,v_2 \text{ distinct}\\ uv_2\in e(h_n)}} (d^n(u))^2(d^n(v_1))^2 d^n(v_2) \le (d^n_{\max})^3 \sum_{uv\in e(h_n)} d^n(u)d^n(v)$$
$$= o(n^{3/2})|\mathcal{N}_n| = o(|\mathcal{M}_n||\mathcal{N}_n|),$$

and

$$\begin{aligned}\sum_{\substack{uv_1,uv_2\in e(h_n)\\ v_1\ne v_2}} d^n(u)d^n(v_1)d^n(v_2) &= \sum_{uv\in e(h_n)} d^n(u)d^n(v)\cdot \sum_{w:uw\in e(h_n)} d^n(w)\\ &\le \sum_{uv\in e(h_n)} d^n(u)d^n(v)\cdot\left(\sup_{u\in[n]}\sum_{w:uw\in e(h_n)} d^n(w)\right)\\ &= |\mathcal{N}_n|\cdot o(n)\\ &= o(|\mathcal{N}_n|^2),\end{aligned}$$

the last bound holding as $\mathcal{N}_n = \Theta(n)$.

To prove the bound on S_2 from (3.30), first notice that

$$\begin{aligned}S_2 &= (|\mathcal{L}_n|)_q(|\mathcal{N}_n|)_s\sum_{k=1}^{r-1}(|\mathcal{M}_n|)_{r-k}\sum_{\ell=0}^{k}(d^n_{\max})^{2\ell}\prod_{i=0}^{q+2r+s-1-(2k-\ell)}\frac{1}{2m-1-2i}\\ &= \frac{(|\mathcal{L}_n|)_q(|\mathcal{N}_n|)_s}{\prod_{i=0}^{q+s-1}2m-1-2i}\sum_{k=1}^{r-1}(|\mathcal{M}_n|)_{r-k}\sum_{\ell=0}^{k}(d^n_{\max})^{2\ell}\prod_{i=q+s}^{q+2r+s-1-(2k-\ell)}\frac{1}{2m-1-2i}.\end{aligned}$$

Hence, it suffices to show the following:

$$\sum_{k=1}^{r-1}(|\mathcal{M}_n|)_{r-k}\sum_{\ell=0}^{k}(d^n_{\max})^{2\ell}\prod_{i=q+s}^{q+2r+s-1-(2k-\ell)}\frac{1}{2m-1-2i}=o\left(\frac{(|\mathcal{M}_n|)_r}{\prod_{i=q+s}^{q+2r+s-1}2m-1-2i}\right).$$

Since r is fixed, we need only show that for arbitrary $k\in[1,r-1]$ and $\ell\in[0,k]$,

$$(|\mathcal{M}_n|)_{r-k}(d^n_{\max})^{2\ell}\prod_{i=q+s}^{q+2r+s-1-(2k-\ell)}\frac{1}{2m-1-2i}=o\left(\frac{(|\mathcal{M}|)_r}{\prod_{i=q+s}^{q+2r+s-1}2m-1-2i}\right).$$

By cancelling out some terms, this follows if we can show that

$$(d^n_{\max})^{2\ell}=o\left(\frac{(|\mathcal{M}_n|-(r-k))_k}{\prod_{i=q+2r+s-(2k-\ell)}^{q+2r+s-1}2m-1-2i}\right).$$

Now since $m=\Theta(n)$, $|\mathcal{M}_n|=\Theta(n^2)$, and r is a constant, this in turn holds, provided that

$$\begin{aligned}(d^n_{\max})^{2\ell}&=o\left(\frac{n^{2k}}{n^{2k-\ell}}\right)\\&=o\left(n^{\ell}\right),\end{aligned}$$

which holds since $d^n_{\max}=o(n^{1/2})$.

Therefore,

$$S_1+S_2=o\left(\frac{(|\mathcal{L}_n|)_q(|\mathcal{M}_n|)_r(|\mathcal{N}_n|)_s}{\prod_{i=0}^{q+2r+s-1}2m-1-2i}\right),$$

which proves that

$$\mathbf{E}\left((L_n)_q(M_n)_r(N_n)_s\right)=\frac{(|\mathcal{L}_n|)_q(|\mathcal{M}_n|)_r(|\mathcal{N}_n|)_s}{\prod_{i=0}^{q+2r+s-1}2m-1-2i}(1+o(1)).$$

Furthermore, since q, r and s are fixed, we obtain that

$$\begin{aligned}
&\frac{(|\mathcal{L}_n|)_q(|\mathcal{M}_n|)_r(|\mathcal{N}_n|)_s}{\prod_{i=0}^{q+2r+s-1} 2m-1-2i}\\
&= \frac{|\mathcal{L}_n|^q|\mathcal{M}_n|^r|\mathcal{N}_n|^s}{(2m)^{q+2r+s}}(1+o(1))\\
&= \left(\frac{|\mathcal{L}_n|}{\sum_{v\in[n]} d^n(v)}\right)^q \left(\frac{|\mathcal{M}_n|}{\left(\sum_{v\in[n]} d^n(v)\right)^2}\right)^r \left(\frac{|\mathcal{N}_n|}{\sum_{v\in[n]} d^n(v)}\right)^s (1+o(1))
\end{aligned}$$

Next, we claim that

$$\frac{|\mathcal{L}_n|}{\sum_{v\in[n]} d^n(v)} \to \nu/2, \frac{|\mathcal{M}_n|}{\left(\sum_{v\in[n]} d^n(v)\right)^2} \to \nu^2/4, \text{ and } \frac{|\mathcal{N}_n|}{\sum_{v\in[n]} d^n(v)} \to \eta.$$

For the first of these three claims, we have

$$\begin{aligned}
\frac{|\mathcal{L}_n|}{\sum_{v\in[n]} d^n(v)} = \frac{\frac{1}{2}\sum_{v\in[n]} d^n(v)(d^n(v)-1)}{\sum_{v\in[n]} d^n(v)} &= \frac{1}{2}\frac{\sum_{v\in[n]}(d^n(v))^2 - \sum_{v\in[n]} d^n(v)}{\sum_{v\in[n]} d^n(v)}\\
&= \frac{1}{2}\left(\frac{\mu_2(p^n)-\mu_1(p^n)}{\mu_1(p^n)}\right) \to \nu/2.
\end{aligned}$$

For the second, we have

$$\begin{aligned}
&\frac{|\mathcal{M}_n|}{\left(\sum_{v\in[n]} d^n(v)\right)^2}\\
&= \frac{\frac{1}{2}\sum_{\{u<v:uv\notin e(h_n)\}} d^n(u)\left(d^n(u)-1\right) d^n(v)\left(d^n(v)-1\right)}{\left(\sum_{v\in[n]} d^n(v)\right)^2}\\
&= \frac{1}{4\left(\sum_{v\in[n]} d^n(v)\right)^2}\Bigg(\left[\sum_{v\in[n]} d^n(v)\left(d^n(v)-1\right)\right]^2 - \sum_{v\in[n]}\left[d^n(v)\left(d^n(v)-1\right)\right]^2\\
&\qquad\qquad - \sum_{uv\in e(h_n)} d^n(u)\left(d^n(u)-1\right) d^n(v)\left(d^n(v)-1\right)\Bigg).
\end{aligned}$$

The second and third terms vanish in the limit since $\mu_1(p) > 0$ and so

$$\left(\sum_{v\in[n]} d^n(v)\right)^2 = \Omega(n^2),$$

and

$$\sum_{v\in[n]} \left[d^n(v)\left(d^n(v)-1\right)\right]^2 \leq (d^n_{\max})^2 \sum_{v\in[n]} (d^n(v))^2 = (d^n_{\max})^2 n\mu_2(p^n) = o(n^2),$$

and

$$\begin{aligned}
\sum_{uv\in e(h_n)} d^n(u)\left(d^n(u)-1\right) d^n(v)\left(d^n(v)-1\right) &\leq (d^n_{\max})^2 \sum_{uv\in e(h_n)} d^n(u)d^n(v) \\
&= (d^n_{\max})^2 |\mathcal{N}_n| \\
&= o(n^2).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{|\mathcal{M}_n|}{\left(\sum_{v\in[n]} d^n(v)\right)^2} &= \frac{1}{4}\frac{\left[\sum_{v\in[n]} d^n(v)\left(d^n(v)-1\right)\right]^2}{\left(\sum_{v\in[n]} d^n(v)\right)^2}(1+o(1)) \\
&\to \frac{1}{4}\frac{\left(\mu_2(p)-\mu_1(p)\right)^2}{\left(\mu_1(p)\right)^2} \\
&= \nu^2/4.
\end{aligned}$$

For the third claim, we have

$$\begin{aligned}
\frac{|\mathcal{N}_n|}{\sum_{v\in[n]} d^n(v)} = \frac{\sum_{uv\in e(h_n)} d^n(u)d^n(v)}{\sum_{v\in[n]} d^n(v)} &= \frac{\sum_{i,j\geq 1} ij\alpha^n(i,j)}{\mu_1(p^n)} \\
&\to \frac{\sum_{i,j\geq 1} ij\alpha(i,j)}{\mu_1(p)} = \eta.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\lim_{n\to\infty} \mathbf{E}\left((L_n)_q(M_n)_r(N_n)_s\right) \\
&= \lim_{n\to\infty} \frac{(|\mathcal{L}_n|)_q(|\mathcal{M}_n|)_r(|\mathcal{N}_n|)_s}{\prod_{i=0}^{q+2r+s-1} 2m-1-2i}(1+o(1)) \qquad (3.32)\\
&= \lim_{n\to\infty} \left(\frac{|\mathcal{L}_n|}{\sum_{v\in[n]} d^n(v)}\right)^q \left(\frac{|\mathcal{M}_n|}{\left(\sum_{v\in[n]} d^n(v)\right)^2}\right)^r \left(\frac{|\mathcal{N}_n|}{\sum_{v\in[n]} d^n(v)}\right)^s (1+o(1)) \\
&= (\nu/2)^q \left(\nu^2/4\right)^r (\eta)^s .
\end{aligned}$$

It then follows, by Theorem 2.6 of [68], that the random variables L_n, M_n and N_n converge to independent Poisson random variables with parameters $\nu/2, \nu^2/4$ and η respectively. This proves the theorem in the case that $p(0)+p(1)<1$ and $\eta>0$.

Lastly we will deal with the cases that arise if $p(0)+p(1)=1$ or if $\eta=0$. First, if $\eta=0$ then $\lim_{n\to\infty}\sum_{i,j\geq 1} ij\alpha^n(i,j)=\sum_{i,j\geq 1} ij\alpha(i,j)=0$. For any two half-edges ui and vj with $u,v\in[n]$, $i\in[d^n(u)]$ and $j\in[d^n(v)]$, we have $\mathbf{P}\left\{\mathbf{1}_{[ui,vj]}=1\right\}=\frac{1}{2m-1}$ from the definition of the configuration model. So by (3.18), since $m=m(n)=n\mu_1(p^n)/2=\Theta(n)$, we have that

$$\begin{aligned}
\mathbf{E}\left(N_n\right) = \sum_{uv\in e(h_n)} \sum_{i\in[d^n(u)]} \sum_{j\in[d^n(v)]} \mathbf{P}\left\{\mathbf{1}_{[ui,vj]}=1\right\} &= \frac{\sum_{uv\in e(h_n)} d^n(u)d^n(v)}{2m-1} \\
&= \frac{n}{2m-1}\sum_{i,j\geq 1} ij\alpha^n(i,j) = o(1).
\end{aligned}$$

Hence, $\lim_{n\to\infty}\mathbf{E}\left(N_n\right)=0$. In this case, if $p(0)+p(1)<1$ then a reprise of the argument leading to (3.32) gives that for all $q,r\geq 1$,

$$\lim_{n\to\infty} \mathbf{E}\left((L_n)_q(M_n)_r\right) = (\nu/2)^q(\nu^2/4)^r,$$

and therefore that L_n and M_n converge to independent Poisson random variables with parameters $\nu/2$ and $\nu^2/4$ respectively.

Lastly, we deal with the case when $p(0)+p(1)=1$. In this case, $\mu_2(p)=\mu_1(p)$, so $\nu=0$. Furthermore,

$$\mathbf{E}\left(L_n\right) = \frac{\frac{1}{2}\sum_{v\in[n]} d^n(v)(d^n(v)-1)}{2m-1} = \frac{n}{4m-2}(\mu_2(p^n)-\mu_1(p^n)).$$

Since $\mu_2(p^n)-\mu_1(p^n) \to \mu_2(p)-\mu_1(p) = 0$, we get that $\lim_{n\to\infty} \mathbf{E}\left(L_n\right) = 0$, and an analogous argument shows that $\lim_{n\to\infty} \mathbf{E}\left(M_n\right) = 0$. In this case, another reprise of the argument leading to (3.32) gives that for all $s \geq 1$,

$$\lim_{n\to\infty} \mathbf{E}\left((N_n)_s\right) = \eta^s,$$

so N_n is asymptotically Poisson(η) distributed. $\square$

Proof of Corollary 3.4.2. First, the fact that $\sum_{k,\ell\geq 0} k\ell \cdot \alpha(k,\ell) < \infty$ appears in (3.8), above, from which it is immediate that $\eta < \infty$. Next, let $G_-(\mathrm{d}^n)$ be the subgraph of $G(\mathrm{d}^n)$ with edge set $\mathrm{e}(G(\mathrm{d}^n))\backslash \mathrm{e}(T(\mathrm{d}^n))$, and let d^n_- be the degree sequence of $G_-(\mathrm{d}^n)$, as defined in Section 3.3 previous to Proposition 3.3.2. Finally, write $L_n = L(G_-(\mathrm{d}^n))$, $M_n = M(G_-(\mathrm{d}^n), T(\mathrm{d}^n))$, and $N_n = N(G_-(\mathrm{d}^n), T(\mathrm{d}^n))$. Our aim is to apply Theorem 3.4.1, with $h_n = T(\mathrm{d}^n)$ and $G_n = G_-(\mathrm{d}^n) = G(\mathrm{d}^n) - T(\mathrm{d}^n)$. Note that with these choices, we have $\alpha^n(k,l) = n^{-1}A^n(k,l)$, where $A^n(k,l)$ is as in Proposition 3.3.2. Moreover, we have

$$\sum_{l\geq 0} A^n(k,l) = \#\{u \in [n] : d^n_-(u) = k\}\,.$$

Conditionally given $T(\mathrm{d}^n)$, the graph $G_-(\mathrm{d}^n)$ is a random graph with degree sequence d^n_-. By Proposition 3.3.2 we know that $\alpha^n(k,l) \xrightarrow{\text{prob}} \alpha(k,l)$ for all $k,l \geq 0$, and that

$$\sum_{k,l\geq 0} kl\alpha^n(k,l) \xrightarrow{\text{prob}} \sum_{k,l\geq 0} kl\alpha(k,l).$$

Moreover, since α defines a probability distribution, it must be that for all k we have

$$p^n_-(k) := \frac{1}{n}\#\{u \in [n] : d^n_-(u) = k\} = \frac{1}{n}\sum_{l\geq 0} A^n(k,l) = \sum_{l\geq 0} \alpha^n(k,l) \xrightarrow{\text{prob}} \sum_{l\geq 0} \alpha(k,l)\,. \quad (3.33)$$

Setting $p_-(k) = \sum_{l\geq 0} \alpha(k,l)$, then (3.33) states that $p^n_-(k) \xrightarrow{\text{prob}} p_-(k)$ for all $k \geq 0$. Moreover, since $p^n_-(k) \leq p^n(k)$, $p^n_-(k) \xrightarrow{\text{prob}} p_-(k)$, and $p^n(k) \xrightarrow{\text{prob}} p(k)$ for all $k \geq 0$, it follows that $\mu_2(p_-) \leq \mu_2(p) < \infty$ and $\mu_2(p^n_-) \xrightarrow{\text{prob}} \mu_2(p_-)$. From these observations, Fact 3.A.1 then implies that

$$\max_{v\in[n]} \deg_{T(\mathrm{d}^n)}(v) \leq \max_{v\in[n]} \deg_{G_n}(v) = o(n^{1/2})\,.$$

Since $\mu_2(p_-^n) \xrightarrow{\text{prob}} \mu_2(p_-)$, we also have $\mu_1(p_-^n) \xrightarrow{\text{prob}} \mu_1(p_-)$; but

$$n\mu_1(p_-^n) = \sum_{i=1}^{n} d_-^n(u) = \left(\sum_{i=1}^{n} d^n(u)\right) - 2(n-1)\,,$$

so $\mu_1(p_-) = \mu_1(p) - 2$. Thus, if $\mu_1(p) > 2$ then $\mu_1(p_-) > 0$, so $p_-(0) < 1$. In this case, applying Theorem 3.4.1, it follows that conditionally given $T(\mathrm{d}^n)$,

$$\|\mathrm{Dist}(L_n, M_n, N_n) - \mathrm{Poi}(\nu/2) \otimes \mathrm{Poi}(\nu^2/4) \otimes \mathrm{Poi}(\eta)\|_{\mathrm{TV}} \xrightarrow{\text{prob}} 0$$

as $n \to \infty$. If (L, M, N) is $\mathrm{Poi}(\nu/2) \otimes \mathrm{Poi}(\nu^2/4) \otimes \mathrm{Poi}(\eta)$-distributed, then we have

$$\mathbf{P}\{L = M = N = 0\} = \exp(-\nu/2 - \nu^2/4 - \eta);$$

since $G(\mathrm{d}^n)$ is simple if and only if $L_n = M_n = N_n = 0$, it follows that

$$\begin{aligned} \mathbf{P}\{G(\mathrm{d}^n) \text{ simple } \mid T(\mathrm{d}^n)\} &= \mathbf{P}\{L_n = M_n = N_n = 0 \mid T(\mathrm{d}^n)\} \\ &\xrightarrow{\text{prob}} \mathbf{P}\{L = M = N = 0\} = \exp(-\nu/2 - \nu^2/4 - \eta)\,, \end{aligned}$$

as required.

It remains to treat the case that $\mu_1(p) = 2$, which implies that $\mu_1(p_-) = 0$ and $p_-(0) = 1$. This case requires a separate argument, which is more involved than one might expect. We note immediately that in this situation, $m = (1 + o(1))n$ as $n \to \infty$.

Write G'_n for the graph obtained from $G_-(\mathrm{d}^n)$ by removing the edge $\Gamma(\mathrm{d}^n)$. Then let $L'_n = L(G'_n)$, $M'_n = M(G'_n, T(\mathrm{d}^n))$, and $N'_n = N(G'_n, T(\mathrm{d}^n))$. Then G'_n is simple precisely if $L'_n + M'_n + N'_n = 0$. We will first prove that $\mathbf{E}\,(L'_n + M'_n + N'_n) \xrightarrow{\text{prob}} 0$, then explain how to deal with the root edge.

By Proposition 3.2.1, we know that the non-root half-edges chosen for $T(\mathrm{d}^n)$ form a uniformly random subset $\mathcal{S}^n$ of $\bigcup_{i=1}^n \{i1, \ldots, i(d^n(i)-1)\}$ of size $n-1$. The half-edges which are paired to form $G_-(\mathrm{d}^n)$ are precisely the edges of $\mathcal{U} := \bigcup_{i=1}^n \{i1, \ldots, i(d^n(i)-1)\} \setminus \mathcal{S}^n$, together with the unique unpaired root half-edge of $T(\mathrm{d}^n)$.

Write $\mathcal{U}$ for the set of half-edges paired to form G'_n. By the observations of the preceding paragraph, $\mathcal{U}$ is a uniformly random subset of $\bigcup_{i=1}^n \{i1, \ldots, i(d^n(i)-1)\}$ of size $2(m-n)$. Moreover, conditionally given $\mathcal{U}$, the pairing of half-edges in $\mathcal{S}$ is uniformly random and independent of $T(\mathrm{d}^n)$. Therefore, we can construct $(G(\mathrm{d}^n), T(\mathrm{d}^n), \Lambda(\mathrm{d}^n))$ as follows. First sample a sequence of $m-n$ disjoint pairs of half-edges from $\bigcup_{i=1}^{n}\{i1,\ldots,i(d^n(i)-1)\}$ uniformly at random and join them to form edges; this determines the set $\mathcal{U}$, and all edges of G_n'. Next, build $T(\mathrm{d}^n)$ via Pitman's additive coalescent applied to $\bigcup_{i=1}^{n}\{i1,\ldots,i(d^n(i)-1)\}\setminus\mathcal{U}$. Finally, pair the root half-edge of T^n with the sole remaining unpaired non-root half-edge. We analyze this construction procedure in order to bound the expected number of loops and multiple edges in G_n'.

Let (h_1,h_2) be a half-edge pair chosen for G_n' in the construction procedure just above. Then h_1 and h_2 are uniform random half-edges chosen from $\bigcup_{i=1}^{n}\{i1,\ldots,i(d^n(i)-1)\}$, and (h_1,h_2) is a loop if $v(h_1)=v(h_2)$. Hence,

$$\begin{aligned}\mathbf{E}\left(L_n'\right) &= |e(G_n')|\,\mathbf{P}\left\{v(h_1)=v(h_2)\right\}\\ &= (m-n)\sum_{i=1}^{n}\mathbf{P}\left\{v(h_1)=v(h_2)=i\right\}\\ &= (m-n)\sum_{i=1}^{n}\frac{(d^n(i)-1)(d^n(i)-2)}{(2m-(n-1))(2m-n)}.\end{aligned}$$

Similarly, edges (h_1,h_2) and (h_3,h_4) form a double edge if $v(h_1)=v(h_3)$ and $v(h_2)=v(h_4)$ or if $v(h_1)=v(h_4)$ and $v(h_2)=v(h_3)$. Hence,

$$\begin{aligned}&\mathbf{E}\left(M_n'\right)\\ &= \left(|e(G_n')|\right)\left(|e(G_n'))|-1\right)\sum_{1\le i<j\le n}\frac{4(d^n(i)-1)(d^n(i)-2)(d^n(j)-1)(d^n(j)-2)}{(2m-n+1))(2m-n)(2m-n-1)(2m-n-2)}\\ &= (m-n)(m-n-1)\sum_{1\le i<j\le n}\frac{4(d^n(i)-1)(d^n(i)-2)(d^n(j)-1)(d^n(j)-2)}{(2m-n+1)(2m-n)(2m-n-1)(2m-n-2)}\\ &\le 2\left((m-n)\sum_{i=1}^{n}\frac{(d^n(i)-1)(d^n(i)-2)}{(2m-n-1)(2m-n-2)}\right)^2.\end{aligned}$$

Since $m=n+o(n)$, we have $\frac{m-n-1}{2m-n-2}=o(1)$. Since also $\sum_{i=1}^{n}(d^n(i)-1)(d^n(i)-2)\le\sum_{i=1}^{n}d^n(i)^2=O(n)=O(2m-n-1)$, it follows from the two preceding displayed inequalities that $\mathbf{E}\left(L_n'+M_n'\right)\xrightarrow{\text{prob}}0$.

