SUCCESSIVE MINIMUM SPANNING TREES

SVANTE JANSON AND GREGORY B. SORKIN

ABSTRACT. In a complete graph K_n with independent uniform$(0, 1)$ (or exponential(1)) edge weights, let T_1 be the MST (minimum-weight spanning tree), and T_k the MST after deleting the edges of all previous trees. We show that each tree's weight $w(T_k)$ converges in probability to a constant γ_k, with $2k - 2\sqrt{k} < \gamma_k < 2k + 2\sqrt{k}$, and we conjecture that $\gamma_k = 2k - 1 + o(1)$. The problem is distinct from Frieze and Johansson's minimum combined weight μ_k of k edge-disjoint spanning trees; indeed, $\mu_2 < \gamma_1 + \gamma_2$.

With an edge of weight w "arriving" at time $t = nw$, Kruskal's algorithm defines forests $F_k(t)$, initially empty and eventually equal to T_k, each edge added to the first possible $F_k(t)$. Using tools of inhomogeneous random graphs we obtain structural results including that the fraction of vertices in the largest component of $F_k(t)$ converges to some $\rho_k(t)$. We conjecture that the functions ρ_k tend to time translations of a single function.

1. INTRODUCTION

1.1. **Problem definition and main results.** Consider the complete graph K_n with edge costs that are i.i.d. random variables, with a uniform distribution $U(0, 1)$ or, alternatively, an exponential distribution Exp(1). A well-known problem is to find the minimum (cost) spanning tree T_1, and its cost or "weight" $w(T_1)$. A famous result by Frieze [11] shows that as $n \to \infty$, $w(T_1)$ converges in probability to $\zeta(3)$, in both the uniform and exponential cases.

Suppose now that we want a second spanning tree T_2, edge-disjoint from the first, and that we do this in a greedy fashion by first finding the minimum spanning tree T_1, and then the minimum spanning tree T_2 using only the remaining edges. (I.e., T_2 is the minimum spanning tree in $K_n \setminus T_1$, meaning the graph with edge set $E(K_n) \setminus E(T_1)$.) We then continue and define T_3 as the minimum spanning tree in $K_n \setminus (T_1 \cup T_2)$, and so on. The main purpose of the present paper is to show that the costs $w(T_2)$, $w(T_3)$, ... also converge in probability to some constants.

Theorem 1.1. *For each $k \geqslant 1$, there exists a constant γ_k such that, as $n \to \infty$, $w(T_k) \xrightarrow{\text{P}} \gamma_k$ (for both uniform and exponential cost distributions).*

The result extends easily to other distributions of the edge costs, see Remark 7.1, but we consider in this paper only the uniform and exponential cases.

A minor technical problem is that T_2 and subsequent trees do not always exist; it may happen that T_1 is a star and then $K_n \setminus T_1$ is disconnected. This happens only with a small probability, and w.h.p. (with high probability, i.e., with probability $1 - o(1)$ as $n \to \infty$) T_k is defined for every fixed k, see Section 7. However, in the main part of the paper we avoid this problem completely by modifying the model: we assume that we have a multigraph, which we denote by K_n^∞, with an infinite number of copies of each edge in K_n, and that each edge's copies' costs are given by the points in a Poisson process with intensity 1 on $[0, \infty)$. (The Poisson processes for different edges are, of course, independent.) Note that when finding T_1, we only care about the cheapest copy of each edge, and its cost has an $\mathrm{Exp}(1)$ distribution, so the problem for T_1 is the same as the original one. However, on K_n^∞ we never run out of edges and we can define T_k for all integers $k = 1, 2, 3, \ldots$. Asymptotically, the three models are equivalent, as shown in Section 7, and Theorem 1.1 holds for any of the models. In particular:

Theorem 1.2. *For each $k \geqslant 1$, as $n \to \infty$, $w(T_k) \xrightarrow{\mathrm{p}} \gamma_k$ also for the multigraph model with Poisson process costs.*

Frieze [11] also proved that the expectation $\mathbb{E}\, w(T_1)$ converges to $\zeta(3)$. For the multigraph model just described, this too extends.

Theorem 1.3. *For the Poisson multigraph model, $\mathbb{E}\, w(T_k) \to \gamma_k$ for each $k \geqslant 1$ as $n \to \infty$.*

Remark 1.4. For the simple graph K_n with, say, exponential costs, as explained above there is a small but positive probability that T_k does not exist for $k \geqslant 2$. Hence, either $\mathbb{E}\, w(T_k)$ is undefined for $k \geqslant 2$, or we define $w(T_k) = \infty$ when T_k does not exist, and then $\mathbb{E}\, w(T_k) = \infty$ for $k \geqslant 2$ and every n. This is no problem for the convergence in probability in Theorem 1.1, but it implies that Theorem 1.3 does not hold for simple graphs, and the multigraph model is essential for studying the expectation. Alternatively, as pointed out by a referee, for the simple graph model one might consider the conditional expectation. We leave this as an open question. $\qquad \square$

Conjecture 1.5. *In the simple graph model with uniformly or exponentially distributed edge weights, $\mathbb{E}\big[w(T_k) \mid T_k \ \text{exists}\big] \to \gamma_k$.*

1.2. **Motivations.** Frieze and Johansson [12] recently considered a related problem, where instead of choosing spanning trees $T_1, T_2, \ldots$ greedily one by one, they choose k edge-disjoint spanning trees with minimum total cost. It is easy to see, by small examples, that selecting k spanning trees greedily one by one does not always give a set of k edge-disjoint spanning trees with minimum cost, so the problems are different.

We show in Theorem 9.3 that, at least for $k = 2$, the two problems also asymptotically have different answers, in the sense that the limiting values of the minimum cost — which exist for both problems — are different. (Also, as discussed in Section 3.1, we improve on the upper bound from [12,

Section 3] on the cost of the net cheapest k trees, since our upper bound (3.1) on the cost of the first k trees is smaller.)

Both our question and that of Frieze and Johansson [12] are natural, both seem generally relevant to questions of robust network design, and both have mathematically interesting answers.

Another reason for interest in T_2 comes from the field of algorithmic mechanism design. Imagine that each edge of $G = K_n$ is owned by a different "agent"; the agent owning edge e values it at $w(e)$, an amount known only to them. We, an "auctioneer", want to buy a spanning tree, at low cost. One "mechanism" for doing so is a sealed-bid auction where each agent posts a price $w'(e)$ for their edge, and we buy the tree that is cheapest according to these prices. Here, agents will naturally inflate their prices, posting prices $w'(e) > w(e)$.

One alternative is a VCG (Vickrey–Clarke–Groves) auction, a generalization of a single-item second-price auction. Here, we again buy the tree that is cheapest according to the posted prices w', but for each edge e purchased, we pay an amount that is a function of w'_{-e}, i.e., of all posted prices *except* that of e; for details see for example [28, Chapter 9]. This means that varying $w'(e)$ affects only whether edge e is purchased, not how much is paid for it if it is, and results in the mechanism being *truthful*: it is in each agent's selfish interest to set $w'(e) = w(e)$. Thus, the tree purchased is simply T_1, the tree cheapest according to the values w. However, the amount paid for it is more than $w(T_1)$, as the mechanism ensures the amount paid for each edge e purchased is at least $w(e)$ and typically more. A central question is the extent of this overpayment, measured by the "frugality ratio" of the VCG cost V (or that of any mechanism) to some benchmark.

The question applies of course to problems other than MSTs, including the purchase of a cheapest path between two given points in a graph, or of a basis in a bridgeless matroid. In any of these contexts, let us continue to use T_1 for the cheapest structure and T_2 for the cheapest structure disjoint from T_1. The cost $w(T_1)$ is not a useful benchmark because $V/w(T_1)$ is unbounded in even the simplest examples (such as buying one of two identical items).

Instead, Talwar [31] and Archer and Tardos [1] propose $w(T_2)$ as the benchmark. (An often-equivalent benchmark, based on a Nash equilibrium, is given by [22] and [28, Chapter 13].) [31] shows that for any bridgeless matroid, $V/w(T_2) \leqslant 1$, and, focusing on the worst case over all weights w, this bound is achieved by some weights (namely weights 0 on T_1, 1 on T_2, and infinity elsewhere). By contrast, for paths the ratio is unbounded. The interpretation, based on worst-case weights, is that this frugality ratio is 1 for amenable problems like MSTs and other matroids, and larger for other problems.

In our setting of an MST in K_n with random weights, though, the frugality ratio is naturally less than its maximum of 1. Specifically, [8] and [20] show that the VCG cost is typically $2w(T_1)$, which by [11] is $2\zeta(3) \doteq 2.4041$. We show here that $w(T_2)$ is typically γ_2, which by Remark 9.5 is at least 2.9683, making the frugality typically at most 0.8099. (We estimate non-rigorously that γ_2 is about 3.0921 — see Table 1 — in which case the frugality ratio is typically about 0.7775.) Specifically, this holds w.h.p. for n large, and also

holds for the ratio between the expected VCG cost and the expected cost $w(T_2)$.

1.3. **Further results, structural properties, and conjectures.** It is well known that the minimum spanning tree (with any given costs, obtained randomly or deterministically) can be found by *Kruskal's algorithm* [23], which processes the edges in order of increasing cost and keeps those that join two different components in the forest obtained so far. (I.e., it keeps each edge that does not form a cycle together with previously chosen edges.) As in many other previous papers on the random minimum spanning tree problem, from [11] on, our proofs are based on analyzing the behavior of this algorithm.

Rescale weight as time, thinking of an edge of weight w as arriving at time $t = nw$. Kruskal's algorithm allows us to construct all trees T_k simultaneously by growing forests $F_k(t)$, with $F_k(0)$ empty and $F_k(\infty) = T_k$: taking the edges of K_n (or K_n^∞) in order of time arrival (increasing cost), an edge is added to the first forest F_k where it does not create a cycle. We will also consider a sequence of graphs $G_k(t) \supseteq F_k(t)$, where when we add an edge to F_k we also add it to all the graphs $G_1, \ldots, G_k$; see Section 2.2 for details.

The proof of Theorem 1.1 is based on a detailed structural characterization of the graphs $G_k(t)$, given by Theorem 2.1 (too detailed to set forth in full here in the introduction), relying heavily on the theory of inhomogeneous random graphs from [5] and related works. Where $C_1(G_k(t))$ denotes the number of vertices in the largest component of $G_k(t)$ (or equivalently of $F_k(t)$, as by construction they have the same components), Theorem 2.1 shows that $C_1(G_k(t))/n$ converges in probability to some function $\rho_k(t)$, uniformly for all times t. Moreover, each G_k has its own giant-component threshold: $\rho_k(t)$ is 0 until some time σ_k, and strictly positive thereafter.

The functions $\rho_k(t)$ are of central interest. For one thing, an edge is rejected from F_k, making it a candidate for F_{k+1}, precisely if its two endpoints are within the same component of F_k, and it is shown (see Corollary 5.9) that this is essentially equivalent to the two endpoints both being within the largest component. This line of reasoning yields the constants γ_k explicitly, see (6.23), albeit not in a form that is easily evaluated. We are able, at least, to re-prove (in Example 6.5) that $\gamma_1 = \zeta(3)$, as first shown in [11].

The functions ρ_k also appear to have a beautiful structure, tending to time-translated copies of a single universal function:

Conjecture 1.6. *There exists a continuous increasing function $\rho_\infty(x)$: $(-\infty, \infty) \to [0, 1)$ such that $\rho_k(2k + x) \to \rho_\infty(x)$ as $k \to \infty$, uniformly in $x \in \mathbb{R}$.*

This suggests, though does not immediately imply, another conjecture.

Conjecture 1.7. *For some δ, as $k \to \infty$, $\gamma_k = 2k + \delta + o(1)$.*

If this conjecture holds, then necessarily $\delta \in [-1, 0]$, see Remark 3.6. A variety of computational results are given in Section 11. They are supportive of Conjecture 1.6 (see Figures 3 and 6) and (see Table 1) a stronger version of Conjecture 1.7 where we take $\delta = -1$:

Conjecture 1.8. *As $k \to \infty$, $\gamma_k = 2k - 1 + o(1)$.*

Although we cannot prove these conjectures, some bounds on γ_k are obtained in Section 3 by a more elementary analysis of the sequence of forests F_k. In particular, Theorem 3.1 and Corollary 3.2 lead to the following, implying that $\gamma_k \sim 2k$ as $k \to \infty$.

Corollary 1.9. *For every $k \geqslant 1$,*

$$2k - 2k^{1/2} < \gamma_k < 2k + 2k^{1/2}. \tag{1.1}$$

See also the related Conjectures 3.5, 10.1 and 11.1.

Remark 1.10. For the minimum spanning tree T_1, various further results are known, including refined estimates for the expectation of the cost $w(T_1)$ [9], a normal limit law [16], and asymptotics for the variance [16; 21; 32]. It seems challenging to show corresponding results for T_2 or later trees. $\qquad\square$

2. Model and main structural results

2.1. **Some notation.** We use := as defining its left-hand side, and $\overset{\text{def}}{=}$ as a reminder that equality of the two sides is by definition. We write $\doteq$ for numerical approximate equality, and $\approx$ for approximate equality in an asymptotic sense (details given where used).

If x and y are real numbers, then $x \vee y := \max(x, y)$ and $x \wedge y := \min(x, y)$. Furthermore, $x_+ := x \vee 0$. These operators bind most strongly, e.g., $t - \tau(i) \vee \tau(j)$ means $t - (\tau(i) \vee \tau(j))$.

We use "increasing" and "decreasing" in their weak senses; for example, a function f is increasing if $f(x) \leqslant f(y)$ whenever $x \leqslant y$.

Unspecified limits are as $n \to \infty$. As said above, w.h.p. means with probability $1 - o(1)$. Convergence in probability is denoted $\overset{\text{p}}{\longrightarrow}$. Furthermore, if X_n are random variables and a_n are positive constants, $X_n = o_{\text{p}}(a_n)$ means, as usual, $X_n/a_n \overset{\text{p}}{\longrightarrow} 0$; this is also equivalent to: for every $\varepsilon > 0$, w.h.p. $|X_n| < \varepsilon a_n$.

Graph means, in general, multigraph. (It is usually clear from the context whether we consider a multigraph or simple graph.) If G is a multigraph, then $\dot{G}$ denotes the simple graph obtained by merging parallel edges and deleting loops. (Loops do not appear in the present paper.) The number of vertices in a graph G is denoted by $|G|$, and the number of edges by $e(G)$.

For a graph G, let $\mathcal{C}_1(G)$, $\mathcal{C}_2(G)$, ... be the largest component, the second largest component, and so on, using any rule to break ties. (If there are less than k components, we define $\mathcal{C}_k(G) = \emptyset$.) Furthermore, let $C_i(G) := |\mathcal{C}_i(G)|$; thus $C_1(G)$ is the the number of vertices in the largest component, and so on. We generally regard components of a graph G as sets of vertices.

2.2. **Model.** We elaborate the multigraph model in the introduction.

We consider (random) (multi)graphs on the vertex set $[n] := \{1, \ldots, n\}$; we usually omit n from the notation. The graphs will depend on time, and are denoted by $G_k(t)$ and $F_k(t)$, where $k = 1, 2, 3, \ldots$ and $t \in [0, \infty]$; they all start as empty at time $t = 0$ and grow as time increases. We will have $G_k(t) \supseteq G_{k+1}(t)$ and $F_k(t) \subseteq G_k(t)$ for all k and t. Furthermore, $F_k(t)$ will be a forest. As $t \to \infty$, $F_k(t)$ will eventually become a spanning tree, $F_k(\infty)$,

which is the kth spanning tree T_k produced by the greedy algorithm in the introduction, operating on the multigraph $G_1(\infty)$.

Since the vertex set is fixed, we may when convenient identify the multigraphs with sets of edges. We begin by defining $G_1(t)$ by letting edges arrive as independent Poisson processes with rate $1/n$ for each pair $\{i, j\}$ of vertices; $G_1(t)$ consists of all edges that have arrived at or before time t. (This scaling of time turns out to be natural and useful. In essence this is because what is relevant is the cheapest edges on each vertex, and these have expected cost $\Theta(1/n)$ and thus appear at expected time $\Theta(1)$.) We define the cost of an edge arriving at time t to be t/n, and note that in $G_1(\infty)$, the costs of the edges joining two vertices form a Poisson process with rate 1. Hence, $G_1(\infty)$ is the multigraph model defined in Section 1.

Thus, for any fixed $t \geqslant 0$, $G_1(t)$ is a multigraph where the number of edges between any two fixed vertices is $\mathrm{Po}(t/n)$, and these numbers are independent for different pairs of vertices. This is a natural multigraph version of the Erdős–Rényi graph $G(n, t)$. (The process $G_1(t)$, $t \geqslant 0$, is a continuous-time version of the multigraph process in e.g. [4] and [17, Section 1], ignoring loops.) Note that $\dot{G}_1(t)$, i.e., $G_1(t)$ with multiple edges merged, is simply the random graph $G(n, p)$ with $p = 1 - e^{-t/n}$.

Next, we let $F_1(t)$ be the subgraph of $G_1(t)$ consisting of every edge that has arrived at some time $s \leqslant t$ and at that time joined two different components of $G_1(s)$. Thus, this is a subforest of $G_1(t)$, as stated above, and it is precisely the forest constructed by Kruskal's algorithm (recalled in the introduction) operating on $G_1(\infty)$, at the time all edges with cost $\leqslant t/n$ have been considered. Hence, $F_1(\infty)$ is the minimum spanning tree T_1 of $G_1(\infty)$.

Let $G_2(t) := G_1(t) \setminus F_1(t)$, i.e., the subgraph of $G_1(t)$ consisting of all edges rejected from $F_1(t)$; in other words $G_2(t)$ consists of the edges that, when they arrive to $G_1(t)$, have their endpoints in the same component.

We continue recursively. $F_k(t)$ is the subforest of $G_k(t)$ consisting of all edges in $G_k(t)$ that, when they arrived at some time $s \leqslant t$, joined two different components in $G_k(s)$. And $G_{k+1}(t) := G_k(t) \setminus F_k(t)$, consisting of the edges rejected from $F_k(t)$.

Hence, the kth spanning tree T_k produced by Kruskal's algorithm equals $F_k(\infty)$, as asserted above.

Note that $F_k(t)$ is a spanning subforest of $G_k(t)$, in other words, the components of $F_k(t)$ (regarded as vertex sets) are the same as the components of $G_k(t)$; this will be used frequently below. Moreover, each edge in $G_{k+1}(t)$ has endpoints in the same component of $G_k(t)$; hence, each component of $G_{k+1}(t)$ is a subset of a component of $G_k(t)$. It follows that an edge arriving to $G_1(t)$ will be passed through $G_2(t), \ldots, G_k(t)$ and to $G_{k+1}(t)$ (and possibly further) if and only if its endpoints belong to the same component of $G_k(t)$, and thus if and only if its endpoints belong to the same component of $F_k(t)$.

2.3. **More notation.** We say that a component $\mathcal{C}$ of a graph G is the *unique giant* of G if $|\mathcal{C}| > |\mathcal{C}'|$ for every other component $\mathcal{C}'$; if there is no such component (i.e., if the maximum size is tied), then we define the unique giant to be $\emptyset$.

We say that a component $\mathcal{C}$ of $F_k(t)$ is the *permanent giant* of $F_k(t)$ (or of $G_k(t)$) if it is the unique giant of $F_k(t)$ and, furthermore, it is a subset of the unique giant of $F_k(u)$ for every $u > t$; if there is no such component then the permanent giant is defined to be $\emptyset$.

Let $\mathfrak{C}_k(t)$ denote the permanent giant of $F_k(t)$. Note that the permanent giant either is empty or the largest component; thus $|\mathfrak{C}_k(t)|$ is either 0 or $C_1(F_k(t)) = C_1(G_k(t))$. Note also that the permanent giant $\mathfrak{C}_k(t)$ is an increasing function of t: $\mathfrak{C}_k(t) \subseteq \mathfrak{C}_k(u)$ if $t \leqslant u$. Furthermore, for sufficiently large t (viz. t such that $G_k(t)$ is connected, and thus $F_k(t)$ is the spanning tree T_k), $\mathfrak{C}_k(t) = \mathfrak{C}_k(\infty) = [n]$.

2.4. **A structure theorem.** The basis of our proof of Theorems 1.1 and 1.2 is the following theorem on the structure of the components of $G_k(t)$. Recall that $F_k(t)$ has the same components as $G_k(t)$, so the theorem applies as well to $F_k(t)$. The proof is given in Section 5.

For $k = 1$, the theorem collects various known results for $G(n, p)$. Our proof includes this case too, making the proof more self-contained.

Theorem 2.1. *With the definitions above, the following hold for every fixed $k \geqslant 1$ as $n \to \infty$.*

(i) *There exists a continuous increasing function $\rho_k : [0, \infty) \to [0, 1)$ such that*
$$C_1(G_k(t))/n \xrightarrow{\text{p}} \rho_k(t), \tag{2.1}$$
uniformly in $t \in [0, \infty)$; in other words, for any $\varepsilon > 0$, w.h.p., for all $t \geqslant 0$,
$$\rho_k(t) - \varepsilon \leqslant C_1(G_k(t))/n \leqslant \rho_k(t) + \varepsilon. \tag{2.2}$$

(ii) $\sup_{t \geqslant 0} C_2(G_k(t))/n \xrightarrow{\text{p}} 0.$

(iii) *There exists a threshold $\sigma_k > 0$ such that $\rho_k(t) = 0$ for $t \leqslant \sigma_k$, but $\rho_k(t) > 0$ for $t > \sigma_k$. Furthermore, ρ_k is strictly increasing on $[\sigma_k, \infty)$.*

(iv) *There exist constants $b_k, B_k > 0$ such that*
$$\rho_k(t) \geqslant 1 - B_k e^{-b_k t}, \qquad t \geqslant 0. \tag{2.3}$$
In particular, $\rho_k(t) \to 1$ as $t \to \infty$.

