MULTI-SOURCE INVASION PERCOLATION ON THE COMPLETE GRAPH

LOUIGI ADDARIO-BERRY AND JORDAN BARRETT

ABSTRACT. We consider invasion percolation on the randomly-weighted complete graph K_n, started from some number $k(n)$ of distinct source vertices. The outcome of the process is a forest consisting of $k(n)$ trees, each containing exactly one source. Let M_n be the size of the largest tree in this forest. Logan, Molloy and Pralat [23] proved that if $k(n)/n^{1/3} \to 0$ then $M_n/n \to 1$ in probability. In this paper we prove a complementary result: if $k(n)/n^{1/3} \to \infty$ then $M_n/n \to 0$ in probability. This establishes the existence of a phase transition in the structure of the invasion percolation forest around $k(n) \asymp n^{1/3}$.

Our arguments rely on the connection between invasion percolation and critical percolation, and on a coupling between multi-source invasion percolation with differently-sized source sets. A substantial part of the proof is devoted to showing that, with high probability, a certain fragmentation process on large random binary trees leaves no components of macroscopic size.

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 $1 \le i < |v(G)| + 1 - |\mathcal{S}|$, 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 $\mathcal{S} \cup \{v_i, 0 \le i < |v(G)| + 1 - |\mathcal{S}|\}$ and edge set $\{e_i, 1 \le i < |v(G)| + 1 - |\mathcal{S}|\}$. 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 [17], and independently (with a slightly different formulation, using vertex rather than edge weights) in [32]. 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 [30].

Invasion percolation has been extensively studied in the probability and statistical physics communities: on lattices [18, 20, 27, 31, 33], on trees [5, 13, 14, 25, 28], and in the mean-field or general graph setting [1, 11, 20, 24, 26]. 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 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 [23]. 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 $F(K_n, [k(n)])$, with $L_{n,i} = 0$ if $F(K_n, [k(n)])$ 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 $(L_{n,i}/n, i \geq 1) \overset{d}{\to} (L_{\infty,i}(c))$ in the sense of finite-dimensional distributions.

1.1. **Overview of the rest of the paper.** In Section 2, we explain several useful connections between invasion percolation, critical percolation, and minimum spanning trees. We then use these connections to prove Theorem 1.1, modulo a key input to the proof. This key input, Proposition 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 2.1, which occupies the bulk of the paper, appears in Section 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 4 proposes some future research directions suggested by the current work.

2. **A sketch proof of Theorem 1.1.**

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* [29] 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 $0 < \varepsilon < \min(\mathrm{w}_e, e \in e(G))$ 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* [21], 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. **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.

2.3. **The critical random graph and the proof of Theorem 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

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

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

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

The continuous-time processes add the same edges as the discrete processes, and in the same order. More strongly, $G(n, \mathcal{S}, U_{e(i)}) = \mathrm{K}_{n,i}^{\mathcal{S}}$ 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 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 2.1. *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 2.1 appears in Section 3. To prove Theorem 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 $T_{v,\lambda}^n$ for the tree of this forest containing v; see Figure 1. Let $q_{n,\lambda}^v = |T_{v,\lambda}^n|/n$ be the proportion of vertices of $F(n,1)$ lying in $T_{v,\lambda}^n$.

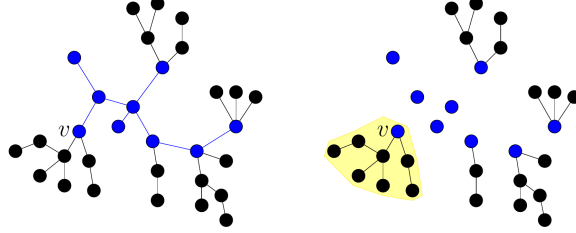FIGURE 1. Left: An instantiation of $F(n,1)$ with $F^1(n, p_{n,\lambda})$ drawn in blue. Right: the forest obtained from $F(n,1)$ by removing the edges of $F^1(n, p_{n,\lambda})$. On the right the tree $T_{v,\lambda}^n$ is highlighted.

Proposition 2.2 ([7], Lemma 4.11). *Write $\Delta_\lambda^n = \max(q_{n,\lambda}^v, v \in F^1(n, p_{n,\lambda}))$. Then for all $\delta > 0$,*

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

Next, let

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

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

Proposition 2.3 ([2], Lemma 6.19). *For every $\lambda \in \mathbb{R}$, conditionally given $\mathcal{F}_{n,\lambda}$, the collection of random variables $(q_{n,\lambda}^v, v \in F^1(n, p_{n,\lambda}))$ is exchangeable.*

The exchangeability in [2, Lemma 6.19] is stated conditionally given $G(n, p_{n,\lambda})$, rather than given $\mathcal{F}_{n,\lambda}$. In other words, in [2] 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 [2] 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 [12].

Proposition 2.4 ([12], Theorem 20.7). *For all $\varepsilon > 0$ 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*

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

See [15, Lemma 7.5] and [16, Lemma 4.9] for quantitative versions of this result. Proposition 2.4 is the last fact we need for the proof of our main result.

Proof of Theorem 1.1. As noted just after the statement of Theorem 1.1, the fact that $M_n/n \to 1$ in probability when $k(n)/n^{1/3} \to 0$ was proved in [23], 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 [22, 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 $(2 + o(1))np_n$ 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 $\delta > 0$,

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

Fix $\varepsilon \in (0, 1)$. Then fix $\delta \in (0, \varepsilon/2)$ small enough that Proposition 2.4 holds for this ε and δ, then let λ be large enough that for all n sufficiently large,

$$\mathbf{P} \left\{ |F^1(n, p_{n,\lambda})|/n^{2/3} \le 1 \right\} < \varepsilon \qquad (2.3)$$

and

$$\mathbf{P} \left\{ \Delta_\lambda^n \ge \delta \right\} < \varepsilon. \qquad (2.4)$$

This is possible by (2.2) and by Proposition 2.2. Since λ is fixed, by Proposition 2.1, we also have that