To show that $\mathbf{E}\left(N_n'\right)\xrightarrow{\text{prob}}0$, we will again use Proposition 3.2.1. Given a set $\mathcal{H}\subseteq\bigcup_{i=1}^{n}\{i1,\ldots,i(d^n(i)-1)\}$, write $\mathcal{H}_i=\mathcal{H}\cap\{i1,\ldots,i(d^n(i)-1)\}$. By (3.12) we know that for any set $\mathcal{H}$ as above with $|\mathcal{H}|=n-1$, for all $1\le i\le n-1$, for any $h\in\mathcal{H}$ and any root half-edge r with $v(r) \neq v(h)$,

$$\mathbf{P}\left\{(r_i, s_i) = (r, h) \mid \mathcal{S}^n = \mathcal{H}\right\} = \mathbf{P}\left\{(r_1, s_1) = (r, h) \mid \mathcal{S}^n = \mathcal{H}\right\}.$$

Moreover, by construction, $T(d^n)$ and G'_n are conditionally independent given $\mathcal{S}^n$, so for any $1 \leq i \leq n-1$,

$$\begin{aligned}
&\mathbf{E}\left\{m_{G'_n}(v(r_i)v(s_i)) \;\middle|\; \mathcal{S}^n = \mathcal{H}\right\} \\
&= \mathbf{E}\left\{m_{G'_n}(v(r_1)v(s_1)) \;\middle|\; \mathcal{S}^n = \mathcal{H}\right\} \\
&= \sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n} \mathbf{P}\left\{v(r_1) = j, v(s_1) = k \mid \mathcal{S}^n = \mathcal{H}\right\} \cdot \mathbf{E}\left\{m_{G'_n}(jk) \;\middle|\; \mathcal{S}^n = \mathcal{H}\right\} \\
&.
\end{aligned}$$

Now, by Proposition 3.2.1,

$$\mathbf{P}\left\{v(r_1) = j, v(s_1) = k \mid \mathcal{S}^n = \mathcal{H}\right\} = \frac{|\mathcal{H}_k|}{(n-1)^2}.$$

Also,

$$\mathbf{E}\left\{m_{G'_n}(jk) \;\middle|\; \mathcal{S}^n = \mathcal{H}\right\} = (m-n) \cdot \frac{(d^n(j) - 1 - |\mathcal{H}_j|)(d^n(k) - 1 - |\mathcal{H}_k|)}{(2m - 2(n-1))(2m - 2(n-1) - 1)}.$$

The term $m-n$ above accounts for the number of edges of G'_n; the fraction is the probability that a uniformly random pair of half-edges from $(\bigcup_{l=1}^{n}\{l1, \ldots, l(d^n(i)-1)\}) \setminus \mathcal{H}$ are incident to vertices j and k. Combining these formulas, we then have that for all $1 \leq i \leq n-1$,

$$\begin{aligned}
&\mathbf{E}\left\{m_{G'_n}(v(r_i)v(s_i)) \;\middle|\; \mathcal{S}^n = \mathcal{H}\right\} \\
&\leq \sum_{j=1}^{n}\sum_{\substack{k=1\\k\neq j}}^{n} \frac{|\mathcal{H}_k|}{(n-1)^2} \cdot (m-n) \cdot \frac{(d^n(j) - 1 - |\mathcal{H}_j|)(d^n(k) - 1 - |\mathcal{H}_k|)}{(2m - 2(n-1))(2m - 2(n-1) - 1)} \\
&= O(1) \cdot \sum_{k=1}^{n} \frac{|\mathcal{H}_k|}{(n-1)^2} \cdot \frac{(d^n(k) - 1 - |\mathcal{H}_k|)}{(2(m-n)+1)} \sum_{\substack{j=1\\j\neq k}}^{n} (d^n(j) - 1 - |\mathcal{H}_j|).
\end{aligned}$$

Now, since $\sum_{j=1}^n |\mathcal{H}_j| = |\mathcal{H}| = n-1$, we have

$$\sum_{\substack{j=1 \\ j\neq k}}^n (d^n(j) - 1 - |\mathcal{H}_j|) = 2m - n - |\mathcal{H}| - (d^n(k) - 1 - |\mathcal{H}_k|) \leq 2(m-n) - 1\,,$$

and it follows that

$$\mathbf{E}\left\{ m_{G'_n}(v(r_i)v(s_i)) \;\middle|\; \mathcal{S}^n = \mathcal{H} \right\} = O(1) \cdot \sum_{k=1}^n \frac{|\mathcal{H}_k|\,(d^n(k) - 1 - |\mathcal{H}_k|)}{2(n-1)^2}.$$

Taking expectation over $\mathcal{S}^n$ in this bound, it follows that

$$\mathbf{E}\left(m_{G'_n}(v(r_i)v(s_i))\right) \leq \frac{1}{(n-1)^2}\mathbf{E}\left(\sum_{k=1}^n |\mathcal{H}_k|\,(d^n(k) - 1 - |\mathcal{H}_k|)\right).$$

We now show that $\mathbf{E}\left(\sum_{k=1}^n |\mathcal{H}_k|\,(d^n(k) - 1 - |\mathcal{H}_k|)\right) = o(n)$. Notice that $|\mathcal{H}_k| \leq d^n(k)$ and that $d^n(k) - 1 - |\mathcal{H}_k| = d^n_-(k)$, unless k is incident to the unpaired root half-edge, in which case $d^n(k) - 1 - |\mathcal{H}_k| = d^n_-(k) - 1 \geq 0$. Letting r be the unpaired root half-edge, it then follows that

$$\sum_{k=1}^n |\mathcal{H}_k|\,(d^n(k) - 1 - |\mathcal{H}_k|) \leq \sum_{k=1}^n \left(d^n(k)\right)^2 \mathbf{1}_{[d^n_-(k)>0]}.$$

Since $|\{k : d^n_-(k) > 0\}| \leq m - n = o(n)$, by the second assertion of Fact 3.A.1 it follows that $\sum_{k=1}^n d^n(k)^2 \mathbf{1}_{[d^n_-(k)>0]} = o(n)$, which combined with the two preceding inequalities yields that

$$\mathbf{E}\left(m_{G'_n}(v(r_i)v(s_i))\right) = o\left(\frac{1}{n}\right).$$

Summing over i, it follows that

$$\mathbf{E}\left(N'_n\right) = \sum_{i=1}^n \mathbf{E}\left(m_{G'_n}(v(r_i)v(s_i))\right) = o(1).$$

At this point we know that $\mathbf{E}\left(L'_n + M'_n + N'_n\right) \to 0$. We now show how to deduce that $L_n + M_n + N_n \to 0$ in probability. We provide full details only in the case that $m - n \to \infty$, i.e., that the number of edges of $G_-(\mathrm{d}^n)$ tends to infinity, and briefly explain the argument in the simpler case that $m - n = O(1)$.

By Proposition 3.2.3, for any pair (g, t) where g is a graph with degree sequence d^n and

t is a spanning tree of g, we have

$$\mathbf{P}\left\{(G(\mathrm{d}^n),T(\mathrm{d}^n))=(g,t)\right\}\propto\frac{\sum_{\gamma\in e(g-t)}2^{\mathbf{1}_{[\gamma\text{ is a loop}]}}\cdot m_{g-t}(\gamma)}{\prod_{i=1}^n 2^{m_{g-t}(ii)}\cdot\prod_{e\in \mathrm{e}(g)}m_{g-t}(e)!}\,, \tag{3.34}$$

and for any edge γ of $g-t$,

$$\mathbf{P}\left\{\Gamma(\mathrm{d}^n)=\gamma\mid\ (G(\mathrm{d}^n),T(\mathrm{d}^n))=(g,t)\right\}\propto 2^{\mathbf{1}_{[\gamma\text{ is a loop}]}}\cdot m_{g-t}(\gamma).$$

If the graph with edge set $e(g)-(e(t)\cup\gamma)$ is simple, then $g-t$ has at most one loop, and no edge with multiplicity more than two, so

$$\sup_{e\in e(g-t)}\mathbf{P}\left\{\Gamma(\mathrm{d}^n)=e\mid\ (G(\mathrm{d}^n),T(\mathrm{d}^n))=(g,t)\right\}\le\frac{2}{|e(g-t)|}.$$

In particular, writing

$$\mathcal{B}(g,t)=\mathcal{L}(g)\cup\mathcal{N}(g,t)\cup\bigcup_{((ui_i,vj_1),(ui_2,vj_2))\in\mathcal{M}(g,t)}\left\{(ui_i,vj_1),(ui_2,vj_2)\right\},$$

and recalling that the half-edges comprising $\Gamma(\mathrm{d}^n)$ are $\Gamma^-(\mathrm{d}^n)$ and $\Gamma^+(\mathrm{d}^n)$, it follows that

$$\begin{aligned}
&\mathbf{P}\left\{(\Gamma^-(\mathrm{d}^n),\Gamma^+(\mathrm{d}^n))\in\mathcal{B}(G_-(\mathrm{d}^n),T(\mathrm{d}^n))\ \middle|\ (G(\mathrm{d}^n),T(\mathrm{d}^n))\right\}\\
&\le\frac{2|\mathcal{B}(G_-(\mathrm{d}^n),T(\mathrm{d}^n))|}{|e(G_-(\mathrm{d}^n))|}\mathbf{1}_{[G'_n\text{ simple}]}+\mathbf{1}_{[G'_n\text{ not simple}]}\\
&\le\frac{4(L(G_-(\mathrm{d}^n))+N(G_-(\mathrm{d}^n),T(\mathrm{d}^n))+M(G_-(\mathrm{d}^n),T(\mathrm{d}^n)))}{|e(G_-(\mathrm{d}^n))|}\mathbf{1}_{[G'_n\text{ simple}]}+\mathbf{1}_{[G'_n\text{ not simple}]}\\
&=\frac{4(L_n+M_n+N_n)}{m-(n-1)}+\mathbf{1}_{[G'_n\text{ not simple}]}\,;
\end{aligned}$$

the last inequality is not tight unless $M(G_-(\mathrm{d}^n),T(\mathrm{d}^n))=0$. Now,

$$L_n+M_n+N_n\le 3(L'_n+M'_n+N'_n+\mathbf{1}_{[(\Gamma^-(\mathrm{d}^n),\Gamma^+(\mathrm{d}^n))\in\mathcal{B}(G_-(\mathrm{d}^n),T(\mathrm{d}^n))]})\,; \tag{3.35}$$

this inequality is never tight but it suffices for our purposes. Using this bound, and taking expectations in the previous conditional probability bound, we obtain that

$$\begin{aligned}
&\mathbf{P}\left\{(\Gamma^-(\mathrm{d}^n), \Gamma^+(\mathrm{d}^n)) \in \mathcal{B}(G_-(\mathrm{d}^n), T(\mathrm{d}^n))\right\} \\
&\leq \frac{12}{m-(n-1)} \mathbf{E}\left(L'_n + M'_n + N'_n + 1\right) + \mathbf{P}\left\{G'_n \text{ not simple}\right\} \\
&= \frac{12}{m-(n-1)} \mathbf{E}\left(L'_n + M'_n + N'_n + 1\right) + \mathbf{P}\left\{L'_n + M'_n + N'_n > 0\right\} .
\end{aligned}$$

Since $\mathbf{E}\left(L'_n + M'_n + N'_n\right) \to 0$, if $m-(n-1) \to \infty$ it follows that

$$\mathbf{P}\left\{(\Gamma^-(\mathrm{d}^n), \Gamma^+(\mathrm{d}^n)) \in \mathcal{B}(G_-(\mathrm{d}^n), T(\mathrm{d}^n))\right\} \xrightarrow{\text{prob}} 0$$

in which case (3.35) and the fact that $L'_n + M'_n + N'_n \xrightarrow{\text{prob}} 0$ together imply that $L_n + M_n + N_n \xrightarrow{\text{prob}} 0$ as required.

Finally, if $m-n = O(1)$, the fact that $L_n + M_n + N_n \xrightarrow{\text{prob}} 0$ can be seen as follows. Let h be a uniformly random half-edge from $\bigcup_{1\leq i\leq n}\{i1, \ldots, i(d^n(i)-1)\}$. Then with high probability, $d^n(v(h)) = O(1)$. Since $|\mathcal{U}| = |\bigcup_{1\leq i\leq n}\{i1, \ldots, i(d^n(i)-1)\} \setminus \mathcal{S}^n| = 2(m-n) = O(1)$, it follows that with high probability $d^n(v(h)) = O(1)$ for all edges of $\bigcup_{1\leq i\leq n}\{i1, \ldots, i(d^n(i)-1)\} \setminus \mathcal{H}$ and that $v(h) \neq v(h')$ for all distinct $h, h' \in \mathcal{U}$. This already implies that $L_n + M_n = 0$ with probability $1 - o(1)$. Finally, by considering Pitman's additive coalescent it is not hard to see that for any vertex v with $d^n(v) = O(1)$, the probability that v is the root of $T^n(v)$ or is adjacent to the root is $o(1)$, and that for any two vertices v, w with $d^n(v) = O(1)$ and $d^n(w) = O(1)$, the probability that v and w are adjacent is $o(1)$. (Verifying the assertions of the last sentence in detail is left to the reader.) This immediately implies that $N_n = 0$ with probability $1 - o(1)$ and so completes the proof. $\square$

Chapter 4

Finding Minimum Spanning Trees via Local Improvements

4.1 Introduction

Local search is the name for an optimization paradigm in which optimal or near-optimal solutions are sought algorithmically, via sequential improvements which are "local" in that at each step, the search space consists only of neighbours (in some sense) of the current solution. Well-known algorithmic examples of this paradigm include simulated annealing, hill climbing, and the Metropolis-Hasting algorithm.

A recent line of research considers the behaviour of local search on *smoothed* optimization problems, in which the input is either fully random or is a random perturbation of a fixed input. The goal in this setting is to characterize the running time of local search and the quality of its output. Problems approached in this vein include *max-cut* [20, 29, 39], for which the allowed "local" improvements consist of moving a single vertex; *max-2CSP* and the *binary function optimization problem* [35], for which the allowed local improvements are bit flips; and Euclidean TSP [54], where the allowed local improvements consist of replacing edge pairs uv, wx with pairs uw, vx (when the result is still a tour).

In the current work, we analyze local search for the *random minimum spanning tree* problem, one of the first and foundational problems in combinatorial optimization. We now briefly describe our results (for more precise statements see Section 4.1.1, below). As input to the problem, we take the randomly-weighted complete graph $\mathbb{K}_n = (K_n, \mathbb{X})$, where $\mathbb{X} = (X_e, e \in \mathrm{E}(K_n))$ are independent copies of a random variable X, and an arbitrary starting graph H_0, which we aim to transform into the minimum-weight spanning tree MST.

We fix a *threshold weight* $\rho > 0$; at step $k \geq 0$, a local improvement consists of choosing a connected induced subgraph of the current MST candidate H_k whose current total weight is at most ρ, and replacing it by the minimum weight spanning tree on the same vertex set.

Suppose that X is non-negative and has a density $f : [0, \infty) \to [0, \infty)$ which is continuous at 0 and satisfies $f(0) > 0$. Then writing $\rho^* = \sup\{x : \mathbb{P}(X > x) > 0\}$, we prove that if $\rho > \rho^*$ then there exist local search paths which output the MST, whereas if $\rho < \rho^*$ then local search cannot reach the MST (and, indeed, with high probability will only achieve an approximation ratio of order $\Theta(n)$).

4.1.1 Detailed statement of the results

Let $\mathbb{G} = (G, w) = (V, E, w)$ be a finite weighted connected graph, where $G = (V, E)$ is a graph and $w : E \to (0, \infty)$ are edge weights; set $\mathrm{V}(\mathbb{G}) = \mathrm{V}(G) = V$ and $\mathrm{E}(\mathbb{G}) = \mathrm{E}(G) = E$. For a subgraph H of G write $w(H) = \sum_{e \in \mathrm{E}(H)} w(e)$ for its weight. A *minimum spanning tree* (MST) of $\mathbb{G}$ is a spanning tree T of G which minimizes $w(T)$ among all spanning trees of G. There is a unique MST provided all edge weights are distinct; we hereafter restrict our attention to weighted graphs $\mathbb{G}$ where all edge weights are distinct (and more strongly where $w(H_1) \neq w(H_2)$ for all pairs of distinct subgraphs $H_1, H_2 \subseteq G$); we call such graphs *generic*. For a generic weighted graph $\mathbb{G}$, we write $\mathrm{MST}(\mathbb{G})$ for the unique minimum spanning tree of $\mathbb{G}$.

For a weighted graph $\mathbb{G} = (V, E, w)$ and a set $S \subset V$, write $G[S]$ for the induced subgraph $G[S] = (S, E|_{S \times S})$ and $\mathbb{G}[S]$ for the induced weighted subgraph $\mathbb{G}[S] = (G[S], w|_{\mathrm{E}(G[S])})$. Now, given a spanning subgraph H of G, define $\Phi(H, S) = \Phi_{\mathbb{G}}(H, S)$ as follows. If $H[S]$ is connected then let $\Phi(H, S)$ be the spanning subgraph with edge set $(\mathrm{E}(H) \setminus \mathrm{E}(H[S])) \cup \mathrm{E}(\mathrm{MST}(\mathbb{G}[S]))$; if $H[S]$ is not connected then let $\Phi(H, S) = H$. In words, to form $\Phi(H, S)$ from H, we replace $H[S]$ by the minimum-weight spanning tree of $\mathbb{G}[S]$, unless $H[S]$ is not connected.

Now suppose we are given a finite weighted connected graph $\mathbb{G} = (V, E, w)$, a spanning subgraph H of G, and a sequence $\mathbb{S} = (S_i, 1 \leq i \leq m)$ of subsets of V. Define a sequence of spanning subgraphs $(H_i, 0 \leq i \leq m)$ as follows. Set $H_0 = H$, and for $1 \leq i \leq m$ let $H_i = \Phi_{\mathbb{G}}(H_{i-1}, S_i)$. Using the previous definition of Φ, this simply corresponds to sequentially replacing the subgraph of H_{i-1} induced by S_i by its corresponding minimum spanning tree (assuming $H_{i-1}[S_i]$ is connected). We refer to $\mathbb{S}$ as an *optimizing sequence* for the pair $(\mathbb{G}, H)$, and call $(H_i, 0 \leq i \leq m)$ the *subgraph sequence corresponding to* $\mathbb{S}$. We say $\mathbb{S}$ is an *MST sequence* for $(\mathbb{G}, H)$ if the final spanning subgraph H_m is the MST of $\mathbb{G}$.

The weight of step i of the sequence $\mathbb{S}$ is defined as

$$\mathrm{wt}(\mathbb{S}, i) = \mathrm{wt}(\mathbb{G}, H, \mathbb{S}, i) := w\left(H_{i-1}[S_i]\right) = \sum_{e \in \mathrm{E}(H_{i-1}[S_i])} w(e)\,,$$

and the weight of the whole sequence is the maximal weight of a single step:

$$\mathrm{wt}(\mathbb{S}) = \mathrm{wt}(\mathbb{G}, H, \mathbb{S}) := \max\Big\{ \mathrm{wt}(\mathbb{S}, i) : 1 \le i \le m \Big\}.$$

The *cost* of the pair $(\mathbb{G}, H)$ is defined as

$$\mathrm{cost}(\mathbb{G}, H) := \min\Big\{ \mathrm{wt}(\mathbb{S}) : \mathbb{S} \text{ is an MST sequence for } (\mathbb{G}, H) \Big\}.$$

The following theorem is the main result of the current work. Write K_n for the complete graph with vertex set $[n] = \{1, \ldots, n\}$, and $\mathbb{K}_n = (K_n, \mathbb{X})$ for the randomly weighted complete graph, where $\mathbb{X} = (X_e, e \in \mathrm{E}(K_n))$ are independent Uniform$[0,1]$ random variables. If $\mathbb{S} = (S_1, \ldots, S_m)$ is an optimizing sequence for $(\mathbb{K}_n, H_n)$ then we write $H_{n,0} = H_n$ and $H_{n,i} = \Phi_{\mathbb{K}_n}(H_{n,i-1}, S_i)$ for $1 \le i \le m$. Finally, we say a sequence $(E_n, n \ge 1)$ of events occurs with high probability if $\mathbb{P}(E_n) \to 1$ as $n \to \infty$.

Theorem 4.1.1. *Fix any sequence $(H_n, n \ge 1)$ of connected graphs with H_n being a spanning subgraph of K_n. Then for any $\varepsilon > 0$, as $n \to \infty$,*

(a) with high probability there exists an MST *sequence $\mathbb{S}$ for $(\mathbb{K}_n, H_n)$ with $\mathrm{wt}(\mathbb{S}) \le 1 + \varepsilon$, and*

(b) there exists $\delta > 0$ such that with high probability, given any optimizing sequence $\mathbb{S} = (S_1, \ldots, S_m)$ for $(\mathbb{K}_n, H_n)$ with $\mathrm{wt}(\mathbb{S}) \le 1 - \varepsilon$, the final spanning subgraph $H_{n,m}$ has weight $w(H_{n,m}) \ge \delta n\, w(\mathrm{MST}(\mathbb{K}_n))$.