(v) *If $t > \sigma_k$, then w.h.p. $G_k(t)$ has a non-empty permanent giant. Hence, for every $t \geqslant 0$,*
$$|\mathfrak{C}_k(t)|/n \xrightarrow{\text{p}} \rho_k(t). \tag{2.4}$$

We note also a formula for the number of edges in $G_k(t)$, and two simple inequalities relating different k.

Theorem 2.2. *For each fixed $k \geqslant 1$ and uniformly for t in any finite interval $[0, T]$,*
$$e(G_k(t))/n \xrightarrow{\text{p}} \frac{1}{2} \int_0^t \rho_{k-1}(s)^2 \, ds. \tag{2.5}$$

Theorem 2.3. $\rho_k(t) \leqslant \rho_{k-1}(t)$ *for every $t \geqslant 0$, with strict inequality when $\rho_{k-1}(t) > 0$ (equivalently, when $t > \sigma_{k-1}$). Furthermore,*
$$\sigma_k \geqslant \sigma_{k-1} + 1. \tag{2.6}$$

Inequality (2.6) is weak in that we conjecture that as $k \to \infty$, $\sigma_k = \sigma_{k-1} + 2 + o(1)$, see Conjecture 10.1.

The limits γ_k in Theorems 1.1–1.3 will be related to the functions $\rho_k(t)$ in (6.23).

3. Bounds on the expected cost

3.1. Total cost of the first k trees.
The following theorem gives lower and upper bounds on the total cost of the first k spanning trees.

Theorem 3.1. *Letting $W_k = \sum_{i=1}^{k} w(T_k)$ be the total cost of the first k spanning trees, for every $k \geqslant 1$,*

$$k^2 \frac{n-1}{n} \leqslant \mathbb{E}\, W_k \leqslant k(k+1)\frac{n-1}{n} < k^2 + k. \tag{3.1}$$

Comparing with Frieze and Johansson [12, Section 3], our upper bound is smaller than their $k^2 + 3k^{5/3}$ despite the fact that they considered a more relaxed minimization problem (see Section 9); as such ours is a strict improvement. In both cases the lower bound is simply the expected total cost of the cheapest $k(n-1)$ edges in G, with (3.2) matching [12, (3.1)].

Proof. The minimum possible cost of the k spanning trees is the cost of the cheapest $k(n-1)$ edges. Since each edge's costs (plural, in our model) are given by a Poisson process of rate 1, the set of all edge costs is given by a Poisson process of rate $\binom{n}{2}$. Recall that in a Poisson process of rate λ, the interarrival times are independent exponential random variables with mean $1/\lambda$, so that the ith arrival, at time Z_i, has $\mathbb{E}\, Z_i = i/\lambda$. It follows in this case that $W_k \geqslant \sum_{i=1}^{k(n-1)} Z_i$ and

$$\mathbb{E}\, W_k \geqslant \sum_{i=1}^{k(n-1)} \frac{i}{\binom{n}{2}} = \frac{(k(n-1))(k(n-1)+1)}{n(n-1)} \geqslant k^2 \frac{n-1}{n}. \tag{3.2}$$

We now prove the upper bound. An arriving edge is rejected from F_i iff both endpoints lie within its "forbidden" set B_i of edges, namely those edges with both endpoints in one component. The nesting property of the components means that $B_1 \supseteq B_2 \supseteq \cdots$. An arriving edge e joins F_k if it is rejected from all previous forests, i.e., $e \in B_{k-1}$ (in which case by the nesting property, e also belongs to all earlier Bs) but can be accepted into F_k, i.e., $e \notin B_k$. The idea of the proof is to show that the first k forests fill reasonably quickly with $n-1$ edges each, and we will do this by coupling the forest-creation process (Kruskal's algorithm) to a simpler, easily analyzable random process.

For an integer $\tau = 0, 1, \ldots$, consider a time just after arrival of the τ'th edge. Let $B_k(\tau)$ denote the forbidden set B_k at this time, let $p_k(\tau) := |B_k(\tau)|/\binom{n}{2}$ denote the rejection probability for F_k, and let $\mathbf{s}(\tau) = \{s_k(\tau)\}_{k=0}^{\infty}$ denote the vector of the sizes (number of edges) of each forest; we may drop the argument τ when convenient.

For any τ, by the nesting property of the components and in turn of the B_k,

$$s_1 \geqslant s_2 \geqslant \cdots \qquad \text{and} \qquad p_1 \geqslant p_2 \geqslant \cdots . \tag{3.3}$$

The MST process can be simulated by using a sequence of i.i.d. random variables $\alpha(\tau) \sim U(0, 1)$, incrementing $s_k(\tau)$ if both $\alpha(\tau) \leqslant p_{k-1}(\tau)$ (so that e is rejected from F_{k-1} and thus from all previous forests too) and $\alpha(\tau) > p_k(\tau)$ (so that e is accepted into F_k). We take the convention that $p_0(\tau) = 1$ for all τ. For intuition, note that when $s_k = 0$ an edge is never rejected from F_k ($p_k = 0$, so $\alpha \sim U(0, 1)$ is never smaller); when $s_k = 1$ it is rejected with probability $p_k = 1/\binom{n}{2}$; and when $s_k = n - 1$ it is always rejected ($|B_k|$ must be $\binom{n}{2}$, so $p_k = 1$).

Given the size $s_k = \sum_{i=1}^{\infty}(C_i(F_k) - 1)$ of the kth forest, $|B_k| = \sum_{i=1}^{\infty}\binom{C_i(F_k)}{2}$ is maximized (thus so is p_k) when all the edges are in one component, i.e.,

$$p_k \leqslant \binom{s_k + 1}{2} \bigg/ \binom{n}{2} \qquad (3.4)$$

$$\leqslant \frac{s_k}{n - 1} =: \bar{p}_k. \qquad (3.5)$$

The size vector $\mathbf{s}(\tau)$ thus determines the values $\bar{p}_k(\tau)$ for all k.

Let $\mathbf{r}(\tau)$ denote a vector analogous to $\mathbf{s}(\tau)$, but with $r_k(\tau)$ incremented if $\hat{p}_k(\tau) < \alpha(\tau) \leqslant \hat{p}_{k-1}(\tau)$, with

$$\hat{p}_k := \frac{r_k}{n - 1}. \qquad (3.6)$$

By construction,

$$r_1 \geqslant r_2 \geqslant \cdots \qquad \text{and} \qquad \hat{p}_1 \geqslant \hat{p}_2 \geqslant \cdots. \qquad (3.7)$$

For intuition, here note that when $r_k = 0$ an arrival is never rejected from r_k ($\bar{p}_k = 0$); when $s_k = 1$ it is rejected with probability $\bar{p}_k = 1/(n-1) > p_k = 1/\binom{n}{2}$; and when $s_k = n - 1$ it is always rejected ($\bar{p}_k = 1$).

Taking each $F_i(0)$ to be an empty forest (n isolated vertices, no edges) and accordingly $\mathbf{s}(0)$ to be an infinite-dimensional 0 vector, and taking $\mathbf{r}(0)$ to be the same 0 vector, we claim that for all τ, $\mathbf{s}(\tau)$ majorizes $\mathbf{r}(\tau)$, which we will write as $\mathbf{s}(\tau) \succeq \mathbf{r}(\tau)$. That is, the prefix sums of $\mathbf{s}$ dominate those of $\mathbf{r}$: for all τ and k, $\sum_{i=1}^{k} s_i(\tau) \geqslant \sum_{i=1}^{k} r_i(\tau)$.

We first prove this; then use it to argue that edge arrivals to the first k forests, i.e., to $\mathbf{s}$, can only precede arrivals to the first k elements of $\mathbf{r}$; and finally analyze the arrival times of all $k(n-1)$ elements to the latter to arrive at an upper bound on the total cost of the first k trees.

We prove $\mathbf{s}(\tau) \succeq \mathbf{r}(\tau)$ by induction on τ, the base case with $\tau = 0$ being trivial. Figure 1 may be helpful in illustrating the structure of this inductive proof. Suppose the claim holds for τ. The probabilities $p_k(\tau)$ are used to determine the increments of the size vector $\mathbf{s}(\tau + 1)$. Consider an intermediate object $\mathbf{s}'(\tau + 1)$, the size vector that would be given by incrementing $\mathbf{s}(\tau)$ using the upper-bound values $\bar{p}_k(\tau)$ taken from $\mathbf{s}(\tau)$ by (3.5). Then, $s_i(\tau + 1)$ receives the increment if $p_{i-1} \geqslant \alpha > p_i$, and $s'_j(\tau + 1)$ receives the increment if $\bar{p}_{j-1} \geqslant \alpha > \bar{p}_j$; hence, from $\bar{p}_{i-1} \geqslant p_{i-1} \geqslant \alpha$ it is immediate that $i \leqslant j$ and thus $\mathbf{s}(\tau + 1) \succeq \mathbf{s}'(\tau + 1)$.

It suffices then to show that $\mathbf{s}'(\tau + 1) \succeq \mathbf{r}(\tau + 1)$. These two vectors are obtained respectively from $\mathbf{s}(\tau)$ and $\mathbf{r}(\tau)$, with $\mathbf{s}(\tau) \succeq \mathbf{r}(\tau)$ by the inductive hypothesis, using probability thresholds $\bar{p}_k(\tau) = f(s_k(\tau))$ and $\hat{p}_k(\tau) = f(r_k(\tau))$ respectively, applied to the common random variable α,

$$\begin{array}{ccc}
\mathbf{F}(\tau) & \xrightarrow{\ p\ } & \mathbf{F}(\tau+1) \\
\downarrow & & \downarrow \\
\mathbf{s}(\tau) & \xrightarrow{\ p\ } & \mathbf{s}(\tau+1) \\
 & \xrightarrow{\bar{p}} & \\
\vdots\ \succeq & & \vdots\ \succeq \\
 & & \mathbf{s}'(\tau+1) \\
 & & \vdots\ \succeq \\
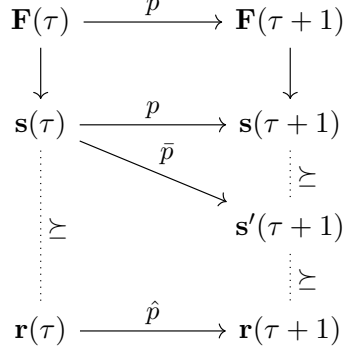\mathbf{r}(\tau) & \xrightarrow{\ \hat{p}\ } & \mathbf{r}(\tau+1)
\end{array}$$

FIGURE 1. Coupling of the forests' sizes $\mathbf{s}(\tau)$ to a simply analyzable random process $\mathbf{r}(\tau)$, showing the structure of the inductive proof (on τ) that $\mathbf{s}(\tau)$ majorizes $\mathbf{r}(\tau)$.

where $f(s) = s/(n-1)$ (but all that is important is that f is a monotone function of s). Suppose that

$$f(s_{i-1}) \geqslant \alpha > f(s_i) \qquad \text{and} \qquad f(r_{j-1}) \geqslant \alpha > f(r_j), \qquad (3.8)$$

so that elements i in $\mathbf{s}$ and j in $\mathbf{r}$ are incremented. If $i \leqslant j$, we are done. (Prefix sums of $\mathbf{s}(\tau)$ dominated those of $\mathbf{r}(\tau)$, and an earlier element is incremented in $\mathbf{s}'(\tau+1)$ than $\mathbf{r}(\tau+1)$, thus prefix sums of $\mathbf{s}'(\tau+1)$ dominate those of $\mathbf{r}(\tau+1)$.) Consider then the case that $i > j$. In both processes the increment falls between indices j and i, so the k-prefix sum inequality continues to hold for $k < j$ and $k \geqslant i$. Thus, for $j \leqslant k < i$,

$$\begin{aligned}
\sum_{\ell=1}^{k} s'_\ell(\tau+1) &= \sum_{\ell=1}^{j-1} s_\ell(\tau) + \sum_{\ell=j}^{k} s_\ell(\tau) \\
\sum_{\ell=1}^{k} r_\ell(\tau+1) &= \sum_{\ell=1}^{j-1} r_\ell(\tau) + 1 + \sum_{\ell=j}^{k} r_\ell(\tau).
\end{aligned} \qquad (3.9)$$

From $j < i$, (3.8), and (3.3) and (3.7) we have that when $j \leqslant \ell \leqslant i-1$,

$$s_\ell \geqslant s_{i-1} \geqslant f^{-1}(\alpha) > r_j \geqslant r_\ell,$$

implying

$$s_\ell \geqslant r_\ell + 1. \qquad (3.10)$$

In (3.9), we have $\sum_{\ell=1}^{i-1} s_\ell(\tau) \geqslant \sum_{\ell=1}^{i-1} r_\ell(\tau)$ from the inductive hypothesis that $\mathbf{s}(\tau) \succeq \mathbf{r}(\tau)$, while using (3.10) gives

$$\sum_{\ell=j}^{k} s_\ell(\tau) \geqslant \sum_{\ell=j}^{k} (1 + r_\ell(\tau)) \geqslant 1 + \sum_{\ell=j}^{k} r_\ell(\tau),$$

from which it follows that $\mathbf{s}'(\tau+1) \succeq \mathbf{r}(\tau+1)$, completing the inductive proof that $\mathbf{s}(\tau) \succeq \mathbf{r}(\tau)$.

Having shown that the vector $\mathbf{s}(\tau)$ of component sizes majorizes $\mathbf{r}(\tau)$, it suffices to analyze the latter. Until this point we could have used (3.4) rather

than (3.5) to define $\bar{p}_k$, $\hat{p}_k$, and the function f, but now we take advantage of the particularly simple nature of the process governing $\mathbf{r}(\tau)$. Recall that a new edge increments r_i for the first i for which the $U(0,1)$ "coin toss" $\alpha(\tau)$ has $\alpha(\tau) > \hat{p}_i \overset{\text{def}}{=} r_i/(n-1)$. Equivalently, consider an array of cells $n-1$ rows high and infinitely many columns wide, generate an "arrival" at a random row or "height" $X(\tau)$ uniform on $1, \ldots, n-1$, and let this arrival occupy the first unoccupied cell i at this height, thus incrementing the occupancy r_i of column i. This is equivalent because if r_i of the $n-1$ cells in column i are occupied, the chance that i is rejected — that $X(\tau)$ falls into this set and thus the arrival moves along to test the next column $i+1$ — is $r_i/(n-1)$, matching (3.6).

Recalling that the cost of an edge arriving at time t is t/n in the original graph problem, the combined cost W_k of the first k spanning trees is $1/n$ times the sum of the arrival times of their $k(n-1)$ edges. The majorization $\sum_{i=1}^{k} s_i(\tau) \geqslant \sum_{i=1}^{k} r_i(\tau)$ means that the ℓ'th arrival to the first k forests comes no later than the ℓ'th arrival to the first k columns of the cell array. Thus, the cost W_k of the first k trees is at most $1/n$ times the sum of the times of the $k(n-1)$ arrivals to the array's first k columns.

The continuous-time edge arrivals are a Poisson process with intensity $1/n$ on each of the $\binom{n}{2}$ edges, thus intensity $(n-1)/2$ in all; it is at the Poisson arrival times that the discrete time τ is incremented and $X(\tau)$ is generated. Subdivide the "X" process into the $n-1$ possible values that X may take on, so that arrivals at each value (row in the cell array) are a Poisson process of intensity $\lambda = \frac{1}{2}$. The sum of the first k arrival times in a row is the sum of the first k arrival times in its Poisson process. The ith such arrival time is the sum of i exponential random variables, and has expectation i/λ. The expected sum of k arrival times of a line is thus $\binom{k+1}{2}/\lambda = k(k+1)$, and (remembering that cost is time divided by n), the expected total cost of all $n-1$ lines is

$$\frac{n-1}{n} k(k+1),$$

yielding the upper bound in (3.1) and completing the proof of the theorem. $\square$

Corollary 3.2. *Let* $\Gamma_k := \sum_{i=1}^{k} \gamma_i$. *Then, for every* $k \geqslant 1$,

$$k^2 \leqslant \Gamma_k = \sum_{i=1}^{k} \gamma_i \leqslant k^2 + k. \qquad (3.11)$$

Proof. Immediate from Theorems 3.1 and 1.3. $\square$

Example 3.3. In particular, Corollary 3.2 gives $1 \leqslant \gamma_1 \leqslant 2$ and $4 \leqslant \gamma_1 + \gamma_2 \leqslant 6$. In fact, we know that $\gamma_1 = \zeta(3) \doteq 1.2021$ [11] and $\gamma_1 + \gamma_2 > 4.1704$ by [12] and Section 9, see Corollary 9.4. A numerical estimate of γ_2 from Table 1 suggests $\gamma_1 + \gamma_2 \doteq 4.2942$. $\square$

3.2. **Corollaries and conjectures for the kth tree.** Turning to individual γ_k instead of their sum Γ_k, we obtain Corollary 1.9, namely that $2k - 2k^{1/2} < \gamma_k < 2k + 2k^{1/2}$.

Proof of Corollary 1.9. For the upper bound, we note that obviously $\gamma_1 \leqslant \gamma_2 \leqslant \ldots$, and thus, for any $\ell \geqslant 1$, using both the upper and lower bound in (3.11),

$$\ell\,\gamma_k \leqslant \sum_{i=k}^{k+\ell-1} \gamma_i = \Gamma_{k+\ell-1} - \Gamma_{k-1} \leqslant (k+\ell-1)(k+\ell) - (k-1)^2$$
$$= \ell^2 + \ell(2k-1) + k - 1 \tag{3.12}$$

and hence

$$\gamma_k \leqslant 2k - 1 + \ell + \frac{k-1}{\ell}. \tag{3.13}$$

Choosing $\ell = \lceil \sqrt{k} \rceil$ gives the upper bound in (1.1).

For the lower bound we similarly have, for $1 \leqslant \ell \leqslant k$,

$$\ell\gamma_k \geqslant \Gamma_k - \Gamma_{k-\ell} \geqslant k^2 - (k-\ell)(k-\ell+1) = -\ell^2 - (2k+1)\ell - k \tag{3.14}$$

and hence

$$\gamma_k \geqslant 2k + 1 - \ell - \frac{k}{\ell}. \tag{3.15}$$

Choosing, again, $\ell = \lceil \sqrt{k} \rceil$ gives the lower bound in (1.1). $\qquad\square$

Remark 3.4. For a specific k, we can improve (1.1) somewhat by instead using (3.13) and (3.15) with $\ell = \lfloor \sqrt{k} \rfloor$ or $\ell = \lceil \sqrt{k} \rceil$. For example, for $k = 2$, taking $\ell = 1$ yields $2 \leqslant \gamma_2 \leqslant 5$. For $k = 3$, taking $\ell = 2$ yields $3.5 \leqslant \gamma_3 \leqslant 8$. $\qquad\square$

Besides these rigorous results, taking increments of the left and right-hand sides of (3.11) also suggests the following conjecture.

Conjecture 3.5. *For $k \geqslant 1$, $2k - 1 \leqslant \gamma_k \leqslant 2k$.*

Remark 3.6. Moreover, if $\gamma_k = 2k + \delta + o(1)$ holds, as conjectured in Conjecture 1.7, then $\Gamma_k = k^2 + k(\delta + 1) + o(k)$, and thus necessarily $\delta \in [-1, 0]$ as a consequence of Corollary 3.2. In fact, the numerical estimates in Section 11, specifically Table 1, suggest that $\delta = -1$; see Conjecture 1.8. $\qquad\square$

3.3. **Improved upper bounds.** The upper bounds in Theorem 3.1 and Corollary 3.2 were proved using the bound (3.5). A stronger, but less explicit, bound can be proved by using instead the sharper (3.4). That is, we consider the random vectors $\mathbf{r}(\tau)$ defined as above but with (3.6) replaced by

$$\hat{p}_k := \binom{r_k + 1}{2} \Big/ \binom{n}{2}. \tag{3.16}$$

As remarked before (3.4), this approximation comes from imagining all edges in each F_k to be in a single component; this overestimates the probability that an arriving edge is rejected from F_k and, as developed in the previous subsection, gives $\mathbf{s}(\tau) \succeq \mathbf{r}(\tau)$ just as when $\hat{p}_k$ was defined by (3.5).

Using for consistency our usual time scaling in which edges arrive at rate $(n-1)/2$, by a standard martingale argument one can show that, for each $k \geqslant 1$,

$$\frac{1}{n} r_k(\lfloor \tfrac{1}{2} nt \rfloor) \xrightarrow{\;\mathrm{p}\;} g_k(t), \qquad \text{uniformly for } t \geqslant 0, \tag{3.17}$$

for some continuously differentiable functions $g_k(t)$ satisfying the differential equations, with $g_0(t) := 1$,

$$g_k'(t) = \tfrac{1}{2}\left(g_{k-1}(t)^2 - g_k(t)^2\right), \qquad g_k(0) = 0, \qquad k \geqslant 1. \tag{3.18}$$

Moreover, using $\mathbf{s}(\tau) \succeq \mathbf{r}(\tau)$ and taking limits, it can be shown that

$$\Gamma_k := \sum_{i=1}^{k} \gamma_i \leqslant \frac{1}{2} \int_0^\infty t\left(1 - g_k(t)^2\right) \mathrm{d}t. \tag{3.19}$$

We omit the details, but roughly, in time dt, $\tfrac{1}{2} n\, dt$ edges arrive, all costing about t/n, and a $g_k(t)^2$ fraction of them pass beyond the first k graphs (to the degree that we are now modeling graphs). Compare (3.19) with (6.19), with reference to (6.3).

For $k = 1$, (3.18) has the solution $g_1(t) = \tanh(t/2)$, and (3.19) yields the bound $\Gamma_1 = \gamma_1 \leqslant 2 \ln 2 \doteq 1.3863$. This is better than the bound 2 given by (3.11), but still far from precise since $\gamma_1 = \zeta(3) \doteq 1.2021$.

For $k \geqslant 2$ we do not know any exact solution to (3.18), but numerical solution of (3.18) and calculation of (3.19) (see Section 11.3) suggests that $\Gamma_k < k^2 + 1$. We leave the proof of this as an open problem. If proved, this would be a marked improvement on $\Gamma_k \leqslant k^2 + k$, which was the exact expectation of the random process given by (3.5) (that part of the analysis was tight). In particular, it would establish that $2k - 2 \leqslant \gamma_k \leqslant 2k$; see Conjecture 11.1.