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

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

$$M_n = \max(|C| : C \text{ is a component of } F(n,[k(n)],1))$$

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

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

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

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

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

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

$$\mathbf{P}\left\{M_n \geq 3\varepsilon n\right\} \leq \mathbf{P}\left\{M_{n,\lambda}(k(n)) > \delta n^{2/3}\right\} + \mathbf{P}\left\{|F^1(n, p_{n,\lambda})| < n^{2/3}\right\}$$

$$+ \mathbf{P}\left\{\max_{1 \leq i \leq m_n}\Big|\sum_{j=1}^{i}\Big(q_{n,\lambda}^{v_j} - \frac{1}{m_n}\Big)\Big| > \varepsilon\right\}$$

$$< 2\varepsilon + \mathbf{P}\left\{\max_{1 \leq i \leq m_n}\Big|\sum_{j=1}^{i}\Big(q_{n,\lambda}^{v_j} - \frac{1}{m_n}\Big)\Big| > \varepsilon\right\}, \qquad (2.6)$$

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

$$\mathbf{P}\left\{\max_{1 \leq i \leq m_n}\Big|\sum_{j=1}^{i}\Big(q_{n,\lambda}^{v_j} - \frac{1}{m_n}\Big)\Big| > \varepsilon\right\}$$

$$= \mathbf{E}\left[\mathbf{P}\left\{\max_{1 \leq i \leq m_n}\Big|\sum_{j=1}^{i}\Big(q_{n,\lambda}^{v_j} - \frac{1}{m_n}\Big)\Big| > \varepsilon \;\Big|\; \mathcal{F}_{n,\lambda}\right\}\right]$$

$$= \mathbf{E}\left[\mathbf{P}\left\{\max_{1 \leq i \leq m_n}\Big|\sum_{j=1}^{i}\Big(q_{n,\lambda}^{v_{\pi(j)}} - \frac{1}{m_n}\Big)\Big| > \varepsilon \;\Big|\; \mathcal{F}_{n,\lambda}\right\}\right],$$

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 2.3.

Recall that $\Delta_\lambda^n := \max(q_{n,\lambda}^v, v \in F^1(n, p_{n,\lambda}))$; then by Proposition 2.4 we have

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

$$\leq \mathbf{P}\left\{\Delta_\lambda^n \geq \delta \;\Big|\; \mathcal{F}_{n,\lambda}\right\} + \mathbf{P}\left\{\max_{1 \leq i \leq m_n}\Big|\sum_{j=1}^{i}\Big(q_{n,\lambda}^{v_{\pi(j)}} - \frac{1}{m_n}\Big)\Big| > \varepsilon \;\Big|\; \mathcal{F}_{n,\lambda}, \Delta_\lambda^n \leq \delta\right\}$$

$$\leq \mathbf{P}\left\{\Delta_\lambda^n \geq \delta \;\Big|\; \mathcal{F}_{n,\lambda}\right\} + \varepsilon,$$

so it follows that

$$\mathbf{P}\left\{\max_{1 \leq i \leq m_n}\Big|\sum_{j=1}^{i}\Big(q_{n,\lambda}^{v_j} - \frac{1}{m_n}\Big)\Big| > \varepsilon\right\} \leq \varepsilon + \mathbf{E}\left[\mathbf{P}\left\{\Delta_\lambda^n \geq \delta \;\Big|\; \mathcal{F}_{n,\lambda}\right\}\right]$$

$$= \varepsilon + \mathbf{P}\left\{\Delta_\lambda^n \geq \delta\right\} < 2\varepsilon,$$

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

3. **Proof of Proposition 2.1.**

Our proof of Proposition 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.

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 $M_{n,\lambda}^i(\mathcal{S})$ 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 $M_{n,\lambda}^i(k(n))$ instead of $M_{n,\lambda}^i([k(n)])$

In this section, we show how Proposition 2.1 is a consequence of the following result.

Proposition 3.1. *Fix $\lambda \in \mathbb{R}$ and $i \in \mathbb{N}$. If $k(n)/n^{1/3} \to \infty$ then $M_{n,\lambda}^i(k(n))/n^{2/3} \to 0$ in probability.*

Proof of Proposition 2.1. Fix $\varepsilon > 0$.

By [10, Corollary 2], there is $j = j(\varepsilon) \in \mathbb{N}$ such that for all $n \in \mathbb{N}$,

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

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

$$\mathbf{P}\left\{M_{n,\lambda}(k(n)) \geq \varepsilon n^{2/3}\right\} \leq \mathbf{P}\left\{\max_{1 \leq \ell \leq j} M_{n,\lambda}^\ell(k(n)) > \varepsilon n^{2/3}\right\}$$

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

$$\leq \varepsilon + \sum_{1 \leq \ell \leq j} \mathbf{P}\left\{M_{n,\lambda}^\ell(k(n)) > \varepsilon n^{2/3}\right\}$$

$$\leq 2\varepsilon$$

for n sufficiently large, the last bound holding due to Proposition 3.1. Since $\varepsilon > 0$ was arbitrary, this proves Proposition 2.1. □

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 $F_i^{G,\mathcal{S}}$ and $F_i^{G,\emptyset}$, on the one hand, and of F_i and G_i, on the other hand, allows us to construct $F_m^{G,\mathcal{S}}$ via a *path-and-cycle-breaking* process starting from G. Recall the definition of the augmented graph $G'_\mathcal{S}$ from Section 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 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 3.3. *For all $\varepsilon > 0$ 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,*

$$\mathbf{E}\left[\max(|C| : C \text{ is a component of } F_q)\right] \leq \varepsilon q.$$

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 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 3.3 and Markov's inequality together imply that for any $\varepsilon > 0$,

$$q^{-1}\mathbf{E}\left[\max(|C| : C \text{ is a component of } F_q)\right] \to 0 \qquad (3.1)$$

as $q \to \infty$.

We prove Proposition 3.3 in Section 3.3, below; before doing so, we use it (or in fact its consequence, (3.1)) to prove Proposition 3.1.

Proof of Proposition 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 [10, 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 $\varepsilon > 0$ there exists $\delta > 0$ such that $\mathbf{P}\left\{Q(n) \geq \delta n^{2/3}\right\} > 1 - \varepsilon$. 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 [10, Corollary 2] we have $\liminf_{n\to\infty} \mathbf{P}\left\{S(n) = s\right\} > 0$, 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 $x > 0$,

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

as $n \to \infty$.