In particular, $\mathrm{cost}(\mathbb{K}_n, H_n) \xrightarrow{\text{prob}} 1$ *as $n \to \infty$.*

We discuss possible refinements of and extensions to Theorem 4.1.1 in the conclusion, Section 4.4. We also explain in that section how to extend Theorem 4.1.1 to more general edge weight distributions than Uniform$[0,1]$, as described just before Section 4.1.1.

4.1.2 Overview of the proof

In this section, we give an overview of the proof of Theorem 4.1.1, while postponing the proofs of the more technical aspects to Sections 4.2 and 4.3 and Appendix 4.A. The lower bound of Theorem 4.1.1 is straightforward, so we provide it in full detail immediately.

Lower bound of Theorem 4.1.1. Fix $\varepsilon > 0$, and let $E_{n,\varepsilon} = \{e \in \mathrm{E}(H_n) : X_e > 1-\varepsilon\}$. The set $E_{n,\varepsilon}$ is a binomial random subset of $\mathrm{E}(H_n)$ in which each edge is present with probability ε, so $\mathbb{P}(|E_{n,\varepsilon}| \geq \varepsilon n/2) \to 1$.

Note that, for any edge $e = uv \in \mathrm{E}(H_n) \setminus \mathrm{E}(\mathrm{MST}(\mathbb{K}_n))$, and any optimizing sequence $\mathbb{S} = (S_1, \ldots, S_m)$ for $(\mathbb{K}_n, H_n)$, if there is no set S_i with $u, v \in S_i$, then $e \in H_{n,m}$. It follows that for any optimizing sequence $\mathbb{S}$ with $\mathrm{wt}(\mathbb{S}) \leq 1 - \varepsilon$, the final spanning subgraph $H_{n,m}$ has $E_{n,\varepsilon} \subset \mathrm{E}(H_{n,m})$ and so on the event that $|E_{n,\varepsilon}| \geq \varepsilon n/2$ we have

$$w(H_{n,m}) \geq n(1-\varepsilon)\varepsilon/2.$$

To conclude, we use that $w(\mathrm{MST}(\mathbb{K}_n)) \to \zeta(3)$ in probability [40]. It follows that with probability tending to 1, both $|E_{n,\varepsilon}| \geq \varepsilon n/2$ and $w(\mathrm{MST}(\mathbb{K}_n)) \leq 2\zeta(3)$, and when both these events occur we have

$$w(H_{n,m}) \geq n(1-\varepsilon)\varepsilon/2 \geq w(\mathrm{MST}(\mathbb{K}_n)) \cdot n(1-\varepsilon)\varepsilon/(4\zeta(3)) .$$

Since this holds for any optimizing sequence with weight at most $1-\varepsilon$, the result follows by taking $\delta = (1-\varepsilon)\varepsilon/(4\zeta(3))$.

Upper bound of Theorem 4.1.1. We now turn to the key ideas underlying our proof of the upper bound. We begin with a deterministic fact.

Fact 4.1.2. *Any connected graph H with vertex set $[n]$ contains an induced subgraph with at least $\frac{1}{2}\sqrt{\log_2 n}$ vertices which is either a clique, a star, or a path.*

We prove the fact immediately since the proof is very short; but its proof can be skipped without consequence for the reader's understanding of what follows.

Proof of Fact 4.1.2. The result is trivial if $n \leq 16$ so assume $n > 16$. Let $m = n^{1/\sqrt{\log_2 n}} \geq 4$. If H has maximum degree less than m then it has diameter at least $\sqrt{\log_2 n} - 1 \geq \frac{1}{2}\sqrt{\log_2 n}$ so it contains a path of length at least $\frac{1}{2}\sqrt{\log_2 n}$. On the other hand, if H has maximum degree at least m then let v be a vertex of H with degree at least m and let N_v be the set of neighbours of v in H. By Ramsey's theorem, and more concretely the diagonal Ramsey upper bound $R(k,k) < 4^k$, the graph $H[N_v]$ contains a set S of size at least

$$\frac{1}{2}\log_2 m = \frac{1}{2}\frac{\log_2 n}{\sqrt{\log_2 n}} = \frac{1}{2}\sqrt{\log_2 n}$$

such that $H[S]$ is either a clique or an independent set. If $H[S]$ is a clique then we are done, and if $H[S]$ is an independent set then $H[S \cup \{v\}]$ is a star of size $|S|+1$ so we are again done. $\square$

Fact 4.1.2 proves to be useful together with the following special case of the upper bound of Theorem 4.1.1, whose proof appears in Section 4.3.

Proposition 4.1.3. *Fix a sequence $(H_n, n \geq 1)$ of connected graphs such that, for all n, H_n is either a clique, a star, or a path with $\mathrm{V}(H_n) = [n]$. Then for all $\varepsilon > 0$, with high probability* $\mathrm{cost}(\mathbb{K}_n, H_n) \leq 1 + \varepsilon$.

We combine Proposition 4.1.3 with Fact 4.1.2 as follows. First, choose $V_n \subset [n]$ with $|V_n| \geq \frac{1}{2}\sqrt{\log n}$ such that $H_n[V_n]$ is a clique, a star or a path, and consider $\mathbb{K}_n[V_n]$, the restriction of the weighted complete graph $\mathbb{K}_n$ to V_n. Let $\mathbb{S}'_n = (S'_0, \ldots, S'_m)$ be an MST sequence for $(\mathbb{K}_n[V_n], H_n[V_n])$ of minimum cost. Now consider using the sequence $\mathbb{S}'_n$ as an optimizing sequence for $(\mathbb{K}_n, H_n)$. In other words, we set $H_{n,i} = \Phi_{\mathbb{K}_n}(H_{n,i-1}, S'_i)$ for $1 \leq i \leq m$. Then $H_{n,m} = \Phi_{\mathbb{K}_n}(H_n, V_n)$, which is to say that $H_{n,m}$ consists of H_n with $H_n[V_n]$ replaced by $\mathrm{MST}(\mathbb{K}_n[V_n])$. Moreover, by Proposition 4.1.3, $\mathrm{wt}(\mathbb{S}'_n) = \mathrm{wt}(\mathbb{K}_n, H_n, \mathbb{S}'_n) = \mathrm{wt}(\mathbb{K}_n[V_n], H_n[V_n], \mathbb{S}'_n) \xrightarrow{\text{prob}} 1$; so with high probability we have transformed a "large" (i.e. whose size is $\geq \frac{1}{2}\sqrt{\log n}$) subgraph of H_n into its minimum spanning tree, using an optimizing sequence of cost at most $1 + o_{\mathbb{P}}(1)$.

The next step is to apply a procedure we call the *eating algorithm*, described in Section 4.2. This algorithm allows us to bound the minimum cost of an MST sequence in terms of the weighted diameters of minimum spanning trees of a growing sequence of induced subgraphs of the input graph, with each graph in the sequence containing one more vertex than its predecessor. In the setting of Theorem 4.1.1, it allows us to find an MST sequence with weight at most $1 + o_{\mathbb{P}}(1)$ provided that the starting graph already contains a large subgraph on which it is equal to the MST. The key result of our analysis of the eating algorithm is summarized in the following proposition.

Proposition 4.1.4. *Fix a sequence $(H_n, n \geq 1)$ of connected graphs with $\mathrm{V}(H_n) = [n]$. Fix any sequence of sets $(V_n, n \geq 1)$ such that $V_n \subset [n]$, $|V_n| \to \infty$ as $n \to \infty$, and $H_n[V_n]$ is* *connected for all* $n \geq 1$*. Let* $H'_n = \Phi_{\mathbb{K}_n}(H_n, V_n)$*, so that* $H'_n[V_n] = \mathrm{MST}(\mathbb{K}_n[V_n])$*. Then for all* $\varepsilon > 0$*, with high probability* $\mathrm{cost}(\mathbb{K}_n, H'_n) \leq 1 + \varepsilon$*.*

The proof of Proposition 4.1.4 appears in Section 4.2. We are now prepared to prove Theorem 4.1.1, modulo the proofs of Proposition 4.1.3 and Proposition 4.1.4.

Proof of Theorem 4.1.1. We already established the lower bound of the theorem, so it remains to show that for all $\varepsilon > 0$,

$$\mathbb{P}\Big(\mathrm{cost}(\mathbb{K}_n, H_n) \leq 1 + \varepsilon \Big) \longrightarrow 1$$

as $n \to \infty$. For the remainder of the proof we fix $\varepsilon > 0$.

Using Fact 4.1.2, let V_n be a subset of $[n]$ with size at least $\frac{1}{2}\sqrt{\log_2 n}$ such that $H_n[V_n]$ is a clique, a star or a path. Write $\mathbb{K}_n^- = \mathbb{K}_n[V_n]$ and $H_n^- = H_n[V_n]$, and let $\mathbb{S}_n^-$ be an MST sequence for $(\mathbb{K}_n^-, H_n^-)$ of minimum cost. By Proposition 4.1.3,

$$\mathbb{P}\Big(\mathrm{wt}(\mathbb{K}_n^-, H_n^-, \mathbb{S}_n^-) \leq 1 + \varepsilon \Big) \longrightarrow 1$$

as $n \to \infty$. Moreover, we have $\mathrm{wt}(\mathbb{K}_n^-, H_n^-, \mathbb{S}_n^-) = \mathrm{wt}(\mathbb{K}_n, H_n, \mathbb{S}_n^-)$: the weight of the sequence $\mathbb{S}_n^-$ is the same with respect to $(\mathbb{K}_n^-, H_n^-) = (\mathbb{K}_n[V_n], H_n[V_n])$ as it is with respect to $(\mathbb{K}_n, H_n)$; this is easily seen be induction. It follows that

$$\mathbb{P}\Big(\mathrm{wt}(\mathbb{K}_n, H_n, \mathbb{S}_n^-) \leq 1 + \varepsilon \Big) \longrightarrow 1\,.$$

Next let $H'_n = \Phi_{\mathbb{K}_n}(H_n, V_n)$, so $H'_n[V_n] = \mathrm{MST}(\mathbb{K}_n[V_n])$. Since $\mathbb{S}_n^-$ is an MST sequence for $(\mathbb{K}_n^-, H_n^-)$, this is also the graph resulting from using $\mathbb{S}_n^-$ as an optimizing sequence for $(\mathbb{K}_n, H_n)$. Now let $\mathbb{S}'_n$ be an MST sequence for H'_n of minimum cost. Since $|V_n| \to \infty$ and $H_n[V_n]$ is connected, it follows from Proposition 4.1.4 that

$$\mathbb{P}\Big(\mathrm{wt}(\mathbb{K}_n, H'_n, \mathbb{S}'_n) \leq 1 + \varepsilon \Big) \longrightarrow 1\,.$$

To conclude, note that the concatenation $\mathbb{S}_n$ of $\mathbb{S}_n^-$ and $\mathbb{S}'_n$ is an MST sequence for $(\mathbb{K}_n, H_n)$, and

$$\mathrm{wt}(\mathbb{K}_n, H_n, \mathbb{S}_n) = \max\Big\{ \mathrm{wt}(\mathbb{K}_n, H_n, \mathbb{S}_n^-), \mathrm{wt}(\mathbb{K}_n, H'_n, \mathbb{S}'_n) \Big\}\,,$$

so $\mathbb{P}(\mathrm{wt}(\mathbb{K}_n, H_n, \mathbb{S}_n) \leq 1 + \varepsilon) \to 1$ and thus $\mathbb{P}(\mathrm{cost}(\mathbb{K}_n, H_n) \leq 1 + \varepsilon) \to 1$, as required. $\square$

The remainder of the paper proceeds as follows. In Section 4.2 we describe the eating algorithm and prove Proposition 4.1.4, modulo the proof of a key technical input (Theorem 4.2.3), an upper tail bound on the weighted diameter of $\mathrm{MST}(\mathbb{K}_n)$, which is postponed to Appendix 4.A. In Section 4.3 we prove Proposition 4.1.3 by using the details of the eating algorithm to generate a well bounded sequence of increasing MSTs that are each built from a clique, a star, or a path. We conclude in Section 4.4 by presenting the generalization of Theorem 4.1.1 to other edge weight distributions, and by discussing avenues for future research.

4.2 The eating algorithm

In this section, we prove Proposition 4.1.4. Informally, we prove this proposition by showing that we can efficiently add vertices to an MST of a large subgraph of K_n, one at a time, via an optimizing sequence which has a low weight, with high probability. For a weighted graph $\mathbb{G} = (V, E, w)$, write $\mathrm{wdiam}(\mathbb{G})$ for the weighted diameter of $\mathbb{G}$,

$$\mathrm{wdiam}(\mathbb{G}) := \max\left\{ \operatorname*{dist}_{\mathbb{G}}(u, v) : u, v \in V \right\},$$

where

$$\operatorname*{dist}_{\mathbb{G}}(u, v) := \min\left\{ w(P) : P \text{ is a path from } u \text{ to } v \text{ in } \mathbb{G} \right\}.$$

It is sometimes convenient to write $\mathrm{wdiam}(G)$ for an unweighted graph G, where the appropriate choice of weights is clear from context. Finally, we also introduce the unweighted diameter

$$\mathrm{diam}(G) := \max\left\{ \min\left\{ \left|\mathrm{E}(P)\right| : P \text{ is a path from } u \text{ to } v \text{ in } G \right\} : u, v \in V \right\},$$

which will be used later in this work (in Section 4.3.2 and in Appendix 4.A).

The key tool to prove Proposition 4.1.4 is the following proposition, which will be applied recursively.

Proposition 4.2.1. *Let $\mathbb{G} = (V, E, w)$ be a generic weighted graph with $V = [n]$ and $\max\{w(e) : e \in E)\} \leq 1$. Suppose that H is a spanning subgraph of G and $H[n-1] = \mathrm{MST}(\mathbb{G}[n-1])$. Then*

$$\mathrm{cost}(\mathbb{G}, H) \leq 1 + \max\left\{ \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{G}[n-1])\right), \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{G})\right) \right\}.$$

The proof of Proposition 4.2.1 occupies the bulk of Section 4.2; it appears below in Sections 4.2.1 and 4.2.2.

Corollary 4.2.2 (The eating algorithm). *Let $\mathbb{G} = (V, E, w)$ be a weighted graph with $V = [n]$ and $\max\{w(e) : e \in E\} \leq 1$. Let H be a spanning subgraph of G and fix a non-empty set $U \subset [n]$ for which $H[U] = \mathrm{MST}(\mathbb{G}[U])$. Let $U = U_0 \subset U_1 \subset \ldots \subset U_k = V$ be any increasing sequence of subsets of V such that, for all $0 \leq i < k$, $U_{i+1} \setminus U_i$ is a singleton and $H[U_i]$ is connected. Then*

$$\mathrm{cost}(\mathbb{G}, H) \leq 1 + \max\Big\{ \mathrm{wdiam}\big(\mathrm{MST}(\mathbb{G}[U_i])\big) : 0 \leq i \leq k \Big\}.$$

Proof. Set $F_0 = H$ and let $\mathbb{S}_1, \ldots, \mathbb{S}_k$ and $F_1, \ldots, F_k$ be constructed inductively as follows. Given F_{i-1}, let $\mathbb{S}_i$ be an MST sequence of minimal weight for the pair $(\mathbb{G}[U_i], F_{i-1}[U_i])$ and let $F_i = \Phi_{\mathbb{G}}(F_{i-1}, U_i)$. Note that $F_i[U_i]$ is the last graph of the subgraph sequence corresponding to $\mathbb{S}_i$.

By using that an optimizing sequence on $(\mathbb{G}[U_i], F_{i-1}[U_i])$ can also be seen as an optimizing sequence on $(\mathbb{G}, F_{i-1})$ of identical weight, we can bound the weight of the global optimizing sequence $\mathbb{S}$ obtained by concatenating $\mathbb{S}_1, \ldots, \mathbb{S}_k$ in that order. Indeed, we have that

$$\mathrm{wt}(\mathbb{G}, H, \mathbb{S}) = \max\Big\{ \mathrm{wt}(\mathbb{G}[U_i], F_{i-1}[U_i], \mathbb{S}_i) : 1 \leq i \leq k \Big\}.$$

Moreover, by the definition of F_{i-1}, we know that $F_{i-1}[U_{i-1}] = \mathrm{MST}(\mathbb{G}[U_{i-1}])$ and by minimality of $\mathbb{S}_i$ along with Proposition 4.2.1, it follows that for all $1 \leq i \leq k$,

$$\mathrm{wt}(\mathbb{G}[U_i], F_{i-1}[U_i], \mathbb{S}_i) \leq 1 + \max\Big\{ \mathrm{wdiam}\big(\mathrm{MST}(\mathbb{G}[U_{i-1}])\big), \mathrm{wdiam}\big(\mathrm{MST}(\mathbb{G}[U_i])\big) \Big\}.$$

Since $\mathrm{cost}(\mathbb{G}, H) \leq \mathrm{wt}(\mathbb{G}, H, \mathbb{S})$, combining the last two results provides us with the desired upper bound for $\mathrm{cost}(\mathbb{G}, H)$. $\square$

The importance of this corollary becomes clear in light of the next theorem, which provides strong tail bounds on the diameter of MSTs of randomly-weighted complete graphs.

Theorem 4.2.3. *Let $\mathbb{K}_n = (K_n, \mathbb{X})$ be the complete graph with vertex set $[n]$, endowed with independent, Uniform$[0,1]$ edge weights $\mathbb{X} = (X_e, e \in \mathrm{E}(K_n))$. Then for all n sufficiently large,*

$$\mathbb{P}\left(\mathrm{wdiam}\big(\mathrm{MST}(\mathbb{K}_n)\big) \geq \frac{7 \log^4 n}{n^{1/10}} \right) \leq \frac{4}{n^{\log n}}.$$

In particular, $\mathrm{wdiam}(\mathrm{MST}(\mathbb{K}_n)) \xrightarrow{\mathbb{P}} 0$ *as* $n \to \infty$.

The proof of Theorem 4.2.3 is postponed to Appendix 4.A. We now use Corollary 4.2.2 and Theorem 4.2.3 to prove Proposition 4.1.4.

Proof of Proposition 4.1.4. Consider any sequence of sets $(V_n, n \geq 1)$ with $V_n \subset [n]$ and $|V_n| \to \infty$ as $n \to \infty$ and such that $H_n[V_n]$ is connected for all $n \geq 1$, and let $H_n' = \Phi_{\mathbb{K}_n}(H_n, V_n)$. Since H_n' is connected, we may list the vertices of $[n] \setminus V_n$ as $v_1, \ldots, v_k$ so that for all $1 \leq i \leq k$, vertex v_i is adjacent to an element of $V_n \cup \{v_1, \ldots, v_{i-1}\}$. Taking $U_0 = V_n$ and $U_i = V_n \cup \{v_1, \ldots, v_i\}$ for $1 \leq i \leq k$, the sequence $U_0, \ldots, U_k$ satisfies the conditions of Corollary 4.2.2 with $\mathbb{G} = \mathbb{K}_n$. It follows that

$$\mathrm{cost}(\mathbb{K}_n, H_n') \leq 1 + \max\Big\{ \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U_i])\right) : 0 \leq i \leq k \Big\}. \tag{4.1}$$

Moreover, since $|V_n| \to \infty$ as $n \to \infty$, for n sufficiently large we may apply Theorem 4.2.3 to $\mathbb{K}_n[U_i]$ for each $0 \leq i \leq k$ and obtain that

$$\begin{aligned}\mathbb{P}\Big(\exists i : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U_i])\right) \geq |U_i|^{-\frac{1}{11}}\Big) &\leq \sum_{i=0}^{k} \mathbb{P}\Big(\mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U_i])\right) \geq |U_i|^{-\frac{1}{11}}\Big) \\ &\leq \sum_{i=0}^{k} \frac{1}{|U_i|^2}\,;\end{aligned}$$

where we have used that $\frac{7\log^4 n}{n^{1/10}} \leq \frac{1}{n^{1/11}}$ and that $\frac{4}{n^{\log n}} < \frac{1}{n^2}$ for n large. Since $|U_i| = |U_0| + i = |V_n| + i$, it follows that for all n sufficiently large,

$$\mathbb{P}\Big(\exists i : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U_i])\right) \geq |U_i|^{-\frac{1}{11}}\Big) \leq \sum_{s=|V_n|}^{n} \frac{1}{s^2} \leq \frac{1}{|V_n| - 1} \longrightarrow 0\,.$$

In view of (4.1), this yields that

$$\mathbb{P}\Big(\mathrm{cost}(\mathbb{K}_n, H_n') \geq 1 + \varepsilon\Big) \leq \mathbb{P}\Big(\exists i : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U_i])\right) \geq \varepsilon\Big) \longrightarrow 0\,,$$

as desired. $\square$

The remainder of Section 4.2 is devoted to proving Proposition 4.2.1.

4.2.1 A special case of Proposition 4.2.1

To prove Proposition 4.2.1, we need to bound $\mathrm{cost}(\mathbb{G}, H)$ when H is a spanning subgraph of G with $H[n-1] = \mathrm{MST}(\mathbb{G}[n-1])$. It is useful to first treat the special case that H only contains one edge which does not lie in $\mathrm{MST}(\mathbb{G}[n-1])$, and more specifically that n is a leaf and H is a tree. We will later use this case as an input to the general argument.

Proposition 4.2.4. *In the setting of Proposition 4.2.1, if H is a tree and n is a leaf of H then*

$$\mathrm{cost}(\mathbb{G}, H) \le 1 + \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{G}[n-1])\right).$$

The next lemma will be useful in the proof of both the special case and the general case; informally, it states that optimizing sequences never remove MST edges that are already present, and that optimizing sequences do not create cycles.

Lemma 4.2.5. *Let $(S_i, 1 \le i \le m)$ be an* MST *sequence for $(\mathbb{G}, H)$ with corresponding spanning subgraph sequence $(H_i, 0 \le i \le m)$. Then*

1. *if $e \in \mathrm{E}(\mathrm{MST}(\mathbb{G}))$ and $e \in \mathrm{E}(H_i)$, then $e \in \mathrm{E}(H_j)$ for all $i \le j \le m$, and*

2. *if H_i is a tree, then H_j is a tree for all $i \le j \le m$.*

Proof. We use the standard fact that if $\mathbb{G} = (V, E, w)$ is a weighted graph with all edge weights distinct, then $e \in \mathrm{E}(\mathrm{MST}(\mathbb{G}))$ if and only if e is not the heaviest edge of any cycle in $\mathbb{G}$.