For $k = 2$, the numerical calculations in Section 11.3 give $\gamma_1 + \gamma_2 \leqslant 4.5542\ldots$ (see Table 2) and thus $\gamma_2 \leqslant 3.3521\ldots$. The same value was also obtained using Maple's numerical differential equation solver, with Maple giving greater precision but the two methods agreeing in the digits shown here.

4. Inhomogeneous random graph preliminaries

Here, we recall various constructs, mostly related to the theory of inhomogeneous random graphs in [5]. They will be used in the proofs that follow, in particular that of Theorem 2.1 in Section 5.

4.1. **Some random graphs.** For a symmetric array $(p_{ij})_{i,j=1}^n$ of probabilities in $[0, 1]$, let $G(n, (p_{ij}))$ be the random (simple) graph on the vertex set $[n] := \{1, \ldots, n\}$ where the edge ij appears with probability p_{ij}, for $i < j$, and these $\binom{n}{2}$ events are independent. We extend this (in a trivial way) by defining $G(n, A) = G(n, (a_{ij})) := G(n, (a_{ij} \wedge 1)_{i,j=1}^n)$ for any symmetric non-negative $n \times n$ matrix $A = (a_{ij})_{i,j=1}^n$. Moreover, the matrix A can be a random, in which case $G(n, A)$ is defined by first conditioning on A. (Hence, edges appear conditionally independently, given A.)

Note that we do not allow loops, so the diagonal entries p_{ii} or a_{ii} are ignored, and may be assumed to be 0 without loss of generality.

4.2. **Susceptibility.** The *susceptibility* $\chi(G)$ of a (deterministic or random) graph G of order $n = |G|$ is defined by

$$\chi(G) := \frac{1}{n} \sum_{i=1}^{\infty} C_i(G)^2. \tag{4.1}$$

$\chi(G)$ can be interpreted as the mean size of the component containing a random vertex, see [19]. We also exclude the first term in the sum and define

$$\widehat{\chi}(G) := \frac{1}{n} \sum_{i=2}^{\infty} C_i(G)^2. \tag{4.2}$$

(This is particularly interesting for a graph G with a single giant component of order $\Theta(n)$, when the sum in (4.1) is dominated by the first term.)

Viewing each term in the sums (4.1)–(4.2) as $(C_i(G)/n)\,C_i(G)$, since $\sum_{i=1}^{\infty} C_i(G)/n = 1$ each sum can be viewed as a weighted sum of the sizes $C_i(G)$ with the weights summing to at most 1, so

$$\chi(G) \leqslant C_1(G), \tag{4.3}$$

$$\widehat{\chi}(G) \leqslant C_2(G). \tag{4.4}$$

Let $\pi(G)$ be the probability that two randomly chosen distinct vertices in G belong to the same component. Then

$$\pi(G) = \frac{\sum_{i=1}^{\infty} C_i(G)\big(C_i(G) - 1\big)}{n(n-1)} = \frac{n\chi(G) - n}{n(n-1)} = \frac{\chi(G) - 1}{n - 1}. \tag{4.5}$$

4.3. **Kernels and an integral operator.** A *kernel*, or *graphon*, is a non-negative symmetric measurable function $\kappa : \mathcal{S}^2 \to [0, \infty)$, where (in a common abuse of notation) $\mathcal{S} = (\mathcal{S}, \mathcal{F}, \mu)$ is a probability space.

Given a kernel κ on a probability space $\mathcal{S}$, let T_κ be the integral operator defined by (for suitable functions f on $\mathcal{S}$),

$$\big(T_\kappa f\big)(x) = \int_{\mathcal{S}} \kappa(x, y) f(y)\,\mathrm{d}\mu(y), \tag{4.6}$$

and let Φ_κ be the non-linear operator

$$\Phi_\kappa f := 1 - e^{-T_\kappa f}. \tag{4.7}$$

In our cases, the kernel κ is bounded, and then T_κ is a compact (in fact, Hilbert–Schmidt) operator on $L^2(\mathcal{S}, \mu)$. Since furthermore $\kappa \geqslant 0$, it follows that there exists an eigenfunction $\psi \geqslant 0$ on $\mathcal{S}$ with eigenvalue $\|T_k\|$, see [5, Lemma 5.15], where $\|T_\kappa\|$ denotes the operator norm of T_k as an operator on $L^2(\mathcal{S}, \mu)$.

4.4. **Branching processes.** Given a kernel κ on a probability space $(\mathcal{S}, \mathcal{F}, \mu)$, as in [5] let $\mathfrak{X}_\kappa(x)$ be the multi-type Galton–Watson branching process with type space $\mathcal{S}$, starting with a single particle of type $x \in \mathcal{S}$, and where in each generation a particle of type y is replaced by its children, consisting of a set of particles distributed as a Poisson process on $\mathcal{S}$ with intensity $\kappa(y, z)\,\mathrm{d}\mu(z)$. Let further $\mathfrak{X}_\kappa$ be the same branching process started with a particle of random type, distributed as μ.

Let

$$\rho_\kappa(x) \quad \text{and} \quad \rho(\kappa) = \rho(\kappa; \mu) \tag{4.8}$$

be the probabilities that $\mathfrak{X}_\kappa(x)$ and $\mathfrak{X}_\kappa$, respectively, survive for ever. Thus,

$$\rho(\kappa) = \int_\mathcal{S} \rho_\kappa(x)\,\mathrm{d}\mu(x). \tag{4.9}$$

By [5, Theorem 6.1], assuming e.g. that κ is bounded as in our cases (much less will do), the following hold.

(i) The function ρ_κ is a fixed point of Φ_κ, i.e., it satisfies the equation

$$\rho_\kappa = \Phi_\kappa \rho_\kappa \overset{\text{def}}{=} 1 - e^{-T_\kappa \rho_\kappa}. \tag{4.10}$$

Moreover, ρ_κ is the largest non-negative solution of this equation.

(ii) If $\|T_\kappa\| \leqslant 1$, then $\rho_\kappa(x) = 0$ for every x, and thus $\rho(\kappa) = 0$.

(iii) If $\|T_\kappa\| > 1$, then $\rho(\kappa) > 0$.

5. Proof of Theorem 2.1

5.1. **Basics, and an approximating inhomogeneous random graph.**
The proof of Theorem 2.1 is based on induction; we assume throughout this section that, for $k \geqslant 1$, Theorem 2.1 holds for $k - 1$ and show that it holds for k.

For convenience, we define $F_0(t) := G_0(t) := K_n$ for every $t \geqslant 0$; this enables us to consider $G_1(t)$ together with $G_k(t)$ for $k > 1$. (Alternatively, we could refer to known results for the random graph process $G_1(t)$.) Note that Theorem 2.1 then trivially holds for $k = 0$, with $\rho_0(t) = 1$ for all t and $\sigma_0 = 0$, except that (iii) has to be modified (since ρ_0 is constant). There are some trivial modifications below in the case $k = 1$ (and also some, more or less important, simplifications); we leave these to the reader.

Thus, fix $k \geqslant 1$, assume that Theorem 2.1 holds for $k - 1$ and consider the evolution of $G_k(t)$. Essentially everything in this proof depends on k, but we often omit it from the notation. (Recall that we also usually omit n.)

We condition on the entire process $(F_{k-1}(s))_{s \geqslant 0}$. For two distinct vertices $i, j \in [n]$, let $\tau(i, j) = \tau_{k-1}(i, j)$ be the time that i and j become members of the same component in $F_{k-1}(t)$. This is the time when edges ij start to be passed to $G_k(t)$, and it follows that, conditionally on $(F_{k-1}(s))_{s \geqslant 0}$, the process $G_k(t)$, $t \geqslant 0$, can be described as $G_1(t)$ above (Section 2), except that for each pair $\{i, j\}$ of vertices, edges appear according to a Poisson process on $(\tau(i, j), \infty)$. In particular, for a fixed time t (a value, independent of n) and conditioned on $(F_{k-1}(s))_{s \geqslant 0}$, in the multigraph $G_k(t)$, the number of edges ij is $\mathrm{Po}\big((t - \tau(i, j))_+/n\big)$, and these numbers are (conditionally) independent for different pairs $\{i, j\}$. Hence, if we merge multiple edges and obtain the simple graph $\dot{G}_k(t)$, we see that

$$\dot{G}_k(t) = G(n, (p_{ij})), \tag{5.1}$$

the random graph defined in Section 4.1 with

$$p_{ij} = p_{ij}(t) := 1 - e^{-(t-\tau(i,j))_+/n} = \frac{(t - \tau(i,j))_+}{n} + O\big(n^{-2}\big) \tag{5.2}$$

when $i \neq j$, and (for completeness) $p_{ii} = 0$. Note that the probabilities p_{ij} depend on $(F_{k-1}(s))_{s \geqslant 0}$ and thus are random, and recall that therefore $G(n, (p_{ij}))$ is defined by first conditioning on $(p_{ij})_{i,j}$ (or on $(F_{k-1}(s))_{s \geqslant 0}$).

Define

$$\tau(i) = \tau_{k-1}(i) := \inf\{t \geqslant 0 : i \in \mathfrak{C}_{k-1}(t)\}, \qquad i = 1, \ldots, n, \qquad (5.3)$$

i.e., the first time that vertex i belongs to the permanent giant of F_{k-1}. Note that

$$\tau(i, j) \leqslant \tau(i) \vee \tau(j), \qquad (5.4)$$

but strict inequality is possible since i and j may both belong to a component of $F_{k-1}(t)$ that is not the permanent giant. We shall see that this does not happen very often, and one of the ideas in the proof is that we may regard the inequality (5.4) as an approximate equality. This is formalized in the following lemma, and leads to a more tractable graph defined in (5.22) and compared with $\dot{G}_k(t)$ in Lemma 5.3.

Lemma 5.1. *For any fixed $t > 0$,*

$$\sum_{i \neq j} \big((t - \tau(i, j))_+ - (t - \tau(i) \vee \tau(j))_+ \big) = o_{\mathrm{p}}(n^2). \qquad (5.5)$$

Proof. Fix $\varepsilon > 0$, let $L := \lceil t/\varepsilon \rceil$ and let $t_\ell := \sigma_{k-1} + \ell\varepsilon$, $\ell = 1, \ldots, L$.
Say that the pair (i, j) is *bad* if

$$\tau(i, j) \leqslant t \qquad \text{and} \qquad \tau(i) \vee \tau(j) - \tau(i, j) > 2\varepsilon. \qquad (5.6)$$

Note that, using (5.4), for any pair (i, j),

$$0 \leqslant (t - \tau(i, j))_+ - (t - \tau(i) \vee \tau(j))_+ \leqslant t \qquad (5.7)$$

and for a good pair (i.e., a pair that is not bad),

$$(t - \tau(i, j))_+ - (t - \tau(i) \vee \tau(j))_+ \leqslant 2\varepsilon. \qquad (5.8)$$

By the induction hypothesis Theorem 2.1(v), w.h.p. $G_{k-1}(t_1)$ has a permanent giant, so we may assume that this holds. (Failures contribute $o_{\mathrm{p}}(n^2)$ to the right-hand side of (5.5).)

If (i, j) is bad, then either $\tau(i, j) \leqslant \sigma_{k-1} - \varepsilon$, or there exists $\ell \in [1, L]$ such that $\tau(i, j) \leqslant t_\ell < \tau(i) \vee \tau(j)$. In the first case, i and j belong to the same component in $G_{k-1}(\sigma_{k-1} - \varepsilon)$, and in the second case they belong to the same component in $G_{k-1}(t_\ell)$, but not to the largest one, since that is assumed to be the permanent giant. Hence, for any t, the number of bad pairs (i, j) is at most, using the definitions (4.1)–(4.2),

$$n\chi\big(G_{k-1}(\sigma_{k-1} - \varepsilon)\big) + \sum_{\ell=1}^{L} n\widehat{\chi}\big(G_{k-1}(t_\ell)\big). \qquad (5.9)$$

By (4.3) and the induction hypothesis (i) and (iii),

$$\chi\big(G_{k-1}(\sigma_{k-1} - \varepsilon)\big) \leqslant C_1\big(G_{k-1}(\sigma_{k-1} - \varepsilon)\big) = n\rho_{k-1}(\sigma_{k-1} - \varepsilon) + o_{\mathrm{p}}(n) = o_{\mathrm{p}}(n) \qquad (5.10)$$

and similarly for every ℓ, by (4.4) and the induction hypothesis (ii),

$$\widehat{\chi}\big(G_{k-1}(t_\ell)\big) \leqslant C_2\big(G_{k-1}(t_\ell)\big) = o_{\mathrm{p}}(n). \qquad (5.11)$$

By (5.9)–(5.11), the number of bad pairs is $o_{\mathrm{p}}(n^2)$. Hence, using (5.7) and (5.8), we obtain

$$\sum_{i \neq j} \big((t - \tau(i, j))_+ - (t - \tau(i) \vee \tau(j))_+ \big) \leqslant 2\varepsilon n^2 + t o_{\mathrm{p}}(n^2), \qquad (5.12)$$

and the result follows since ε is arbitrary. □

We use the machinery and notation in Bollobás, Janson and Riordan [5, in particular Section 2], and make the following definitions:

- $\mathcal{S}$ is the set
$$\mathcal{S} := [0, \infty). \tag{5.13}$$

- μ_{k-1} is the probability measure on $\mathcal{S}$ with distribution function
$$\mu_{k-1}([0, x]) = \rho_{k-1}(x). \tag{5.14}$$

- $\mathbf{x}_n := (x_1, \ldots, x_n)$ where
$$x_i = \tau_{k-1}(i). \tag{5.15}$$

- ν_n is a probability measure given by
$$\nu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}, \tag{5.16}$$

  where δ_x is the point mass (Dirac delta) at x. (In other words, ν_n is the empirical distribution of $\{x_1, \ldots, x_n\}$. Put yet another way, for any set $A \subset \mathcal{S}$, $\nu_n(A) := \frac{1}{n} |\{i \colon x_i \in A\}|$.)

Note that $\mathbf{x}_n$ and ν_n are random, and determined by $(F_{k-1}(s))_{s \geqslant 0}$.

Lemma 5.2. $\nu_n \xrightarrow{\mathrm{p}} \mu_{k-1}$ in the space $\mathcal{P}(\mathcal{S})$ of probability measures on $\mathcal{S}$.

Proof. The claim is equivalent to
$$\nu_n[0, x] \xrightarrow{\mathrm{p}} \mu_{k-1}[0, x] \tag{5.17}$$

for every continuity point x of μ_{k-1}, see e.g. [5, Lemma A.2 and Remark A.3] or [2, Section 3]. (In our case, for $k > 1$, μ_{k-1} is a continuous measure, by (5.14) and (i), so we should consider all x.) However, (5.17) follows from (5.3) and the induction hypothesis (v), which yield, for any $x \geqslant 0$,

$$\nu_n[0, x] = \frac{1}{n} |\{i : \tau_{k-1}(i) \leqslant x\}| = \frac{1}{n} |\mathfrak{C}_{k-1}(x)| \xrightarrow{\mathrm{p}} \rho_{k-1}(x) = \mu_{k-1}[0, x]. \tag{5.18}$$

□

In the terminology of [5, Section 2], $(\mathcal{S}, \mu_{k-1})$ is a *ground space* and, by Lemma 5.2,
$$\mathcal{V} := (\mathcal{S}, \mu_{k-1}, (\mathbf{x}_n)_{n \geqslant 1}) \tag{5.19}$$
is a *vertex space*, meaning that the number of vertices x_i appearing by time t is governed by μ_{k-1}, as made precise by (5.18). We define also, for every $t \geqslant 0$, the *kernel*

$$\kappa_t(x, y) := (t - x \vee y)_+ = (t - x)_+ \wedge (t - y)_+, \qquad x, y \in \mathcal{S} = [0, \infty). \tag{5.20}$$

Note that, for fixed t, the kernel κ_t is bounded and continuous; hence κ_t is a *graphical kernel* [5, Definition 2.7, Remark 2.8 and Lemma 8.1]. Furthermore, κ_t is strictly positive, and thus irreducible, on $[0, t) \times [0, t)$, and 0 on the complement $\mathcal{S}^2 \setminus [0, t)^2$; hence, κ_t is *quasi-irreducible* [5, Definition 2.11], provided $t > \sigma_{k-1}$ so $\rho_{k-1}(t) > 0$. (If $t \leqslant \sigma_{k-1}$, then $\mu_{k-1}[0, t] = \rho_{k-1}(t) = 0$, and thus $\kappa_t = 0$ μ_{k-1}^2-a.e. on $\mathcal{S}^2$.)

As detailed in [5, Section 2], specifically near its (2.3), these ingredients define a random graph

$$G^{\mathcal{V}}(n, \kappa_t). \qquad (5.21)$$

Recall that in our case the kernel κ_t is given by (5.20) while the vertex space $\mathcal{V}$ is given by (5.19), in turn with $\mathcal{S}$ and μ_{k-1} given by (5.13) and (5.14), and $\mathbf{x}_n$ given by (5.15).

In general, $G^{\mathcal{V}}(n, \kappa_t)$ denotes a random graph with vertices arriving at random times $\mathbf{x}_n$, vertices i and j joined with probability $\kappa_t(x_i, x_j)/n$, independently for all pairs $\{i, j\}$, and [5] describes the behavior of such an inhomogeneous random graph. It suffices to think of $G^{\mathcal{V}}(n, \kappa_t)$ in terms of $\mathcal{S}$, μ_{k-1}, and κ_t, because as shown in [5] the particulars of $\mathbf{x}_n$ are irrelevant as long as $\mathbf{x}_n$ is consistent with μ_{k-1} in the sense of (5.18) and (5.16), this consistency following from the fact that $\mathcal{V}$ is a vertex space (see (5.19) and the line following it).

Here, $G^{\mathcal{V}}(n, \kappa_t)$ is the random graph alluded to after (5.4), a proxy for $G_k(t)$ with the difference that it is based on the times $\tau(i)$ of vertices joining the permanent giant of F_{k-1}, rather than the more complicated two-variable times $\tau(i, j)$ of two vertices first belonging to a common component. Concretely,

$$G^{\mathcal{V}}(n, \kappa_t) \stackrel{\text{def}}{=} G(n, (p_{ij}^-)), \qquad (5.22)$$

the right-hand side being the random graph defined in Section 4.1 with (recalling (5.20) and (5.15))

$$p_{ij}^- := \frac{1}{n}\kappa_t(x_i, x_j) = \frac{1}{n}\big(t - x_i \vee x_j\big)_+ = \frac{1}{n}\big(t - \tau(i) \vee \tau(j)\big)_+ \qquad (5.23)$$

when $i \neq j$, and (for completeness) $p_{ii}^- = 0$. We assume throughout that $n \geqslant t$, so that $p_{ij}^- \in [0, 1]$; this is not an issue since t is fixed while $n \to \infty$. Note that by (5.2) and (5.5),

$$\sum_{i,j} |p_{ij} - p_{ij}^-| = \sum_{i \neq j} \frac{(t - \tau(i, j))_+ - (t - \tau(i) \vee \tau(j))_+}{n} + O(1) = o_{\mathrm{p}}(n). \qquad (5.24)$$

Recall that both p_{ij} and p_{ij}^- depend on $(F_{k-1}(s))_{s \geqslant 0}$, and thus are random. By (5.1) and (5.22), $\dot{G}_k(t) = G(n, (p_{ij}))$ and $G^{\mathcal{V}}(n, \kappa_t) = G(n, (p_{ij}^-))$, so by making the obvious maximal coupling of $G(n, (p_{ij}))$ and $G(n, (p_{ij}^-))$ conditionally on $(F_{k-1}(s))_{s \geqslant 0}$, we obtain a coupling of $\dot{G}_k(t)$ and $G^{\mathcal{V}}(n, \kappa_t)$ such that if $e\big(\dot{G}_k(t) \triangle G^{\mathcal{V}}(n, \kappa_t)\big)$ is the number of edges that are present in one of the graphs but not in the other, then

$$\mathbb{E}\big(e\big(\dot{G}_k(t) \triangle G^{\mathcal{V}}(n, \kappa_t)\big) \mid (F_{k-1}(s))_{s \geqslant 0}\big) = \sum_{i < j} |p_{ij} - p_{ij}^-|. \qquad (5.25)$$

Lemma 5.3. *For every fixed $t \geqslant 0$,*

$$e\big(\dot{G}_k(t) \triangle G^{\mathcal{V}}(n, \kappa_t)\big) = o_{\mathrm{p}}(n). \qquad (5.26)$$

Proof. Let $\varepsilon > 0$. For convenience, let $X_n := e\big(\dot{G}_k(t) \triangle G^{\mathcal{V}}(n, \kappa_t)\big)$ and $Y_n := \mathbb{E}\big(X_n \mid (F_{k-1}(s))_{s \geqslant 0}\big)$. Then, using Markov's inequality, for any

$\delta > 0$,

$$\mathbb{P}\big(X_n > \varepsilon n\big) \leqslant \mathbb{P}(Y_n > \delta n) + \mathbb{E}\,\mathbb{P}(X_n > \varepsilon n \mid Y_n \leqslant \delta n) \leqslant \mathbb{P}(Y_n > \delta n) + \frac{\delta n}{\varepsilon n}.$$
(5.27)

But $Y_n = o_{\mathrm{p}}(n)$ by (5.25) and (5.24), so $\mathbb{P}(Y_n > \delta n) = o(1)$. Hence $\mathbb{P}(X_n > \varepsilon n) \leqslant \delta/\varepsilon + o(1)$. Since δ is arbitrary, this shows $\mathbb{P}(X_n > \varepsilon n) = o(1)$, i.e., $0 \leqslant X_n \leqslant \varepsilon n$ w.h.p., which completes the proof. $\square$

5.2. **Towards part (i) of Theorem 2.1.** The following lemma establishes (2.1) of Theorem 2.1(i) for any fixed $t \geqslant 0$; doing so uniformly for all $t \geqslant 0$, as the theorem states, follows later. Here we rely on [5, Theorem 3.1], which, roughly speaking, relates the size of the largest component of a random graph $G^{\mathcal{V}}(n, \kappa_t)$, to the survival probability of the branching process defined by the same kernel κ_t and the measure (here μ_{k-1}) comprised by the vertex space $\mathcal{V}$. By Lemma 5.3, the graph $\dot{G}_k(t)$ of interest differs from $G^{\mathcal{V}}(n, \kappa_t)$ in only $o_{\mathrm{p}}(n)$ edges, and the stability theorem [5, Theorem 3.9] shows that the size of the largest component of $\dot{G}_k(t)$ is about the same as that of $G^{\mathcal{V}}(n, \kappa_t)$.

