Two crossover scales above the critical window for weighted random connected subgraphs

Scott Schubr
Luminary Oddities

Abstract

We study a random nonempty connected subgraph of the labelled complete graph K_n, with weight λ per edge, and the corresponding measure restricted to trees, just above the critical window around $\lambda = 1/(en)$. For $\lambda = e^{-1+s/\sqrt{n}}/n$, uniformly as $s \to \infty$ with $s = o(n^{1/6})$, we prove an estimate that separates the finite critical contribution from a large-size saddle at $k_* = s\sqrt{n}$ and remains relative when the saddle's contribution is small. Because the saddle's first moment exceeds its mass by the factor k_*, two crossover scales appear. Near $s^2 = \frac{1}{2}\log n + \frac{3}{2}\log\log n$ the mean size gains a positive excess while the weak size limit is unchanged, and the adjacent-edge ratio $\mathbb{P}(e, f)/(\mathbb{P}(e)\mathbb{P}(f))$, which for trees is a ratio of size moments, has an explicit limit after division by n. Near $s^2 = \frac{3}{2}\log n + \frac{5}{2}\log\log n$ the size divided by k_* has a two-point limit. Conditional on a large size, the fluctuations are Gaussian throughout the range, including when that event is rare. Using surplus estimates of Liu and Slade and a marked-forest deletion argument, we transfer these asymptotics to connected subgraphs, including the probabilities of every fixed prescribed forest.

Keywords: random subtrees; connected subgraphs; complete graph; finite-size scaling; edge correlations; Cayley's formula.
MSC 2020: 05C80, 60C05, 05C05, 05C30.

1 Introduction

Several natural random-subgraph models on a complete graph select a single connected object. Here the probability of a nonempty connected subgraph G of K_n is proportional to $\lambda^{|E(G)|}$; singleton vertices are included. Restricting this measure to acyclic objects gives the random-subtree model. Neither measure is the law of the full collection of components of an independent-edge random graph. Rooting a selected object also changes the size bias and hence the observable.

Random-subtree measures form part of the broader collection of models surveyed by Fredes and Marckert [2]. For uniform random subtrees of a complete graph, Luo, Xu and Tian [4] studied size distributions and edge dependence, including the adjacent-edge ratio. Those results concern the uniform model and its near-spanning behavior. Here the common activity tends to zero at order $1/n$, and the moving saddle occupies $o(n)$ vertices. An earlier note [6] showed that, for every edge weight $\lambda > 0$ and every $n \geq 3$, the conditional probability of an edge given an adjacent edge exceeds three quarters of its unconditional probability, proving the lower-bound conjecture of Luo, Xu and Tian and extending it to all $\lambda > 0$, and that for each fixed λ the corresponding tree ratio is $3/4 + 3/(4e\lambda n) + o(1/n)$. When λ is of order $1/n$ and at most $1/(en)$, the ratio is instead of order n. Just above the critical window it decreases from that order, and the present paper locates the two scales at which this happens. Chin, Gordon, MacPhee and Vincent [1] and Wagner [5] studied the probability that a random subtree of a complete graph is spanning, again a near-spanning regime.

The finite-size critical profile for both trees and connected subgraphs was established by Liu and Slade [3]. Their edge weight on V vertices is $p/(eV)$, and their scaling theorems fix s in $p = 1 + sV^{-1/2}$. Their generating functions have rooted or marked interpretations. In the present unrooted measure, the partition function and its first size moment can have different crossover scales. We keep both contributions in a uniform estimate rather than infer growing-parameter behavior by substitution into a fixed-parameter limit.

Our parameterization is $p_{\mathrm{LS}} = e^{s/\sqrt{n}}$, so $\sqrt{n}(p_{\mathrm{LS}} - 1) = s + O(s^2/\sqrt{n})$ and our s agrees with theirs to first order. We prove estimates uniformly on every interval $a_n \le s \le b_n$ with $a_n \to \infty$ and $b_n = o(n^{1/6})$. Within this range the large-size saddle lies at $k_* = \sqrt{n}\, s = o(n)$. Its mass and its first moment differ by the factor k_*. This yields a first-moment window centered at

$$s^2 = \tfrac{1}{2}\log n + \tfrac{3}{2}\log\log n + O(1)$$

and a later mass window centered at

$$s^2 = \tfrac{3}{2}\log n + \tfrac{5}{2}\log\log n + O(1).$$

Each window is a crossover scale of width $O(1)$ in s^2, above the critical window of bounded s. At the first one the mean size, and with it the adjacent-edge ratio, changes while the weak size limit does not: the size law fails to be uniformly integrable.

For distinct adjacent edges e, f we use the ratio $\mathbb{P}(e, f)/(\mathbb{P}(e)\mathbb{P}(f))$, not a Pearson correlation coefficient. Theorem 3.1 gives its uniform tree profile together with the three required raw moments. Corollaries 4.1 and 4.2 identify the windows, the escaping first moment, the mass mixture and the conditional Gaussian limit. The final section transfers these conclusions to connected subgraphs. Although cycles have vanishing probability in this regime, that observation alone does not control the relatively rare prescribed-edge events. A separate deletion and rooted-forest argument supplies the marked comparison. It applies to every fixed forest pattern, including disconnected ones; the higher-moment estimates also give their explicit probability profiles.