Fix $e \in \mathrm{E}(\mathrm{MST}(\mathbb{G}))$ and suppose that $e \in \mathrm{E}(H_i)$. If the endpoints of e do not both lie in S_{i+1} then clearly $e \in \mathrm{E}(H_{i+1})$ since H_i and H_{i+1} agree except on S_{i+1}. If the endpoints of e both lie in S_{i+1} then since e is not the heaviest edge of any cycle in $\mathbb{G}$, it is not the heaviest edge of any cycle in $\mathbb{G}[S_{i+1}]$. Thus $e \in \mathrm{E}(\mathrm{MST}(\mathbb{G}[S_{i+1}]))$, and so again $e \in \mathrm{E}(H_{i+1})$. It follows by induction that $e \in \mathrm{E}(H_j)$ for all $i \le j \le m$.

The second claim of the lemma is immediate from the the fact that if T is any tree, S is a subset of $\mathrm{V}(T)$ such that $T[S]$ is a tree, and T' is another tree with $\mathrm{V}(T) = S$, then the graph with vertices $\mathrm{V}(T')$ and edges $(\mathrm{E}(T) \setminus \mathrm{E}(T[S])) \cup \mathrm{E}(T')$ is again a tree. □

We now assume $\mathbb{G}$ and H are as in Proposition 4.2.4. Define an optimizing sequence $\mathbb{S} = (S_i, 1 \le i \le n-1)$ for $(\mathbb{G}, H)$ as follows. Let S_1 be the set of vertices on the path from n to 1 in $H_0 = H$, and let $H_1 = \Phi_{\mathbb{G}}(H_0, S_1)$. Then, inductively, for $1 < i \le n-1$ let S_i be the set of vertices on the path from n to i in H_{i-1} and let $H_i = \Phi_{\mathbb{G}}(H_{i-1}, S_i)$. Since $H = H_0$ is a tree, by point 2 of Lemma 4.2.5 it follows that H_i is a tree for all i, so the paths S_i are uniquely determined and the sequence $\mathbb{S}$ is well-defined.

Proposition 4.2.4 is now an immediate consequence of the following two lemmas.

Lemma 4.2.6. $\mathbb{S}$ *is an* MST *sequence for* $(\mathbb{G}, H)$.

Proof. Since H_m is a tree, it suffices to show that $\mathrm{MST}(\mathbb{G})$ is a subtree of H_m. Let $e \in \mathrm{E}(\mathrm{MST}(\mathbb{G}))$. Then either $e \in \mathrm{E}(H_0[n-1])$ or $e = in$ for some $i \in [n-1]$. If $e \in \mathrm{E}(H_0[n-1])$ then $e \in \mathrm{E}(H_0)$ meaning that, by point 1 of Lemma 4.2.5, we have $e \in \mathrm{E}(H_m)$. Otherwise, if $e = in$ for some $i \in [n-1]$, then $e \in \mathrm{E}(\mathbb{G}[S_i])$ since S_i is the set of vertices on a path from n to i. Hence, $e \in \mathrm{E}(\mathrm{MST}(\mathbb{G}[S_i]))$, meaning that $e \in \mathrm{E}(H_i)$. Once again, by point 1 of Lemma 4.2.5, this implies that $e \in \mathrm{E}(H_m)$, proving that $\mathrm{MST}(\mathbb{G})$ is a subtree of H_m. □

Lemma 4.2.7. $\mathrm{wt}(\mathbb{S}) \leq 1 + \mathrm{wdiam}(\mathrm{MST}(\mathbb{G}[n-1])$

Proof. Let $i \in [n-1]$. Notice that the path from n to i in H_{i-1} contains a single edge from n to $[n-1]$. Hence, the weight of this path is bounded from above by $1 + \mathrm{wdiam}(H_{i-1}[n-1])$. To prove the lemma it therefore suffices to show that $\mathrm{E}(H_i[n-1]) \subseteq \mathrm{E}(H_0[n-1]) = \mathrm{E}(\mathrm{MST}(\mathbb{G}[n-1]))$.

We prove this by induction on i, the base case $i = 0$ being automatic. For $i > 0$, suppose that $\mathrm{E}(H_{i-1}[n-1]) \subseteq \mathrm{E}(H_0[n-1])$. Fix any vertices $u, v \in S_i \cap [n-1]$ with $uv \notin \mathrm{E}(H_{i-1})$ and let P be the path from u to v in H_{i-1}. Then P is a subpath of $H_{i-1}[S_i]$, and so by induction it is also a subpath of H_0. Since $H_0[n-1] = \mathrm{MST}(\mathbb{G}[n-1])$ it follows that P is a subpath of $\mathrm{MST}(\mathbb{G}[n-1])$. This yields that uv is the edge with highest weight on the cycle created by closing P, and all the vertices of this cycle lie in S_i; so $uv \notin \mathrm{E}(\mathrm{MST}(\mathbb{G}[S_i]))$ and thus $uv \notin \mathrm{E}(H_i)$. This shows that $\mathrm{E}(H_i[S_i \setminus \{n\}]) \subseteq \mathrm{E}(H_{i-1}[S_i \setminus \{n\}]) \subseteq \mathrm{E}(H_0[S_i \setminus \{n\}])$. Since the rest of $H_{i-1}[n-1]$ and $H_i[n-1]$ are identical, it follows that $\mathrm{E}(H_i[n-1]) \subseteq \mathrm{E}(H_0[n-1])$, as required. □

4.2.2 The general case of Proposition 4.2.1

We now lift the assumption that H is a tree; in this case, $\mathrm{E}(H) \setminus \mathrm{E}(H[n-1])$ could contain up to $n-1$ edges. As a result, the MST sequence previously defined in Section 4.2.1 does not provide us with the desired cost, since a path from n to $i \in [n-1]$ might contain additional edges with n as an endpoint, increasing the weight of the sequence. Thus, we require a more careful method. Informally, our approach is to first apply the method from the previous section to a sequence of subgraphs of $H[n-1]$, each of which is only joined to the vertex n by a single edge, but together which contain all the edges from n to $[n-1]$. We show that this yields a graph which contains the MST of G. We then prove that any cycles in the resulting graph can be removed at a low cost.

Let $\mathbb{G} = (V, E, w)$ be a generic weighted graph with $V = [n]$ and let H be a spanning subgraph of G with $H[n-1] = \mathrm{MST}(\mathbb{G}[n-1])$. Let $\{v_1 n, \ldots, v_k n\} \subseteq \mathrm{E}(H)$ be the set of edges in H with n as an endpoint, and for $1 \le i \le k$ let

$$V_i = \left\{ v \in [n-1] : \operatorname*{dist}_{(H[n-1],w)}(v_i, v) = \min\left\{ \operatorname*{dist}_{(H[n-1],w)}(v_j, v) : 1 \le j \le k \right\} \right\}.$$

That is to say, $(V_i, 1 \le i \le k)$ is the Voronoi partition of $[n-1]$ in $H[n-1]$ with respect to the vertices $v_1, \ldots, v_k$; it is indeed a partition since $\mathbb{G}$ is generic.

Note that since $H[n-1] = \mathrm{MST}(\mathbb{G}[n-1])$ it follows that $H[V_i] = \mathrm{MST}(\mathbb{G}[V_i])$ for any $1 \le i \le k$. Moreover, vertex n has degree one in $H[V_i \cup \{n\}]$. Using Proposition 4.2.4, let $\mathbb{S}_i = (S_{i,j}, 1 \le j \le m_i)$ be an MST sequence for $(\mathbb{G}[V_i \cup \{n\}], H[V_i \cup \{n\}])$ with weight at most $1 + \mathrm{wdiam}(\mathrm{MST}(\mathbb{G}[V_i])) \le 1 + \mathrm{wdiam}(\mathrm{MST}(\mathbb{G}[n-1]))$, and write $(H_{i,j}, 0 \le j \le m_i)$ for the corresponding subgraph sequence. Now set $m = m_1 + \ldots + m_k$ and let $\mathbb{S}^* = (S_1^*, \ldots, S_m^*)$ be formed by concatenating $\mathbb{S}_1, \ldots, \mathbb{S}_k$, so

$$\mathbb{S}^* = (S_{1,1}, \ldots, S_{1,m_1}, \ldots, S_{k,1}, \ldots, S_{k,m_k}),$$

and let $(H_0^*, \ldots, H_m^*)$ be the subgraph sequence corresponding to $\mathbb{S}^*$.

Lemma 4.2.8. *We have* $\mathrm{MST}(\mathbb{G}) \subseteq H_m^*$, *and* $\mathrm{wt}(\mathbb{S}^*) \le 1 + \mathrm{diam}(\mathrm{MST}(\mathbb{G}[n-1]))$.

Proof. First, by assumption, $H_0[n-1] = \mathrm{MST}(\mathbb{G}[n-1])$. Since $\mathrm{MST}(\mathbb{G})[n-1]$ is a subgraph of $\mathrm{MST}(\mathbb{G}[n-1])$, point 1 of Lemma 4.2.5 implies that $\mathrm{MST}(\mathbb{G})[n-1]$ is a subgraph of H_i^* for all i, so in particular of H_m^*.

Next, since $V_1, \ldots, V_k$ are disjoint, we have $S_{i,j} \cap S_{i',j'} \subseteq \{n\}$ whenever $i \ne i'$, and it follows that $H^*_{m_1+\ldots+m_{i-1}}[V_i \cup \{n\}] = H[V_i \cup \{n\}]$ for all $1 \le i \le k$. This implies that $H^*_{m_1+\ldots+m_{i-1}+j}[V_i \cup \{n\}] = H_{i,j}$ for each $1 \le j \le m_i$, so in particular $H^*_{m_1+\ldots+m_i}[V_i \cup \{n\}] = \mathrm{MST}(\mathbb{G}[V_i \cup \{n\}])$.

Now fix any edge vn of $\mathrm{MST}(\mathbb{G})$. Then $v \in V_i$ for some $1 \le i \le k$, so $vn \in \mathrm{E}(\mathrm{MST}(\mathbb{G}[V_i \cup \{n\}]))$. It follows that $vn \in H^*_{m_1+\ldots+m_i}$, and thus by point 1 of Lemma 4.2.5 that vn is an edge of H_m^*. Therefore all edges of $\mathrm{MST}(\mathbb{G})$ are edges of H_m^*, as required.

Finally, the bound on the weight of the sequence is immediate by the definition of $\mathbb{S}^*$ and by using that $\mathrm{wt}(\mathbb{G}[V_i \cup \{n\}], H[V_i \cup \{n\}], \mathbb{S}_i) = \mathrm{wt}(\mathbb{G}, H, \mathbb{S}_i)$. $\square$

We are now left to deal with the edges $\mathrm{E}(H_m^*) \setminus \mathrm{E}(\mathrm{MST}(\mathbb{G}))$. This is taken care of in the following lemma.

Lemma 4.2.9. *Let $\mathbb{G} = (V, E, w)$ be a generic weighted graph with $V = [n]$ and with all edge weights at most 1, and let H be a subgraph of G such that $\mathrm{MST}(\mathbb{G})$ is a subgraph of H. Write $k = |\mathrm{E}(H)| - (n-1)$. Then there exists an MST sequence $\mathbb{S}' = (S_1', \ldots, S_k')$ with*

$$\mathrm{wt}(\mathbb{S}') \leq 1 + \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{G})\right).$$

Proof. If H is a tree then there is nothing to prove, so assume H contains at least one cycle (so $k \geq 1$). In this case there exist vertices u, v which are not adjacent in $\mathrm{MST}(\mathbb{G})$ but are joined by an edge in H; choose such u and v so that the length (number of edges) on the path P from u to v in $\mathrm{MST}(\mathbb{G})$ is as small as possible. Let $S = \mathrm{V}(P)$ be the set of vertices of the path P; then $H[S]$ is a cycle (by the minimality of the length of P), and uv is the edge with largest weight on $H[S]$. It follows that $\mathrm{MST}(\mathbb{G}[S]) = P$, so $\Phi_{\mathbb{G}}(H, S)$ has edge set $E = \mathrm{E}(H) \setminus \{uv\}$. Moreover, since P is a path of $\mathrm{MST}(\mathbb{G})$, it follows that

$$w(H[S]) = w(uv) + \mathrm{wt}(P) \leq 1 + \mathrm{wdiam}(\mathrm{MST}(\mathbb{G})).$$

Since $\Phi_{\mathbb{G}}(H, S)$ contains $\mathrm{MST}(G)$ but has one fewer edge than H, the result follows by induction. $\square$

We now combine Lemmas 4.2.8 and 4.2.9 to conclude the proof of Proposition 4.2.1.

Proof of Proposition 4.2.1. Let $\mathbb{S}^* = (S_1^*, \ldots, S_m^*)$ be the optimization sequence defined above Lemma 4.2.8, and let $(H_0^*, \ldots, H_m^*)$ be the corresponding subgraph sequence. By that lemma, $\mathrm{MST}(\mathbb{G})$ is a subgraph of H_m^* and $\mathrm{wt}(\mathbb{S}^*) \leq 1 + \mathrm{wdiam}(\mathrm{MST}(\mathbb{G}[n-1]))$.

Next let $\mathbb{S}' = (S_1', \ldots, S_k')$ be an MST sequence for $(\mathbb{G}, H_m^*)$ of weight at most $1 + \mathrm{wdiam}(\mathrm{MST}(\mathbb{G}))$; the existence of such a sequence is guaranteed by Lemma 4.2.9. Then the concatenation

$$\mathbb{S} = (S_1^*, \ldots, S_m^*, S_1', \ldots, S_k')$$

of $\mathbb{S}^*$ and $\mathbb{S}'$ is an MST sequence for $(\mathbb{G}, H)$, of weight at most

$$\mathrm{wt}(\mathbb{G}, H, \mathbb{S}) \leq 1 + \max\Big\{ \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{G}[n-1])\right), \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{G})\right) \Big\},$$

and the desired bound on $\mathrm{cost}(\mathbb{G}, H)$ follows. $\square$

4.3 MST sequences for the clique, the star, and the path

This section is aimed at proving Proposition 4.1.3. We start by proving the result in the case of the clique, since it is straightforward using the result of Lemma 4.2.9. After that, the case of the star and the path are covered together; the proof in those cases uses the eating algorithm, Corollary 4.2.2, to find adequate sequences of increasing subsets on which to build increasing sequences of MSTs.

Proof of Proposition 4.1.3 (Case of the clique). Using Lemma 4.2.9, since $\mathrm{MST}(\mathbb{K}_n)$ is a subgraph of $H_n = K_n$, it follows that

$$\mathrm{cost}(\mathbb{K}_n, H_n) \leq 1 + \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n)\right).$$

By Theorem 4.2.3 we have $\mathrm{wdiam}(\mathrm{MST}(\mathbb{K}_n)) \xrightarrow{\mathrm{prob}} 0$, and the result follows. $\square$

4.3.1 MST sequences for the star and the path

In this section, we assume that H_n is either a star or a path. If H_n is a star, then by relabeling we may assume H_n has center n, so has edge set $\{e_1, \ldots, e_{n-1}\}$ with $e_i = in$; call this star S_n. If H_n is a path, then by relabeling we may assume H_n is the path $P_n = 12\ldots n$, so has edge set $\{e_i, \ldots, e_{n-1}\}$ with $e_i = i(i+1)$. In either case, with this edge labeling, for any $1 \leq i < j \leq n-1$, the set $V(i,j)$ defined as the endpoints in $\{e_i, \ldots, e_{j-1}\}$ is connected in H_n. Note that $V(i,j) = \{i, \ldots, j-1\} \cup \{n\}$ when H_n is a star and $V(i,j) = \{1, \ldots, j\}$ when H_n is a path, and in both cases $|V(i,j)| = j - i + 1$. For the remainder of the section, it might be helpful to imagine that H_n is the path, $12\ldots n$.

Recall that $\mathbb{X} = (X_e, e \in \mathrm{E}(K_n))$ is a set of independent Uniform$[0,1]$ random variables. For $W \in (0,1)$ and $2 \leq L < n-1$, let

$$\mathrm{I} = \mathrm{I}(W, L) = (n-L) \wedge \min\left\{i : \forall i \leq j < i+L, X_{e_j} \leq W\right\}. \tag{4.2}$$

Note that I is a function of $\mathbb{X}$ and more precisely that

$$\{\mathrm{I} \leq k\} \in \sigma\Big(\big\{X_{e_i} \leq W\big\}, 1 \leq i < k+L\Big)\,,$$

where $\sigma(X)$ is the σ-algebra generated by X.

Next, let $\mathbb{U} = \mathbb{U}(\mathrm{I}) = (U_i, 0 \leq i < n-L)$ be the sequence of sets defined as follows.

$$(U_0, \ldots, U_{n-L-1}) = \Big(V(\mathrm{I}, \mathrm{I}+L), \ldots, V(\mathrm{I}, n), V(\mathrm{I}-1, n), \ldots, V(1, n)\Big)\,. \tag{4.3}$$

In words, U_0 is the set of vertices that belong to the edges $e_{\mathrm{I}}, \ldots, e_{\mathrm{I}+L-1}$ (that is $V(\mathrm{I}, \mathrm{I}+L)$); then we sequentially build $U_1, \ldots, U_{n-L-1}$ by first adding the vertices belonging to $e_{\mathrm{I}+L}, \ldots, e_{n-1}$, then adding the vertices belonging to $e_{\mathrm{I}-1}, \ldots, e_1$; see Figure 4.1 for a representation of I and $\mathbb{U}$.

We now use the sequence $\mathbb{U}$ to bound the cost of $(\mathbb{K}_n, H_n)$ when H_n is a star or a path. The following lemma gives a first bound on the cost using $\mathbb{U}$.

Lemma 4.3.1. *Let H_n be the star S_n or path P_n. Then, conditionally given that* $\mathrm{I}(W, L) < n - L$*, we have*

$$\mathrm{cost}(\mathbb{K}_n, H_n) \leq \max\left\{WL, 1 + \max\Big\{\,\mathrm{wdiam}\,\big(\,\mathrm{MST}(\mathbb{K}_n[U_i])\big) : 0 \leq i < n-L\Big\}\right\}.$$

Proof. This result almost directly follows from Corollary 4.2.2. Indeed, let $H_n' = \Phi(H_n, U_0)$. Then the sets $U_0, \ldots, U_{n-L-1}$ satisfy the condition of Corollary 4.2.2 with $H = H_n'$, implying that

$$\mathrm{cost}(\mathbb{K}_n, H_n') \leq 1 + \max\Big\{\,\mathrm{wdiam}\,\big(\,\mathrm{MST}(\mathbb{K}_n[U_i])\big) : 0 \leq i < n-L\Big\}\,.$$

But now, by concatenating any minimal weight MST sequence for $(\mathbb{K}_n[U_0], H_n[U_0])$ and any minimal weight MST sequence for $(\mathbb{K}_n, H_n')$, it follows that

$$\mathrm{cost}(\mathbb{K}_n, H_n) \leq \max\Big\{\,\mathrm{cost}\,\big(\mathbb{K}_n[U_0], H_n[U_0]\big), \mathrm{cost}\,\big(\mathbb{K}_n, H_n'\big)\Big\}\,.$$

In order to complete the proof of the lemma, note that, conditionally given $\mathrm{I} < n - L$,

$$w(H_n[U_0]) = \sum_{e \in \mathrm{E}(H_n[U_0])} X_e \leq WL\,.$$

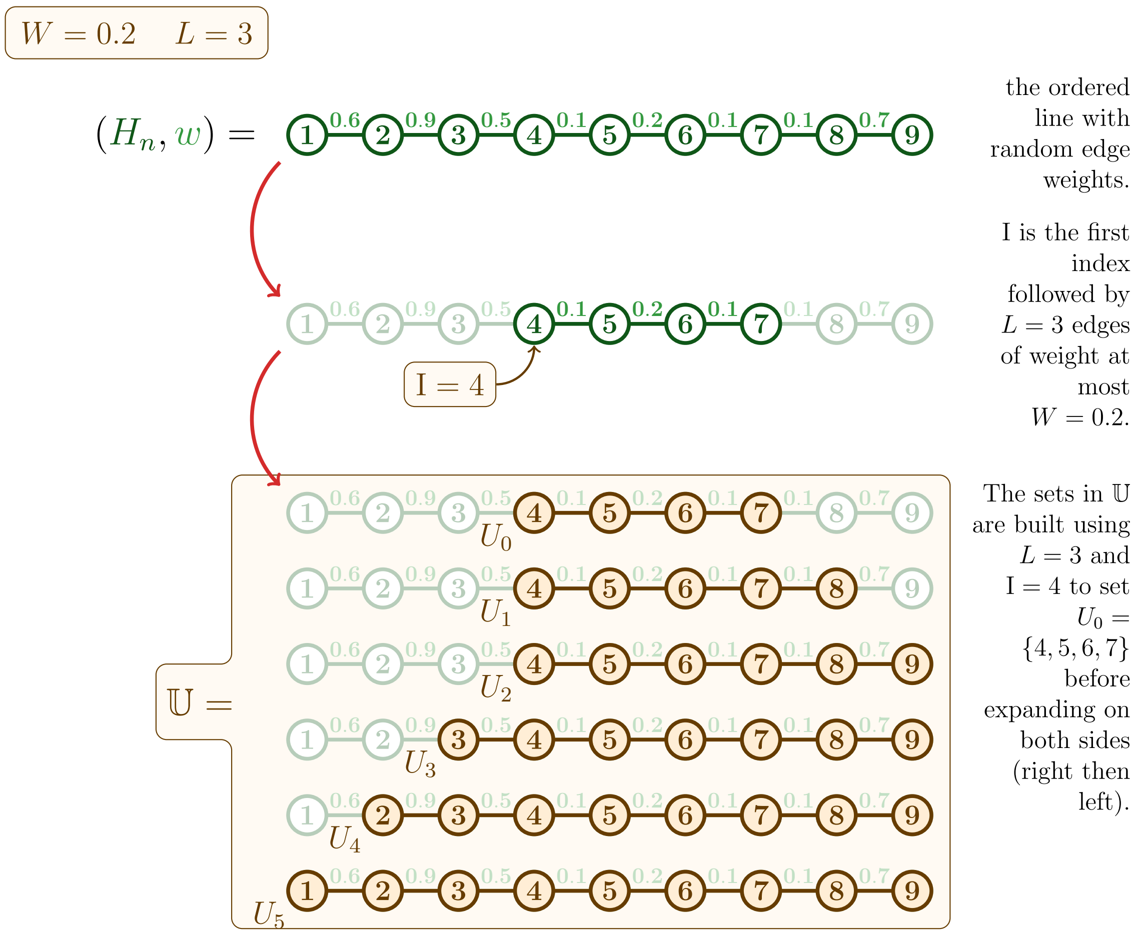

Figure 4.1: An example of I and $\mathbb{U}$ for an instance of the weighted ordered line (H_n, w), with $W = 0.2$ and $L = 3$. First, I is set to be the first sequence of $L = 3$ consecutive edges with weights less than $W = 0.2$. In this example, $\mathrm{I} = 4$. Then, given I, set $U_0 = V(\mathrm{I}, \mathrm{I} + L) = \{4, 5, 6, 7\}$ and expand first to the right and then to left to obtain $U_1, \ldots, U_{n-L-1}$. In other words, in order to obtain U_1, U_2, U_3, U_4, and U_5, we sequentially add 8, 9, 3, 2, and 1 to U_0.