Let $\mathfrak{X}_t = \mathfrak{X}_{t,k} := \mathfrak{X}_{\kappa_t}$ be the branching process defined in Section 4.4 for the kernel κ_t and the measure μ_{k-1}, and (recalling (4.8)) let $\rho(\kappa_t) \overset{\text{def}}{=} \rho(\kappa_t; \mu_{k-1})$ be its survival probability.

Lemma 5.4. *For every fixed $t \geqslant 0$, (2.1) holds with*

$$\rho_k(t) := \rho(\kappa_t; \mu_{k-1}) \overset{\text{def}}{=} \rho(\kappa_t), \tag{5.28}$$

the survival probability of the branching process $\mathfrak{X}_t$.

Do not confuse ρ_k with ρ_κ, respectively the ρ and ρ_κ of (4.8).

Proof. Fix $t \geqslant 0$. First, if $t \leqslant \sigma_{k-1}$, then, by the induction hypothesis, $C_1(G_{k-1}(t))/n \overset{\mathrm{p}}{\longrightarrow} \rho_{k-1}(t) = 0$. Since each component of $G_k(t)$ is a subset of a component of $G_{k-1}(t)$, we have $C_1(G_k(t)) \leqslant C_1(G_{k-1}(t))$, and thus also $C_1(G_k(t))/n \overset{\mathrm{p}}{\longrightarrow} 0$, which shows (2.1) with $\rho_k(t) = 0$. The survival probability $\rho(\kappa_t) = 0$ here as well, establishing (5.28), as indeed no particle has any children: recalling the definitions in Section 4.4, the total number of children of a particle of any type y is $\mathrm{Po}\left(\int_0^\infty \kappa_t(y, z)\,\mathrm{d}\mu_{k-1}(z)dz\right) = \mathrm{Po}(0)$, since for $z < \sigma_{k-1}$ we have $\mu_{k-1}(z) = 0$ thus $\mathrm{d}\mu_{k-1}(z) = 0$, while for $z \geqslant \sigma_{k-1} \geqslant t$ we have $\kappa_t(y, z) = 0$.

Hence we may in the rest of the proof assume $t > \sigma_{k-1}$ and thus $\mu_{k-1}(t) = \rho_{k-1}(t) > 0$. As noted after (5.20) above, the kernel κ_t then is quasi-irreducible. Hence, it follows from [5, Theorem 3.1] that

$$C_1(G^{\mathcal{V}}(n, \kappa_t))/n \overset{\mathrm{p}}{\longrightarrow} \rho(\kappa_t). \tag{5.29}$$

We have shown in Lemma 5.3 that $\dot{G}_k(t)$ differs from $G^{\mathcal{V}}(n, \kappa_t)$ by only $o_{\mathrm{p}}(n)$ edges, and we appeal to the stability theorem [5, Theorem 3.9] to show that largest components of these two graphs have essentially the same size. (Alternatively, we could use [6, Theorem 1.1].) A minor technical problem is that this theorem is stated for irreducible kernels, while κ_t is only quasi-irreducible. We can extend the theorem (in a standard way) by considering only the vertices i with $x_i = \tau_{k-1}(i) \leqslant t$, i.e., the vertices i in the permanent giant $\mathfrak{C}_{k-1}(t)$ of $G_{k-1}(t)$, see (5.3). This defines a *generalized vertex space* [5,

Section 2] $\mathcal{V}' = (\mathcal{S}', \mu'_{k-1}, (\mathbf{x}'_n)_{n\geqslant 1})$, where $\mathcal{S}' := [0, t]$, μ'_{k-1} is the restriction of μ_{k-1} to $\mathcal{S}'$, and $\mathbf{x}'_n$ is the subsequence of $\mathbf{x}_n = (x_1, \ldots, x_n)$ consisting of all $x_i \in \mathcal{S}'$. The kernel κ_t is strictly positive a.e. on $\mathcal{S}' \times \mathcal{S}'$, and is thus irreducible.

Thus, we may take G_n in [5, Theorem 3.9] to be

$$G_n := G^{\mathcal{V}'}(n, \kappa_t), \tag{5.30}$$

which may be thought of as the restriction of $G^{\mathcal{V}}(n, \kappa_t)$ to $\mathfrak{C}_{k-1}(t)$. Take the theorem's G'_n to be

$$G'_n := \dot{G}_k(t)\left[\mathfrak{C}_{k-1}(t)\right], \tag{5.31}$$

the restriction of $\dot{G}_k(t)$ to $\mathfrak{C}_{k-1}(t)$. For any $\delta > 0$, from Lemma 5.3, w.h.p. $e\big(\dot{G}_k(t) \triangle G^{\mathcal{V}}(n, \kappa_t)\big) \leqslant \delta n$. Restricting each of these graphs to $\mathfrak{C}_{k-1}(t)$, it follows that w.h.p.

$$e(G'_n \triangle G_n) \leqslant \delta n. \tag{5.32}$$

Thus, G_n and G'_n fulfill the theorem's hypotheses. For any $\varepsilon > 0$, we may choose $\delta > 0$ per the theorem's hypotheses, and it follows from the theorem and (5.29) that w.h.p.

$$\big(\rho(\kappa_t) - \varepsilon\big)n \leqslant C_1(G'_n) \leqslant \big(\rho(\kappa_t) + \varepsilon\big)n. \tag{5.33}$$

Our aim is to establish (2.2), which is (5.33) with $C_1(G_k(t))$ in lieu of $C_1(G'_n)$. Each component $\mathcal{C}$ of $G_k(t)$ (or equivalently of $\dot{G}_k(t)$) is a subset of some component of $G_{k-1}(t)$, either $\mathcal{C}_1(G_{k-1}(t))$ or some other component. Since $t > \sigma_{k-1}$ and by the induction hypothesis (v) of Theorem 2.1, w.h.p. $\mathfrak{C}_{k-1}(t) \neq \emptyset$ and thus $\mathcal{C}_1(G_{k-1}(t)) = \mathfrak{C}_{k-1}(t)$. Thus, components of $G_k(t)$ contained in $\mathcal{C}_1(G_{k-1}(t))$ are also contained in G'_n, and the largest such component is governed by (5.33). Components of $G_k(t)$ contained in a smaller component of $G_{k-1}(t)$ have size at most $C_2(G_{k-1}(t))$, which by the induction hypothesis (ii) is w.h.p. smaller than any constant times n, and thus smaller than the component described by (5.33). Consequently, w.h.p. $C_1(G_k(t)) = C_1(G'_n)$, and thus (5.33) implies (2.2) w.h.p., for every $\varepsilon > 0$ (for the fixed t, rather than for all $t \geqslant 0$ as in Theorem 2.1). This is equivalent to (2.1) (again, for the fixed t). □

5.3. **Towards part (ii) of Theorem 2.1.** The next lemma establishes something like Theorem 2.1 (ii), but only for any fixed $t \geqslant 0$; extending this to the supremum follows later.

Lemma 5.5. *For every fixed $t \geqslant 0$, $C_2(G_k(t)) = o_p(n)$.*

Proof. We use the notation of the proof of Lemma 5.4, specifically (5.30) and (5.31). Let $G^{\dagger}_n$ be the graph G'_n with a single edge added such that the two largest components $\mathcal{C}_1(G'_n)$ and $\mathcal{C}_2(G'_n)$ are joined and let $\varepsilon > 0$. (If $\mathcal{C}_2(G'_n) = \emptyset$, let $G^{\dagger}_n := G'_n$.) Since w.h.p. the analog of (5.32) holds also for $G^{\dagger}_n$, [5, Theorem 3.9] applies also to G_n and $G^{\dagger}_n$ and shows that w.h.p.

$$C_1(G'_n) + C_2(G'_n) = C_1(G^{\dagger}_n) \leqslant \big(\rho(\kappa_t) + \varepsilon\big)n. \tag{5.34}$$

This and (5.33) imply that w.h.p.

$$C_2(G'_n) \leqslant 2\varepsilon n. \tag{5.35}$$

Furthermore, as shown in the proof of Lemma 5.4, w.h.p. every component of $G_k(t)$ that is *not* part of G'_n has size at most $C_2(G_{k-1}(t))$, which w.h.p. is $\leqslant \varepsilon n$ by the induction hypothesis. Consequently, w.h.p.

$$C_2(G_k(t)) \leqslant 2\varepsilon n, \tag{5.36}$$

which completes the proof. □

Let $T_t = T_{t,k} := T_{\kappa_t}$ be the integral operator defined by (4.6) with the measure μ_{k-1}. We regard T_t as an operator on $L^2(\mathcal{S}, \mu_{k-1})$, and recall that (since κ_t is bounded) T_t is a bounded and compact operator for every $t \geqslant 0$.

Lemma 5.6. *The operator norm $\|T_t\|$ is a continuous function of $t \geqslant 0$. Furthermore,*

(a) $\|T_t\| = 0 \iff t \leqslant \sigma_{k-1}$
(b) $\|T_t\|$ *is strictly increasing on* $[\sigma_{k-1}, \infty)$.
(c) $\|T_t\| \to \infty$ *as* $t \to \infty$.

Proof. If $0 \leqslant t \leqslant u < \infty$, then by (5.20), for any $x, y \in \mathcal{S} = [0, \infty)$,

$$0 \leqslant \kappa_u(x, y) - \kappa_t(x, y) \leqslant u - t, \tag{5.37}$$

and consequently, with $\|\cdot\|_{\mathrm{HS}}$ the Hilbert–Schmidt norm,

$$\big| \|T_t\| - \|T_u\| \big| \leqslant \|T_t - T_u\| \leqslant \|T_t - T_u\|_{\mathrm{HS}} = \|\kappa_t - \kappa_u\|_{L^2(\mathcal{S}^2, \mu_{k-1}^2)} \leqslant |t - u|. \tag{5.38}$$

Hence, $t \mapsto \|T_t\|$ is continuous.

(a): By (5.20), $\kappa_t(x, y) > 0 \iff (x, y) \in [0, t)^2$. If $t \leqslant \sigma_{k-1}$, then $\mu_{k-1}[0, t) = \rho_{k-1}(t) = 0$, and thus $\kappa_t = 0$ a.s., so $T_t = 0$.

Conversely, if $t > \sigma_{k-1}$, then $\kappa_t > 0$ on a set of positive measure, and thus $T_t 1 > 0$ on a set of positive measure (where 1 denotes the function that is constant 1); hence $\|T_t\| \geqslant \|T_t 1\|_{L^2} > 0$.

(b): Assume $u > t > \sigma_{k-1}$. Then, as just shown, $\|T_t\| > 0$. Moreover, as said in Section 4.3, there exists an eigenfunction $\psi_t \geqslant 0$ with eigenvalue $\|T_t\|$, which we can assume is normalized: $\|\psi_t\|_{L^2} = 1$. Since $\kappa_t > 0$ on $[0, t) \times [0, t)$, and 0 elsewhere, it follows from $\psi_t = T_t \psi_t / \|T_t\|$ and (4.6) that $\psi_t = 0$ a.e. on $[t, \infty)$, and thus nonzero somewhere on $[0, t)$, and thus $\psi_t > 0$ a.e. on $[0, t)$. As $u > t$, then $\kappa_u(x, y) > \kappa_t(x, y)$ on $[0, t) \times [0, t)$, and it follows that for $x \in [0, t)$,

$$(T_u \psi_t)(x) = \int_0^\infty \kappa_u(x, y) \psi_t(y) \, \mathrm{d}\mu_{k-1}(y)$$
$$> \int_0^\infty \kappa_t(x, y) \psi_t(y) \, \mathrm{d}\mu_{k-1}(y) = T_t \psi_t(x) = \|T_t\| \psi_t(x). \tag{5.39}$$

Thus,

$$\|T_u\| \geqslant \|T_u \psi_t\|_{L^2} > \|T_t\| \, \|\psi_t\|_{L^2} = \|T_t\|. \tag{5.40}$$

Consequently, (b) holds.

(c): By (5.20), $\kappa_t(x, y) \nearrow \infty$ as $t \to \infty$ for every $x, y \in [0, \infty)$. Hence, using monotone convergence, $T_t 1(x) = \int_0^\infty \kappa_t(x, y) \, \mathrm{d}\mu_{k-1}(y) \nearrow \infty$ for every $x \in [0, \infty)$, and thus by monotone convergence again, $\|T_t 1\|_{L^2} \nearrow \infty$. Consequently, $\|T_t\| \geqslant \|T_t 1\|_{L^2} \to \infty$. □

5.4. **Proofs of parts (iii) and (iv) of Theorem 2.1.**

Proof of Theorem 2.1(iii). By Lemma 5.6, there exists a unique $\sigma_k > 0$ such that

$$\|T_{\sigma_k}\| = 1. \tag{5.41}$$

Furthermore, $\|T_t\| < 1$ if $t < \sigma_k$ and $\|T_t\| > 1$ if $t > \sigma_k$. Thus,

$$t > \sigma_k \iff \|T_t\| > 1 \iff \rho_k(t) > 0, \tag{5.42}$$

where the last equivalence follows from [5, Theorem 3.1], establishing $\|T_t\| > 1$ as a necessary and sufficient condition for the existence of a giant component, and providing its size. In order to see that ρ_k is strictly increasing on $[\sigma_k, \infty)$, let $\sigma_k < t < u$. Since $\kappa_u(x, y) \geqslant \kappa_t(x, y)$ for all $x, y \in \mathcal{S}$, we may couple the branching processes $\mathfrak{X}_t = \mathfrak{X}_{\kappa_t}$ and $\mathfrak{X}_u = \mathfrak{X}_{\kappa_u}$ such that $\mathfrak{X}_u$ is obtained from $\mathfrak{X}_t$ by adding extra children to some individuals. (Each individual of type x gets extra children of type y distributed as a Poisson process with intensity $(\kappa_u(x, y) - \kappa_t(x, y)) \, \mathrm{d}\mu_{k-1}(y)$, independent of everything else.) Then clearly $\mathfrak{X}_u$ survives if $\mathfrak{X}_t$ does, so $\rho_k(u) := \rho(\kappa_u) \geqslant \rho(\kappa_t) = \rho_k(t)$. (See [5, Lemma 6.3].) Moreover, there is a positive probability that $\mathfrak{X}_t$ dies out but $\mathfrak{X}_u$ survives, for example because the initial particle has no children in $\mathfrak{X}_t$ but at least one in $\mathfrak{X}_u$, and this child starts a surviving branching process. Hence $\rho_k(u) > \rho_k(t)$. $\qquad\square$

We next prove Theorem 2.1(iv). A simple lemma will be useful here and subsequently.

Consider the process defined in Section 2 of all graphs $G_j(t)$, $j \geqslant 1$ and $t \geqslant 0$, under some edge-arrival process; consider also a similar set of graphs $G'_j(t)$ coming from a second arrival process thicker than the first (i.e., containing the same arrivals and possibly others).

Lemma 5.7. *The thicker process yields larger graphs, i.e., $G_j(t) \subseteq G'_j(t)$ for all j and t. Also, any edge e present in both arrival processes, if contained in $F'_1(t) \cup \cdots \cup F'_j(t)$, is also contained in $F_1(t) \cup \cdots \cup F_j(t)$.*

Proof. It is easy to see that adding edges can only make G_1 larger, i.e., that $G'_1(t) \supseteq G_1(t)$. Thus any edge originally passed on to $G_2(t)$ will still be passed on, plus perhaps some others; by induction on j, any $G_j(t)$ can only increase, i.e., $G'_j(t) \supseteq G_j(t)$. This proves the first assertion. The second assertion follows from the first. If e is not contained in $F_1(t) \cup \cdots \cup F_j(t)$ then it is passed on to $G_{j+1}(t)$, and hence, as just shown, it belongs also to $G'_{j+1}(t)$ and therefore not to $F'_1(t) \cup \cdots \cup F'_j(t)$. $\qquad\square$

Let $q_n(t)$ be the probability that two fixed, distinct, vertices in $G_k(t)$ belong to the same component. By symmetry, this is the same for any pair of vertices, and thus also for a random pair of distinct vertices. Hence, recalling (4.5),

$$q_n(t) = \mathbb{E} \, \pi(G_k(t)). \tag{5.43}$$

Lemma 5.8. *There exist constants $b_k, B_k > 0$ such that, for every $n \geqslant 2$,*

$$q_n(t) \geqslant 1 - B_k e^{-b_k t}, \qquad t \geqslant 0. \tag{5.44}$$

Proof. Fix some $t_0 > \sigma_k$; thus $\rho_k(t_0) > 0$ by (iii). Then, cf. (4.5), writing $C_i := C_i(G_k(t_0))$,

$$q_n(t_0) = \mathbb{E}\frac{\sum_{i=1}^{\infty} C_i(C_i - 1)}{n(n-1)} \geqslant \mathbb{E}\frac{C_1(C_1 - 1)}{n(n-1)}. \qquad (5.45)$$

By Lemma 5.4, $C_1/n \xrightarrow{\text{p}} \rho_k(t_0)$ as $n \to \infty$, and thus $C_1(C_1 - 1)/(n(n-1)) \xrightarrow{\text{p}} \rho_k(t_0)^2$. Hence, by (5.45) and dominated convergence (see e.g. [15, Theorems 5.5.4 and 5.5.5]),

$$\liminf_{n\to\infty} q_n(t_0) \geqslant \lim_{n\to\infty} \mathbb{E}\frac{C_1(C_1 - 1)}{n(n-1)} = \rho_k(t_0)^2 > 0. \qquad (5.46)$$

Let $q := \rho_k(t_0)^2/2$, say. Then (5.46) shows that if n is large enough, $q_n(t_0) > q$. By reducing q, if necessary, we may assume that this holds for every n, since obviously $q_n(t_0) > 0$ for every fixed $n \geqslant 2$.

For an integer $m \geqslant 0$, consider the process defined in Section 2 of all graphs $G_j(t)$, $j \geqslant 1$ and $t \geqslant 0$, but erase all edges and restart at mt_0; denote the resulting random graphs by $G_j^{(m)}(t)$ and note that $G_j^{(m)}(t + mt_0) \overset{\text{d}}{=} G_j(t)$. In particular, let $G_{k,m} := G_k^{(m)}((m+1)t_0)$. Then $G_{k,m} \overset{\text{d}}{=} G_k(t_0)$; furthermore, the random graphs $G_{k,m}$, $m = 0, 1, \dots$, are independent, since they depend on edges arriving in disjoint time intervals.

Consider the process at times $i\,t_0$ for integers i. By Lemma 5.7, $G_k(i\,t_0)$ dominates what it would have been had no edges arrived by $(i-1)t_0$, which, for $i \geqslant 2$, is simply an independent copy of $G_k(t_0)$ (that is, independent of $G_k(t_0)$ but identically distributed). Consequently, for any integer M, vertices x and y can be in different components of $G_k(Mt_0)$ only if they are in different components in each of the M copies of $G_k(t_0)$. Thinking of all values $i \geqslant 1$ at once, these copies of $G_k(t_0)$ are all independent, as they depend on edge arrivals in disjoint time intervals $\big((i-1)t_0, i\,t_0\big]$. Thus,

$$1 - q_n(Mt_0) \leqslant (1 - q_n(t_0))^M \leqslant (1 - q)^M \leqslant e^{-qM}. \qquad (5.47)$$

Thus, for any $t \geqslant 0$, taking $M := \lfloor t/t_0 \rfloor$,

$$q_n(t) \geqslant q_n(Mt_0) \geqslant 1 - e^{-q\lfloor t/t_0\rfloor} \geqslant 1 - e^{q-(q/t_0)t}, \qquad (5.48)$$

which shows (5.44). (In fact, we get $B_k = e^q < e$; we can take q arbitrarily small and thus B_k arbitrarily close to 1, at the expense of decreasing b_k.) $\square$

Proof of Theorem 2.1(iv). Let $\mathcal{C}^1(t)$ be the component of $G_k(t)$ that contains vertex 1. Then, by Lemma 5.8,

$$\mathbb{E}\,C_1(G_k(t)) \geqslant \mathbb{E}\,|\mathcal{C}^1(t)| = 1 + (n-1)q_n(t) \geqslant (n-1)\big(1 - B_k e^{-b_k t}\big). \quad (5.49)$$

Furthermore, by Lemma 5.4 and dominated convergence, $\mathbb{E}\,C_1(G_k(t))/n \to \rho_k(t)$ as $n \to \infty$. Hence, (5.49) implies $\rho_k(t) \geqslant 1 - B_k e^{-b_k t}$, which is (2.3).
$\square$

5.5. **Proofs of parts (i), (ii), and (v) of Theorem 2.1.**

Proof of Theorem 2.1(i). If t_n is a sequence such that either $t_n \nearrow t$ or $t_n \searrow t$, then $\kappa_{t_n}(x,y) \to \kappa_t(x,y)$ for all x and y by (5.20), and thus by [5, Theorem 6.4], recalling (5.28), $\rho_k(t_n) = \rho(\kappa_{t_n}) \to \rho(\kappa_t) = \rho_k(t)$. Hence, ρ_k is continuous. That ρ_k is increasing was shown above in Theorem 2.1(iii).