Next, recall that

$$M_{n,\lambda}^i(k(n))$$

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

and write $M_{n,\lambda}^i = M_{n,\lambda}^i(k(n))$ for succinctness. Conditionally given $Q(n)$, $R(n)$ and $S(n)$, the random variable $M_{n,\lambda}^i$ has the same distribution as $\max(|C| : C \text{ is a component of } F_{Q(n)})$, 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 (3.1) and Markov's inequality that for any $\varepsilon > 0$,

$$\mathbf{P}\left\{M_{n,\lambda}^i > \varepsilon|F^i(n, p_{n,\lambda})| \mid S(n) = s\right\} = \mathbf{P}\left\{Q(n)^{-1}M_{n,\lambda}^i > \varepsilon \mid S(n) = s\right\}$$
$$\to 0,$$

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

this yields that for any $\varepsilon > 0$,

$$\limsup_{n \to \infty} \mathbf{P}\left\{ M_{n,\lambda}^i > \varepsilon |F^i(n, p_{n,\lambda})| \right\}$$

$$\leq \limsup_{n \to \infty} \max_{1 \leq s \leq s_0} \mathbf{P}\left\{ Q(n)^{-1} M_{n,\lambda}^i > \varepsilon \mid S(n) = s \right\} + \limsup_{n \to \infty} \mathbf{P}\left\{ S(n) > s_0 \right\}$$

$$\leq \varepsilon \,.$$

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

3.3. **Step 3: partitioning a component.** The goal of this section is to prove Proposition 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 3.3 as follows.

Proposition 3.4. *For all $\varepsilon > 0$, 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,*

$$\mathbf{E}\left[\max(|C| : C \text{ is a component of } F_q) \right] \leq \varepsilon q \,.$$

In Section 3.3.1 we prove Proposition 3.4, establishing the case $s = 0$ of Proposition 3.3. We then use Proposition 3.4 to handle the cases when $s \geq 1$, completing the proof of Proposition 3.3, in Section 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).

3.3.1. *Proof of Proposition 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 $X, Y \in_u [q]$,

$$\mathbf{P}\left\{X \xleftrightarrow{F_q} Y\right\} = \mathbf{E}\left[\mathbf{P}\left\{X \xleftrightarrow{F_q} Y \;\middle|\; F_q\right\}\right]$$

$$= \mathbf{E}\left[\sum_{C \text{ is a component of } F_q} \frac{|v(C)|^2}{q^2}\right]$$

$$\geq \mathbf{E}\left[\frac{\max(|v(C)| : C \text{ is a component of } F_q)^2}{q^2}\right]$$

$$\geq \frac{\mathbf{E}\left[\max(|v(C)| : C \text{ is a component of } F_q)\right]^2}{q^2}. \qquad (3.2)$$

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 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 $U \xleftrightarrow{F_q} W$ if and only if $U \xleftrightarrow{F'_q} W$ 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. $\qquad \square$

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. 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) $1 \leq |e(P_i)| \leq \sqrt{q}$.

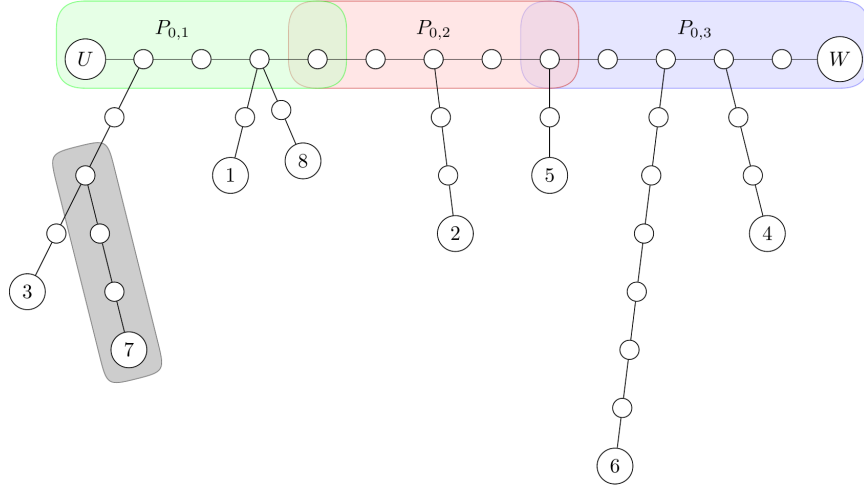Additionally, write $\mathcal{P}_j = \{P_i : i \in \mathcal{U}_j\}$.

FIGURE 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 $|e(P_6(T))| = 7 > \sqrt{48}$.

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 3.7), write $P_i = P_i(T'_q)$ for $0 \le i \le r$, 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 3.6. *If $U \overset{F'_q}{\longleftrightarrow} W$ and both $|\mathcal{U}_1| \ge 1$ and $|\mathcal{U}_3| \ge 1$, 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 $1 \le t \le M_q$. 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 $e_t \in P_{0,2}$ 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. $\qquad\square$

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

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

Corollary 3.7. *Let $E_q(p, u)$ be the event that $|e(P_{0,2})| \ge p$, $|\mathcal{U}_1| \ge u$ and $|\mathcal{U}_3| \ge u$. Then for $u \ge 1$,*

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

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 3.6, if U and W are connected in F'_q and both $\mathcal{U}_1$ and $\mathcal{U}_3$ are nonempty, then $s \ge |\mathcal{S}| - 3$.