Taking $\mathbb{S} = (U_0)$, this yields

$$\mathrm{cost}\left(\mathbb{K}_n[U_0], H_n[U_0]\right) \leq \mathrm{wt}\left(\mathbb{K}_n[U_0], H_n[U_0], \mathbb{S}\right) = w\big(H_n[U_0]\big) \leq WL\,.$$

This proves the desired upper bound and concludes the proof of the lemma. □

The next two results, combined with Lemma 4.3.1, will allow us to give the full proof of Proposition 4.1.3 when H_n is either a star or a path.

Proposition 4.3.2. *For any $\varepsilon > 0$, for $W = \frac{1}{\log n}$ and $L = \lfloor \log\log n \rfloor$, as $n \to \infty$ we have*

$$\mathbb{P}\Big(\exists U \in \mathbb{U}(\mathrm{I}(W,L)) : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon\Big) \longrightarrow 0\,.$$

Lemma 4.3.3. *Let $W = \frac{1}{\log n}$ and $L = \lfloor \log\log n \rfloor$. Then, for any $a > 0$, as $n \to \infty$ we have*

$$\mathbb{P}\big(\mathrm{I}(W,L) \geq n^a\big) \longrightarrow 0\,.$$

Lemma 4.3.3 is straightforward and we prove it immediately. On the other hand, Proposition 4.3.2 is quite technical and we dedicate Section 4.3.2 below to proving it.

Proof of Lemma 4.3.3. For any integer $k \geq 1$, by the definition of I,

$$\begin{aligned}\mathbb{P}\big(\mathrm{I} \geq kL+1\big) &= \mathbb{P}\Big(\forall i < kL+1, \exists j \in \{i, \dots, i+L-1\} : X_{e_j} > W\Big) \\ &\leq \mathbb{P}\Big(\forall i \in \big\{1, 1+L, \dots, 1+(k-1)L\big\}, \exists j \in \{i, \dots, i+L-1\} : X_{e_j} > W)\Big)\,.\end{aligned}$$

But then, by independence of the weights of $\mathbb{X}$, we have

$$\begin{aligned}\mathbb{P}\big(\mathrm{I} \geq kL+1\big) &= \prod_{i=0}^{k-1} \mathbb{P}\Big(\exists j \in \{1+iL, \dots, 1+(i+1)L-1\} : X_{e_j} > W\Big) \\ &= \prod_{i=0}^{k-1}\Big(1 - W^L\Big) \leq e^{-kW^L}\,,\end{aligned}$$

where the last inequality follows from the convexity of the exponential. Applying this result with $k = \lfloor \frac{n^a - 1}{L} \rfloor$, we obtain

$$\mathbb{P}\big(\mathrm{I} \geq n^a\big) \leq \mathbb{P}\big(\mathrm{I} \geq kL+1\big) \leq \exp\left(-\left\lfloor \frac{n^a-1}{L}\right\rfloor \cdot W^L\right)\,,$$

and the final expression tends to 0 as $n \to \infty$. $\square$

Proof of Proposition 4.1.3 (Case of the star and the path). Let $W = \frac{1}{\log n}$ and $L = \lfloor \log\log n \rfloor$. Fixing $\varepsilon > 0$, we have

$$\mathbb{P}\Big(\operatorname{cost}(\mathbb{K}_n, H_n) > 1+\varepsilon \Big) \le \mathbb{P}\Big(\operatorname{cost}(\mathbb{K}_n, H_n) > 1+\varepsilon \;\Big|\; \mathrm{I} < n-L \Big) + \mathbb{P}\Big(\mathrm{I} = n-L \Big).$$

Applying Lemma 4.3.3 with any $a < 1$, for large enough n we have

$$\mathbb{P}\Big(\mathrm{I} = n-L \Big) \le \mathbb{P}\Big(\mathrm{I} \ge n^a \Big) \longrightarrow 0 .$$

Hence, we have

$$\mathbb{P}\Big(\operatorname{cost}(\mathbb{K}_n, H_n) > 1+\varepsilon \Big) = \mathbb{P}\Big(\operatorname{cost}(\mathbb{K}_n, H_n) > 1+\varepsilon \;\Big|\; \mathrm{I} < n-L \Big) + o(1) .$$

Since $WL \to 0$, combining the previous bound with Lemma 4.3.1 leads to

$$\begin{aligned}
&\mathbb{P}\Big(\operatorname{cost}(\mathbb{K}_n, H_n) > 1+\varepsilon \Big) \\
&\quad \le \mathbb{P}\left(\max\Big\{ WL, 1 + \max\Big\{ \operatorname{wdiam}\big(\mathrm{MST}(\mathbb{K}_n[U_i])\big) \Big\}\Big\} > 1+\varepsilon \;\middle|\; \mathrm{I} < n-L \right) + o(1) \\
&\quad = \mathbb{P}\left(\max\Big\{ \operatorname{wdiam}\big(\mathrm{MST}(\mathbb{K}_n[U])\big) : U \in \mathbb{U} \Big\} > \varepsilon \;\middle|\; \mathrm{I} < n-L \right) + o(1) .
\end{aligned}$$

The upper bound now follows from Proposition 4.3.2, once again since $\mathbb{P}(\mathrm{I} < n-L) \to 0$. $\square$

4.3.2 Proof of Proposition 4.3.2

In this section, we prove Proposition 4.3.2, which concludes the proof of Proposition 4.1.3. Before doing so, we state a proposition which is an important input to the proof.

Proposition 4.3.4. *Let $\mathbb{G} = (G, w)$ be a weighted graph. Let T be a subtree (not necessarily spanning) of $\mathbb{G}$ and let $\mathbb{G}^* = (G, w^*)$ be a weighted graph such that $w^*(e) \le w(e)$ for $e \in \mathrm{E}(T)$ and $w^*(e) = w(e)$ otherwise. Then*

$$\operatorname{wdiam}\big(\mathrm{MST}(\mathbb{G}^*)\big) \le w^*(T) + |\mathrm{V}(T)| \times \operatorname{wdiam}\big(\mathrm{MST}(\mathbb{G})\big) .$$

Moreover, if T is a subtree of $\mathrm{MST}(\mathbb{G}^*)$*, then*

$$\operatorname{wdiam}\big(\mathrm{MST}(\mathbb{G}^*)\big) \le w^*(T) + 2 \times \operatorname{wdiam}\big(\mathrm{MST}(\mathbb{G})\big)\,.$$

Proof. Let us try to understand the relation between $\mathrm{MST}(\mathbb{G})$ and $\mathrm{MST}(\mathbb{G}^*)$. First note that

$$\mathrm{E}(\mathrm{MST}(\mathbb{G}^*)) \subset \mathrm{E}\big(\mathrm{MST}(\mathbb{G})\big) \cup \mathrm{E}(T)\,. \tag{4.4}$$

Indeed, any edge $e \notin \mathrm{E}(T)$ has the same weight with respect to w and w^*. Then, for any $e \in \mathrm{E}(\mathrm{MST}(\mathbb{G}^*)) \setminus \mathrm{E}(T)$, no cycle has e as the heaviest edge with respect to w^*, which implies that no cycle has e as the heaviest edge with respect to w, and thus $e \in \mathrm{E}(\mathrm{MST}(\mathbb{G}))$.

Consider now a path P contained in $\mathrm{MST}(\mathbb{G}^*)$. Using (4.4), we have

$$\mathrm{E}(P) \subseteq \mathrm{E}\big(\mathrm{MST}(\mathbb{G})\big) \cup \mathrm{E}(T)\,,$$

so we may uniquely decompose P into pairwise edge-disjoint paths $P_0, \ldots, P_{2k}$, where $k \ge 1$, and P_i is a subpath of T for i odd and of $\mathrm{MST}(\mathbb{G})$ for i even (it is possible that either or both of P_0, P_{2k} consists of a single vertex). Since $P_1, P_3, \ldots, P_{2k-1}$ are disjoint subpaths of T, it follows that $k \le |\mathrm{E}(T)|$ and that $\sum_{i \text{ odd}} w^*(P_i) \le w^*(T)$. Moreover, each of the paths $P_0, P_2, \ldots, P_{2k}$ have weight at most $\operatorname{wdiam}\big(\mathrm{MST}(\mathbb{G})\big)$, so

$$\begin{aligned}\sum_{i \text{ even}} w^*(P_i) = \sum_{i \text{ even}} w(P_i) &\le (k+1) \times \operatorname{wdiam}\big(\mathrm{MST}(\mathbb{G})\big) \\ &\le \big(|\mathrm{E}(T)| + 1\big) \times \operatorname{wdiam}\big(\mathrm{MST}(\mathbb{G})\big) \\ &= |\mathrm{V}(T)| \times \operatorname{wdiam}\big(\mathrm{MST}(\mathbb{G})\big)\,.\end{aligned} \tag{4.5}$$

The first bound of the proposition follows since

$$w^*(P) = \sum_{i \text{ even}} w^*(P_i) + \sum_{i \text{ odd}} w^*(P_i)\,.$$

To establish the second bound, note that if T is a subtree of $\mathrm{MST}(\mathbb{G}^*)$ then in the above decomposition of P we must have $k = 1$; a path in $\mathrm{MST}(\mathbb{G}^*)$ may enter T and then leave it, after which it can never reenter T. In this case the first inequality of (4.5) becomes

$$\sum_{i \text{ even}} w(P_i) \leq 2 \times \text{wdiam}\big(\,\text{MST}(\mathbb{G})\big)\,,$$

so we obtain

$$w^*(P) = \sum_{i \text{ even}} w^*(P_i) + \sum_{i \text{ odd}} w^*(P_i) \leq w^*(T) + 2 \times \text{wdiam}(\text{MST}(\mathbb{G}))\,,$$

as required. $\square$

For the remainder of this section we assume $W = \frac{1}{\log n}$ and $L = \lfloor \log\log n \rfloor$ and write $\mathrm{I} = \mathrm{I}(W, L)$. Consider the partition $\mathbb{U} = \mathbb{U}_r^- \cup \mathbb{U}_r^+ \cup \mathbb{U}_\ell$ where $\mathbb{U}_r^- = \mathbb{U}_r^-(\mathrm{I}) = (U_i, 0 \leq i \leq \min(L^{20}, n - \mathrm{I} - L))$, $\mathbb{U}_r^+ = \mathbb{U}_r^+(\mathrm{I}) = (U_i, \min(L^{20}, n - \mathrm{I} - L) < i \leq n - \mathrm{I} - L)$, and $\mathbb{U}_\ell = \mathbb{U}_\ell(\mathrm{I}) = (U_i, n - \mathrm{I} - L < i \leq n - L - 1)$. Then, in the case where $\mathrm{I} < n - L - L^{20}$, $\mathbb{U}_r^-$ corresponds to adding the first L^{20} vertices on the right of U_0, $\mathbb{U}_r^+$ corresponds to adding all remaining vertices on the right, and $\mathbb{U}_\ell$ corresponds to adding the vertices on the left of U_0. We aim to prove tail bounds similar to that of Proposition 4.3.2 for each of the sets $\mathbb{U}_r^-$, $\mathbb{U}_r^+$, and $\mathbb{U}_\ell$, and we start with an important lemma regarding the distribution of $\mathbb{G}$ conditioned on the value of I.

Lemma 4.3.5. *Fix $k < n - L$ and let $\mathbb{K}_n^* = (K_n, \mathbb{X}^*)$ have the law of $\mathbb{K}_n$ conditioned on the event that $\mathrm{I}(W, L) = k$. Then for any $e \in \{e_i, k \leq i < k+L\}$, X_e^* is a* $\text{Uniform}[0, W]$*; for any $e \notin \{e_i : 1 \leq i < k+L\}$, $X_{e_i}^*$ is a random* $\text{Uniform}[0, 1]$*, and the edge weights $\big(X_e^*, e \in \mathrm{E}(K_n) \setminus \{e_i, 1 \leq i < k\}\big)$ are mutually independent and independent of $(X_e^*, e \in \{e_i, 1 \leq i < k\})$. It follows that there exists a coupling between $\mathbb{K}_n^* = (K_n, \mathbb{X}^*)$ and $\mathbb{K}_n' = (K_n, \mathbb{X}')$ where $\mathbb{X}'$ is a set of independent* $\text{Uniform}[0, 1]$*, such that $X_e^* \leq X_e'$ if $e \in \{e_i : k \leq i < k + L\}$, and $X_e^* = X_e'$ if $e \in \mathrm{E}(K_n) \setminus \{e_i : 1 \leq i < k + L\}$.*

Proof. Using the definition of I, we know that

$$\big\{\mathrm{I} = k\big\} \in \sigma\Big(\big\{X_{e_i} \leq W : 1 \leq i < k + L\big\}\Big)\,,$$

from which it directly follows that the distribution of X_e is a $\text{Uniform}[0, 1]$ for any $e \notin \{e_{n,i} : 1 \leq i < k + L\}$. Furthermore, for any $e \in \{e_i : k \leq i < k + L\}$, X_e conditioned on $\mathrm{I} = k$ is the same as X_e conditioned on $X_e \leq W$. Since X_e is uniformly distributed, it follows that

X_e conditioned on $\mathrm{I} = k$ is a Uniform$[0, W]$. Finally, note that

$$\big\{\mathrm{I} = k\big\} = \big\{X_{e_i} \leq W : k \leq i < k + L\big\} \cap \bigcap_{j=1}^{k-1} \Big\{\exists j \leq i < \min\{j + L, k\} : X_{e_i} > W\Big\},$$

from which we see that the edges of $\mathrm{E}(K_n) \setminus \{e_i, 1 \leq i < k\}$ are conditionally independent of $\{e_i, 1 \leq i < k\}$ given that $\mathrm{I} = k$. It follows that all the edges in $\mathrm{E}(\mathbb{K}_n) \setminus \{e_i, 1 \leq i < k\}$ have independent weights in $\mathbb{K}_n^*$. The existence of the coupling asserted in the lemma is then an immediate consequence. □

We now split the proof of Proposition 4.3.2 into proving analogous statements for the three different sets $\mathbb{U}_r^-$, $\mathbb{U}_r^+$, and $\mathbb{U}_\ell$.

First right set $\mathbb{U}_r^-$.

Lemma 4.3.6. *For any $\varepsilon > 0$, we have*

$$\mathbb{P}\Big(\exists U \in \mathbb{U}_r^- : \operatorname{wdiam}\big(\operatorname{MST}(\mathbb{K}_n[U])\big) > \varepsilon\Big) \longrightarrow 0\,.$$

Proof. Fix $0 < a < 1$ and assume n is large enough so that $n^a < n - L - L^{20}$. Then, by Lemma 4.3.3, we have

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U}_r^- : \operatorname{wdiam}\big(\operatorname{MST}(\mathbb{K}_n[U])\big) > \varepsilon\Big) \\
&\leq \mathbb{P}\Big(\exists U \in \mathbb{U}_r^- : \operatorname{wdiam}\big(\operatorname{MST}(\mathbb{K}_n[U])\big) > \varepsilon \;\Big|\; \mathrm{I} < n - L - L^{20}\Big) + \mathbb{P}\big(\mathrm{I} \geq n - L - L^{20}\big) \\
&= \mathbb{P}\Big(\exists U \in \mathbb{U}_r^- : \operatorname{wdiam}\big(\operatorname{MST}(\mathbb{K}_n[U])\big) > \varepsilon \;\Big|\; \mathrm{I} < n - L - L^{20}\Big) + o(1)\,.
\end{aligned}$$

Next fix $k < n - L - L^{20}$ and condition on the event $\mathrm{I} = k$. Under this conditioning, $\mathbb{U}_r^- = \mathbb{U}_r^-(\mathrm{I}) = \mathbb{U}_r^-(k)$ is a deterministic sequence of sets. Further recall from (4.3) that $U_0 = V(k, k+L)$ consists of the endpoints of the edges $e_k, \ldots, e_{k+L-1}$, so equals $\{k, \ldots, k+L\}$ if H_n is the path P_n and equals $\{k, \ldots, k + L - 1, n\}$ if H_n is the star S_n. Let $T = H_n[U_0]$. Since $\mathrm{I} = k < n - L$, all edges in T have weight less than W. Now, suppose that all other edges of $\mathbb{K}_n[U_{L^{20}}]$ have weight larger than W. In this case, T is a subtree of $\operatorname{MST}(\mathbb{K}_n[U_{L^{20}}])$, from which it follows that T is a subtree of $\operatorname{MST}(\mathbb{K}_n[U_i])$ for any $0 \leq i \leq L^{20}$ (since $U_i \subset U_{L^{20}}$ for such U_i). Now, using that $\{\mathrm{I} = k\} \in \sigma(\{X_{e_i} : 1 \leq i < k + L\})$, we have

$$\mathbb{P}\Big(\forall e \in \mathrm{E}\big(\mathbb{K}_n[U_{L^{20}}]\big) \setminus \mathrm{E}(T), X_e > W \;\Big|\; \mathrm{I} = k\Big) = \big(1 - W\big)^{\binom{L^{20}}{2} - L}\,.$$

Since $W = \frac{1}{\log n}$, we have $1 - W \geq \exp(-2W)$ for n large, so

$$\begin{aligned}
\mathbb{P}\Big(\mathrm{E}(T) \subset \mathrm{E}\big(\,\mathrm{MST}\left(\mathbb{K}_n[U_{L^{20}}]\right)\big) \;\Big|\; \mathrm{I} = k\Big) &\geq \left(1 - W\right)^{\binom{L^{20}}{2} - L} \\
&\geq \exp\left(-2W\left(\binom{L^{20}}{2} - L\right)\right) \\
&\geq \exp\left(-WL^{40}\right) \\
&\geq 1 - \frac{(\log\log n)^{40}}{\log n}\,,
\end{aligned}$$

the last inequality holding since $W = \frac{1}{\log n}$, $L = \lfloor \log\log n \rfloor$, and $e^{-x} \geq 1 - x$ for $x \geq 0$. Hence,

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U}_r^- : \mathrm{wdiam}\left(\,\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon \;\Big|\; \mathrm{I} = k\Big) \\
&\quad \leq \mathbb{P}\Big(\exists U \in \mathbb{U}_r^- : \mathrm{wdiam}\left(\,\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon, \mathrm{E}(T) \subset \mathrm{E}\big(\,\mathrm{MST}\left(\mathbb{K}_n[U_{L^{20}}]\right)\big) \;\Big|\; \mathrm{I} = k\Big) \\
&\qquad + \frac{(\log\log n)^{40}}{\log n}\,.
\end{aligned} \tag{4.6}$$

Let $(\mathbb{K}_n^*, \mathbb{K}_n')$ be as in Lemma 4.3.5. By the definition of $\mathbb{K}_n^*$ and (4.6), we have that

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U}_r^- : \mathrm{wdiam}\left(\,\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon \;\Big|\; \mathrm{I} = k\Big) \\
&\quad \leq \mathbb{P}\Big(\exists U \in \mathbb{U}_r^-(k) : \mathrm{wdiam}\left(\,\mathrm{MST}(\mathbb{K}_n^*[U])\right) > \varepsilon, \mathrm{E}(T) \subset \mathrm{E}\big(\,\mathrm{MST}\left(\mathbb{K}_n^*[U_{L^{20}}]\right)\big)\Big) \\
&\qquad + \frac{(\log\log n)^{40}}{\log n}\,.
\end{aligned}$$

Note that if $\mathrm{E}(T) \subset \mathrm{E}(\mathrm{MST}(\mathbb{K}_n^*[U_{L^{20}}]))$, then for any $U \in \mathbb{U}_r^-(k)$, $\mathrm{E}(T) \subset \mathrm{E}(\mathrm{MST}(\mathbb{K}_n^*[U]))$, since $\mathrm{MST}(\mathbb{K}_n^*[U_{L^{20}}])[U]$ is a subgraph of $\mathrm{MST}(\mathbb{K}_n^*[U])$. Applying Proposition 4.3.4 to $\mathrm{MST}(\mathbb{K}_n^*[U])$ and $\mathrm{MST}(\mathbb{K}_n'[U])$, it follows that

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U}_r^-(k) : \mathrm{wdiam}\left(\,\mathrm{MST}(\mathbb{K}_n^*[U])\right) > \varepsilon, \mathrm{E}(T) \subset \mathrm{E}\big(\,\mathrm{MST}\left(\mathbb{K}_n^*[U_{L^{20}}]\right)\big)\Big) \\
&\quad \leq \mathbb{P}\Big(\exists U \in \mathbb{U}_r^-(k) : w^*(T) + 2 \times \mathrm{wdiam}\left(\,\mathrm{MST}(\mathbb{K}_n'[U])\right) > \varepsilon, \mathrm{E}(T) \subset \mathrm{E}\big(\,\mathrm{MST}\left(\mathbb{K}_n^*[U_{L^{20}}]\right)\big)\Big) \\
&\quad \leq \mathbb{P}\Big(\exists U \in \mathbb{U}_r^-(k) : w^*(T) + 2 \times \mathrm{wdiam}\left(\,\mathrm{MST}(\mathbb{K}_n'[U])\right) > \varepsilon\Big)\,.
\end{aligned}$$