Let $\varepsilon > 0$ and let N be an integer with $N \geqslant \varepsilon^{-1}$. Since ρ_k is continuous, and $\rho_k(t) \to 1$ as $t \to \infty$ by Theorem 2.1(iv), we can choose $t_0 = 0 < t_1 < \cdots < t_{N-1} < t_N = \infty$ with

$$\rho_k(t_j) = j/N. \tag{5.50}$$

By Lemma 5.4, w.h.p.

$$\rho_k(t_j) - \varepsilon \leqslant C_1\big(G_k(t_j)\big)/n \leqslant \rho_k(t_j) + \varepsilon \tag{5.51}$$

for every $j \leqslant N$. (The case $j = N$ is trivial, since $G_k(\infty)$ a.s. is connected.) Then, for every $j = 1, \ldots, N$ and every $t \in [t_{j-1}, t_j]$,

$$C_1\big(G_k(t)\big)/n \leqslant C_1\big(G_k(t_j)\big)/n \leqslant \rho_k(t_j) + \varepsilon \leqslant \rho_k(t) + \frac{1}{N} + \varepsilon \leqslant \rho_k(t) + 2\varepsilon, \tag{5.52}$$

which together with a similar lower bound shows that w.h.p. $|C_1\big(G_k(t)\big)/n - \rho_k(t)| \leqslant 2\varepsilon$ for all $t \geqslant 0$. Since ε is arbitrary, this shows (2.2). $\qquad\square$

Proof of Theorem 2.1(ii). Let ε, N, and t_j, $j = 0, \ldots, N$, be as in the proof of Theorem 2.1(i) above. Again, w.h.p., (5.51) holds for every $j \leqslant N$. Moreover, by Lemma 5.5, for every $j \leqslant N - 1$, and trivially when $j = N$, w.h.p.

$$C_2\big(G_k(t_j)\big) \leqslant \varepsilon n. \tag{5.53}$$

Assume (5.51) and (5.53) for every $j \leqslant N$, and also that $C_2\big(G_k(t)\big) > 3\varepsilon n$ for some $t \geqslant 0$. Choose j with $1 \leqslant j \leqslant N$ such that $t \in [t_{j-1}, t_j]$. If $\mathcal{C}_2(G_k(t))$ has *not* merged with $\mathcal{C}_1(G_k(t))$ by time t_j, then

$$C_2\big(G_k(t_j)\big) \geqslant C_2\big(G_k(t)\big) > 3\varepsilon n, \tag{5.54}$$

which contradicts (5.53). If on the other hand these two components *have* merged, then, using (5.51) and (from (5.50)) that $\rho_k(t_{j-1}) \geqslant \rho_k(t_j) - \varepsilon$,

$$C_1\big(G_k(t_j)\big) \geqslant C_1\big(G_k(t)\big) + C_2\big(G_k(t)\big) > C_1\big(G_k(t_{j-1})\big) + 3\varepsilon n$$
$$\geqslant \rho_k(t_{j-1})n - \varepsilon n + 3\varepsilon n \geqslant \rho_k(t_j)n + \varepsilon n, \tag{5.55}$$

which contradicts (5.51). Consequently, w.h.p. $\sup_t C_2\big(G_k(t)\big) \leqslant 3\varepsilon n$. $\qquad\square$

Proof of Theorem 2.1(v). If $t > \sigma_k$, then $\rho_k(t) > 0$ by Theorem 2.1(iii). Let $\delta = \rho_k(t)/2$. Then, by (i) and (ii), w.h.p. $C_1(G_k(t)) > \delta n$, and, simultaneously for every $u \geqslant 0$, $C_2(G_k(u)) < \delta n$. Assume that these inequalities hold. Then, in particular, the largest component of $G_k(t)$ is a unique giant. (Recall the definition from Section 2.3.) Moreover, for every $u \geqslant t$, the component $\mathcal{C}$ of $G_k(u)$ that contains the largest component of $G_k(t)$ then satisfies

$$|\mathcal{C}| \geqslant C_1(G_k(t)) > \delta n > C_2(G_k(u)), \tag{5.56}$$

showing that $\mathcal{C}$ is the unique giant of $G_k(u)$. Hence, the largest component of $G_k(t)$ is w.h.p. a permanent giant.

Consequently, if $t > \sigma_k$, then w.h.p. $|\mathfrak{C}_k(t)| = C_1(G_k(t))$ and (2.4) follows from (2.1). On the other hand, if $t \leqslant \sigma_k$, then (2.1) and (iii) yield

$$|\mathfrak{C}_k(t)|/n \leqslant C_1(G_k(t))/n \xrightarrow{\mathrm{p}} \rho_k(t) = 0, \tag{5.57}$$

and (2.4) follows in this case too. $\qquad\square$

This completes the proof of Theorem 2.1.

5.6. **A corollary, and proof of Theorem 2.3.** We note the following corollary.

Corollary 5.9. *For $k \geqslant 1$, uniformly for all $t \in [0, \infty)$,*

$$\chi\big(G_k(t)\big)/n \xrightarrow{\text{p}} \rho_k(t)^2, \tag{5.58}$$

$$\widehat{\chi}\big(G_k(t)\big)/n \xrightarrow{\text{p}} 0. \tag{5.59}$$

Proof. First, (5.59) follows immediately from (4.4) and Theorem 2.1(ii).
   Next, by the definitions (4.1)–(4.2),

$$\chi\big(G_k(t)\big)/n = \widehat{\chi}\big(G_k(t)\big)/n + C_1\big(G_k(t)\big)^2/n^2, \tag{5.60}$$

and (5.58) follows by (5.59) and (2.1). □

Remark 5.10. Using [19, Theorem 4.7 and Lemma 2.2] together with results above (in particular (5.1)–(5.2) and Lemmas 5.1 and 5.2), it is not difficult to prove the much stronger results that if $t < \sigma_k$ is fixed, then there exists a finite constant $\chi_k(t)$ such that $\chi(G_k(t)) \xrightarrow{\text{p}} \chi_k(t)$, and if $t \neq \sigma_k$ is fixed, then there exists a finite constant $\widehat{\chi}_k(t)$ such that $\widehat{\chi}(G_k(t)) \xrightarrow{\text{p}} \widehat{\chi}_k(t)$. Furthermore, these limits can be calculated from the branching process $\mathfrak{X}_t = \mathfrak{X}_{\kappa_t}$ on $(\mathcal{S}, \mu_{k-1})$: if we let $|\mathfrak{X}_t|$ be the total population of the branching process, then $\chi_k(t) = \mathbb{E}(|\mathfrak{X}_t|)$ and $\widehat{\chi}_k(t) = \mathbb{E}(|\mathfrak{X}_t| \mathbf{1}\{|\mathfrak{X}_t| < \infty\})$. We omit the details. □

Proof of Theorem 2.3. Since $\kappa_t(x, y) = 0$ for every y when $x \geqslant t$, a particle of type $x \geqslant t$ will not get any children at all in the branching process $\mathfrak{X}_{t,k} = \mathfrak{X}_{\kappa_t}$, hence has survival probability $\rho_\kappa(x) = 0$. Thus, recalling (4.9) and (5.14), the survival probability

$$\rho_k(t) = \int_0^\infty \rho_\kappa(x) d\mu_{k-1}(x) \leqslant \int_0^t 1 \, d\mu_{k-1}(x) = \mu_{k-1}[0, t) = \rho_{k-1}(t). \tag{5.61}$$

Moreover, even if $x < t$, there is a positive probability that x has no children in $\mathfrak{X}_{t,k}$, and thus there is strict inequality in (5.61) whenever $\rho_{k-1}(t) > 0$.
   Turning to the threshold, we note that $\mu[0, \sigma_{k-1}) = \rho_{k-1}(\sigma_{k-1}) = 0$, and thus μ_{k-1}-a.e. x satisfies $x \geqslant \sigma_{k-1}$, in which case $\kappa_t(x, y) \leqslant (t - \sigma_{k-1})_+$. In particular, if $t < \sigma_{k-1} + 1$, then $\|T_t\| \leqslant \|T_t\|_{\text{HS}} = \|\kappa_t\|_{L^2} \leqslant (t - \sigma_{k-1})_+ < 1$ and hence $t < \sigma_k$, see after (5.41). Consequently, $\sigma_k \geqslant \sigma_{k-1} + 1$.
   An alternative view of the last part is that, asymptotically, no edges arrive in $G_k(t)$ until $t = \sigma_{k-1}$, and even if all edges were passed on to $G_k(t)$ from that instant, $G_k(t)$ would thenceforth evolve as a simple Erdős–Rényi random graph, developing a giant component only 1 unit of time later, at $t = \sigma_{k-1} + 1$. □

6. Proofs of Theorems 1.2, 1.3 and 2.2

For a and b with $0 \leqslant a < b \leqslant \infty$, let $N_k(a, b)$ be the number of edges that arrive to $G_1(t)$ during the interval $(a, b]$ and are *not* passed on to $G_{k+1}(t)$; furthermore, let $W_k(a, b)$ be their total cost. In other words, we consider the edges, arriving in $(a, b]$, that end up in one of $T_1 = F_1(\infty), \ldots, T_k = F_k(\infty)$.

In particular, for $0 \leqslant t \leqslant \infty$,

$$N_k(0,t) = \sum_{i=1}^{k} e(F_i(t)), \tag{6.1}$$

$$W_k(0,t) = \sum_{i=1}^{k} w(F_i(t)) \tag{6.2}$$

and thus

$$W_k(0,\infty) = \sum_{i=1}^{k} w(T_i). \tag{6.3}$$

Since an edge arriving at time t has cost t/n, we have

$$\frac{a}{n} N_k(a,b) \leqslant W_k(a,b) \leqslant \frac{b}{n} N_k(a,b). \tag{6.4}$$

Lemma 6.1. *Let $0 \leqslant a < b \leqslant \infty$ and $k \geqslant 1$. For any $\varepsilon > 0$, w.h.p.*

$$\tfrac{1}{2}(b-a)\big(1 - \rho_k(b)^2 - \varepsilon\big)n \leqslant N_k(a,b) \leqslant \tfrac{1}{2}(b-a)\big(1 - \rho_k(a)^2 + \varepsilon\big)n. \tag{6.5}$$

Proof. Let $\mathcal{F}_t$ be the σ-field generated by everything that has happened up to time t. At time t, the fraction of edges arriving to $G_1(t)$ that are rejected by all of $F_1(t), \ldots, F_k(t)$ is simply the fraction lying within a component of $F_k(t)$, namely $\pi(G_k(t))$ (see (4.5)). Since edges arrive to $G_1(t)$ at a total rate $\frac{1}{n}\binom{n}{2} = \frac{n-1}{2}$, conditioned on $\mathcal{F}_t$, edges are added to $F_1(t) \cup \cdots \cup F_k(t)$ at a rate, using (4.5),

$$r_k(t) := \frac{n-1}{2}\big(1 - \pi(G_k(t))\big) = \frac{n - \chi(G_k(t))}{2}. \tag{6.6}$$

By Corollary 5.9, for every fixed t,

$$r_k(t)/n \xrightarrow{\mathrm{p}} \big(1 - \rho_k(t)^2\big)/2. \tag{6.7}$$

Condition on the event $r_k(a) \leqslant \big(1 - \rho_k(a)^2 + \varepsilon\big)n/2$, which by (6.7) occurs w.h.p. Then, since $r_k(t)$ is a decreasing function of t, the process of edges that are added to $F_1(t) \cup \cdots \cup F_k(t)$ can for $t \geqslant a$ be coupled with a Poisson process with constant intensity $\big(1 - \rho_k(a)^2 + \varepsilon\big)n/2$ that is thicker (in the sense defined just before Lemma 5.7). Thus, letting Z be the number arriving in the latter process in $(a,b]$, we have w.h.p.

$$N_k(a,b) \leqslant Z \sim \mathrm{Po}\big((b-a)(1 - \rho_k(a)^2 + \varepsilon)n/2\big). \tag{6.8}$$

Furthermore, by the law of large numbers, w.h.p.

$$Z \leqslant (b-a)\big(1 - \rho_k(a)^2 + 2\varepsilon\big)n/2. \tag{6.9}$$

Combining (6.8) and (6.9) yields the upper bound in (6.5) (with 2ε in place of ε).

For the lower bound, we stop the entire process as soon as $r_k(t) < \frac{1}{2}\big(1 - \rho_k(b)^2 - \varepsilon\big)n$. Since $r_k(t)$ is decreasing, if the stopping condition does not hold at time $t = b$ then it also does not hold at any earlier time, so by (6.7), w.h.p. we do not stop before b. As long as we have not stopped, we can couple with a Poisson process with constant intensity $\big(1 - \rho_k(b)^2 - \varepsilon\big)n/2$

that is thinner (i.e., opposite to thicker), and we obtain the lower bound in (6.5) in an analogous way as the upper bound. □

Lemma 6.2. *If* $0 < b < \infty$, *then*

$$W_k(0, b) \xrightarrow{\text{p}} \frac{1}{2} \int_0^b \big(1 - \rho_k(t)^2\big) t \, \mathrm{d}t. \tag{6.10}$$

Proof. Let $N \geqslant 1$ and define $t_j := jb/N$. By (6.4) and Lemma 6.1, for every $j \in [N]$, w.h.p.

$$W_k(t_{j-1}, t_j) \leqslant t_j \tfrac{1}{2}(t_j - t_{j-1})\big(1 - \rho_k(t_{j-1})^2 + N^{-1}\big) = \int_{t_{j-1}}^{t_j} f_N(t) \, \mathrm{d}t, \tag{6.11}$$

where we define the piecewise-constant function f_N by

$$f_N(t) := \tfrac{1}{2}t_j\big(1 - \rho_k(t_{j-1})^2 + N^{-1}\big), \qquad t_{j-1} \leqslant t < t_j. \tag{6.12}$$

Consequently, w.h.p.,

$$W_k(0, b) = \sum_{j=1}^N W_k(t_{j-1}, t_j) \leqslant \sum_{j=1}^N \int_{t_{j-1}}^{t_j} f_N(t) \, \mathrm{d}t = \int_0^b f_N(t) \, \mathrm{d}t. \tag{6.13}$$

Now let $N \to \infty$. The functions $f_N(t)$ are uniformly bounded on $[0, b]$, and $f_N(t) \to f(t) := \tfrac{1}{2}t\big(1 - \rho_k(t)^2\big)$ as $N \to \infty$ for every t by (6.12) and the continuity of $\rho_k(t)$. Hence, dominated convergence yields $\int_0^b f_N(t) \, \mathrm{d}t \to \int_0^b f(t) \, \mathrm{d}t$. Given $\varepsilon > 0$, we may thus choose N such that $\int_0^b f_N(t) \, \mathrm{d}t < \int_0^b f(t) \, \mathrm{d}t + \varepsilon$, and then (6.13) shows that w.h.p.

$$W_k(0, b) < \int_0^b f(t) \, \mathrm{d}t + \varepsilon. \tag{6.14}$$

We obtain a corresponding lower bound similarly, using the lower bounds in (6.4) and (6.5). Consequently, $W_k(0, b) \xrightarrow{\text{p}} \int_0^b f(t) \, \mathrm{d}t$, which is (6.10). □

We want to extend Lemma 6.2 to $b = \infty$. This will be Lemma 6.4, but to prove it we need the following lemma.

Lemma 6.3. *For any* $k \geqslant 1$ *there exist constants* $b_k', B_k' > 0$ *such that, for all* $t \geqslant 0$,

$$\mathbb{E}\, W_k(t, \infty) \leqslant B_k' e^{-b_k' t}. \tag{6.15}$$

Proof. For any t, recalling that N_k counts edges arriving at rate $r_k(t)$ and that $r_k(t)$ is a decreasing function, we obtain by (6.6), (5.43), and Lemma 5.8,

$$\mathbb{E}\, N_k(t, t+1) \leqslant \mathbb{E}\, r_k(t) = \frac{n-1}{2}\big(1 - q_n(t)\big) \leqslant n\, B_k e^{-b_k t}. \tag{6.16}$$

Thus, by (6.4), for $b_k' := b_k/2$ and some $B_k'' < \infty$,

$$\mathbb{E}\, W_k(t, t+1) \leqslant \frac{t+1}{n} \mathbb{E}\, N_k(t, t+1) \leqslant (t+1)B_k e^{-b_k t} \leqslant B_k'' e^{-b_k' t}. \tag{6.17}$$

Hence, for some $B_k' < \infty$ and all $t \geqslant 0$,

$$\mathbb{E}\, W_k(t, \infty) = \sum_{j=0}^{\infty} \mathbb{E}\, W_k(t+j, t+j+1) \leqslant B_k' e^{-b_k' t}. \tag{6.18}$$

□

Lemma 6.4.

$$W_k(0, \infty) \xrightarrow{\mathrm{P}} \frac{1}{2} \int_0^\infty \big(1 - \rho_k(t)^2\big)\, t\, \mathrm{d}t \, < \, \infty. \qquad (6.19)$$

Proof. First, $1 - \rho_k(t)^2 \leqslant 2(1 - \rho_k(t)) \leqslant 2B_k e^{-b_k t}$ by Theorem 2.1(iv), establishing that the integral converges.

Let $\varepsilon > 0$. We may choose $b < \infty$ such that

$$\frac{1}{2} \int_0^b \big(1 - \rho_k(t)^2\big) t\, \mathrm{d}t > \frac{1}{2} \int_0^\infty \big(1 - \rho_k(t)^2\big) t\, \mathrm{d}t - \varepsilon, \qquad (6.20)$$

and then Lemma 6.2 shows that w.h.p.

$$W_k(0, \infty) \geqslant W_k(0, b) \geqslant \frac{1}{2} \int_0^\infty \big(1 - \rho_k(t)^2\big) t\, \mathrm{d}t - \varepsilon. \qquad (6.21)$$

For an upper bound, we note that by Lemma 6.3, given $\delta, \varepsilon > 0$, we may choose b such that $\mathbb{E}\, W_k(b, \infty) < \delta\varepsilon$ (for all n), and thus with probability $\geqslant 1 - \delta$, $W_k(b, \infty) < \varepsilon$. It then follows from Lemma 6.2 that with probability $\geqslant 1 - \delta - o(1)$ (failure probabilities at most δ for the first inequality and $o(1)$ for the second),

$$W_k(0, \infty) < W_k(0, b) + \varepsilon \leqslant \frac{1}{2} \int_0^b \big(1 - \rho_k(t)^2\big) t\, \mathrm{d}t + 2\varepsilon$$

$$\leqslant \frac{1}{2} \int_0^\infty \big(1 - \rho_k(t)^2\big) t\, \mathrm{d}t + 2\varepsilon. \qquad (6.22)$$

Since δ is arbitrary, (6.22) holds w.h.p., which together with (6.21) shows (6.19). $\qquad \square$

Proof of Theorem 1.2. By (6.3) and Lemma 6.4 (with $W_0(0, \infty) = 0$),

$$w(T_k) = W_k(0, \infty) - W_{k-1}(0, \infty) \xrightarrow{\mathrm{P}} \gamma_k := \frac{1}{2} \int_0^\infty \big(\rho_{k-1}(t)^2 - \rho_k(t)^2\big) t\, \mathrm{d}t. \qquad (6.23)$$
$$\square$$

Example 6.5. The limit γ_k in Theorem 1.2 is thus given by the integral in (6.23). Unfortunately, we do not know how to calculate this, even numerically, for $k \geqslant 2$. However, we can illustrate the result with the case $k = 1$. In this case, $\rho_1(t)$ is the asymptotic relative size of the giant component in $G(n, t/n)$, and as is well-known, and follows from (5.28) and (4.10) noting that $\kappa_t(x, y) = t$, $\sigma_1 = 1$ and for $t > 1$, $\rho_1(t) = 1 - e^{-t\rho_1(t)}$. The latter function has the inverse $t(\rho) = -\log(1 - \rho)/\rho$, $\rho \in (0, 1)$. Hence, by an integration by parts and two changes of variables, with $\rho = 1 - e^{-x}$,

$$\gamma_1 = \frac{1}{2} \int_0^\infty \big(1 - \rho_1(t)^2\big)\, t\, \mathrm{d}t = \frac{1}{4}\big[t^2(1 - \rho_1(t)^2)\big]_0^\infty + \frac{1}{2} \int_0^\infty t^2 \rho_1(t)\, \mathrm{d}\rho_1(t)$$

$$= \frac{1}{2} \int_0^1 t(\rho)^2 \rho\, \mathrm{d}\rho = \frac{1}{2} \int_0^1 \frac{\log^2(1 - \rho)}{\rho}\, \mathrm{d}\rho$$

$$= \frac{1}{2} \int_0^\infty \frac{x^2 e^{-x}}{1 - e^{-x}}\, \mathrm{d}x = \zeta(3), \qquad (6.24)$$

where the final integral can be evaluated using a series expansion. Hence we recover the limit $\zeta(3)$ found by Frieze [11]. $\qquad \square$

Remark 6.6. An argument similar to the proofs of Lemmas 6.2 and 6.4 shows that

$$N_k(0,\infty)/n \xrightarrow{\text{p}} \frac{1}{2}\int_0^\infty \left(1 - \rho_k(t)^2\right) \mathrm{d}t. \tag{6.25}$$

However, since T_k has $n-1$ edges, we trivially have $N_k(0,\infty) = k(n-1)$ a.s. Hence, for any $k \geqslant 1$,

$$\frac{1}{2}\int_0^\infty \left(1 - \rho_k(t)^2\right) \mathrm{d}t = k. \tag{6.26}$$

(This is easily verified for the case $k = 1$, by calculations similar to (6.24).) Equivalently, for any $k \geqslant 1$ (since (6.26) holds trivially for $k = 0$ too),

$$\frac{1}{2}\int_0^\infty \left(\rho_{k-1}(t)^2 - \rho_k(t)^2\right) \mathrm{d}t = 1. \tag{6.27}$$

$\square$

Equation (6.27) is weakly supportive of Conjecture 1.6.

Proof of Theorem 1.3. It follows from (6.1) that $N_k(0,t) \leqslant k(n-1)$ and thus, using also (6.4), $W_k(0,b) \leqslant kb$. Consequently, Lemma 6.2 and dominated convergence yield, for every $b < \infty$,

$$\mathbb{E}\,W_k(0,b) \to \frac{1}{2}\int_0^b \left(1 - \rho_k(t)^2\right)t\,\mathrm{d}t. \tag{6.28}$$

Lemma 6.3 shows that $\mathbb{E}\,W_k(0,b) = \mathbb{E}\,W_k(0,\infty) - \mathbb{E}\,W_k(b,\infty) \to \mathbb{E}\,W_k(0,\infty)$ uniformly in n as $b \to \infty$. Hence, (6.28) holds for $b = \infty$ too by the following routine three-epsilon argument: We have

$$\mathbb{E}\,W_k(0,\infty) - \frac{1}{2}\int_0^\infty \left(1 - \rho_k(t)^2\right)t\,\mathrm{d}t$$

$$= \left(\mathbb{E}\,W_k(0,b) - \frac{1}{2}\int_0^b \left(1 - \rho_k(t)^2\right)t\,\mathrm{d}t\right)$$

$$+ \mathbb{E}\,W_k(b,\infty) - \frac{1}{2}\int_b^\infty \left(1 - \rho_k(t)^2\right)t\,\mathrm{d}t, \tag{6.29}$$

where, for any $\varepsilon > 0$, we can make all three terms on the right-hand side less than ε (in absolute value) by choosing first b and then n large enough.