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 $1 \le j \le m$ 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 $0 \le j \le m$,

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

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

$$\hat{\mathbf{P}}\left\{ s \geq |\mathcal{S}| - 3 \mid E_q^=(p, u) \right\}$$

$$= \hat{\mathbf{P}}\left\{ P_{0,2} \notin \{P^{(1)}, \ldots, P^{(m-3)}\} \;\middle|\; E_q^=(p, u) \right\}$$

$$= \prod_{j=0}^{m-4} \hat{\mathbf{P}}\left\{ P_{0,2} \neq P^{(j+1)} \;\middle|\; E_q^=(p, u), P_{0,2} \notin \{P^{(1)}, \ldots, P^{(j)}\} \right\}$$

$$\leq \prod_{j=0}^{2u-3} \left(1 - \frac{p}{p + \sqrt{q}(2u - j)} \right). \tag{3.3}$$

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 (3.3) by the law of total probability. $\qquad\square$

In order to make use of Corollary 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 [9]. 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 $\mathrm{Exp}(1)$ 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 [9, 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 3.8. *As $q \to \infty$,*

$$(q^{-1/2}\mathrm{dist}_{T_q'(r)}(x, y) : x, y \in \{U, W\} \cup \{1, \ldots, r\})$$

$$\xrightarrow{\mathrm{d}} (\mathrm{dist}_{T_\infty(r)}(x, y) : x, y \in \{U, W\} \cup \{1, \ldots, r\}).$$

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 $T_\infty^\phi(r)$ 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 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 3.8.

Corollary 3.9. *For all $\varepsilon > 0$ and positive integers r, for large enough q,*

$$\mathbf{P}\left\{|e(P_{0,2}(T_q'))| < \varepsilon\sqrt{q}\right\} \leq 5\varepsilon^2.$$

Proof. By Proposition 3.8, we have that

$$\mathbf{P}\left\{|e(P_0(T_q'))| < 3\varepsilon\sqrt{q}\right\} = (1 + o_q(1))\mathbf{P}\left\{\sqrt{2}(E_1)^{1/2} < 3\varepsilon\right\}$$

$$= (1 + o_q(1))\mathbf{P}\left\{E_1 < \frac{9\varepsilon^2}{2}\right\}$$

$$= (1 + o_q(1))\left(1 - e^{-\frac{9\varepsilon^2}{2}}\right)$$

$$\leq (1 + o_q(1))\frac{9\varepsilon^2}{2}.$$

The result now follows from the fact that $|e(P_{0,2}(T_q'))| \geq \frac{1}{3}|e(P_0(T_q'))| - 1$. $\square$

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

Lemma 3.10. *For all $\varepsilon > 0$ sufficiently small, there exists a positive integer r_0 such that for all $r \geq r_0$, for all q sufficiently large, with $T'_q = T'_q(r, \mathbf{e})$,*

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

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

The proof of Lemma 3.10 is somewhat involved, so before proving it we first use it together with the preceding results of the section to prove Proposition 3.4.

Proof of Proposition 3.4. In this proof, for readability we omit insignificant floors and ceilings. Fix $\varepsilon > 0$ small enough that Lemma 3.10 applies, let r_0 be as in Lemma 3.10, and fix $r \geq r_0$ large enough that $\sum_{j=4}^{\varepsilon r^{1/2}} \frac{1}{j} \geq \varepsilon^{-1}\log(\varepsilon^{-1})$. For the duration of the proof, write $P_j = P_j(T'_q)$ for all $0 \leq j \leq r$ 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 3.7 that $E_q(p, u)$ is the event that $|e(P_{0,q})| \geq p$, $|\mathcal{U}_1| \geq u$ and $|\mathcal{U}_3| \geq u$. Taking $p = \varepsilon q^{1/2}$ and $u = \varepsilon r^{1/2}/2$, by that corollary and Bayes' formula we have

$$\hat{\mathbf{P}}\left\{U \xleftrightarrow{F'_q} W\right\} \leq \hat{\mathbf{P}}\left\{U \xleftrightarrow{F'_q} W \;\middle|\; E_q(p, u)\right\} + \hat{\mathbf{P}}\left\{E_q(p, u)^c\right\}$$

$$\leq \prod_{j=0}^{2u-3}\left(1 - \frac{p}{p + q^{1/2}(2u - j)}\right) + \hat{\mathbf{P}}\left\{E_q(p, u)^c\right\}$$

$$= \prod_{j=0}^{\varepsilon r^{1/2}-3}\left(1 - \frac{\varepsilon}{\varepsilon(1 + r^{1/2}) - j}\right) + \hat{\mathbf{P}}\left\{E_q(p, u)^c\right\}$$

$$\leq \exp\left(-\varepsilon\sum_{j=0}^{\varepsilon r^{1/2}-3} \frac{1}{\varepsilon(1 + r^{1/2}) - j}\right) + \hat{\mathbf{P}}\left\{E_q(p, u)^c\right\}$$

$$\leq \exp\left(-\varepsilon\sum_{j=4}^{\varepsilon r^{1/2}} \frac{1}{j}\right) + \hat{\mathbf{P}}\left\{E_q(p, u)^c\right\}$$

$$\leq \varepsilon + \hat{\mathbf{P}}\left\{E_q(p, u)^c\right\} \,,$$

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

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

By Corollary 3.9 we have $\mathbf{P}\left\{|e(P_{0,2})| \leq \varepsilon q^{1/2}\right\} \leq 5\varepsilon^2$ and by Lemma 3.10 we have $\mathbf{P}\left\{\min(|\mathcal{U}_1|, |\mathcal{U}_3|) \leq \varepsilon r^{1/2}/2\right\} \leq 46\varepsilon$, so $\mathbf{P}\left\{E_q(p, u)^c\right\} \leq 5\varepsilon^2 + 46\varepsilon$.

Also, by Lemma 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

$$\mathbf{P}\left\{U \xleftrightarrow{F_q} W\right\} \le 47\varepsilon + 5\varepsilon^2\,.$$

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

$$\mathbf{P}\left\{X \xleftrightarrow{F_q} Y\right\} \le \mathbf{P}\left\{U \xleftrightarrow{F_q} W\right\} + \mathbf{P}\left\{X = Y\right\} + \mathbf{P}\left\{X \in [r]\right\} + \mathbf{P}\left\{Y \in [r]\right\}$$

$$\le 47\varepsilon + 5\varepsilon^2 + \frac{2r+1}{q}\,.$$

Finally, by (3.2), we have

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

and so

$$\mathbf{E}\left[\max\left(|C| : C \text{ is a component of } F_q\right)\right] \le q\mathbf{P}\left\{X \xleftrightarrow{F_q'} X\right\}^{1/2}$$

$$\le q\left(47\varepsilon + 5\varepsilon^2 + \frac{2r+1}{q}\right)^{1/2}\,.$$

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

The remainder of the section is devoted to proving Lemma 3.10. We use Proposition 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 $|P_i(t)| = \text{dist}_t(i, \alpha(i))$. 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 $\text{dist}_t(\alpha_i(t), U) \le \text{dist}_t(U, W)/3$, and
   (3) $\text{dist}_t(i, \alpha_i(t)) \le 1$.