Using that $w^*(T) \le WL$ and combining the two previous inequalities yields the bound

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U}_r^- : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon \;\Big|\; \mathrm{I} = k\Big) \qquad (4.7)\\
&\quad \le \mathbb{P}\Big(\exists U \in \mathbb{U}_r^-(k) : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n'[U])\right) > (\varepsilon - WL)/2\Big) + \frac{(\log\log n)^{40}}{\log n}\,.
\end{aligned}$$

We can now replace $\mathbb{K}_n'$ by $\mathbb{K}_n$ since they are identically distributed. Furthermore, recall that Theorem 4.2.3 states that, for n sufficiently large, we have

$$\mathbb{P}\left(\mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n)\right) \ge \frac{7\log^4 n}{n^{1/10}}\right) \le \frac{4}{n^{\log n}}\,.$$

Since $L \to \infty$ and $WL \to 0$ as $n \to \infty$, and since any set $U \in \mathbb{U}_r^-$ has size $|U| \ge |U_0| = L+1$, we can choose n large enough so that, for any set $U \in \mathbb{U}_r^-$, we have $7\log^4|U|/|U|^{1/10} \le (\varepsilon - WL)/2$. It follows that

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U}_r^-(k) : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n'[U])\right) > (\varepsilon - WL)/2\Big)\\
&\quad \le \mathbb{P}\left(\exists U \in \mathbb{U}_r^-(k) : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) \ge \frac{7\log^4|U|}{|U|^{1/10}}\right)\\
&\quad \le \sum_{U \in \mathbb{U}_r^-(k)} \frac{4}{|U|^{\log|U|}}
\end{aligned}$$

The final step of the proof is to use that $\mathbb{U}_r^-(k) = (U_i, 0 \le i \le L^{20})$ where $|U_i| = |U_0| + i = L + i + 1$, along with the fact that $4/n^{\log n} \le 1/n^2$ for n large enough, to obtain that

$$\mathbb{P}\Big(\exists U \in \mathbb{U}_r^-(k) : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n'[U])\right) > (\varepsilon - WL)/2\Big) \le \sum_{k=L+1}^{L+L^{20}+1} \frac{1}{k^2} \le \frac{1}{L} \le \frac{2}{\log\log n}\,.$$

Plugging this into (4.7), it follows that

$$\mathbb{P}\Big(\exists U \in \mathbb{U}_r^- : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon \;\Big|\; \mathrm{I} = k\Big) \le \frac{(\log\log n)^{40}}{\log n} + \frac{2}{\log\log n}\,.$$

Finally, since the previous inequality holds for any $k < n - L - L^{20}$, we have

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U}_r^- : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon\Big) \\
&\quad = \mathbb{P}\Big(\exists U \in \mathbb{U}_r^- : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon \;\Big|\; \mathrm{I} < n - L - L^{20}\Big) + o(1) \\
&\quad \le \frac{(\log\log n)^{40}}{\log n} + \frac{1}{(\log\log n)^{20}} + o(1) \longrightarrow 0\,,
\end{aligned}$$

which is the desired result. □

Second right set $\mathbb{U}_r^+$.

Lemma 4.3.7. *For any $\varepsilon > 0$, we have*

$$\mathbb{P}\Big(\exists U \in \mathbb{U}_r^+ : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon\Big) \longrightarrow 0\,.$$

Proof. Fix $0 < a < 1$ and assume n is large enough so that $n^a < n - L - L^{20}$. Then, by Lemma 4.3.3, we have

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U}_r^+ : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon\Big) \\
&\quad \le \mathbb{P}\Big(\exists U \in \mathbb{U}_r^+ : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon \;\Big|\; \mathrm{I} < n - L - L^{20}\Big) + \mathbb{P}\big(\mathrm{I} \ge n - L - L^{20}\big) \\
&\quad = \mathbb{P}\Big(\exists U \in \mathbb{U}_r^+ : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon \;\Big|\; \mathrm{I} < n - L - L^{20}\Big) + o(1)\,.
\end{aligned}$$

Fix now $k < n - L - L^{20}$ and condition on the event $\mathrm{I} = k$. Let $T = H_n[U_0]$ and let $(\mathbb{K}'_n, \mathbb{K}^*_n)$ be given by the coupling in Lemma 4.3.5. Then, by Proposition 4.3.4,

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U}_r^+ : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon \;\Big|\; \mathrm{I} = k\Big) \\
&\quad = \mathbb{P}\Big(\exists U \in \mathbb{U}_r^+(k) : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}^*_n[U])\right) > \varepsilon\Big) \\
&\quad \le \mathbb{P}\Big(\exists U \in \mathbb{U}_r^+(k) : w^*(T) + |\mathrm{V}(T)| \times \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}'_n[U])\right) > \varepsilon\Big) \\
&\quad \le \mathbb{P}\Big(\exists U \in \mathbb{U}_r^+(k) : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > (\varepsilon - WL)/(L+1)\Big)\,,
\end{aligned}$$

where the last step follows from the fact that $w^*(T) \le WL$ conditionally given that $\mathrm{I} < n - L$, that $|\mathrm{V}(T)| = L+1$, and that $\mathbb{K}'_n$ is distributed as $\mathbb{K}_n$. Since $x \mapsto \frac{\log^4 x}{x^{1/10}}$ is a decreasing function for large enough x, since any set $U \in \mathbb{U}_r^+$ has size $|U| \ge |U_{L^{20}}| = L + L^{20} + 1$, and since $L = \lfloor \log\log n \rfloor \to \infty$ and $WL = \lfloor \log\log n \rfloor / \log n \to 0$, we can choose n large enough so that, for any $U \in \mathbb{U}_r^+$

$$\frac{7\log^4|U|}{|U|^{1/10}} \le \frac{7\log^4(L^{20})}{(L^{20})^{1/10}} = \frac{7\cdot 20^4\log^4\cdot(L)}{L^2} \le \frac{\varepsilon - WL}{L+1}\,.$$

Then, recalling that $\mathbb{U}_r^+(k) = (U_i, L^{20} < i \le n-k-L)$ where $|U_i| = |U_0| + i = L + i + 1$, Theorem 4.2.3 gives us

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U}_r^+(k) : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > (\varepsilon - WL)/(L+1)\Big)\\
&\quad\le \mathbb{P}\left(\exists U \in \mathbb{U}_r^+(k) : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \frac{7\log^4|U|}{|U|^{1/10}}\right)\\
&\quad\le \sum_{U\in\mathbb{U}_r^+(k)} \frac{4}{|U|^{\log|U|}}\\
&\quad\le \frac{1}{L+L^{20}}\,,
\end{aligned}$$

where the last inequality uses that $x^{\log x} \ge 4x^2$ for x large enough, along with the fact that $|U_i| = L + i + 1$. Therefore,

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U}_r^+ : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon\Big)\\
&\quad\le \mathbb{P}\Big(\exists U \in \mathbb{U}_r^+(k) : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > (\varepsilon - WL)/(L+1)\Big) + o(1)\\
&\quad\le \frac{1}{L+L^{20}} + o(1) \longrightarrow 0\,,
\end{aligned}$$

concluding the proof of the lemma. □

Left set $\mathbb{U}_\ell$.

Lemma 4.3.8. *For any $\varepsilon > 0$, we have*

$$\mathbb{P}\Big(\exists U \in \mathbb{U}_\ell : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon\Big) \longrightarrow 0\,.$$

Proof. Fix $a < \frac{1}{4}$. Thanks to Lemma 4.3.3, we know that $\mathbb{P}(\mathrm{I} \ge n^a) \to 0$. Moreover, note that under this event, any set $U \in \mathbb{U}_\ell$ has size $|U| \ge n - k \ge n - n^a$. Our strategy now is to prove that, due to the large size of these sets, conditioning on the event $\{\mathrm{I} < n^a\}$ does not notably affect the structure of $\mathrm{MST}(\mathbb{K}_n[U])$.

Let us try to understand how the edge weights $\{e_1, \ldots, e_{n-1}\}$ behave given that $\mathrm{I} < n^a$;

call $\mathbb{K}_n^a$ the random weighted graph corresponding to the distribution of $\mathbb{K}_n$ conditionally given that $\mathrm{I} < n^a$. Recall that $\{\mathrm{I} < n^a\} \in \sigma(\{X_{e_i} : 1 \leq i < \lceil n^a \rceil + L\})$ and write $m = \lceil n^a \rceil + L - 1$ (note that $e_1, \ldots, e_m$ are the only edges affected when we condition on $\mathrm{I} < n^a$). Let $\mathbf{A} = \{i \leq m : X_{e_i} \leq W\}$ and let $\mathcal{A}$ be the collection of sets $A \subset [m]$ such that there exists $i < n^a$ with $\{i, \ldots, i + L - 1\} \subset A$. Then, by definition, $\{\mathbf{A} \in \mathcal{A}\} = \{\mathrm{I} < n^a\}$. Now, for any $A \in \mathcal{A}$, conditionally given that $\mathbf{A} = A$, the weights of $e_1, \ldots, e_m$ are independent of each other and are distributed as Uniform$[0, W]$ or Uniform$[W, 1]$, according to whether or not the index i of the edge e_i lies in A. This means that for any $x_1, \ldots, x_m \in [0, 1]$, and any $A \in \mathcal{A}$, we have

$$\begin{aligned}
&\mathbb{P}\Big(\forall i \in [m] : X_{e_i} \leq x_i \;\Big|\; \mathbf{A} = A, \mathrm{I} < n^a\Big) \\
&\quad = \mathbb{P}\Big(\forall i \in [m] : X_{e_i} \leq x_i \;\Big|\; \mathbf{A} = A\Big) \\
&\quad = \left(\prod_{i \in A} \frac{\min\{x_i, W\}}{W}\right) \left(\prod_{i \in [m] \setminus A} \frac{\max\{x_i, W\} - W}{1 - W}\right).
\end{aligned}$$

Now, using that $\frac{\max\{x_i, W\} - W}{1 - W} \leq \frac{\min\{x_i, W\}}{W}$, it follows that

$$\mathbb{P}\Big(\forall i \in [m] : X_{e_i} \leq x_i \;\Big|\; \mathbf{A} = A, \mathrm{I} < n^a\Big) \leq \mathbb{P}\Big(\forall i \in [m] : X'_{e_i} \leq x_i\Big),$$

where $(X'_{e_1}, \ldots, X'_{e_m})$ are independent Uniform$[0, W]$. This implies that there exists a generic weighted graph $\mathbb{K}'_n = (K_n, \mathbb{X}')$ with independent weights, where X'_e is a Uniform$[0, 1]$ if $e \notin \{e_1, \ldots, e_m\}$ and a Uniform$[0, W]$ otherwise, and a coupling between $\mathbb{K}'_n$ and $\mathbb{K}_n^a$ such that $X'_e \leq X_e^a$ for any $e \in \mathrm{E}(K_n)$. We now use this coupling to prove the lemma.

Consider the event

$$E' = \Big\{\forall k < n^a, \forall i \in [m], e_i \notin \mathrm{E}\big(\,\mathrm{MST}(\mathbb{K}'_n[V(k, n)])\big)\Big\}$$

By using two union bounds, we have that

$$\mathbb{P}(E') \geq 1 - \sum_{k < n^a} \sum_{i \in [m]} \mathbb{P}\Big(e_i \in \mathrm{E}\big(\,\mathrm{MST}(\mathbb{K}'_n[V(k, n)])\big)\Big).$$

For k and i as in the above sum, if there exists $j \in V(k, n) \setminus e_i$ such that the weight of e_i is larger than the weight of the two other edges in the triangle $\Delta_{i,j}$ formed by e_i and j, then

e_i is not in the MST of $\mathbb{K}'_n[V(k,n)]$. This means that

$$\begin{aligned}
&\mathbb{P}\Big(e_i \in \mathrm{E}\big(\,\mathrm{MST}(\mathbb{K}'_n[V(k,n)])\big) \;\Big|\; X'_{e_i}\Big) \\
&\quad \le \mathbb{P}\Big(\forall j \in V(k,n) \setminus e_i, \max(X'_e : e \in \Delta_{i,j}) > X'_{e_i} \;\Big|\; X'_{e_i}\Big) \\
&\quad = \big(1 - (X'_{e_i})^2\big)^{|V(k,n)|-2}
\end{aligned}$$

Using that X'_{e_i} is uniformly distributed over $[0,W]$ and that $|V(k,n)| = n-k+1$, it follows that

$$\begin{aligned}
\mathbb{P}\Big(e_i \in \mathrm{E}\big(\,\mathrm{MST}(\mathbb{K}'_n[V(k,n)])\big)\Big) &\le \frac{1}{W}\int_0^W (1-x^2)^{n-k-1}dx \\
&\le \frac{1}{W}\int_0^\infty e^{-(n-k-1)x^2}dx \\
&= \frac{\sqrt{\pi}}{2W\sqrt{n-k-1}}\,,
\end{aligned}$$

from which we obtain

$$\mathbb{P}(E') \ge 1 - \sum_{k<n^a}\sum_{i\in[m]} \frac{\sqrt{\pi}}{2W\sqrt{n-k-1}} \ge 1 - \frac{\sqrt{\pi}}{2}\frac{n^a m}{W\sqrt{n-n^a-1}} \longrightarrow 1\,,$$

where the last convergence follows from $W = \frac{1}{\log n}$, $m = \lceil n^a \rceil + L - 1 = \lceil n^a \rceil + \lfloor \log\log n \rfloor - 1$, and $a < \frac{1}{4}$.

Combining the fact that $\mathbb{P}(\mathrm{I} < n^a) \to 1$ with the definitions of $\mathbb{K}_n^a$ and $\mathbb{U}_\ell$, we now have that

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U}_\ell : \mathrm{wdiam}\,\big(\,\mathrm{MST}(\mathbb{K}_n[U])\big) > \varepsilon\Big) \\
&\quad = \mathbb{P}\Big(\exists U \in \mathbb{U}_\ell : \mathrm{wdiam}\,\big(\,\mathrm{MST}(\mathbb{K}_n[U])\big) > \varepsilon \;\Big|\; \mathrm{I} < n^a\Big) + o(1) \\
&\quad \le \mathbb{P}\Big(\exists k < n^a : \mathrm{wdiam}\,\big(\,\mathrm{MST}(\mathbb{K}_n^a[V(k,n)])\big) > \varepsilon\Big) + o(1)\,,
\end{aligned} \tag{4.8}$$

where the last inequality comes from the definition of $\mathbb{K}_n^a$, and is due to $\mathbb{U}_\ell = (V(\mathrm{I}-1,n),\ldots,V(1,n)) \subset (V(n^a-1,n),\ldots,V(1,n)))$ whenever $\mathrm{I} < n^a$. Note that the coupling between $\mathbb{K}_n^a$ and $\mathbb{K}'_n$ only reduces the weight of the edges $e_1,\ldots,e_m$ in $\mathbb{K}'_n$ relative to $\mathbb{K}_n^a$, from which it follows that, if $e_i \notin \mathrm{E}(\mathrm{MST}(\mathbb{K}'_n[V(k,n)]))$ for some $i \in [m]$, then $e_i \notin \mathrm{E}(\mathrm{MST}(\mathbb{K}_n^a[V(k,n)]))$. This implies that, conditionally given E', the trees $\mathrm{MST}(\mathbb{K}_n^a[V(k,n)])$

and $\mathrm{MST}(\mathbb{K}_n'[V(k,n)])$ are equal. Using that $\mathbb{P}(E') \to 1$, we thus obtain

$$\begin{aligned}
&\mathbb{P}\Big(\exists k < n^a : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n^a[V(k,n)])\right) > \varepsilon\Big) \qquad (4.9)\\
&\quad = \mathbb{P}\Big(\exists k < n^a : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n^a[V(k,n)])\right) > \varepsilon \;\Big|\; E'\Big) + o(1)\\
&\quad = \mathbb{P}\Big(\exists k < n^a : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n'[V(k,n)])\right) > \varepsilon \;\Big|\; E'\Big) + o(1)\,.
\end{aligned}$$

Finally, consider a coupling between $\mathbb{K}_n'$ and $\mathbb{K}_n$ where $X_e' \leq X_e$ for any $e \in \mathrm{E}(K_n)$ and such that $X_e' = X_e$ whenever $e \notin \{e_1, \ldots, e_m\}$. By using that $\mathrm{MST}(\mathbb{K}_n') = \mathrm{MST}(\mathbb{K}_n)$ whenever E' holds, it follows that

$$\begin{aligned}
&\mathbb{P}\Big(\exists k < n^a : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n'[V(k,n)])\right) > \varepsilon \;\Big|\; E'\Big) \qquad (4.10)\\
&\quad = \mathbb{P}\Big(\exists k < n^a : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[V(k,n)])\right) > \varepsilon \;\Big|\; E'\Big)\\
&\quad = \mathbb{P}\Big(\exists k < n^a : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[V(k,n)])\right) > \varepsilon\Big) + o(1)\,,
\end{aligned}$$

where we used that $\mathbb{P}(E') \to 1$ for the last equality. Now, using Theorem 4.2.3 similarly as before, we obtain that

$$\mathbb{P}\Big(\exists k < n^a : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[V(k,n)])\right) > \varepsilon\Big) \longrightarrow 0\,.$$

The proof of this lemma now follows by combining (4.8), (4.9), and (4.10). □

With the above lemmas in hand, the proof of Proposition 4.3.2 is routine.

Proof of Proposition 4.3.2. Fix $\varepsilon > 0$ and let $W = \frac{1}{\log n}$ and $L = \lfloor \log\log n \rfloor$. Then

$$\begin{aligned}
&\mathbb{P}\Big(\exists U \in \mathbb{U} : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon\Big)\\
&\quad = \mathbb{P}\Big(\exists U \in \mathbb{U}_r^- \cup \mathbb{U}_r^+ \cup \mathbb{U}_\ell : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon\Big)\\
&\quad \leq \mathbb{P}\Big(\exists U \in \mathbb{U}_r^- : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon\Big)\\
&\qquad + \mathbb{P}\Big(\exists U \in \mathbb{U}_r^+ : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon\Big)\\
&\qquad + \mathbb{P}\Big(\exists U \in \mathbb{U}_\ell : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n[U])\right) > \varepsilon\Big)\,,
\end{aligned}$$

and the right hand side converges to 0 by Lemma 4.3.6, 4.3.7, and 4.3.8, proving the proposition. □

4.4 Conclusion

4.4.1 More general weight distributions

The extension of Theorem 4.1.1 from Uniform$[0, 1]$ to more general weight distributions is quite straightforward. Fix a probability density function $f : [0, \infty) \to [0, \infty)$, and let $\rho^* = \sup(x : \int_0^x f(y)dy < 1)$. Let $\mathbb{X}' = (X'_e, e \in \mathrm{E}(K_n))$ be independent random variables with density f, and let $\mathbb{K}'_n = (K_n, \mathbb{X}')$.

Theorem 4.4.1. *Suppose that $f(0) > 0$, that f is continuous at zero, and that $\rho^* < \infty$. Fix any sequence $(H_n, n \geq 1)$ of connected graphs with H_n being a spanning subgraph of K_n. Then for any $\varepsilon > 0$, as $n \to \infty$,*

(a) with high probability there exists an MST *sequence $\mathbb{S}$ for $(\mathbb{K}'_n, H_n)$ with $\mathrm{wt}(\mathbb{S}) \leq \rho^* + \varepsilon$, and*

(b) there exists $\delta > 0$ such that with high probability, given any optimizing sequence $\mathbb{S} = (S_1, \ldots, S_m)$ for $(\mathbb{K}'_n, H_n)$ with $\mathrm{wt}(\mathbb{S}) \leq \rho^* - \varepsilon$, the final spanning subgraph $H_{n,m}$ has weight $w(H_{n,m}) \geq \delta n \, w(\mathrm{MST}(\mathbb{K}_n))$.

In particular, $\mathrm{cost}(\mathbb{K}'_n, H_n) \xrightarrow{\text{prob}} \rho^*$ *as $n \to \infty$.*

The proof is very similar to that of Theorem 4.1.1, so we only describe the changes that are required to prove the more general version.

The proof of the lower bound, part (b), proceeds just as in the case of Uniform$[0, 1]$ edge weights: for any $\varepsilon > 0$, any optimizing sequence $\mathbb{S} = (S_0, \ldots, S_m)$ for $(\mathbb{K}'_n, H_n)$ with $\mathrm{wt}(\mathbb{S}) \leq \rho^* - \varepsilon$ leaves edges of weight greater than $\rho^* - \varepsilon$ untouched, so all such edges appear in the final subgraph $H_{n,m}$. The number of such edges is $\mathrm{Binomial}\big(|\mathrm{E}(H_n)|, \int_{\rho^*-\varepsilon}^{\rho^*} f(x)dx\big)$-distributed, so with high probability there are a linear number of such edges. On the other hand, $w(\mathrm{MST}\, \mathbb{K}_n) \to \zeta(3)/f(0)$ in probability [40], and the lower bound follows.

For the upper bound, note that the bounds on the total cost of the optimizing sequences we construct essentially all have the form $A + B$ where A is the greatest weight of a single edge, and B is the weighted diameter of the minimum spanning tree of some subgraph of K_n. In order to prove Theorem 4.1.1, we used that $A \leq 1$, and proved using Theorem 4.2.3 and Proposition 4.3.2 that we could take B as close to zero as we wished (by a careful choice of optimizing sequence). For the edge weights $\mathbb{X}'$, we can simply replace the bound $A \leq 1$ by the bound $A \leq \rho^*$. To show that we can make B as close to zero as we like, we can carry through the same proof as in the Uniform$[0,1]$ case, provided that versions of Theorem 4.2.3 and Proposition 4.3.2 are still available to us.