The result follows since $w(T_k) = W_k(0,\infty) - W_{k-1}(0,\infty)$, cf. (6.23). $\square$

We can now prove Theorem 2.2.

Proof of Theorem 2.2. Let $0 \leqslant a < b \leqslant \infty$, and let $N(a,b)$ be the total number of edges arriving to $G_1(s)$ in the interval $s \in (a,b]$. Then $N(a,b) \sim$ $\mathrm{Po}\left(\binom{n}{2}\frac{1}{n}(b-a)\right)$, and by the law of large numbers, for any $\varepsilon > 0$, w.h.p.

$$\tfrac{1}{2}(b-a-\varepsilon)n \leqslant N(a,b) \leqslant \tfrac{1}{2}(b-a+\varepsilon)n. \tag{6.30}$$

The number of edges passed to $G_k(s)$ in $(a,b]$ is $N(a,b) - N_{k-1}(a,b)$, and thus it follows from (6.30) and (6.5) that for any $\varepsilon > 0$, w.h.p.

$$e(G_k(b)) - e(G_k(a)) = N(a,b) - N_{k-1}(a,b) \begin{cases} \geqslant \tfrac{1}{2}(b-a)\left(\rho_{k-1}(a)^2 - \varepsilon\right)n, \\ \leqslant \tfrac{1}{2}(b-a)\left(\rho_{k-1}(b)^2 + \varepsilon\right)n. \end{cases}$$

For any fixed $t > 0$, we obtain (2.5) by partitioning the interval $[0,t]$ into small subintervals and taking limits as in the proof of Lemma 6.2. Uniform

convergence for $t \in [0, T]$ then follows as in the proof of Theorem 2.1(i), cf. (5.52). $\square$

7. Simple graphs and proof of Theorem 1.1

In the *Poisson (process) model* studied so far, we have a multigraph with an infinite number of parallel edges (with increasing costs) between each pair of vertices.

It is also of interest to consider the simple graph K_n with a single edge (with random cost) between each pair of vertices, with the costs i.i.d. random variables. We consider two cases, the *exponential model* with costs $\text{Exp}(1)$ and the *uniform model* with costs $U(0, 1)$. When necessary, we distinguish the three models by superscripts P, E, and U.

We use the standard coupling of the exponential and uniform models: if $X_{ij}^{\text{E}} \sim \text{Exp}(1)$ is the cost of edge ij in the exponential model, then the costs

$$X_{ij}^{\text{U}} := 1 - \exp(-X_{ij}^{\text{E}}) \tag{7.1}$$

are i.i.d. and $U(0, 1)$, and thus yield the uniform model. Since the mapping $X_{ij}^{\text{E}} \mapsto X_{ij}^{\text{U}}$ is monotone, the Kruskal algorithm selects the same set of edges for both models, and thus the trees $T_1, T_2, \ldots$ (as long as they exist) are the same for both models; the edge costs are different, but since we select edges with small costs, $X_{ij}^{\text{U}} \approx X_{ij}^{\text{E}}$ for all edges in T_k and thus $w(T_k^{\text{U}}) \approx w(T_k^{\text{E}})$; see Lemma 7.4 for a precise statement.

Remark 7.1. We can in the same way couple the exponential (or uniform) model with a model with i.i.d. edge costs with any given distribution. It is easily seen, by the proof below and arguments as in Frieze [11] or Steele [30] for T_1, that Theorem 1.1 extends to any edge costs X_{ij} that have a continuous distribution on $[0, \infty)$ with the distribution function $F(x)$ having a right derivative $F'(0+) = 1$ (for example, an absolutely continuous distribution with a density function $f(x)$ that is right-continuous at 0 with $f(0+) = 1$); if $F'(0+) = a > 0$, we obtain instead $w(T_k) \xrightarrow{\text{p}} \gamma_k/a$. This involves no new arguments, so we confine ourselves to the important models above as an illustration, and leave the general case to the reader. $\square$

Moreover, we obtain the exponential model from the Poisson model by keeping only the first (cheapest) edge for each pair of vertices. We assume throughout the section this coupling of the two models. We regard also the exponential model as evolving in time, and define $G_k^{\text{E}}(t)$ and $F_k^{\text{E}}(t)$ recursively as we did $G_k(t)$ and $F_k(t)$ in Section 2, starting with $G_1^{\text{E}}(t) := \dot{G}_1(t)$, the simple subgraph of K_n obtained by merging parallel edges and giving the merged edge the smallest cost of the edges (which is the same as keeping just the first edge between each pair of vertices).

Recall from the introduction that while in the Poisson model every T_k exists a.s., in the exponential and uniform models there is a positive probability that T_k does not exist, for any $k \geqslant 2$ and any $n \geqslant 2$. (In this case we define $w(T_k) := \infty$.) The next lemma shows, in particular, that this probability is $o(1)$ as $n \to \infty$. (The estimates in this and the following lemma are not best possible and can easily be improved.)

Lemma 7.2. *In any of the three models and for any fixed $k \geqslant 1$, w.h.p. T_k exists and, moreover, uses only edges of costs $\leqslant 2k \log n/n$.*

Proof. Consider the exponential model; the result for the other two models is an immediate consequence by the couplings above (or by trivial modifications of the proof). The result then says that w.h.p. $G_k^{\mathrm{E}}(2k \log n)$ is connected.

By induction, we may for $k \geqslant 1$ assume that the result holds for $k-1$. Thus, w.h.p., $G_{k-1}^{\mathrm{E}}\big(2(k-1) \log n\big)$ is connected, and then all later edges are passed to $G_k^{\mathrm{E}}(t)$.

Consider now the edges arriving in $(2(k-1) \log n, 2k \log n]$. They form a random graph $G(n, p)$ with

$$p = e^{-2(k-1) \log n/n} - e^{-2k \log n/n} = e^{-2k \log n/n}\big(e^{2 \log n/n} - 1\big)$$
$$= \big(2 + o(1)\big) \log n/n. \tag{7.2}$$

As is well-known since the beginning of random graph theory [10], see e.g. [3], such a random graph $G(n, p)$ is w.h.p. connected. We have also seen that w.h.p. this graph $G(n, p)$ is a subgraph of $G_k^{\mathrm{E}}(2k \log n)$. Hence, $G_k^{\mathrm{E}}(2k \log n)$ is w.h.p. connected, which completes the induction. $\square$

Lemma 7.3. *For each fixed k,*

$$0 \leqslant w(T_k^{\mathrm{E}}) - w(T_k^{\mathrm{P}}) = O_{\mathrm{p}}\Big(\frac{\log^3 n}{n}\Big). \tag{7.3}$$

Proof. Since the exponential model is obtained from the Poisson model by deleting some edges, we have by Lemma 5.7 that every edge contained in both processes and contained in $F_1(t) \cup \cdots \cup F_k(t)$ is also contained in $F_1^{\mathrm{E}}(t) \cup \cdots \cup F_k^{\mathrm{E}}(t)$; the only edges "missing" from the latter are those that were repeat edges in the Poisson model.

At time $t_k := 2k \log n$, how many repeat edges are there? For two given vertices i and j, the number of parallel edges is $\mathrm{Po}(t_k/n)$, so the probability that it is two or more is $p_2(t_k) := \mathbb{P}(\mathrm{Po}(t_k/n) \geqslant 2) \leqslant (t_k/n)^2/2$. (We use that the kth factorial moment of $\mathrm{Po}(\lambda)$ is λ^k, and Markov's inequality.) Hence, the number of pairs $\{i, j\}$ with more than one edge is $\mathrm{Bi}(\binom{n}{2}, p_2(t_k))$, which is stochastically smaller than $\mathrm{Bi}(n^2, (t_k/n)^2)$, which by Chebyshev's inequality w.h.p. is $\leqslant 2t_k^2 = 8k^2 \log^2 n$. Similarly, the probability that i and j have three or more parallel edges is $\leqslant (t_k/n)^3/6$ and thus w.h.p. there are no triple edges in $G_1(t_k)$.

By Lemma 7.2, w.h.p. $T_1^{\mathrm{P}} \cup \cdots \cup T_k^{\mathrm{P}} = F_1(t_k) \cup \cdots \cup F_k(t_k)$, and we have just established that w.h.p. all but at most $2t_k^2$ of the edges in $F_1(t_k) \cup \cdots \cup F_k(t_k)$ are also in $F_1^{\mathrm{E}}(t_k) \cup \cdots \cup F_k^{\mathrm{E}}(t_k) \subseteq T_1^{\mathrm{E}} \cup \cdots \cup T_k^{\mathrm{E}}$. Since each spanning tree has exactly $n-1$ edges, the missing edges are replaced by the same number of other edges, which by Lemma 7.2 w.h.p. also have cost $\leqslant t_k/n$ each, thus total cost at most $2t_k^3 = 16k^3 \log^3 n/n$. Consequently, w.h.p.,

$$w\big(T_1^{\mathrm{E}} \cup \cdots \cup T_k^{\mathrm{E}}\big) \leqslant w\big(T_1^{\mathrm{P}} \cup \cdots \cup T_k^{\mathrm{P}}\big) + 16k^3 \log^3 n/n. \tag{7.4}$$

Having additional edges can never hurt (in this matroidal context), so

$$w(T_j^{\mathrm{P}}) \leqslant w(T_j^{\mathrm{E}}), \qquad j \geqslant 1. \tag{7.5}$$

This yields the first inequality in (7.3), while the second follows from (7.4) together with (7.5) for $j \leqslant k - 1$. $\qquad\square$

Lemma 7.4. *For each fixed k,*

$$0 \leqslant w(T_k^{\mathrm{E}}) - w(T_k^{\mathrm{U}}) = O_{\mathrm{p}}\Big(\frac{\log^2 n}{n}\Big). \qquad (7.6)$$

Proof. As said above, T_k^{E} and T_k^{U} consist of the same edges, with edge costs related by (7.1). Since (7.1) implies $0 \leqslant X_{ij}^{\mathrm{E}} - X_{ij}^{\mathrm{U}} \leqslant \frac{1}{2}(X_{ij}^{\mathrm{E}})^2$, it follows that, using Lemma 7.2, w.h.p.

$$0 \leqslant w\big(T_k^{\mathrm{E}}\big) - w\big(T_k^{\mathrm{U}}\big) \leqslant \sum_{e \in E(T_k)} (X_e^{\mathrm{E}})^2 \leqslant n\Big(\frac{2k \log n}{n}\Big)^2. \qquad (7.7)$$

$\qquad\square$

Proof of Theorem 1.1. It follows from Theorem 1.2 and Lemma 7.3 that for each fixed k, $w(T_k^{\mathrm{E}}) \xrightarrow{\mathrm{P}} \gamma_k$, and then from Lemma 7.4 that $w(T_k^{\mathrm{U}}) \xrightarrow{\mathrm{P}} \gamma_k$, which is Theorem 1.1. $\qquad\square$

Recall that the corresponding statement for the expectation is false, as $\mathbb{E}\, w(T_k^{\mathrm{E}}) = \mathbb{E}\, w(T_k^{\mathrm{U}}) = \infty$ for $k \geqslant 2$; see Remark 1.4.

8. The second threshold

As noted in Example 6.5 we do not know how to calculate the limit γ_2. However, we can find the threshold σ_2. In principle, the method works for σ_k for any $k \geqslant 2$, provided we know ρ_{k-1}, so we will explain the method for general k. However, we will assume the following:

$\rho_{k-1}(x)$ is continuously differentiable on (σ_{k-1}, ∞), with $\rho_{k-1}'(x) > 0$,

and $\rho_{k-1}'(x)$ is continuous as $x \searrow \sigma_{k-1}$. $\qquad (8.1)$

This is, we think, not a serious restriction, for the following reasons. First, (8.1) is easily verified for $k = 2$, since we know ρ_1 explicitly (see Example 6.5), so the calculation of σ_2 is rigorous. Second, we conjecture that (8.1) holds for all $k \geqslant 2$, although we have not proved this. (Cf. what we have proved in Theorem 2.1.) Third, even if this conjecture is wrong and (8.1) does not hold for some k, we believe that the result below is true, and can be shown by suitable modifications of the argument and perhaps replacing ρ_{k-1} by suitable approximations.

By (5.41), σ_k is defined by $\|T_{\sigma_k}\| = 1$. As said in Section 4.3, since κ_{σ_k} is bounded, there exists an eigenfunction $\psi(x) \geqslant 0$ of T_{σ_k} with eigenvalue $\|T_{\sigma_k}\| = 1$, i.e.,

$$\psi(x) = T_{\sigma_k}\psi(x) \overset{\text{def}}{=} \int_0^\infty \kappa_{\sigma_k}(x, y)\psi(y)\, \mathrm{d}\mu_{k-1}(y). \qquad (8.2)$$

By (5.20), $\kappa_{\sigma_k}(x, y) = 0$ when $x \geqslant \sigma_k$, and thus $\psi(x) = 0$ when $x \geqslant \sigma_k$; furthermore, we can write (8.2) as

$$\psi(x) = \int_0^{\sigma_k} \big(\sigma_k - x \vee y\big)\psi(y)\, \mathrm{d}\mu_{k-1}(y), \qquad x \in [0, \sigma_k]. \qquad (8.3)$$

Since $\psi \in L^2(\mu_{k-1})$, it follows from (8.3) that ψ is bounded and continuous on $[0, \sigma_k]$, and thus on $[0, \infty)$. Moreover, it is easily verified by dominated convergence that we can differentiate under the integral sign in (8.3); thus $\psi(x)$ is differentiable on $(0, \sigma_k)$ with, using the definition (5.14) of μ_{k-1},

$$\psi'(x) = -\int_0^x \psi(y) \, \mathrm{d}\mu_{k-1}(y) = -\int_0^x \psi(y) \, \mathrm{d}\rho_{k-1}(y), \qquad x \in (0, \sigma_k). \quad (8.4)$$

We now use our assumption in (8.1) that $\rho_{k-1}(x)$ is continuously differentiable on (σ_{k-1}, ∞). Then (8.4) yields, for $x \in (\sigma_{k-1}, \sigma_k)$,

$$\psi''(x) = -\rho'_{k-1}(x)\psi(x), \qquad (8.5)$$

Furthermore, we have the boundary conditions $\psi(\sigma_k) = 0$ and, by (8.4), $\psi'(\sigma_{k-1}) = 0$. We can thus find σ_k by solving the Sturm–Liouville equation (8.5) for $x \geqslant \sigma_{k-1}$ with the initial conditions $\psi(\sigma_{k-1}) = 1$, say, and $\psi'(\sigma_{k-1}) = 0$; then σ_k is the first zero of ψ in (σ_{k-1}, ∞). (Note that the solution ψ of (8.5) with these initial conditions is unique, again using (8.1).)

We can transform the solution further as follows, which is advantageous for $k = 2$, when we know the inverse of ρ_{k-1} explicitly. Since ρ_{k-1} is strictly increasing on $[\sigma_{k-1}, \infty)$, there exists an inverse

$$\varphi := \rho_{k-1}^{-1} : [0, 1) \to [\sigma_{k-1}, \infty). \qquad (8.6)$$

Let

$$s_k := \rho_{k-1}(\sigma_k) \in (0, 1) \qquad (8.7)$$

and, for $x \in [0, s_k]$, let

$$h(x) := \psi(\varphi(x)) \qquad (8.8)$$

and

$$H(x) := \int_0^x h(y) \, \mathrm{d}y. \qquad (8.9)$$

Then h is continuous on $[0, s_k]$, and on the interior $(0, s_k)$,

$$H'(x) = h(x). \qquad (8.10)$$

The assumption (8.1) implies that φ is differentiable on $(0, 1)$. Thus, for $x \in (0, s_k)$, (8.10) and (8.8) yield

$$H''(x) = h'(x) = \psi'(\varphi(x))\varphi'(x). \qquad (8.11)$$

We now make a change of variables in (8.4). In (8.12) we use that $\rho_{k-1}(y) = 0$ for $y \leqslant \sigma_{k-1}$ from Theorem 2.1(iii); in (8.13) that $\varphi(\rho_{k-1}(y)) = y$ for $y \geqslant \sigma_{k-1}$ by definition (8.6); and in (8.14) again that $\rho_{k-1} = \varphi^{-1}$, as well as $\rho_{k-1}(\sigma_{k-1}) = 0$.

$$\psi'(\varphi(x)) = -\int_0^{\varphi(x)} \psi(y) \, \mathrm{d}\rho_{k-1}(y)$$

$$= -\int_{\sigma_{k-1}}^{\varphi(x)} \psi(y) \frac{\mathrm{d}\rho_{k-1}(y)}{\mathrm{d}y} \, \mathrm{d}y \qquad (8.12)$$

$$= -\int_{\sigma_{k-1}}^{\varphi(x)} \psi(\varphi(\rho_{k-1}(y)))\rho'_{k-1}(y) \, \mathrm{d}y \qquad (8.13)$$

$$= -\int_0^x \psi(\varphi(\rho)) \, \mathrm{d}\rho \qquad (8.14)$$

$$= -\int_0^x h(\rho)\,\mathrm{d}\rho = -H(x). \tag{8.15}$$

By (8.11) and (8.15), for $x \in (0, s_k)$,

$$H''(x) = -\varphi'(x)H(x). \tag{8.16}$$

Note that $H(0) = 0$ by (8.9) while, by (8.10), (8.8), (8.7), (8.6), and (8.3),

$$H'(s_k) \stackrel{\text{def}}{=} h(s_k) \stackrel{\text{def}}{=} \psi(\varphi(s_k)) = \psi(\varphi(\rho_{k-1}(\sigma_k))) = \psi(\sigma_k) = 0. \tag{8.17}$$

Furthermore, (8.3) yields $\psi(x) > 0$ for $x \in [0, \sigma_k)$, and thus (8.8) and (8.6)–(8.7) yield

$$h(x) \stackrel{\text{def}}{=} \psi(\varphi(x)) > 0 \text{ for } x \in [0, s_k), \tag{8.18}$$

as $\varphi(0) = 0$ while $\varphi(s_k) = \sigma_k$.

Consequently, if we solve the Sturm–Liouville equation (8.16) with

$$H(0) = 0 \quad \text{and} \quad H'(0) = 1 \tag{8.19}$$

(the former, repeating, by (8.9), the latter by (8.10), (8.18), and convenient scaling, going back to the arbitrary scale of ψ in (8.2)), then s_k is the smallest positive root of $H'(s_k) = 0$, and by (8.7) and (8.6), $\sigma_k = \varphi(s_k)$.

We can reduce (8.16) to a first-order equation by the substitution

$$H = R\sin\theta \quad \text{and} \quad H' = R\cos\theta, \tag{8.20}$$

which yields

$$R'(x)\sin\theta(x) + R(x)\theta'(x)\cos\theta(x) = H'(x) = R(x)\cos\theta(x), \tag{8.21}$$

$$R'(x)\cos\theta(x) - R(x)\theta'(x)\sin\theta(x) = H''(x) = -\varphi'(x)R(x)\sin\theta(x). \tag{8.22}$$

Multiplying the first equation by $\cos\theta(x)$ and the second by $\sin\theta(x)$ and subtracting yields, after division by $R(x)$,

$$\theta'(x) = \cos^2(\theta(x)) + \varphi'(x)\sin^2(\theta(x)). \tag{8.23}$$

Since $\varphi'(x) > 0$ on $(0, 1)$, by (8.6) and the assumption $\rho'_{k-1} > 0$ on (σ_{k-1}, ∞) in (8.1), it follows from (8.23) that $\theta'(x) > 0$, and thus the function $\theta(x)$ is strictly increasing and thus invertible. Moreover, taking reciprocals, (8.23) shows that the inverse function $x(\theta)$ satisfies

$$\frac{\mathrm{d}x}{\mathrm{d}\theta} = \frac{1}{\cos^2\theta + \varphi'(x(\theta))\sin^2\theta}. \tag{8.24}$$

Now, $H(0) = 0$ and $H'(0) = 1$ (see (8.19)), so from (8.20) $\theta(0)$ is a multiple of 2π, and we take $\theta(0) = 0$. Also, by (8.18) H is increasing up to s_k and thus $H(s_k) > 0$, while $H'(s_k) = 0$ by (8.17), so (8.20) yields $\theta(s_k) = \pi/2$. Thus

$$s_k = x(\pi/2), \tag{8.25}$$

where $x(\theta)$ is the solution of (8.24) with $x(0) = 0$ (since $\theta(0) = 0$ as just shown).

For $k = 2$, as said in Example 6.5 we have $\rho_1(t) = 1 - e^{-t\rho_1(t)}$ and thus the inverse $\varphi(x) = -\log(1 - x)/x$. A numerical solution of (8.24) (with Maple) yields from (8.25) that $s_2 \doteq 0.915\,11$ and thus

$$\sigma_2 = \varphi(s_2) \doteq 2.695\,21. \tag{8.26}$$

9. A related problem by Frieze and Johansson

As said in the introduction, Frieze and Johansson [12] recently considered the problem of finding the minimum total cost of k edge-disjoint spanning trees in K_n, for a fixed integer $k \geqslant 2$. (They used random costs with the uniform model, see Section 7; we may consider all three models used above.) We denote this minimum cost by mst_k, following [12] (which uses $\mathrm{mst}_k(K_n, \mathbf{X})$ for the random variable, where $\mathbf{X}$ is the vector of random edge costs, and uses $\mathrm{mst}_k(K_n)$ for its expectation). Trivially,

$$\mathrm{mst}_k \leqslant \sum_{i=1}^{k} w(T_i), \tag{9.1}$$

and as said in the introduction, it is easy to see that strict inequality may hold when $k \geqslant 2$, i.e., that our greedy procedure of choosing $T_1, T_2, \ldots$ successively does not yield the minimum cost set of k disjoint spanning trees.