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

The convergence in Proposition 3.8 implies that for any $r \ge 1$, as $q \to \infty$ we have $(|\mathcal{U}_1(T_q', r, q)|, |\mathcal{U}_3(T_q', r, q)| \xrightarrow{\text{d}} (|\mathcal{U}_1(T_\infty(r))|, |\mathcal{U}_3(T_\infty(r))|)$. 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 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 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

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

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

Proof of Lemma 3.10. As mentioned earlier, the convergence in distribution from Proposition 3.8 implies that for any $r \geq 1$, as $q \to \infty$ we have $(\mathcal{U}_1(T_q', r, q), \mathcal{U}_3(T_q', r, q) \overset{\mathrm{d}}{\to} (\mathcal{U}_1(T_\infty(r)), \mathcal{U}_3(T_\infty(r)))$. To prove the lemma it thus suffices to show that for all $\varepsilon > 0$ there exists r_0 such that for all $r \geq r_0$,

$$\mathbf{P}\left\{|\mathcal{U}_1(T_\infty(r))| < \varepsilon\sqrt{r}\right\} < 22\varepsilon. \tag{3.4}$$

(The same bound then holds for $\mathcal{U}_3(T_\infty(r))$ by symmetry.) The remainder of the proof is thus devoted to establishing (3.4). In what follows we write $\mathcal{U}_1 = \mathcal{U}_1(T_\infty(r))$.

By Lemma 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

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

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 $r \geq 2$,

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

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
$$\mathbf{P}\left\{ r \in \mathcal{U}_1 \mid E_1 \right\} = \mathbf{E}\left\{ R\mathbf{1}_{[\pi_r \le 1]} \mid E_1 \right\} \tag{3.5}$$
and
$$\mathbf{P}\left\{ r-1 \in \mathcal{U}_1, r \in \mathcal{U}_1 \mid E_1 \right\} = \mathbf{E}\left\{ R'\mathbf{1}_{[\pi_{r-1}, \pi_r \le 1]} \mid E_1 \right\} \tag{3.6}$$

Now fix $\varepsilon \in (0, 1/2)$ small and let $A(\varepsilon, r)$ be the event that
$$(1-\varepsilon)r \le E_2 + \ldots + E_{r-1} \le E_2 + \ldots + E_r \le (1+\varepsilon)r$$
and that $E_r \le r^{1/2}$ and $E_{r+1} \le r^{1/2}$. On $A(\varepsilon, r)$, using the bound $(a + x)^{1/2} - a^{1/2} \le x/(2a^{1/2})$ for all $a, x > 0$, we have that
$$\pi_r = \sqrt{2}\left((E_1 + \ldots + E_{r+1})^{1/2} - (E_1 + \ldots + E_r)^{1/2} \right)$$
$$\le \sqrt{2}\left((E_2 + \ldots + E_{r+1})^{1/2} - (E_2 + \ldots + E_r)^{1/2} \right)$$
$$\le \sqrt{2}\frac{E_{r+1}}{2((1-\varepsilon)r)^{1/2}} < 1,$$
and likewise $\pi_{r-1} < 1$ on $A(\varepsilon, r)$.

For r large enough, $\mathbf{P}\left\{ E_r \ge r^{1/2} \right\} = e^{-r^{1/2}} \le 1/r^2$ and $\mathbf{P}\left\{ E_{r+1} \ge r^{1/2} \right\} \le 1/r^2$, and by Chebyshev's inequality,
$$\mathbf{P}\left\{ E_2 + \ldots + E_{r-1} \le (1-\varepsilon)r \right\} \le \frac{2}{(\varepsilon r)^2}$$
and
$$\mathbf{P}\left\{ E_2 + \ldots + E_r \ge (1+\varepsilon)r \right\} \le \frac{2}{(\varepsilon r)^2},$$
and hence
$$\mathbf{P}\left\{ A(\varepsilon, r)^c \right\} \le \frac{2}{r^2} + \frac{4}{(\varepsilon r)^2} \le \frac{6}{(\varepsilon r)^2}.$$
Moreover, $A(\varepsilon, r)$ is independent of E_1, so $\mathbf{P}\left\{ A(\varepsilon, r) \mid E_1 \right\} = \mathbf{P}\left\{ A(\varepsilon, r) \right\}$.

With the preceding bounds at hand, we have enough information to control (3.5) and (3.6). Writing $A = A(\varepsilon, r)$, we have

$$
\begin{aligned}
\mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\} &= \mathbf{E}\left\{R\mathbf{1}_{[\pi_r \leq 1]} \mid E_1\right\} \\
&\geq \mathbf{E}\left\{R\mathbf{1}_{[\pi_r \leq 1]}\mathbf{1}_{[A]} \mid E_1\right\} \\
&= \mathbf{E}\left\{R\mathbf{1}_{[A]} \mid E_1\right\} \\
&= \mathbf{E}\left\{R \mid A, E_1\right\}\mathbf{P}\left\{A \mid E_1\right\} \\
&\geq \frac{1}{3}\frac{E_1^{1/2}}{(E_1 + (1+\varepsilon)r)^{1/2}}\left(1 - \frac{6}{(\varepsilon r)^2}\right),
\end{aligned} \tag{3.7}
$$

and

$$
\begin{aligned}
\mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\} &\leq \mathbf{E}\left\{R\mathbf{1}_{[\pi_r \leq 1]} \mid A, E_1\right\} + \mathbf{P}\left\{A^c \mid E_1\right\} \\
&\leq \mathbf{E}\left\{R \mid A, E_1\right\} + \frac{6}{(\varepsilon r)^2} \\
&\leq \frac{1}{3}\frac{E_1^{1/2}}{((1-\varepsilon)r)^{1/2}} + \frac{6}{(\varepsilon r)^2}.
\end{aligned} \tag{3.8}
$$