To see that Theorem 4.2.3 and Proposition 4.3.2 do essentially carry over to the setting of $\mathbb{K}'_n = (K_n, \mathbb{X}')$, we make use of the following coupling. For $t \in [0, \rho^*]$ let $g(t) = \mathbb{P}(X' \leq t)$, so that $g(X')$ is Uniform$[0,1]$-distributed. We can thus couple the random weights $\mathbb{X}'$ to independent Uniform$[0,1]$ weights $\mathbb{X} = (X_e, e \in \mathrm{E}(K_n))$ by taking $X_e = g(X'_e)$, and thereby couple $\mathbb{K}'_n = (K_n, \mathbb{X}')$ to $\mathbb{K}_n = (K_n, \mathbb{X})$. The edge weights $\mathbb{X}' = (X'_e, e \in \mathrm{E}(K_n))$ are almost surely pairwise distinct, and on this event, the ordering of $\mathrm{E}(K_n)$ in increasing order of weight is the same for the weights $\mathbb{X}$ and $\mathbb{X}'$ and thus $\mathrm{MST}(\mathbb{K}'_n) = \mathrm{MST}(\mathbb{K}_n)$.

Since $f(0) > 0$ and f is continuous, for all u sufficiently small we have $f(u) > f(0)/2$ and $g(u) \geq uf(0)/2$. It follows in particular that if $X_e \leq uf(0)/2$ then $X'_e \leq 2X_e/f(0) \leq u$. This observation implies that, under the above coupling between $\mathbb{K}_n$ and $\mathbb{K}'_n$, if $\mathrm{wdiam}(\mathrm{MST}(\mathbb{K}_n)) \leq uf(0)/2$ then $\mathrm{wdiam}(\mathrm{MST}(\mathbb{K}'_n)) \leq u$, and Theorem 4.2.3 thus yields that for all n sufficiently large,

$$\mathbb{P}\left(\mathrm{wdiam}(\mathrm{MST}(\mathbb{K}'_n)) \geq \frac{2}{f(0)} \frac{7\log^4 n}{n^{1/10}}\right) \leq \mathbb{P}\left(\mathrm{wdiam}(\mathrm{MST}(\mathbb{K}'_n)) \geq \frac{7\log^4 n}{n^{1/10}}\right) \leq \frac{4}{n^{\log n}}\,. \tag{4.11}$$

Similarly, Proposition 4.3.2 implies that (in the notation of that proposition), for all $\varepsilon > 0$

$$\mathbb{P}\Big(\exists U \in \mathbb{U} : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}'_n[U])\right) > 2\varepsilon/f(0)\Big) \longrightarrow 0$$

as $n \to \infty$. But since $\varepsilon > 0$ was arbitrary, this implies that also

$$\mathbb{P}\Big(\exists U \in \mathbb{U} : \mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}'_n[U])\right) > \varepsilon\Big) \longrightarrow 0 \tag{4.12}$$

for all $\varepsilon > 0$.

All the remaining ingredients of the proof of Theorem 4.1.1 use only information about the graph-theoretic structure of $\mathrm{MST}(\mathbb{K}_n)$, not its weights, and so carry over to the setting of non-uniform weights (using the fact that $\mathrm{MST}(\mathbb{K}'_n)$ and $\mathrm{MST}(\mathbb{K}_n)$ have the same distributions as unweighted graphs – indeed, they are equal under the above coupling). By running the proof of Theorem 4.1.1 but replacing all expressions of the form $1 + \mathrm{wdiam}(F)$ by $\rho^* + \mathrm{wdiam}(F)$, and when needed invoking (4.11) and (4.12) in place of Theorem 4.2.3 and Proposition 4.3.2, respectively, we obtain Theorem 4.4.1.

Before concluding this subsection, we note that if $\rho^* = \infty$ then for any $r > 0$, the probability that at least one edge of H_n has weight at least r tends to 1, so $\mathbb{P}(\mathrm{cost}(\mathbb{K}'_n, H_n) >$

$r) \to 1$ as $n \to \infty$. Thus, in this case we also have $\mathrm{cost}(\mathbb{K}'_n, H_n) \xrightarrow{\text{prob}} \rho^*$.

4.4.2 Open questions and future directions

This work introduces the notion of local minimum spanning tree searches and proves a weak law of large numbers for the cost of such local searches. The framework naturally suggests several directions for future research, some of which we now highlight.

- Our main results concern low-weight MST sequences $\mathbb{S}$ for randomly weighted complete graphs, where $\mathrm{wt}(\mathbb{S})$ is measured in the L_∞ sense: it is the maximum weight of any single step of the optimizing sequence. However, one may wish to vary the norm used to measure the weights of optimizing sequences. The other L_p norms are natural alternatives, and correspond to studying the values

  $$\mathrm{cost}_{\,p}(\mathbb{G}, H) = \min\Big\{ \mathrm{wt}_{\,p}(\mathbb{S}) : \mathbb{S} \text{ is an MST sequence for } (\mathbb{G}, H)\Big\}$$

  where

  $$\mathrm{wt}_{\,p}(\mathbb{S}) = \left(\sum_{i=1}^{m} \big(\mathrm{wt}(\mathbb{S}, i)\big)^p\right)^{\frac{1}{p}}$$

  is the L^p norm of $(\mathrm{wt}(\mathbb{S}, i), 1 \le i \le m)$.

  At first sight, using L^1 weights may seem very natural, as it corresponds to the total weight of all the subgraphs modified by the sequence. Mathematically, however, in the setting considered in this paper the L^1 cost is quite easy to understand. Indeed, for $\mathbb{K}_n$ and H_n as in Theorem 4.1.1, by considering the sequence $\mathbb{S} = ([n])$ which simply replaces H_n by $\mathrm{MST}(\mathbb{K}_n)$ in one step, we obtain that

  $$\mathrm{cost}_{\,1}(\mathbb{K}_n, H_n) \le w(H_n)\,.$$

  Conversely, since any edge of $e \in \mathrm{E}(H_n) \setminus \mathrm{E}(\mathrm{MST}(\mathbb{K}_n))$ must be removed in order to form the MST, for any MST sequence $\mathbb{S} = (S_1, \dots, S_m)$, there must exist $i \in [m]$ such that $e \in \mathrm{E}(H_{n,i-1}[S_i])$. This implies that

  $$\mathrm{wt}_{\,1}(\mathbb{S}) \ge \sum_{i \in [m]} w\big(H_{n,i-1}[S_i]\big) \ge \sum_{e \in \mathrm{E}(H_n) \setminus \mathrm{E}(\mathrm{MST}(\mathbb{K}_n))} X_e = \big(1 + o_{\mathbb{P}}(1)\big) w(H_n)\,,$$

where the final asymptotic follows from the fact that H_n is chosen independently of $\mathbb{X}$ and that any fixed edge belongs to the MST with probability $(n-1)/\binom{n}{2} = o_{\mathbb{P}}(1)$. Since the lower bound $\sum_{e \in \mathrm{E}(H_n) \setminus \mathrm{E}(\mathrm{MST}(\mathbb{K}_n))} X_e$ does not depend on the choice of MST sequence $\mathbb{S}$, it is also a lower bound on $\mathrm{cost}_1(\mathbb{K}_n, H_n)$, and thus

$$\frac{\mathrm{cost}_1(\mathbb{K}_n, H_n)}{w(H_n)} \longrightarrow 1$$

in probability. When $p < 1$, this argument can be adapted to prove the same convergence result for $\mathrm{cost}_p(\mathbb{K}_n, H_n)$. However, when $p > 1$ it is less clear what behaviour to expect, and in particular it is unclear whether the dependence on the initial spanning subgraph H_n will play a more complicated role.

- Another natural modification of the setting is to measure the cost of a step by the *size*, rather than the weight, of the subgraph which is replaced by its MST. That is, we may define

$$\mathrm{wt}'(\mathbb{G}, H, \mathbb{S}) := \max\left\{ \left|\mathrm{E}(H_{i-1}[S_i])\right| : 1 \le i \le m \right\},$$

and study

$$\mathrm{cost}'(\mathbb{G}, H) = \min\left\{ \mathrm{wt}'(\mathbb{G}, H, \mathbb{S}) : \mathbb{S} \text{ is an MST sequence for } (\mathbb{G}, H) \right\}.$$

For this notion of cost, even the behaviour of $\mathrm{cost}'(\mathbb{K}_n, K_n)$ is unclear to us; how $\mathrm{cost}'(\mathbb{K}_n, H_n)$ will depend on the starting graph H_n is likewise unclear. However, at a minimum we expect that $\mathrm{cost}'(\mathbb{K}_n, H_n) \to \infty$ in probability, provided that the initial spanning subgraphs H_n are chosen independently of the weights.

- Our result proves the *existence* of MST sequences of weight at most $(\rho^* + \varepsilon)$ (where ρ^* is as in Theorem 4.4.1), with high probability. However, our construction does not yield insight into the ubiquity of such sequences, and it would be interesting to know whether low-weight MST sequences can be found easily and without using "non-local" information. For example, suppose that at each step we choose a subgraph to optimize uniformly at random over all subgraphs of weight at most w. For which values of w will the resulting sequence be an MST sequence with high probability?

- What is the asymptotic behaviour of $\mathrm{cost}(\mathbb{K}_n, H_n) - \rho^*$? In particular, is there a sequence a_n such that $a_n(\mathrm{cost}(\mathbb{K}_n, H_n) - \rho^*)$ converges in distribution to a non-trivial random variable?

- What happens if $(K_n, \mathbb{X}_n)$ is replaced by a different fixed connected, weighted graph $\mathbb{G}_n = (G_n, \mathbb{X}_n)$? How does the asymptotic behaviour of $\text{cost}(\mathbb{G}_n, H_n)$ depend on G_n?

- What happens if the iid structure of the edge weights of $\mathbb{K}_n$ is modified? For example, one might generate $\mathbb{X}_n$ by first taking n independent, uniformly random points $P_1, \ldots, P_n \in [0,1]^d$, then letting $X_{ij} = |P_i - P_j|$ be the Euclidean distance between i and j.

4.A Bounds on the weighted diameter

In this section, we prove Theorem 4.2.3. The proof exploits Kruskal's algorithm for constructing minimum spanning trees. We first recall a very useful connection between Kruskal's algorithm run on the complete graph with independent Uniform$[0,1]$ edge weights $\mathbb{X} = (X_e, e \in E(K_n))$ and the Erdős-Rényi random graph process. In this setting, Kruskal's algorithm may be phrased as follows. Write $N = \binom{n}{2}$.

- Order the edges of $\mathrm{E}(\mathbb{K}_n)$ in increasing order of weight as $e_1, \ldots, e_N$.

- Let $F_0 = ([n], \emptyset)$ be the forest with vertex set n and no edges.

- For $1 \leq i \leq N$, if e_i joins distinct connected components of F_{i-1} then let $\mathrm{E}(F_i) = \mathrm{E}(F_{i-1}) \cup \{e_i\}$; otherwise let $F_i = F_{i-1}$.

The final forest F_N is $\mathrm{MST}(\mathbb{K}_n)$.

The Erdős-Rényi random graph process can be described very similarly:

- Order the edges of $\mathrm{E}(\mathbb{K}_n)$ in increasing order of weight as $e_1, \ldots, e_N$.

- Let $G_0 = ([n], \emptyset)$ be the graph with vertex set n and no edges.

- For $1 \leq i \leq N$, let $\mathrm{E}(G_i) = \mathrm{E}(G_{i-1}) \cup \{e_i\}$.

It is straightforward to see by induction that F_i and G_i always have the same connected components and, more strongly, that F_i is the minimum spanning forest of G_i (in that each tree of F_i is the minimum spanning tree of the corresponding connected component of G_i).

We also take $G(n,p)$ to be the subgraph of $\mathbb{K}_n$ with edge set $\{e \in \mathrm{E}(\mathbb{K}_n) : X_e \leq p\}$. Since we ordered the edges in increasing order of weight as $e_1, \ldots, e_N$, the edge set of $G(n,p)$ is thus $\{e_1, \ldots, e_m\}$, where $m = m(p)$ is maximal so that $X_{e_m} \le p$. We likewise let $F(n,p)$ be the subgraph of F_N consisting of all edges of F_N with weight at most p, and note that $F(n,p) = F_{m(p)}$.

With this coupling in hand, we next explain our approach to bounding the weighted diameter of $\mathrm{MST}(\mathbb{K}_n)$. Our bound has two parts. Fix $p \in (0,1)$, and let $T^{\max}_{n,p}$ be the largest connected component of $F(n,p)$, with ties broken lexicographically. Note that $T^{\max}_{n,p}$ is a subgraph of $\mathrm{MST}(\mathbb{K}_n)$. Further write $L_{n,p}$ for the greatest number of edges in any path of $\mathrm{MST}(\mathbb{K}_n)$ which has exactly one vertex lying in $T^{\max}_{n,p}$. Finally, write W_n for the greatest weight of any edge of $\mathrm{MST}(\mathbb{K}_n)$.

Proposition 4.A.1. *For any $p \in (0,1)$,*

$$\mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n)\right) \le p\big(|T^{\max}_{n,p}| - 1\big) + 2W_n L_{n,p}\,.$$

Proof. Fix any path P in $\mathrm{MST}(\mathbb{K}_n)$. Then the set of vertices of P contained in $T^{\max}_{n,p}$ form a subpath of P, since otherwise $\mathrm{MST}(\mathbb{K}_n)$ would contain a cycle; call this subpath P_0. Then P_0 contains at most $|T^{\max}_{n,p}|$ vertices, so at most $|T^{\max}_{n,p}| - 1$ edges, and each such edge has weight at most p. Moreover, the edges of P not lying in P_0 form at most two subpaths of P. Each of these subpaths has at most $L_{n,p}$ edges, so the number of edges of P which are not edges of P_0 is at most $2L_{n,p}$; and the edges of P which are not edges of P_0 all have weight at most W_n. $\square$

To exploit this bound and prove Theorem 4.2.3, we must bound $|T^{\max}_{n,p}|$ and $L_{n,p}$, for some well chosen value of p, and bound W_n. The latter bound is the easiest, and we take care of it first. We will need the following bound on the probability of connectedness of $G(n,p)$. We believe we have seen this bound in the literature, but were unable to find a reference, so we have included its short proof.

Lemma 4.A.2. *Let $G \sim G(n,p)$. Then*

$$\mathbb{P}\left(G \text{ is not connected}\right) \le e^{ne^{-\frac{np}{2}}} - 1$$

Proof. Let S be a subset of $[n]$ such that $S \neq \emptyset$ and $S \neq [n]$. Then

$$\mathbb{P}\Big(S \text{ is not connected to } S^c \text{ in } G\Big) = (1-p)^{|S|(n-|S|)}\,.$$

This implies that

$$\begin{aligned}\mathbb{P}\big(G \text{ is not connected}\big) &= \mathbb{P}\big(\exists S \subseteq [n] : 1 \le |S| \le n/2 \text{ and } S \text{ is not connected to } S^c \text{ in } G\big) \\ &\le \sum_{S\subseteq[n]:1\le|S|\le n/2} \mathbb{P}\Big(S \text{ is not connected to } S^c \text{ in } G\Big)\,.\end{aligned}$$

Combined with the previous result, this leads to

$$\mathbb{P}\big(G \text{ is not connected}\big) = \sum_{S\subseteq[n]:1\le|S|\le n/2} (1-p)^{|S|(n-|S|)} \le \sum_{1\le k\le n/2} \binom{n}{k}(1-p)^{k(n-k)}\,.$$

Use now that $(n-k) \ge n/2$ along with the fact that $1-p \ge 0$ to obtain that

$$\mathbb{P}\big(G \text{ is not connected}\big) \le \sum_{1\le k\le n} \binom{n}{k}(1-p)^{kn/2} = \big(1+(1-p)^{n/2}\big)^n - 1\,.$$

Finally, by using twice the convexity of exponential, we have

$$\mathbb{P}\big(G \text{ is not connected}\big) \le \Big(1+e^{-\frac{pn}{2}}\Big)^n - 1 \le e^{ne^{-\frac{pn}{2}}} - 1\,,$$

which is the desired result. $\square$

Fact 4.A.3. *For all n sufficiently large, it holds that $\mathbb{P}(W_n > 3\log^2 n/n) \le 1/n^{\log n}$.*

Proof. Under the above coupling, $F(n,p)$ and $G(n,p)$ have the same connected components, so

$$\begin{aligned}\mathbb{P}\big(W_n > 3\log^2 n/n\big) &= \mathbb{P}\Big(F(n, 3\log^2 n/n) \text{ is not connected}\Big) \\ &= \mathbb{P}\Big(G(n, 3\log^2 n/n) \text{ is not connected}\Big).\end{aligned}$$

Use now the bound from Lemma 4.A.2 to obtain that

$$\mathbb{P}\Big(G(n, 3\log^2 n/n) \text{ is not connected}\Big) \le \exp(ne^{-(3/2)\log^2 n}) - 1 = e^{n/(n^{\log n})^{3/2}} - 1 \le 1/n^{\log n}$$

the final bound holding for all n sufficiently large. $\square$

Proof of Theorem 4.2.3. We prove the theorem by bounding $|T^{\max}_{n,p}|$ and $L_{n,p}$, for a carefully chosen value of p (spoiler: we will take $p = 1/n + 1/n^{11/10}$), then applying Proposition 4.A.1.

Our arguments lean heavily on results from [12], and we next introduce those results (and the terminology necessary to do so).

For $c > 0$, let $\alpha(c)$ be the largest real solution of $e^{-cx} = 1 - x$ (the quantity $\alpha(c)$ is the survival probability of a Poisson(c) branching process). The key to the proof is the fact that the size of the largest component of $G(n,p)$ is with high probability close to $n\,\alpha(np)$ when $p = (1+o(1))/n$. We now provide a precise and quantitative version of this statement, with error bounds.

By [3, Exercise 21 (d)], for $\varepsilon \geq 0$ we have

$$2\varepsilon(1 - o(1)) \leq \alpha(1+\varepsilon) \leq 2\varepsilon\,,$$

the first inequality holding as $\varepsilon \to 0$. In particular,

$$(3/2)\varepsilon \leq \alpha(1+\varepsilon) \leq 2\varepsilon \tag{4.13}$$

for all $\varepsilon \geq 0$ sufficiently small.

For the remainder of the proof, fix $p = 1/n + 1/n^{11/10}$ and write $s^+ = n\,\alpha(n\log(1/(1-p)) + n^{3/4}$ and $s^- = n\,\alpha(n\log(1/(1-p)) - 2n^{3/4}$. (Aside: for the careful reader who is verifying the connections to the results from [3], note that $s^+ = t^+$ but $s^- \neq t^-$, where t^+, t^- are defined in [3, Proof of Theorem 4.4, Case 2]). By [3, Exercise 23 (a)], for all n sufficiently large we have

$$n\,\alpha(np) \leq n\,\alpha(n\log(1/(1-p)) \leq n\,\alpha(np) + \frac{2n^{1/2}}{1-p},$$

and using the above bounds on α, this yields

$$n^{9/10} \leq s^- \leq s^+ \leq 3n^{9/10}\,,$$

for n sufficiently large.

Let $\mathcal{C}^{\max}$ be the largest connected component of $G(n,p)$, and let $\mathcal{C}^{\text{runnerup}}$ be its second largest component. Using the previous inequality on s^- and s^+, by [3, (4.7)] we have

$$\mathbb{P}\big(|\mathcal{C}^{\max}| \geq 3n^{9/10}\big) \leq \mathbb{P}\big(|\mathcal{C}^{\max}| \geq s^+\big) \leq ne^{-(25/2)n^{1/10}}\,; \tag{4.14}$$

moreover, by [3, (4.10)], we have

$$\mathbb{P}\big(|\mathcal{C}^{\max}| \leq n^{9/10}\big) \leq \mathbb{P}\big(|\mathcal{C}^{\max}| \leq s^-\big) \leq 2ne^{-(25/2)n^{1/10}}\,; \tag{4.15}$$

finally, by [3, (4.10) and(4.11)], we have

$$\mathbb{P}\big(|\mathcal{C}^{\mathrm{runnerup}}| \geq n^{4/5}\big) \leq 5ne^{-(25/2)n^{1/10}}\,. \tag{4.16}$$

Furthermore, under the coupling between $G(n,p)$ and $F(n,p)$, we have $|\mathcal{C}^{\max}| = |T^{\max}_{n,p}|$, so (4.14) immediately gives us that for all n sufficiently large,

$$\mathbb{P}\big(|T^{\max}_{n,p}| \geq 3n^{9/10}\big) \leq ne^{-(25/2)n^{1/10}}\,. \tag{4.17}$$

It remains to bound $L_{n,p}$. For this, we use (4.15) and a Prim's-algorithm-type construction to control the greatest number of connected components of $G(n,p)$ that any path of $\mathrm{MST}(\mathbb{K}_n)$ lying outside $T^{\max}_{n,p}$ passes through, and use (4.16) to bound the size of those components.