We assume in this section that $n \geqslant 2k$; then k edge-disjoint spanning trees exist and thus $\mathrm{mst}_k < \infty$. (Indeed, K_{2k} can be decomposed into k Hamilton paths, as shown in 1892 by Lucas [25, pp. 162–164] using a construction he attributes to Walecki.[1])

Remark 9.1. As observed by Frieze and Johansson [12], the problem is equivalent to finding the minimum cost of a basis in the matroid $\mathcal{M}_k$, defined as the union matroid of k copies of the cycle matroid of K_n. This means that the elements of $\mathcal{M}_k$ are the edges in K_n, and a set of edges is independent in $\mathcal{M}_k$ if and only if it can be written as the union of k forests, see e.g. [34, Chapter 8.3]. (Hence, the bases, i.e., the maximal independent sets, are precisely the unions of k edge-disjoint spanning trees. For the multigraph version in the Poisson model, of course we use instead the union matroid of k copies of the cycle matroid of K_n^∞; we use the same notation $\mathcal{M}_k$.) We write r_k for rank in this matroid. □

Kruskal's algorithm, recapitulated in the introduction, is valid for finding a minimum cost basis in any matroid; see e.g. [34, Chapter 19.1]. In the present case it means that we process the edges in order of increasing cost and keep the ones that are not dependent (in $\mathcal{M}_k$) on the ones already selected; equivalently, we keep the next edge e if $r_k(S \cup \{e\}) > r_k(S)$, where r_k is the rank function in $\mathcal{M}_k$ and S is the set of edges already selected.

Remark 9.2. It follows that the largest individual edge cost for the optimal set of k edge-disjoint spanning trees is at most the largest edge cost for any given set of k edge-disjoint spanning trees. Hence, it follows from Lemma 7.2 that for the random models studied here, the optimal k spanning trees w.h.p. use only edges of cost $\leqslant 2k \log n/n$. It follows, with only minor modifications of the proofs, that analogues of Lemmas 7.3 and 7.4 hold for mst_k for the

---

[1]Lucas introduces the problem as one of "Les Jeux de Demoiselles", namely "Les Rondes Enfantines", a game of children holding hands in a circle repeatedly, never repeating a partner. The conversion between the Hamilton cycles of the game and the Hamilton paths serving as our spanning trees is simple, and Walecki's construction is more naturally viewed in terms of Hamilton paths. For much stronger recent results on Hamilton decompositions, see for example [24].

three different models. Hence, for limits in probability, the three models are equivalent for mst_k too.

Moreover, one can similarly show that for any $b > 0$ there is a constant B such that with probability at least $1 - n^{-b}$, the optimal k spanning trees w.h.p. use only edges of cost $\leqslant Bk \log n/n$. One can then argue as for the minimum spanning tree, see e.g. [13], [26, Section 4.2.3] or [7, Example 3.15], and obtain strong concentration of mst_k for any of the three models; in particular $\mathrm{Var}(\mathrm{mst}_k) = o(1)$, and thus convergence of the expectation $\mathbb{E}(\mathrm{mst}_k)$ is equivalent to convergence in probability of mst_k.

Frieze and Johansson [12] stated their results for the expectation $\mathbb{E}\,\mathrm{mst}_k$ (for the uniform model), but the results thus hold also for convergence in probability (and for any of the three models). □

For $k = 2$, Frieze and Johansson [12] show that the expectation

$$\mathbb{E}\,\mathrm{mst}_2 \to \mu_2 \doteq 4.1704. \tag{9.2}$$

As noted in Example 3.3, using $\gamma_1 = \zeta(3)$ and the estimate of γ_2 from Table 1 suggests that the cost of two edge-disjoint spanning trees chosen successively is strictly larger, namely $\gamma_1 + \gamma_2 \doteq 4.2942$. This would show that choosing minimum spanning trees one by one is not optimal, even asymptotically, except that our estimate is not rigorous. The following theorem is less precise but establishes rigorously (subject to the numerical solution to (8.24) giving σ_2 as in (8.26)) that the values are indeed different.

Theorem 9.3. *There exists $\delta > 0$ such that, for any of the three models, w.h.p. $w(T_1) + w(T_2) \geqslant \mathrm{mst}_2 + \delta$.*

With μ_2 defined by the limit in (9.2), this can be restated in the following equivalent form.

Corollary 9.4. $\gamma_1 + \gamma_2 > \mu_2$.

Proof. The equivalence of the statements in Theorem 9.3 and Corollary 9.4 is immediate since $w(T_1) \xrightarrow{\mathrm{p}} \gamma_1$ and $w(T_2) \xrightarrow{\mathrm{p}} \gamma_2$ by Theorem 1.1 or 1.2 (depending on the choice of model), and $\mathrm{mst}_2 \xrightarrow{\mathrm{p}} \mu_2$ by [12] and Remark 9.2. □

Remark 9.5. Numerically,

$$\gamma_2 > 2.9683. \tag{9.3}$$

This is immediate from Corollary 9.4, (9.2), and (by [11]) $\gamma_1 = \zeta(3)$. □

The proof of Theorem 9.3 is based on the fact that many edges are rejected from T_1 and T_2 starting at time σ_2, but none is rejected from the union matroid before the threshold c_3 for appearance of a 3-core in a random graph, and $\sigma_2 < c_3$.

We begin with three elementary lemmas that are deterministic, and do not assume any particular distribution of edge costs; nevertheless, we use the same scaling of time as before, and say that an edge with cost w is born at time nw. (Lemma 9.6 has been used in several works, including [12], in the study of minimum spanning trees.)

Lemma 9.6. *Suppose that we select N edges $e_1, \ldots, e_N$, by any procedure, and that e_i has cost w_i. Let $N(t) := |\{i : w_i \leqslant t/n\}|$, the number of selected edges born at or before time t. Then the total cost is*

$$\sum_{i=1}^{N} w_i = \frac{1}{n} \int_0^\infty (N - N(t)) \, \mathrm{d}t. \tag{9.4}$$

Proof.

$$\sum_{i=1}^{N} w_i = \sum_{i=1}^{N} \frac{1}{n} \int_0^\infty \mathbf{1}\{t/n < w_i\} \, \mathrm{d}t = \frac{1}{n} \int_0^\infty \sum_{i=1}^{N} \left(1 - \mathbf{1}\{w_i \leqslant t/n\}\right) \mathrm{d}t$$

$$= \frac{1}{n} \int_0^\infty (N - N(t)) \, \mathrm{d}t. \qquad \square$$

For the next lemma, recall from Remark 9.1 that r_k is rank in the union matroid $\mathcal{M}_k$. We consider several (multi)graphs with the same vertex set $[n]$, and we define the intersection $G \cap H$ of two such graphs by $E(G \cap H) := E(G) \cap E(H)$. (We regard the multigraphs as having labelled edges, so parallel edges are distinguishable.) Note too that the trees $\overline{T}_i$ in the lemma are arbitrary, not necessarily the trees T_i defined in Section 2.2.

Lemma 9.7. *Consider K_n^∞ with any costs $w_e \geqslant 0$. Suppose that $\overline{T}_1, \ldots, \overline{T}_k$ are any k edge-disjoint spanning trees. For $t \geqslant 0$, let $G(t)$ be the graph with edge set $\{e \in E(K_n^\infty) : w_e \leqslant t/n\}$, and let $N(t) := e\big(G(t) \cap (\overline{T}_1 \cup \cdots \cup \overline{T}_k)\big)$. Then, $N(t) \leqslant r_k(G(t))$ for every t, and*

$$\sum_{i=1}^{k} w(\overline{T}_i) - \mathrm{mst}_k = \frac{1}{n} \int_0^\infty \big(r_k(G(t)) - N(t)\big) \, \mathrm{d}t. \tag{9.5}$$

Proof. First, $N(t)$ is by definition the number of edges in $E\big(G(t) \cap (\overline{T}_1 \cup \cdots \cup \overline{T}_k)\big)$, an independent (with respect to $\mathcal{M}_k$) subset of $E(G(t))$, and thus $N(t) \leqslant r_k(G(t))$, as asserted.

Now apply Lemma 9.6, taking $N = k(n-1)$, taking the edges $e_1, \ldots, e_N$ to be the N edges in $\overline{T}_1 \cup \cdots \cup \overline{T}_k$, and noting that the definition of $N(t)$ in Lemma 9.6 matches that here. This yields

$$\sum_{i=1}^{k} w(\overline{T}_i) = \frac{1}{n} \int_0^\infty (N - N(t)) \, \mathrm{d}t. \tag{9.6}$$

Next, as a special case, consider a collection of k spanning trees $\hat{T}_1, \ldots, \hat{T}_k$ with minimum total cost. (Since we are in a deterministic setting, such a collection may not be unique.) We may assume that they are found by Kruskal's algorithm, and thus, for every t, the set of edges in $G(t) \cap (\hat{T}_1 \cup \cdots \cup \hat{T}_k)$ is a maximal set of independent edges in $G(t)$ (independent with respect to $\mathcal{M}_k$), hence the number of these edges is $N(t) = r_k(G(t))$. Consequently, Lemma 9.6 yields

$$\mathrm{mst}_k = \sum_{i=1}^{k} w(\hat{T}_i) = \frac{1}{n} \int_0^\infty \big(N - r_k(G(t))\big) \, \mathrm{d}t. \tag{9.7}$$

The result (9.5) follows by subtracting (9.6) from (9.7). $\qquad \square$

Lemma 9.8. *Let the multigraph G be a subgraph of K_n^∞ and assume that the $(k+1)$-core of G is empty for some $k \geqslant 1$. Then the edge set $E(G)$ is a union of k disjoint forests. In other words, $r_k(G) = e(G)$.*

The properties in two last sentences are equivalent by Remark 9.1.

Proof. We use induction on $|G|$; the base case $|G| = 1$ is trivial.

If $|G| > 1$ and $|G|$ has an empty $(k+1)$-core, then there exists a vertex v in G of degree $d(v) \leqslant k$. Let G' be G with v and its incident edges deleted. By the induction hypothesis, $E(G')$ is the union of k edge-disjoint forests $F_1', \ldots, F_k'$. These forests do not contain any edge with v as an endpoint, so we may simply add the first edge of v to F_1', the second to F_2', and so on, to obtain the desired decomposition of $E(G)$.

Alternatively, the lemma follows easily from a multigraph version of a theorem of Nash-Williams [27], appearing also as [34, Theorem 8.4.4]; specifically, the matroidal proof in [34] extends to multigraphs. This theorem hypothesizes that G is "sparse", meaning that for every vertex subset A, $e(G[A]) \leqslant k(|A| - 1)$, but this follows from our hypothesis. If G has empty core, so does $G[A]$, thus $G[A]$ has a vertex v of degree $\leqslant k$, whose deletion leaves another such vertex, and so on until there are no edges, showing that $e(G[A]) \leqslant k(|A| - 1)$. $\qquad\square$

Proof of Theorem 9.3. By Lemmas 7.3–7.4 and Remark 9.2, the choice of model does not matter; for convenience we again take the Poisson model.

We use Lemma 9.7, with the first and second minimum spanning trees, i.e., $k = 2$ and $\overline{T}_j = T_j$, $j = 1, 2$. Then the lemma's $G(t) = G_1(t)$ as defined in Section 2.2 and used throughout, and

$$N(t) \stackrel{\text{def}}{=} e\big(G_1(t) \cap (T_1 \cup T_2)\big) = \sum_{i=1}^{2} e\big(G_1(t) \cap T_i\big) = e(F_1(t)) + e(F_2(t)). \quad (9.8)$$

Now, by Pittel, Spencer and Wormald [29], see also [18], the 3-core threshold of a random graph is $c_3 := \min_{\lambda > 0} \left(\lambda / \mathbb{P}(\text{Po}(\lambda) \geqslant 2)\right) \doteq 3.3509$, so that for any $b < c_3$, w.h.p. the 3-core of $G(n, b/n)$ is empty. In our context this says that for all $t \leqslant b$, $\dot{G}_1(t)$ has an empty 3-core. This holds for the multigraph $G_1(t)$ too, e.g. by the proof in [18] which uses random multigraphs. (Alternatively, we can use $\dot{G}_1(t)$ in (9.9) below, with a negligible error.) Hence, Lemma 9.8 shows that w.h.p., for all $t \leqslant b$, $G(t) = G_1(t)$ has full rank, i.e., $r_2(G(t)) = e(G(t))$.

Furthermore, by (8.26), $\sigma_2 < c_3$. Choosing any a and b with $\sigma_2 < a < b < c_3$, by (9.5) and (9.8), w.h.p.,

$$
\begin{aligned}
w(T_1) + w(T_2) - \text{mst}_2 &\geqslant \frac{1}{n} \int_0^b \big(r_k(G(t)) - N(t)\big)\,\mathrm{d}t \\
&= \frac{1}{n} \int_0^b \big(e(G(t)) - e(F_1(t)) - e(F_2(t))\big)\,\mathrm{d}t \\
&= \frac{1}{n} \int_0^b e(G_3(t))\,\mathrm{d}t \geqslant \frac{1}{n}(b - a)e(G_3(a)). \quad (9.9)
\end{aligned}
$$

But by Theorem 2.2,

$$(b-a)e(G_3(a))/n \xrightarrow{\text{p}} (b-a) \cdot \frac{1}{2} \int_0^a \rho_2(s)^2 \, ds =: 2\delta, \qquad (9.10)$$

where $\delta > 0$ since $\rho_2(s) > 0$ for $s > \sigma_2$ by Theorem 2.1. Thus the theorem is established by (9.9) and (9.10). $\qquad \square$

10. Conjectured asymptotics of $\rho_k(t)$

As discussed in Section 3, $\gamma_k \sim 2k$ for large k, see for example Corollary 1.9. Moreover, simulations (see Section 11) suggest that the functions $\rho_k(t)$ converge, after suitable translations. If so, and assuming suitable tail bounds, (6.26) implies that the translations should be by $2k$, up to an arbitrary constant plus $o(1)$; this is formalized in Conjecture 1.6.

It is easy to see that this, together with suitable tail bounds justifying dominated convergence, by (6.23) and (6.27) would imply

$$\begin{aligned}
\gamma_k - 2k &= \frac{1}{2} \int_0^\infty \left(\rho_{k-1}(t)^2 - \rho_k(t)^2 \right) (t - 2k) \, dt \\
&= \frac{1}{2} \int_{-2k}^\infty \left(\rho_{k-1}(x+2k)^2 - \rho_k(x+2k)^2 \right) x \, dx \\
&\to \frac{1}{2} \int_{-\infty}^\infty \left(\rho_\infty(x+2)^2 - \rho_\infty(x)^2 \right) x \, dx, \qquad (10.1)
\end{aligned}$$

which would show Conjecture 1.7, with $\delta = \frac{1}{2} \int_{-\infty}^\infty \left(\rho_\infty(x+2)^2 - \rho_\infty(x)^2 \right) x \, dx$ (and necessarily $\delta \in [-1, 0]$, see Remark 3.6).

Recall that $\rho_k(t)$ is given by Lemma 5.4 as the survival probability of the branching process $\mathfrak{X}_t$ defined in Section 4.4 with kernel $\kappa_t(x, y)$ on the probability space $(\mathbb{R}_+, \mu_{k-1})$ where μ_{k-1} has the distribution function $\rho_{k-1}(t)$. More generally, we could start with any distribution function $F(t)$ on $\mathbb{R}_+$ and the corresponding probability measure μ and define a new distribution function $\Psi(F)(t)$ as the survival probability $\rho(\kappa_t; \mu)$. This defines a map from the set of distribution functions (or probability measures) on $[0, \infty)$ into itself, and we have $\rho_k = \Psi(\rho_{k-1})$. If one could show that Ψ is a contraction for some complete metric (perhaps on some suitable subset of distribution functions), then Banach's fixed point theorem would imply the existence of a unique fixed point ρ_∞, and convergence of ρ_k to it. However, the mapping Ψ is quite complicated, and we leave the possible construction of such a metric as an open problem.

Recall also that $t = \sigma_k$ is where $\rho_k(t)$ becomes non-zero, see Theorem 2.1(iii). Hence, Conjecture 1.6 suggests also the following, related conjecture.

Conjecture 10.1. *There exists a real constant σ_∞ such that as $k \to \infty$,*

$$\sigma_k = 2k + \sigma_\infty + o(1). \qquad (10.2)$$

In particular,

$$\sigma_k - \sigma_{k-1} \to 2. \qquad (10.3)$$

11. Computational results

11.1. **Naive simulations.** For intuition and as a sanity check on all calculations, we first directly simulate the problem described in the introduction's Poisson edge-weight model. Specifically, we take a graph with n vertices and random edge weights (i.i.d.exponential random variables with mean 1), find the MST, add fresh exponentials to the weights of the MST edges, and repeat to get the second and subsequent MSTs. For each MST, we plot each edge's rank within the MST, divided by n (so, $1/n$ for the first edge, up to $(n-1)/n$ for the last) on the vertical axis, against the edge's weight (multiplied by n in accordance with our time scaling) on the horizontal axis. The results are shown in Figure 2. The corresponding estimates of γ_k, for k
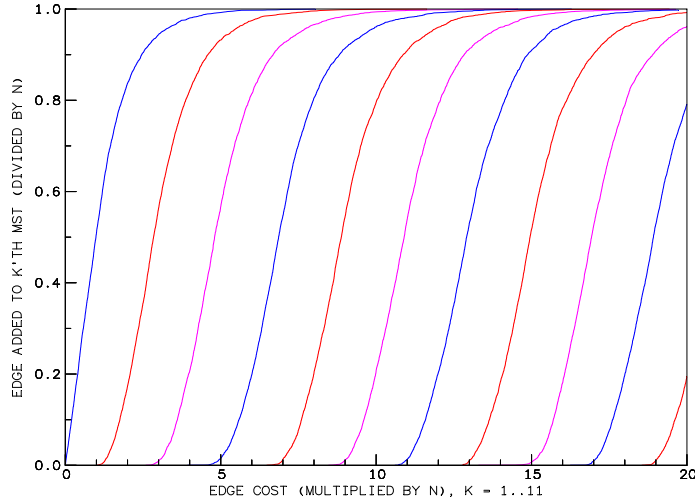
FIGURE 2. Size of kth MST (divided by n) plotted against the weight of the next edge added (multiplied by n), with $n = 4000$, for $k = 1, \ldots, 11$.

up to 5, are $1.197, 3.055, 5.035, 7.086, 9.100$. This was done for just a single graph with $n = 4000$, not averaged over several graphs. For a sense of the limited accuracy of the estimates, remember that $\gamma_1 = \zeta(3) \doteq 1.2021$.

11.2. **Better simulations.** Better simulations can be done with reference to the model introduced in Section 2.2 and used throughout. We begin with k empty graphs of order n. At each step we introduce a random edge e and, in the first graph G_i for which e does not lie within a component, we merge the two components given by its endpoints. (If this does not occur within the k graphs under consideration, we do nothing, just move on to the next edge.) For each graph we simulate only the components (i.e., the sets of vertices comprised by each component); there is no need for any more detailed structure. The edge arrivals should be regarded as occurring as a Poisson process of intensity $(n-1)/2$, but to eliminate one source of randomness we treat the edges as arriving at times $2/(n-1)$, $4/(n-1)$, etc., and thus as having weights $2/(n(n-1))$, $4/(n(n-1))$, etc.
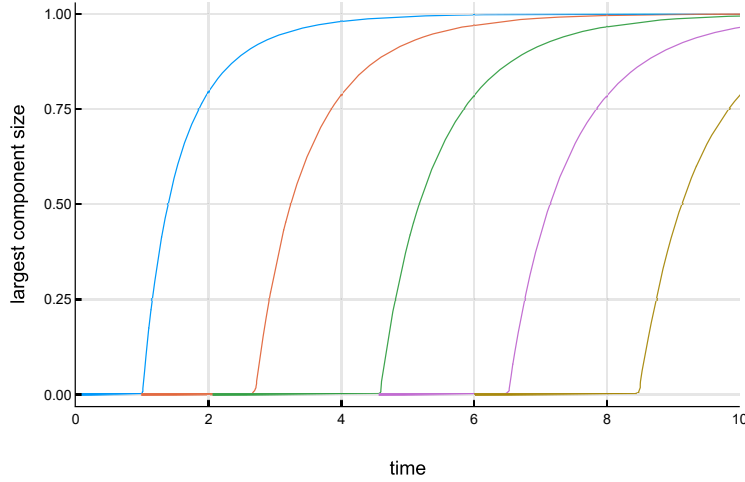
FIGURE 3. Largest component sizes, as a fraction of n, for graphs $G_1, \ldots, G_5$, based on a single simulation with $n = 1\,000\,000$.

Figure 3 depicts the result of a single such simulation with $n = 1\,000\,000$, showing for each k from 1 to 5 the size of the largest component of G_k (as a fraction of n) against time.

A larger experiment supports Conjecture 1.8 that $\gamma_k = 2k - 1 + o(1)$. Here we performed $N = 100$ simulations each with $n = 10\,000\,000$, running each simulation until the first 20 trees were complete. The results are shown in Table 1. Where x_i indicates one observation of a value of interest, the table's sample mean is $\bar{x} = \sum x_i / N$ and its standard error is $s/\sqrt{N}$, where s is the sample standard deviation, with $s^2 = \sum (x_i - \bar{x})^2/(N-1)$.

11.3. **Estimates of the improved upper bound.** The differential equation system (3.18), giving the improved upper bound of Section 3.3, is easy to solve numerically. We did so as a discrete-time approximation, setting $g_k(t + \Delta t) = g_k(t) + \frac{1}{2}\Delta\big(g_{k-1}(t)^2 - g_k(t)^2\big)$, using $\Delta t = 0.000\,01$ and considering k up to 50.