We similarly have

$$
\begin{aligned}
\mathbf{P}\left\{r-1 \in \mathcal{U}_1, r \in \mathcal{U}_1 \mid E_1\right\} &= \mathbf{E}\left\{R'\mathbf{1}_{[\pi_{r-1}, \pi_r \leq 1]} \mid E_1\right\} \\
&\leq \mathbf{E}\left\{R'\mathbf{1}_{[\pi_{r-1}, \pi_r \leq 1]} \mid E_1, A\right\} + \mathbf{P}\left\{A^c \mid E_1\right\} \\
&\leq \mathbf{E}\left\{R' \mid A, E_1\right\} + \frac{6}{(\varepsilon r)^2} \\
&\leq \frac{1}{9}\frac{E_1}{(1-\varepsilon)r} + \frac{6}{(\varepsilon r)^2}.
\end{aligned} \tag{3.9}
$$

We use the above bounds in the following formulas, which are immediate consequences of Lemma 3.11 (using the fact that conditioning on π_0 and on E_1 is equivalent):

$$
\mathbf{E}\left\{|\mathcal{U}_1| \mid E_1\right\} = r\mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\} \tag{3.10}
$$

and

$$
\begin{aligned}
\mathbf{Var}&\left\{|\mathcal{U}_1|\mid E_1\right\} \\
&= r(r-1)\left(\mathbf{P}\left\{r-1, r \in \mathcal{U}_1 \mid E_1\right\} - \mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\}^2\right) \\
&\quad + r\left(\mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\} - \mathbf{P}\left\{r \in \mathcal{U}_1 \mid E_1\right\}^2\right).
\end{aligned} \tag{3.11}
$$

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

$$\frac{1}{3} \frac{E_1^{1/2}}{(E_1 + (1+\varepsilon)r)^{1/2}} \left(1 - \frac{6}{(\varepsilon r)^2}\right) \geq \frac{1}{3} \frac{E_1^{1/2}}{((1+2\varepsilon)r)^{1/2}},$$

$$\frac{1}{3} \frac{E_1^{1/2}}{((1-\varepsilon)r)^{1/2}} + \frac{6}{(\varepsilon r)^2} \leq \frac{1}{3} \frac{E_1^{1/2}}{((1-2\varepsilon)r)^{1/2}}, \text{ and}$$

$$\frac{1}{9} \frac{E_1}{(1-\varepsilon)r} + \frac{6}{(\varepsilon r)^2} \leq \frac{1}{9} \frac{E_1}{(1-2\varepsilon)r}.$$

Using the first of these bounds together with (3.7) in (3.10) gives that on the event $\{E_1 \in (\varepsilon, \varepsilon^{-1})\}$,

$$\mathbf{E}\{|\mathcal{U}_1| \mid E_1\} \geq \frac{r^{1/2}}{3} \frac{E_1^{1/2}}{(1+2\varepsilon)^{1/2}}, \qquad (3.12)$$

and using all three bounds together with (3.7),(3.8) and (3.9) in (3.11), we obtain that on the event $\{E_1 \in (\varepsilon, \varepsilon^{-1})\}$,

$$\mathbf{Var}\{|\mathcal{U}_1| \mid E_1\}$$

$$\leq r(r-1)\left(\frac{1}{9} \frac{E_1}{(1-2\varepsilon)r} - \frac{1}{9} \frac{E_1}{(1+2\varepsilon)r}\right) + r\left(\frac{1}{3} \frac{E_1^{1/2}}{((1-2\varepsilon)r)^{1/2}}\right)$$

$$= (r-1)\frac{E_1}{9} \frac{4\varepsilon}{1-4\varepsilon^2} + r^{1/2} \frac{E_1^{1/2}}{3} \frac{1}{(1-2\varepsilon)^{1/2}}$$

$$< 5\varepsilon \mathbf{E}\{|\mathcal{U}_1| \mid E_1\}^2,$$

the final bound holding by (3.12) for ε sufficiently small ($\varepsilon < 1/8$ is enough), and still provided that and $r \geq r_0$.

The lower bound in (3.12) is at least $(rE_1)^{1/2}/4$ provided ε is small enough, so it now follows by the conditional Chebyshev inequality that

$$\mathbf{P}\left\{|\mathcal{U}_1| \leq \frac{(rE_1)^{1/2}}{8}, E_1 \in (\varepsilon, \varepsilon^{-1}) \,\middle|\, E_1\right\}$$

$$\leq \frac{\mathbf{Var}\{|\mathcal{U}_1| \mid E_1\} \mathbf{1}_{[E_1 \in (\varepsilon, \varepsilon^{-1})]}}{(\mathbf{E}\{|\mathcal{U}_1| \mid E_1\}/2)^2}$$

$$\leq 20\varepsilon \mathbf{1}_{[E_1 \in (\varepsilon, \varepsilon^{-1})]}.$$

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

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

the last bound holding since $\mathbf{P}\left\{E_1 \notin \varepsilon, \varepsilon^{-1}\right\} < 2\varepsilon$ for ε small. This establishes (3.4) and completes the proof. $\qquad\square$

3.3.2. *Proof of Proposition 3.3.* Having already proved Proposition 3.4, to complete the proof of Proposition 3.3 it remains to handle the cases when $s > 0$. So fix $\varepsilon > 0$ 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 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

$$
\mathbf{P}\left\{X \stackrel{F_q}{\longleftrightarrow} Y\right\} < \varepsilon\,.
$$