Condition on the graph $G(n,p)$, and fix a connected component $\mathcal{C}_1$ of $G(n,p)$ different from $\mathcal{C}^{\max}$. Let $f_1 = u_1v_1$ be the smallest-weight edge with exactly one endpoint in $\mathcal{C}_1$, and let p_1 be its weight. Then $p_1 > p$, and f_1 is a cut-edge of $G(n,p_1)$. It follows that f_1 is an edge of $\mathrm{MST}(\mathbb{K}_n)$. Moreover, by the exchangeability of the edge weights, the endpoint v_1 of f_1 not lying in $\mathcal{C}_1$ is uniformly distributed over the remainder of the vertices, so

$$\mathbb{P}\Big(v_1 \notin \mathcal{C}^{\max} \;\Big|\; G(n,p)\Big) \leq 1 - \frac{|\mathcal{C}^{\max}|}{n - |\mathcal{C}_1|} < 1 - \frac{|\mathcal{C}^{\max}|}{n}.$$

If v_1 is not in $\mathcal{C}^{\max}$, then it lies in another connected component $\mathcal{C}_2$. Let $f_2 = u_2v_2$ be the smallest-weight edge leaving $\mathcal{C}_1 \cup \mathcal{C}_2$, and let p_2 be its weight. Then f_2 is an edge of $\mathrm{MST}(\mathbb{K}_n)$; to see this, note that any path γ connecting u_2 and v_2 which is not just the edge f_2 contains some edge e of weight strictly greater than p_2, meaning that f_2 is never the heaviest edge of any cycle. Moreover, the endpoint v_2 of f_2 not lying in $\mathcal{C}_1 \cup \mathcal{C}_2$ is uniformly distributed over the remainder of the graph, so once again

$$\mathbb{P}\Big(v_2 \notin \mathcal{C}^{\max} \;\Big|\; G(n,p), v_1 \notin \mathcal{C}^{\max}\Big) < 1 - \frac{|\mathcal{C}^{\max}|}{n}.$$

Continuing this process, we construct a sequence $\mathcal{C}_1, \ldots, \mathcal{C}_K$ of distinct connected components of $G(n,p)$ and a sequence $f_1, \ldots, f_K$ of edges of $\mathrm{MST}(\mathbb{K}_n)$, where where $f_i = u_iv_i$ is the smallest-weight edge from $\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_i$ to the remainder of the graph, $\mathcal{C}_1, \ldots, \mathcal{C}_K$ are all connected components of $G(n,p)$ different from $\mathcal{C}^{\max}$, and $v_K \in \mathcal{C}^{\max}$. To bound the length K of the sequences, we use that at each step of the construction, the conditional probability that $f_j = u_i v_j$ has an endpoint in $\mathcal{C}^{\max}$ given $G(n,p)$ and given that $e_1, \ldots, e_{i-1}$ do not have an endpoint in $\mathcal{C}^{\max}$, is greater than $|\mathcal{C}^{\max}|/n$, and so

$$\begin{aligned}
\mathbb{P}\Big(K > k \;\Big|\; G(n,p)\Big) &= \mathbb{P}\Big(v_k \notin \mathcal{C}^{\max} \;\Big|\; G(n,p)\Big) \\
&= \prod_{i=1}^{k} \mathbb{P}\Big(v_i \notin \mathcal{C}^{\max} \;\Big|\; G(n,p), v_1, \ldots, v_{i-1} \notin \mathcal{C}^{\max}\Big) \\
&\leq \left(1 - \frac{|\mathcal{C}^{\max}|}{n}\right)^k
\end{aligned}$$

Note now that any path in $\mathrm{MST}(\mathbb{K}_n)$ with one endpoint in $\mathcal{C}_1$ and the other endpoint in $\mathcal{C}^{\max}$ passes through $\mathcal{C}_1, \ldots, \mathcal{C}_K$ and edges $f_1, \ldots, f_K$. Since each of the components $\mathcal{C}_1, \ldots, \mathcal{C}_K$ has size at most that of $\mathcal{C}^{\text{runnerup}}$, it follows that the greatest number of edges in any path with one endpoint in $\mathcal{C}_1$ which only intersects $\mathcal{C}^{\max}$ in one vertex is at most $K|\mathcal{C}^{\text{runnerup}}|$. Taking a union bound over the possible choices for $\mathcal{C}_1$ among all components of $G(n,p)$ different from $\mathcal{C}^{\max}$ (there are less than n of them), it follows that

$$\begin{aligned}
\mathbb{P}\Big(L_{n,p} > k|\mathcal{C}^{\text{runnerup}}| \;\Big|\; G(n,p)\Big) &\leq n\mathbb{P}\Big(K > k \;\Big|\; G(n,p)\Big) \\
&\leq n\left(1 - \frac{|\mathcal{C}^{\max}|}{n}\right)^k.
\end{aligned}$$

Recall now the tail bounds for $|\mathcal{C}^{\max}|$ and $|\mathcal{C}^{\text{runnerup}}|$ from (4.15) and (4.16) and use that

$$\begin{aligned}
&\mathbb{P}\Big(L_{n,p} > n^{9/10}\log^2 n \;\Big|\; G(n,p), |\mathcal{C}^{\max}| > n^{9/10}, |\mathcal{C}^{\text{runnerup}}| < n^{4/5}\Big) \\
&\quad= \mathbb{P}\Big(L_{n,p} > (n^{1/10}\log^2 n)\cdot n^{4/5} \;\Big|\; G(n,p), |\mathcal{C}^{\max}| > n^{9/10}, |\mathcal{C}^{\text{runnerup}}| < n^{4/5}\Big) \\
&\quad\leq n\left(1 - \frac{n^{9/10}}{n}\right)^{n^{1/10}\log^2 n} \leq ne^{-\log^2 n}
\end{aligned}$$

to obtain

$$\mathbb{P}\Big(L_{n,p} > n^{9/10}\log^2 n\Big) \leq 7ne^{-(25/2)n^{1/10}} + ne^{-\log^2 n} \leq \frac{2}{n^{\log n}}, \tag{4.18}$$

the last bound holding for n large enough.

We can now conclude the proof of Theorem 4.2.3. By Fact 4.A.3, for n sufficiently large,

$\mathbb{P}(W_n > 3\log^2 n/n) \le 1/n^{\log n}$. Combined with (4.18), this implies that

$$\mathbb{P}\left(2W_n L_{n,p} > \frac{6\log^4 n}{n^{1/10}}\right) \le \frac{3}{n^{\log n}}\,.$$

Using the bound of Proposition 4.A.1 and combining it with the previous inequality and (4.17), we obtain that

$$\begin{aligned}
&\mathbb{P}\left(\mathrm{wdiam}\left(\mathrm{MST}(\mathbb{K}_n)\right) > 3pn^{9/10} + \frac{6\log^4 n}{n^{1/10}}\right) \\
&\quad \le \mathbb{P}\left(\left|T^{\max}_{n,p}\right| \ge 3n^{9/10}\right) + \mathbb{P}\left(2W_n L_{n,p} > \frac{6\log^4 n}{n^{1/10}}\right) \\
&\quad \le ne^{-(25/2)n^{1/10}} + \frac{3}{n^{\log n}} \le \frac{4}{n^{\log n}}\,,
\end{aligned}$$

the last inequality holding when n is large. Finally, since $p = 1/n + 1/n^{11/10}$, for n large we have $3pn^{9/10} + \frac{6\log^4 n}{n^{1/10}} < \frac{7\log^4 n}{n^{1/10}}$, so the bound of Theorem 4.2.3 follows.

□

Chapter 5

Conclusion

This thesis explored combinatorial and algorithmic approaches to random trees and graphs. We studied constructive methods for generating random graphs, and we drew from a massive pool of existing results on simple constructions to aid in understanding more complex variations of said constructions. In Chapter 2 we studied a constructive model for competitive networks closely related to the Erdős-Rényi process and showed how (with high probability) no giant components can form in the model if there are sufficiently many vertices that refuse to connect to one another during the construction. In Chapter 3 we studied random simple tree-weighted graphs and showed that the trees that emerge asymptotically behave like uniform labelled trees. We gave a constructive method, built from Pitman's additive coalescent and the configuration model, and showed that results on this construction translate to results on random simple tree-weighted graphs. Finally, in Chapter 4 we presented an algorithm that can (with high probability) transform an arbitrary spanning subgraph of the weighted complete graph into the minimum weight spanning tree while, at every step, only modifying subgraphs with total weight bounded by $1 + \epsilon$.

Each of the papers presented in this thesis generate a wealth of open problems and invite new avenues for future research. Although we have already discussed many open problems during the conclusions of Chapters 2 and 4, there are two more specific problems that I would like to mention here.

Recall the random process $(G(n, k(n), p), 0 \leq p \leq 1)$ presented in Chapter 2 and recall that M_n is size of the largest component of $G(n, k(n), 1)$. The behaviour of M_n is still unknown when $k(n)/n^{1/3} \to c$ for some constant $c > 0$. In this case, the number of special vertices from $[k(n)]$ that appear in the $\Theta(n^{2/3})$ sized components of $G(n, 1/n)$ is, in expectation, a positive constant dependent on c. In this case, the components would be chopped up into only a few pieces, and perhaps more than one of those pieces could grow into a $\Theta(n)$ sized component in $G(n, k(n), 1)$. In that case, M_n/n would asymptotically approach a constant bounded away from both 0 and 1. Understanding this critical regime could lead to a beautiful conclusion about the change in behaviour of M_n with respect to $k(n)$.

In Chapter 3 we have shown that if $\mathrm{d}^n = (d^n(i), 1 \leq i \leq n)$ is a degree sequence and (G_n, T_n, Γ_n) is a random simple tree-weighted graph with degree sequence d^n then a rescaling of T_n converges in distribution to the Brownian continuum random tree under certain conditions on d^n. The conditions required for our result ensure that G_n is a sparse graph. However, it is straightforward to show that our result holds if $d^n(i) = n - 1$ for all $1 \leq i \leq n$; in this case, G_n is a complete graph, and so T_n is a uniformly random element from the set of trees on vertex set $[n]$. A natural question then arises: does our convergence result still hold under weaker conditions on d^n? We of course require *some* conditions on d^n since, for example, if $\mathrm{d}^n = (n-1, 1, \ldots, 1)$ then $T_n = G_n$ is a star graph for all $n \geq 1$. However, it seems reasonable to guess that our result holds when G_n is a dense graph, and it is interesting to consider what happens when G_n is neither sparse nor dense.

This thesis places itself as part of the literature on random graphs and complex networks, and on scaling limits and optimization algorithms for such networks. I hope that this collection of papers spawns new avenues of research for future generations.

Bibliography

[1] Romain Abraham, Jean-François Delmas, and Patrick Hoscheit. A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces. *Electron. J. Probab.*, 18: no. 14, 21, 2013.

[2] Louigi Addario-Berry. The local weak limit of the minimum spanning tree of the complete graph. arXiv:1301.1667 [math.PR], January 2013.

[3] Louigi Addario-Berry. Partition functions of discrete coalescents: from Cayley's formula to Frieze's $\zeta(3)$ limit theorem. *XI Symposium on Probability and Stochastic Processes*, pages 1–45. Springer, 2015.

[4] Louigi Addario-Berry and Jordan Barrett. Random Tree-Weighted Graphs. *Stochastic Analysis, Random Fields and Integrable Probability, Adv. Studies in Pure Math*, 87, pages 1–59. MSJ, 2021.

[5] Louigi Addario-Berry and Jordan Barrett. Multi-source Invasion Percolation on the Complete Graph. Submitted.

[6] Louigi Addario-Berry, Jordan Barrett and Benoît Corsini. Finding Minimum Spanning Trees via Local Improvements. Submitted.

[7] Louigi Addario-Berry, Anna Brandenberger, Jad Hamdan, and Céline Kerriou. Universal height and width bounds for random trees. arXiv:2105.03195 [math.PR], May 2021.

[8] Louigi Addario-Berry, Nicolas Broutin, and Christina Goldschmidt. The continuum limit of critical random graphs. *Probab. Theory Related Fields*, 152(3-4):367–406, (2012).

[9] Louigi Addario-Berry, Nicolas Broutin, and Christina Goldschmidt. Critical Random Graphs: Limiting Constructions and Distributional Properties. *Electronic Journal of Probability*, 15 (none):741 – 775, 2010.

[10] Louigi Addario-Berry, Nicolas Broutin, Christina Goldschmidt, and Grégory Miermont. The scaling limit of the minimum spanning tree of the complete graph. *Ann. Probab.*, 45(5):3075–3144, 2017.

[11] Louigi Addario-Berry, Nicolas Broutin, and Cecilia Holmgren. Cutting down trees with a Markov chainsaw. *Ann. Appl. Probab.*, 24(6):2297–2339, 2014.

[12] Louigi Addario-Berry, Nicolas Broutin and Bruce Reed. The diameter of the minimum spanning tree of a complete graph. *Discrete Mathematics and Theoretical Computer Science*, pages 237–248. 2006.

[13] Louigi Addario-Berry, Simon Griffiths, and Ross J. Kang. Invasion percolation on the Poisson-weighted infinite tree. *Ann. Appl. Probab.*, 22(3):931–970, 2012.

[14] Louigi Addario-Berry and Sanchayan Sen. Geometry of the minimal spanning tree of a random 3-regular graph. *Probab. Theory Related Fields*, 180(3-4):553–620, 2021.

[15] Reka Albert, Hawoong Jeong and Albert-Laszlo Barabasi. The Diameter of the World Wide Web. *Nature*, 401, pages 130–131, 1999.

[16] David Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991.

[17] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25(2):812–854, 1997.

[18] David Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985.

[19] David Aldous and J. Michael Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 1–72. Springer, Berlin, 2004.

[20] Omer Angel, Sébastien Bubeck, Yuval Peres, and Fan Wei. Local max-cut in smoothed polynomial time. *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 429–437. ACM, New York, 2017

[21] Omer Angel, Jesse Goodman, Frank den Hollander, and Gordon Slade. Invasion percolation on regular trees. *Ann. Probab.*, 36(2):420–466, 2008.

[22] Omer Angel, Jesse Goodman, and Mathieu Merle. Scaling limit of the invasion percolation cluster on a regular tree. *Ann. Probab.*, 41(1):229–261, 2013.

[23] Albert-Laszlo Barabási, Réka Albert and Hawoong Jeong. Mean-Field Theory for Scale-Free Random Networks. *Physica A: Statistical Mechanics and its Applications*, 272, pages 173–187, 2000.

[24] Nathanaël Berestycki, Benoît Laslier, and Gourab Ray. Critical exponents on Fortuin-Kasteleyn weighted planar maps. *Comm. Math. Phys.*, 355 (2): 427–462, 2017.

[25] Olivier Bernardi. Bijective counting of tree-rooted maps and shuffles of parenthesis systems. *Electron. J. Combin.*, 14 (1): Research Paper 9, 36, 2007.

[26] Olivier Bernardi. Tutte polynomial, subgraphs, orientations and sandpile model: new connections via embeddings. *Electron. J. Combin.*, 15(1):Research Paper 109, 53, 2008.

[27] Shankar Bhamidi, Remco van der Hofstad, and Sanchayan Sen. The multiplicative coalescent, inhomogeneous continuum random trees, and new universality classes for critical random graphs. *Probab. Theory Related Fields*, 170(1-2):387–474, 2018.

[28] Shankar Bhamidi and Sanchayan Sen. Geometry of the vacant set left by random walk on random graphs, Wright's constants, and critical random graphs with prescribed degrees. *Random Structures Algorithms*, 56(3):676–721, 2020.

[29] Ali Bibak, Charles Carlson and Karthekeyan Chandrasekaran. Improving the smoothed complexity of FLIP for max cut problems. *ACM Transactions on Algorithms*, 17(3), Art. 19, 38, 2021.

[30] Béla Bollobás. The Evolution of Random Graphs. *Transactions of the American Mathematical Society*, 286(1), pages 257–274, 1984.

[31] Otakar Borůvka. O Jistém Problému minimálním. *Práce Mor. Přírodově D. Spol. V. Brně*, 3, pages 37–58, 1926

[32] Mireille Bousquet-Mélou and Julien Courtiel. Spanning forests in regular planar maps. *J. Combin. Theory Ser. A*, 135:1–59, 2015.

[33] Nicolas Broutin and Jean-François Marckert. Asymptotics of trees with a prescribed degree sequence and applications. *Random Structures Algorithms*, 44(3):290–316, 2014.

[34] Richard Chandler, Joel Koplik, Kenneth Lerman, and Jorge F. Willemsen. Capillary displacement and percolation in porous media. *Journal of Fluid Mechanics*, 119:249–267, 1982.

[35] Xi Chen, Chenghao Guo, Emmanouil V. Vlatakis-Gkaragkounis, Mihalis Yannakakis and Xinzhi Zhang. *Smoothed Complexity of Local Max-Cut and Binary Max-CSP*, pages 1052-1065. Association for Computing Machinery, New York, USA, 2020.

[36] Michael Damron and Artëm Sapozhnikov. Limit theorems for 2D invasion percolation. *Ann. Probab.*, 40(3):893–920, 2012.

[37] Paul Erdős and Alfréd Rényi. On Random Graphs I. *Publicationes Mathematicae*, 6, pages 290–297, 1959.

[38] Paul Erdős and Alfréd Rényi. On the Evolution of Random Graphs. *Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, 5, pages 17–61, 1960.

[39] Michael Etscheid and Heiko Röglin. Smoothed Analysis of Local Search for the Maximum-Cut Problem. *ACM Transactions on Algorithms*, 13(2), 2017.

[40] Alan M. Frieze. On the value of a random minimum spanning tree problem. *Discrete Applied Mathematics*, 10(1), pages 47–56. Elsevier, 1985.

[41] Christophe Garban, Gábor Pete, and Oded Schramm. The scaling limits of near-critical and dynamical percolation. *J. Eur. Math. Soc. (JEMS)*, 20(5):1195–1268, 2018.

[42] Christophe Garban, Gábor Pete, and Oded Schramm. The scaling limits of the minimal spanning tree and invasion percolation in the plane. *Ann. Probab.*, 46(6):3501–3557, 2018.

[43] Catherine Greenhill, Matthew Kwan, and David Wind. On the number of spanning trees in random regular graphs. *Electron. J. Combin.*, 21(1):Paper 1.45, 26, 2014.

[44] Catherine Greenhill, Mikhail Isaev, Matthew Kwan, and Brendan D. McKay. The average number of spanning trees in sparse graphs with given degrees. *European J. Combin.*, 63:6–25, 2017.

[45] Ewain Gwynne, Nina Holden, and Xin Sun. A mating-of-trees approach for graph distances in random planar maps. *Probab. Theory Related Fields*, 177(3-4):1043–1102, 2020.

[46] Ewain Gwynne, Adrien Kassel, Jason Miller, and David B. Wilson. Active spanning trees with bending energy on planar maps and SLE-decorated Liouville quantum gravity for $\kappa > 8$. *Comm. Math. Phys.*, 358(3):1065–1115, 2018.

[47] Ewain Gwynne, Jason Miller and Scott Sheffield. The Tutte embedding of the mated-CRT map converges to Liouville quantum gravity. Annals of Probability, 49(6), pages 2732–2829, 2021.

[48] Nina Holden and Xin Sun. SLE as a mating of trees in Euclidean geometry. *Comm. Math. Phys.*, 364(1):171–201, 2018.

[49] Joseph B. Kruskal, Jr. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Amer. Math. Soc.*, 7:48–50, 1956.

[50] Jean-François Le Gall. Random Trees and Applications. *Problab. Surveys*, 2, pages 245–311, 2005.

[51] Yiting Li, Xin Sun, and Samuel S. Watson. Schnyder woods, SLE(16), and Liouville quantum gravity. arXiv:1705.03573, May 2017.

[52] Tomasz Łuczak. Component behavior near the critical point of the random graph process. *Random Structures Algorithms*, 1(3):287–310, 1990.

[53] Adam Logan, Mike Molloy, and Pawel Pralat. A variant of the Erdos-Renyi random graph process. arXiv:1806.10975 [math.CO], 2018.

[54] Bodo Manthey. Smoothed Analysis of Local Search Algorithms. *Algorithms and Data Structures*, pages 518–527. Springer International Publishing, 2015.

[55] Colin McDiarmid, Theodore Johnson, and Harold S. Stone. On finding a minimum spanning tree in a network with random weights. *Random Structures Algorithms*, 10(1-2):187–204, 1997.

[56] Brendan D. McKay. Spanning trees in random regular graphs. In *Proceedings of the Third Caribbean Conference on Combinatorics and Computing (Bridgetown, 1981)*, pages 139–143. Univ. West Indies, Cave Hill Campus, Barbados, 1981.

[57] Marcus Michelen, Robin Pemantle, and Josh Rosenberg. Invasion percolation on Galton-Watson trees. *Electron. J. Probab.*, 24:Paper No. 31, 35, 2019.

[58] Jason Miller and Scott Sheffield. Liouville quantum gravity spheres as matings of finite-diameter trees. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55(3):1712–1750, 2019.

[59] J. W. Moon. *Counting labelled trees*, volume 1969 of *From lectures delivered to the Twelfth Biennial Seminar of the Canadian Mathematical Congress (Vancouver*. Canadian Mathematical Congress, Montreal, Que., 1970.

[60] R. C. Mullin. On the enumeration of tree-rooted maps. *Canadian J. Math.*, 19:174–183, 1967.

[61] C. M. Newman and D. L. Stein. Random walk in a strongly inhomogeneous environment and invasion percolation. *Ann. Inst. H. Poincaré Probab. Statist.*, 31(1):249–261, 1995.

[62] C. M. Newman and D. L. Stein. Ground-state structure in a highly disordered spin-glass model. *J. Statist. Phys.*, 82(3-4):1113–1132, 1996.

[63] Bernie Nickel and David Wilkinson. Invasion percolation on the Cayley tree: exact solution of a modified percolation model. *Phys. Rev. Lett.*, 51(2):71–74, 1983.

[64] Jim Pitman. Coalescent random forests. *J. Combin. Theory Ser. A*, 85(2):165–193, 1999.

[65] R. C. Prim. Shortest connection networks and some generalizations. *The Bell System Technical Journal*, 36(6):1389–1401, 1957.

[66] Colin P. Stark. An invasion percolation model of drainage network evolution. *Nature*, 352 (6334):423–425, 1991.

[67] Jacob van den Berg, Antal A. Járai, and Bálint Vágvölgyi. The size of a pond in 2D invasion percolation. *Electron. Comm. Probab.*, 12:411–420, 2007.

[68] Remco van der Hofstad. *Random graphs and complex networks. Vol. 1.* Cambridge Series in Statistical and Probabilistic Mathematics, [43]. Cambridge University Press, Cambridge, 2017.

[69] T. Walsh and A. B. Lehman. Counting rooted maps by genus. II. *J. Combinatorial Theory Ser. B*, 13:122–141, 1972.

[70] David Wilkinson and Jorge F. Willemsen. Invasion percolation: a new form of percolation theory. *J. Phys. A*, 16(14):3365–3376, 1983.

[71] Yu Zhang. The fractal volume of the two-dimensional invasion percolation cluster. *Comm. Math. Phys.*, 167(2):237–254, 1995.