Figure 4 shows the results up to time $t = 10$. Because $g_k(t)$ pertains to a model in which all edges of F_k are imagined to be in a single component, this plot is comparable both to that in Figure 2 (which counts all edges) and to those in Figure 3 (which counts edges in the largest component) and Figure 6 (the theoretical giant-component size).

Table 2 and Figure 5 show the corresponding upper bounds on Γ_k. Specifically, the bound on Γ_k from (3.19), call it $\overline{\Gamma}_k$, is estimated as $\overline{\Gamma}_k \doteq \frac{1}{2}\Delta \sum_{t \in T} t(1 - g_k(t)^2)$ where $T = \{0, \Delta, 2\Delta, \ldots\}$. Since we cannot sum to infinity, we terminate when the final g_k under consideration is judged sufficiently close to 1, specifically within $0.000\,000\,1$ of 1. It appears experimentally that the gap $1 - g_k(t)$ decreases exponentially fast (very plausible in light of (2.3)) so termination should not be a large concern; see also (6.15).

Figure 5 suggests that the gaps $\overline{\Gamma}_k - k^2$ level off at about 0.743. (Beyond about $k = 25$ the gaps decrease, but using $\Delta t = 0.0001$ they continued to

k	sample mean	standard error
1.00	1.20203	0.00002
2.00	3.09211	0.00002
3.00	5.04782	0.00002
4.00	7.02665	0.00002
5.00	9.01553	0.00002
6.00	11.00938	0.00002
7.00	13.00580	0.00003
8.00	15.00364	0.00003
9.00	17.00231	0.00003
10.00	19.00150	0.00002
11.00	21.00099	0.00003
12.00	23.00062	0.00003
13.00	25.00042	0.00003
14.00	27.00028	0.00003
15.00	29.00017	0.00002
16.00	31.00014	0.00003
17.00	33.00010	0.00002
18.00	35.00007	0.00002
19.00	37.00007	0.00002
20.00	39.00005	0.00003

TABLE 1. Estimates of $\gamma_1, \ldots, \gamma_{20}$ from 100 simulations each with $n = 10\,000\,000.00$. Compare with Conjecture 1.8 that $\gamma_k = 2k - 1 + o(1)$.
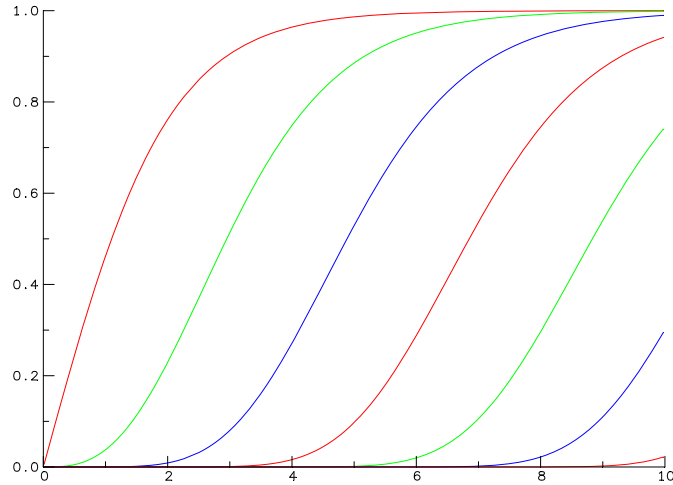
FIGURE 4. Values $g_k(t)$ plotted against t. The function $g_7(t)$ is just rising from 0 within the plot range; the values of $g_k(t)$ for larger k are too close to 0 to be seen.

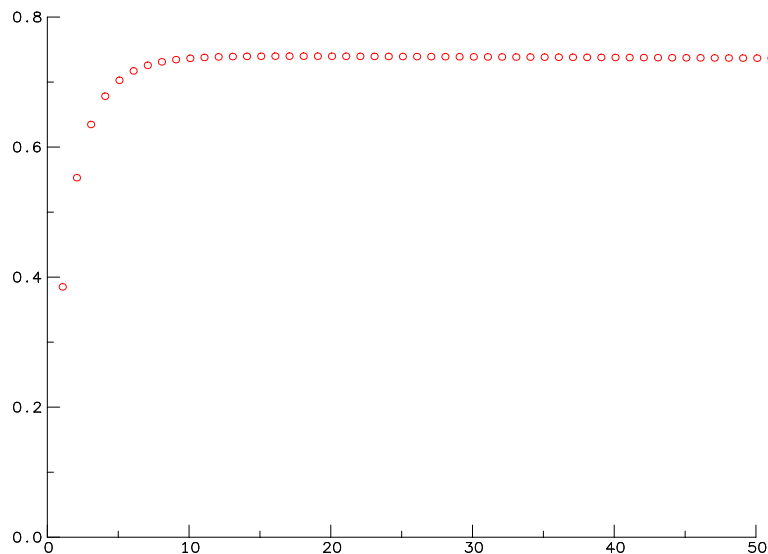increase, and in either case the degree of change is comparable with Δt and

FIGURE 5. Upper bounds from (3.19) on $\Gamma_k - k^2$, for k from 1 to 50, based on numerical solution to the differential equations (3.18).

k	1	2	3	4	5	6	7	8	9
Γ_k	1.3863	4.5542	9.6362	16.6799	25.7045	36.7189	49.7277	64.7331	81.7365

TABLE 2. Upper bounds on Γ_k obtained from numerical solution of (3.19).

thus numerically unreliable.) This suggests the following conjecture. (Recall from (3.11) that $\Gamma_k \geqslant k^2$.)

Conjecture 11.1. *For every $k \geqslant 1$, $\Gamma_k \leqslant \overline{\Gamma}_k \leqslant k^2 + \bar{\delta}$ for some constant $\bar{\delta}$.*

We established in Section 3.3 that $\Gamma_k \leqslant \overline{\Gamma}_k$, so only $\overline{\Gamma}_k \leqslant k^2 + \bar{\delta}$ is conjectural. If the conjecture holds, then it follows, using also (3.11), that $\gamma_k = \Gamma_k - \Gamma_{k-1} \geqslant (k^2) - ((k-1)^2 + \bar{\delta}) = 2k - 1 - \bar{\delta}$ and $\gamma_k = \Gamma_k - \Gamma_{k-1} \leqslant (k^2 + \bar{\delta}) - (k-1)^2 = 2k - 1 + \bar{\delta}$. Hence, the conjecture would imply

$$2k - 1 - \bar{\delta} \leqslant \gamma_k \leqslant 2k - 1 + \bar{\delta}. \tag{11.1}$$

In particular, if Conjecture 11.1 holds with $\bar{\delta} \leqslant 1$ as it appears, then $2k - 2 \leqslant \gamma_k \leqslant 2k$.

11.4. **Estimates of the fixed-point distributions** ρ_k**.** We also numerically estimated the distributions ρ_k; recall from Theorem 2.1 that $C_1(G_k(t))/n \xrightarrow{\text{p}} \rho_k(t)$. We may begin with either $\rho_0(t)$, which is 0 for $t < 0$ and 1 for $t \geqslant 0$, or with $\rho_1(t)$, which as described in Example 6.5 is the inverse function of $-\log(1 - \rho)/\rho$. (Both choices gave similar results, the latter being slightly preferable numerically.) We use ρ_{k-1} to obtain ρ_k, following the branching process described in Section 4.4. The survival probability $\rho_t(x)$ at time t of a particle born at time x in the branching process equivalent of G_k is given by the function $\rho_t = \rho_\kappa$ which (see (4.10)) is the largest fixed

point of

$$\rho_t = \Phi_\kappa \rho_t \stackrel{\text{def}}{=} 1 - e^{-T_\kappa \rho_t} \qquad (11.2)$$

(the time t is implicit in the kernel $\kappa = kk_t$ thus in the operators Φ_κ and T_κ), where (see (4.6)) T_κ is given by $(T_\kappa f)(x) = \int_{\mathcal{S}} \kappa(x,y) f(y) \, d\mu(y)$. With reference to the kernel κ defined in (5.20), $(T_\kappa f)(x) = 0$ for $x > t$, while otherwise, with $\mu = \rho_{k-1}$ as in (5.14),

$$
\begin{aligned}
(T_\kappa f)(x) &= \int_0^t (t - x \vee y) f(y) \, d\rho_{k-1}(y) \\
&= \int_0^x (t - x) f(y) \, d\rho_{k-1}(y) + \int_x^t (t - y) f(y) \, d\rho_{k-1}(y). \qquad (11.3)
\end{aligned}
$$

Given t, to find ρ_t numerically we iterate (11.2), starting with some f known to be larger than ρ_t and repeatedly setting $f(x)$ equal to $(\Phi_\kappa f)(x)$; this gives a sequence of functions that converges to the desired largest fixed point ρ_t, cf. [5, Lemma 5.6]. We will estimate ρ_t for times $i \, \Delta t$, for Δt some small constant and $i = 0, \ldots, I$, with $I \Delta t$ judged to be sufficient time to observe all relevant behavior. We initialize with $f \equiv 1$ to find $\rho_{I\Delta t}$, then iteratively initialize with $f = \rho_{i\Delta t}$ to find $\rho_{(i-1)\Delta t}$. Since the branching process is monotone in t — each vertex can only have more children by a later time t — so is the survival probability, thus $\rho_{i\Delta t}$ is larger than $\rho_{(i-1)\Delta t}$ and therefore a suitable starting estimate. In practice we find that the process converges in 20 iterations or so even for $\rho_{I\Delta t}$, and less for subsequent functions $\rho_{i\Delta t}$, with convergence defined as two iterates differing by at most 10^{-8} for any x.

For each k in turn, we do the above for all times t, whereupon the desired function $\rho_k(t) \stackrel{\text{def}}{=} \rho(\kappa) = \rho(\kappa_t)$ is given by (see (4.9) and (5.28))

$$\rho_k(t) = \int_0^\infty \rho_t(x) \, d\rho_{k-1}(x). \qquad (11.4)$$

Do not confuse ρ_t of (11.2) and ρ_k of (11.4), respectively the ρ_κ and ρ of (4.8); see also (5.28) and the comment following it.

All the calculations were performed with time (t, x, and y) discretized to multiples of $\Delta t = 0.01$ and restricted to the interval $[0, 10]$. For a fixed t, the calculation in (11.3) can be done efficiently for all x. The derivative of (11.3) with respect to x is $-\int_0^x f(y) \, d\rho_{k-1}(y)$ (cf. (8.3) and (8.4)). So, given the value of (11.3) for some x, that at the next discrete x is the discrete sum corresponding to this integral, and in one pass we can compute these integrals (discretized to summations) for all x. Each computed $\rho_k(t)$ is translated by $2k$ to keep the functions' interesting regimes within the time range $[0, 10]$, before doing the computations for $k+1$, but these translations are reversed before interpreting the results.

The first observation is that the estimates of ρ_k are consistent with Conjecture 1.6. As shown in Figure 6, even the first few functions ρ_k have visually very similar forms.

To make a more precise comparison, we time-shift each function ρ_k so that it reaches the value $1 - e^{-1}$ at time $t = 4$ (arbitrarily chosen). Figure 7
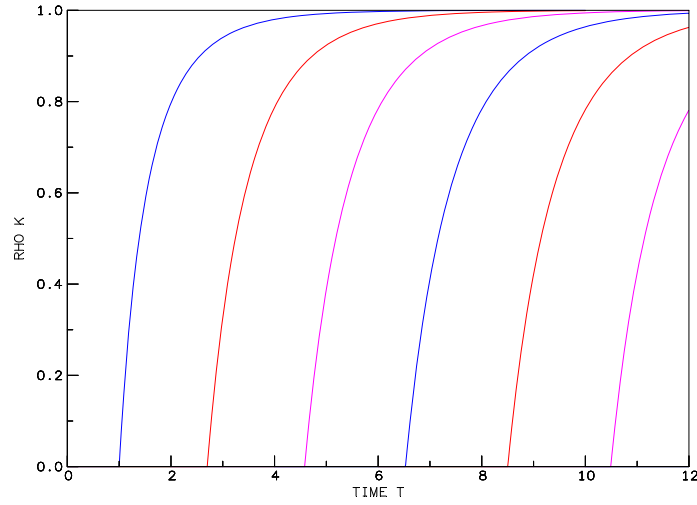
FIGURE 6. Estimates of ρ_k for $k = 1, \ldots, 6$, already suggest-
ing a limiting ρ_∞ (up to translation).

shows the thus-superposed curves for ρ_1, ρ_2, and ρ_{1000}; the curve ρ_5 (not
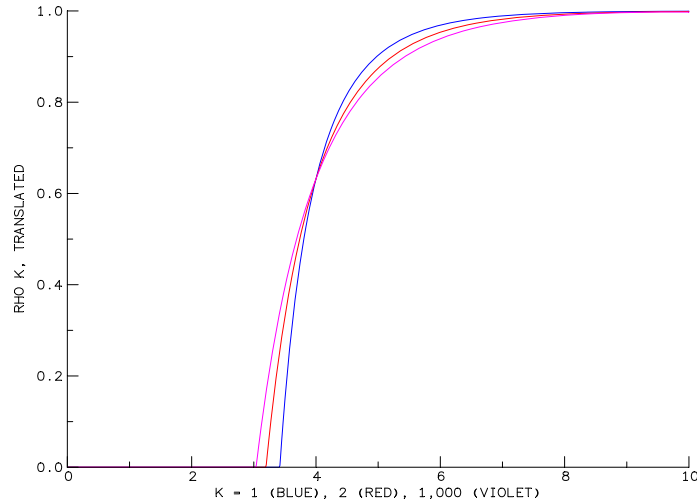shown) is already visually indistinguishable from ρ_{1000}.

FIGURE 7. Functions ρ_k time-shifted to coincide as nearly
as possible, each shifted so as to make $\rho_k(4) = 1 - e^{-1}$, for
$k = 1, 2,$ and 1000.

Estimates for γ_k, obtained via (6.23) from those for ρ_k, are shown in Ta-
ble 3. Estimates of γ_k for large k were deemed numerically unreliable for
two reasons. First, discretization of time to intervals of size $\Delta t = 0.01$ is
problematic: the timing of ρ_1 is uncertain to this order, that of ρ_2 addi-
tionally uncertain by the same amount, and so on, and translation of ρ_k
directly affects the corresponding estimate of γ_k. Second, the time range

$t \in [0, 10]$ (translated as appropriate) used in the computations proved to be too narrow, in that for large k the maximum value of ρ_k observed was only about 0.9975, and the gap between this and 1 may be enough to throw off the estimates of γ_k perceptibly.

For the same reason, we believe the results in Table 1 are more accurate than those in Table 3, despite the former coming from simulations and the latter from deterministic calculations.

k	1	2	3	4	5	6	7	8	9	10
γ_k	1.202	3.095	5.057	7.043	9.038	11.039	13.042	15.047	17.052	19.058

TABLE 3. Estimates of γ_k from (6.23).

12. Open questions

We would be delighted to confirm the various conjectures above, in particular Conjectures 1.5–1.8, and to get a better understanding of (and ideally a closed form for) ρ_∞ (provided it exists).

It is also of natural interest to ask this kth-minimum question for structures other than spanning trees. Subsequent to this work, the length X_k of the kth shortest s–t path in a complete graph with random edge weights has been studied by Gerke, Mezei, and Sorkin [14]. They show that $X_k/(2k/n + \ln n/n) \overset{\mathrm{p}}{\longrightarrow} 1$ for all k from 1 to $n-1$. In particular, the first few paths all cost nearly identical amounts, quite different from the situation for successive MSTs.

The "random assignment problem" is to determine the cost of a minimum-cost perfect matching in a complete bipartite graph with random edge weights. A great deal is known about it, by a variety of methods; for one relatively recent work, with references to others, see Wästlund [33]. It would be interesting to understand the kth cheapest matching.

It could also be interesting to consider other variants of all these questions. Frieze and Johansson [12] considered the k disjoint structures which together have the smallest possible total cost, where we consider disjoint structures generated successively. In either case, instead of asking for disjoint structures, we could require structures which are merely distinct, or perhaps which differ in some adversarially specified elements.

Acknowledgments

References

[1] Aaron Archer and Éva Tardos, Frugal Path Mechanisms, *ACM Transactions on Algorithms*, **3**(1) (2007), 3:1–3:22.

[2] Patrick Billingsley, *Convergence of Probability Measures*. Wiley, New York, 1968.

[3] B. Bollobás, *Random Graphs*, 2nd ed., Cambridge Univ. Press, Cambridge, 2001.

[4] Béla Bollobás and Alan M. Frieze, On matchings and Hamiltonian cycles in random graphs. *Random graphs '83 (Poznań, 1983)*, 23–46, *Ann. Discrete Math.*, **28**, North-Holland, Amsterdam, 1985.

[5] Béla Bollobás, Svante Janson and Oliver Riordan, The phase transition in inhomogeneous random graphs. *Random Struct. Alg.* **31** (2007), 3–122.

[6] Béla Bollobás, Svante Janson and Oliver Riordan, The cut metric, random graphs, and branching processes. *J. Statistical Physics* **140** (2010), 289–335.

[7] Stéphane Boucheron, Gábor Lugosi and Pascal Massart, *Concentration Inequalities. A Nonasymptotic Theory of Independence.* Oxford University Press, Oxford, 2013.

[8] Prasad Chebolu, Alan Frieze, Páll Melsted, and Gregory B. Sorkin, Average-case analyses of Vickrey costs. *Approximation, Randomization, and Combinatorial Optimization*, in Lecture Notes in Comput. Sci. volume 5687, 434–447, 2009.

[9] Colin Cooper, Alan Frieze, Nate Ince, Svante Janson and Joel Spencer, On the length of a random minimum spanning tree. *Combin. Probab. Comput.* **25** (2016), no. 1, 89–107.

[10] Paul Erdős and Alfréd Rényi, On random graphs I. *Publ. Math. Debrecen* **6** (1959), 290–297.

[11] Alan M. Frieze, On the value of a random minimum spanning tree problem, *Discrete Applied Mathematics* **10** (1985), 47–56.

[12] Alan Frieze and Tony Johansson, On edge disjoint spanning trees in a randomly weighted complete graph. *Combin. Probab. Comput.* **27** (2018), no. 2, 228–244.

[13] A. M. Frieze and C. J. H. McDiarmid, On random minimum length spanning trees. *Combinatorica* **9** (4) (1989), 363–374.

[14] Stefanie Gerke, Balázs Mezei and Gregory B. Sorkin, Successive shortest paths. Manuscript in preparation.

[15] Allan Gut, *Probability: A Graduate Course*, 2nd ed., Springer, New York, 2013.

[16] Svante Janson, The minimal spanning tree in a complete graph and a functional limit theorem for trees in a random graph. *Random Struct. Alg.* **7** (1995), 337–355.

[17] Svante Janson, Donald E. Knuth, Tomasz Łuczak and Boris Pittel, The birth of the giant component. *Random Struct. Alg.* **4** (1993), no. 3, 231–358.

[18] Svante Janson and Malwina J. Luczak, A simple solution to the k-core problem. *Random Struct. Alg.* **30** (2007), no. 1-2, 50–62.

[19] Svante Janson and Oliver Riordan, Susceptibility in inhomogeneous random graphs. *Electronic J. Combin.* **19**(1) (2012), #P31.

[20] Svante Janson and Gregory B. Sorkin, VCG auction mechanism cost expectations and variances. Preprint, 2013. `arXiv:1310.1777`

[21] Svante Janson and Johan Wästlund, Addendum to "The minimal spanning tree in a complete graph and a functional limit theorem for trees

in a random graph", *Random Struct. Alg.* **28** (2006), no. 4, 511–512.

[22] Anna R. Karlin, David Kempe, and Tami Tamir, Beyond VCG: Frugality of Truthful Mechanisms, *Proceedings of 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS'05)*, 615–624, IEEE, 2005.

[23] Joseph B. Kruskal, Jr., On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Amer. Math. Soc.* **7** (1956), 48–50.

[24] Daniela Kühn and Deryk Osthus, Hamilton decompositions of regular expanders: a proof of Kelly's conjecture for large tournaments. *Adv. Math.* **237** (2013), 62–146.

[25] Édouard Lucas, *Récréations Mathématiques*, Vol. 2. Gauthier–Villars, Paris, 1892. (A facsimile of the 1896 second printing is accessible at `https://archive.org/details/recretionmatedou02lucarich`.)

[26] Colin McDiarmid, Concentration. *Probabilistic Methods for Algorithmic Discrete Mathematics*, eds. M. Habib, C. McDiarmid, J. Ramirez and B. Reed, Springer, Berlin, 1998, pp. 195–248.

[27] C. St. J. A. Nash-Williams, Decomposition of finite graphs into forests. *J. London Math. Soc.* **39** (1964), 12.

[28] Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay V. Vazirani, editors, *Algorithmic Game Theory*, Cambridge Univ. Press, New York, 2007.

[29] Boris Pittel, Joel Spencer and Nicholas Wormald, Sudden emergence of a giant k-core in a random graph. *J. Combin. Theory Ser. B* **67** (1996), no. 1, 111–151.

[30] J. Michael Steele, On Frieze's $\zeta(3)$ limit for lengths of minimal spanning trees. *Discrete Appl. Math.* **18** (1987), no. 1, 99–103.

[31] Kunal Talwar, The price of truth: frugality in truthful mechanisms, *Proceedings of STACS 2003*, in Lecture Notes in Comput. Sci. volume 2607, 608–619, 2003.

[32] Johan Wästlund, Evaluation of Janson's constant for the variance in the random minimum spanning tree problem. *Linköping Studies in Mathematics* No. 7 (2005).

[33] Johan Wästlund, An easy proof of the $\zeta(2)$ limit in the random assignment problem, *Elect. Comm. in Probab.* volume 14, 261-–269, 2009.

[34] Dominic J. A. Welsh, *Matroid theory*. Academic Press, New York, 1976. Reprinted, Dover, Mineola, New York, 2010.

(Svante Janson) Department of Mathematics, Uppsala University, PO Box 480, SE-751 06 Uppsala, Sweden
Email address: `svante.janson@math.uu.se`
URL: `http://www.math.uu.se/svante-janson`

(Gregory B. Sorkin) Department of Mathematics, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, England
Email address: `g.b.sorkin@lse.ac.uk`
URL: `http://personal.lse.ac.uk/sorkin/`