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 3.4. We next turn to defining the necessary decomposition. The definitions of the next four paragraphs are illustrated in Figure 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 nonstandard). 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 $\text{core}(G_q)$. We then define the kernel $K(G_q)$ to be the multigraph obtained from G_q by contracting each maximal path or cycle of $\text{core}^+(G_q)$ 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

$$\mathbf{V}_q = \{V_q(a), a \in v(K(G_q)) \cup e(K(G_q))\} \tag{3.13}$$

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).
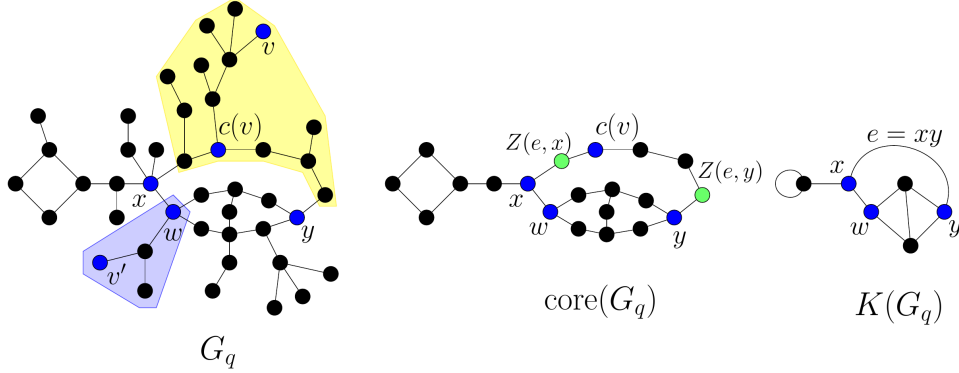

FIGURE 3. Left: An instantiation of graph G_q; here $q = 43$ and $s = 3$. Center: the graph $\text{core}(G_q)$. Right: the kernel $K(G_q)$. In the graph G_q, the vertex v has $\kappa(v) = e = xy$ since $c(v)$ lies on the path of $\text{core}(G_q)$ which is contracted to form e. The vertex v' has $\kappa(v') = w$ 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.

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 (3.13), 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 (3.13). 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$(\alpha_1, \ldots, \alpha_k)$ 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 3.12 ([3] Theorem 22, [4] 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 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

$$\mathbf{P}\left\{X \xleftrightarrow{F_q} Y\right\} = \mathbf{E}\left[\mathbf{P}\left\{E_q \mid F_q\right\}\right]$$

$$\geq \mathbf{E}\left[\frac{\max(|C| : C \text{ is a component of } F_q)^2}{q^2}\right]$$

$$\geq \frac{(\mathbf{E}\left[\max(|C| : C \text{ is a component of } F_q)\right])^2}{q^2},$$

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

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

such a value δ exists by Proposition 3.12. Then by Bayes' formula, for q sufficiently large,

$$\mathbf{P}\left\{\min(|V_q(e)|, e \in e(K(G_q))) < \delta q \mid B\right\} < \varepsilon/6 \tag{3.14}$$

and

$$\mathbf{P}\left\{\min(|V_q(e)|, e \in e(K(G_q))) < \delta q \mid C\right\} < \varepsilon/6. \tag{3.15}$$

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

$$\mathbf{P}\left\{X \xleftrightarrow{F_q} Y \mid \mathbf{V}_q, B\right\}$$

$$\leq \mathbf{P}\left\{X \xleftrightarrow{T_q(e) \cap F_q} Y \mid \mathbf{V}_q, B\right\} + \mathbf{P}\left\{X \xleftrightarrow{T_q(e) \cap F_q} Z(e, u) \mid \mathbf{V}_q, B\right\}$$

$$+ \mathbf{P}\left\{X \xleftrightarrow{T_q(e) \cap F_q} Z(e, v) \mid \mathbf{V}_q, B\right\}$$

$$= 3\mathbf{P}\left\{X \xleftrightarrow{T_q(e) \cap F_q} Y \mid \mathbf{V}_q, B\right\},$$

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

$$\mathbf{P}\left\{X \xleftrightarrow{F_q} Y \mid \mathbf{V}_q, B\right\} \leq 3\mathbf{P}\left\{X \xleftrightarrow{T_q(e) \cap F_q} Y \mid \mathbf{V}_q, B\right\}$$

$$\leq 3\mathbf{P}\left\{X \xleftrightarrow{F_q(e)'} Y \mid \mathbf{V}_q, B\right\}. \tag{3.16}$$

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 $\mathbf{P}\left\{X \xleftrightarrow{G} Y\right\} \leq \frac{c}{q}$. If G is instead random, then this bound and the tower law give that

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

It thus follows from Proposition 3.4 there is q_0 such that if $|V_q(e)| \geq q_0$ then the conditional probability on the right of (3.16) is less than $\varepsilon/12$, so we have

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

which together with (3.14) yields that for q large enough (and in particular large enough that $\delta q > q_0$),

$$\mathbf{P}\left\{X \xleftrightarrow{F_q} Y \mid B\right\} = \mathbf{E}\left[\mathbf{P}\left\{X \xleftrightarrow{F_q} Y \mid \mathbf{V}_q, B\right\} \mid B\right]$$
$$(\varepsilon/4)\mathbf{P}\left\{V_q(\kappa(X)) \geq q_0 \mid B\right\} + 3\mathbf{P}\left\{V_q(\kappa(X)) < q_0 \mid B\right\}$$
$$< 3\varepsilon/4.$$

A nearly identical proof, but using (3.15) in place of (3.14), shows that $\mathbf{P}\left\{X \xleftrightarrow{F_q} Y \mid C\right\} < 3\varepsilon/4$ 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 $X \xleftrightarrow{T_q(e)\cap F_q} Y$ does not appear.) Since if A occurs then either B or C must occur, it follows that

$$\mathbf{P}\left\{X \xleftrightarrow{F_q} Y\right\} \leq \mathbf{P}\left\{X \xleftrightarrow{F_q} Y, B\right\} + \mathbf{P}\left\{X \xleftrightarrow{F_q} Y, C\right\} + \mathbf{P}\left\{A^c\right\}$$
$$\leq \frac{3\varepsilon}{4} + \mathbf{P}\left\{A\right\} < \varepsilon$$

the last two inequalities holding for all q sufficiently large. This completes the proof of Proposition 3.3 in the case $s > 0$.

4. CONCLUSION

In addition to the conjectures raised directly after the statement of Theorem 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 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 [19]) 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 [6]. 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.

5. Acknowledgements

---

[1]We follow the terminological suggestion of [8], using the term "Bienaymé trees" rather than "Galton-Watson trees" for the family trees of branching processes.

References

[1] Louigi Addario-Berry. The local weak limit of the minimum spanning tree of the complete graph. arXiv:1301.1667 [math.PR], January 2013.

[2] 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. ISSN 0178-8051. doi: 10.1007/s00440-021-01071-3. URL https://doi.org/10.1007/s00440-021-01071-3.

[3] 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). ISSN 0178-8051. doi: 10.1007/s00440-010-0325-4. URL http://dx.doi.org.dianus.libr.tue.nl/10.1007/s00440-010-0325-4.

[4] 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. doi: 10.1214/EJP.v15-772. URL https://doi.org/10.1214/EJP.v15-772.

[5] 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. ISSN 1050-5164. doi: 10.1214/11-AAP761. URL https://doi.org/10.1214/11-AAP761.

[6] Louigi Addario-Berry, Nicolas Broutin, and Cecilia Holmgren. Cutting down trees with a Markov chainsaw. *Ann. Appl. Probab.*, 24(6):2297–2339, 2014. ISSN 1050-5164. doi: 10.1214/13-AAP978. URL https://doi.org/10.1214/13-AAP978.

[7] 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. ISSN 0091-1798. doi: 10.1214/16-AOP1132. URL https://doi.org/10.1214/16-AOP1132.

[8] 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.

[9] David Aldous. The continuum random tree. I. *Ann. Probab.*, 19(1):1–28, 1991. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(199101)19:1<1:TCRTI>2.0.CO;2-B&origin=MSN.

[10] David Aldous. Brownian excursions, critical random graphs and the multiplicative coalescent. *Ann. Probab.*, 25(2):812–854, 1997. ISSN 0091-1798. doi: 10.1214/aop/1024404421. URL https://doi.org/10.1214/aop/1024404421.

[11] 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. doi: 10.1007/978-3-662-09444-0_1. URL https://doi.org/10.1007/978-3-662-09444-0_1.

[12] David J. 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. doi: 10.1007/BFb0099421. URL https://doi.org/10.1007/BFb0099421.

[13] Omer Angel, Jesse Goodman, Frank den Hollander, and Gordon Slade. Invasion percolation on regular trees. *Ann. Probab.*, 36(2):420–466, 2008. ISSN 0091-1798. doi: 10.1214/07-AOP346. URL https://doi.org/10.1214/07-AOP346.

[14] 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. ISSN 0091-1798. doi: 10.1214/11-AOP731. URL https://doi.org/10.1214/11-AOP731.

[15] 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. ISSN 1042-9832. doi: 10.1002/rsa.20880. URL https://doi.org/10.1002/rsa.20880.

[16] 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. ISSN 0178-8051. doi: 10.1007/s00440-017-0760-6. URL https://doi.org/10.1007/s00440-017-0760-6.

[17] 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. doi: 10.1017/S0022112082001335.

[18] Michael Damron and Artëm Sapozhnikov. Limit theorems for 2D invasion percolation. *Ann. Probab.*, 40(3):893–920, 2012. ISSN 0091-1798. doi: 10.1214/10-AOP641. URL https://doi.org/10.1214/10-AOP641.

[19] 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. ISSN 1435-9855. doi: 10.4171/JEMS/786. URL https://doi.org/10.4171/JEMS/786.

[20] 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. ISSN 0091-1798. doi: 10.1214/17-AOP1252. URL https://doi.org/10.1214/17-AOP1252.

[21] 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. ISSN 0002-9939. doi: 10.2307/2033241. URL https://doi.org/10.2307/2033241.

[22] Tomasz Ł uczak. Component behavior near the critical point of the random graph process. *Random Structures Algorithms*, 1(3):287–310, 1990. ISSN 1042-9832. doi: 10.1002/rsa.3240010305. URL https://doi.org/10.1002/rsa.3240010305.

[23] Adam Logan, Mike Molloy, and Pawel Pralat. A variant of the Erdos-Renyi random graph process. arXiv:1806.10975 [math.CO], 2018.

[24] 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. ISSN 1042-9832.

[25] Marcus Michelen, Robin Pemantle, and Josh Rosenberg. Invasion percolation on Galton-Watson trees. *Electron. J. Probab.*, 24:Paper No. 31, 35, 2019. doi: 10.1214/19-EJP281. URL `https://doi.org/10.1214/19-EJP281`.

[26] 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. ISSN 0246-0203. URL `http://www.numdam.org/item?id=AIHPB_1995__31_1_249_0`.

[27] 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. ISSN 0022-4715. doi: 10.1007/BF02179805. URL `https://doi.org/10.1007/BF02179805`.

[28] 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. ISSN 0031-9007. doi: 10.1103/PhysRevLett.51.71. URL `https://doi.org/10.1103/PhysRevLett.51.71`.

[29] R. C. Prim. Shortest connection networks and some generalizations. *The Bell System Technical Journal*, 36(6):1389–1401, 1957. doi: 10.1002/j.1538-7305.1957.tb01515.x.

[30] Colin P. Stark. An invasion percolation model of drainage network evolution. *Nature*, 352(6334):423–425, 1991. doi: 10.1038/352423a0. URL `https://doi.org/10.1038/352423a0`.

[31] 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. ISSN 1083-589X. doi: 10.1214/ECP.v12-1327. URL `https://doi.org/10.1214/ECP.v12-1327`.

[32] David Wilkinson and Jorge F. Willemsen. Invasion percolation: a new form of percolation theory. *J. Phys. A*, 16(14):3365–3376, 1983. ISSN 0305-4470. URL `http://stacks.iop.org/0305-4470/16/3365`.

[33] Yu Zhang. The fractal volume of the two-dimensional invasion percolation cluster. *Comm. Math. Phys.*, 167(2):237–254, 1995. ISSN 0010-3616. URL `http://projecteuclid.org/euclid.cmp/1104271992`.

DEPARTMENT OF MATHEMATICS AND STATISTICS, McGILL UNIVERSITY, MONTRÉAL, CANADA

Email address: `louigi.addario@mcgill.ca`

URL: `http://problab.ca/louigi/`

Email address: `jordan.barrett@mail.mcgill.ca`

URL: `https://www.math.mcgill.ca/jbarrett/`