Cayley counts, tree generating functions and Gaussian saddle analysis are classical ingredients. Liu–Slade's connected-subgraph surplus bounds are an explicit external input. The contribution here is the uniform coexistence of finite critical mass and a moving saddle in different moments, its two-window consequences, and the marked comparison in that growing-parameter regime.

2 Exact tree weights and correlation moments

We work on the labelled complete graph K_n, $n \ge 3$. A subtree means a nonempty connected acyclic subgraph, not necessarily induced or spanning. In particular each singleton vertex is included. Give a subtree T weight $\lambda^{|E(T)|}$, and divide by the partition function $N_n(\lambda)$ to obtain its probability. Write $K = |V(T)|$. Cayley's formula gives the exact weights

$$w_{n,1} = n, \qquad w_{n,k} = \binom{n}{k} k^{k-2} \quad (2 \le k \le n), \qquad N_n(\lambda) = \sum_{k=1}^{n} w_{n,k}\lambda^{k-1}. \tag{1}$$

Conditional on $K = k$, the selected vertex set is uniform and its spanning tree is uniform. The classical extension count for a forest with component sizes $\sigma_1, \dots, \sigma_j$ is $k^{j-2}\prod_i \sigma_i$. Consequently, for fixed distinct incident edges e, f,

$$\mathbb{P}(e \in T \mid K = k) = \frac{2(k-1)}{n(n-1)}, \qquad \mathbb{P}(e, f \in T \mid K = k) = \frac{3(k-1)(k-2)}{kn(n-1)(n-2)}. \tag{2}$$

The expressions are zero at the small sizes where the corresponding edges cannot occur. Put

$$h(k) = k - 3 + \frac{2}{k}, \quad B_n = \sum_k (k-1) w_{n,k} \lambda^{k-1}, \quad A_n = \sum_k h(k) w_{n,k} \lambda^{k-1}.$$

Averaging the conditional probabilities gives the exact identity

$$R_{n,\lambda} := \frac{\mathbb{P}(e, f \in T)}{\mathbb{P}(e \in T)\mathbb{P}(f \in T)} = \frac{3n(n-1)}{4(n-2)} \frac{N_n A_n}{B_n^2} = \frac{3n(n-1)}{4(n-2)} \frac{\mathbb{E}h(K)}{\big(\mathbb{E}(K-1)\big)^2}. \tag{3}$$

The denominator is the square of the mean, rather than a second moment. Cayley enumeration and the labelled-tree generating functions below are classical. The complete-graph finite-size scaling framework and its extension to connected subgraphs are studied by Liu–Slade [3]. We retain that ancestry while proving the growing-parameter estimates required here.

3 A uniform composite estimate

For $s > 0$ set

$$\eta = \frac{s}{\sqrt{n}}, \quad c = e^{-1} e^{\eta}, \quad \lambda = \frac{c}{n}, \quad k_* = \sqrt{n}\, s, \qquad q = 2 n^{-1/4} s^{-3/2} e^{s^2/2}, \quad p = \frac{q}{k_*}. \tag{4}$$

The saddle size k_* is a real number; thresholds involving it are rounded when used for integer sums.

Theorem 3.1 (Uniform tree composite law)**.** *Let $a_n \to \infty$ and $b_n = o(n^{1/6})$, with $a_n \le b_n$ eventually. Uniformly for $a_n \le s \le b_n$, with the parameters in* (4)*,*

$$\frac{N_n}{n} = \frac{e}{2}(1+p)(1+o(1)), \tag{5}$$

$$\frac{B_n}{n} = \frac{e}{2}(1+q)(1+o(1)), \tag{6}$$

$$\frac{A_n}{n} = \frac{e}{2}\left(\frac{2}{3} + q\right)(1+o(1)). \tag{7}$$

In particular,

$$\frac{R_{n,\lambda}}{n} = \frac{3}{4} \frac{(1+p)(2/3+q)}{(1+q)^2} (1+o(1)). \tag{8}$$

Every error displayed here is relative and uniform on the stated interval; p and q need not remain bounded.

We first record the constants from the small-size contribution. The classical rooted and unrooted labelled-tree series satisfy

$$T(z) = \sum_{k \ge 1} \frac{k^{k-1}}{k!} z^k = z e^{T(z)}, \qquad U(z) = \sum_{k \ge 1} \frac{k^{k-2}}{k!} z^k = T - \frac{T^2}{2}.$$

The rooted identity follows by decomposing at the root into an unordered set of rooted branches; integrating $T(z)/z$ gives U. A further integration gives

$$V(z) = \sum_{k \ge 1} \frac{k^{k-3}}{k!} z^k = T - \frac{3}{4} T^2 + \frac{1}{6} T^3.$$

Indeed $dz/z = (1-T)dT/T$, so integrating $U(z)/z$ gives this last polynomial. On the nonnegative branch at $z = 1/e$ one has $T = 1$, $U = 1/2$, and $V = 5/12$. The convergent critical sums therefore are

$$\sum_{k\geq 1} \frac{e^{1-k}k^{k-2}}{k!} = \frac{e}{2}, \quad \sum_{k\geq 1}(k-1)\frac{e^{1-k}k^{k-2}}{k!} = \frac{e}{2}, \quad \sum_{k\geq 1} h(k)\frac{e^{1-k}k^{k-2}}{k!} = \frac{e}{3}. \tag{9}$$

The corresponding sum with factor k is e.

Theorem 3.1 rests on the following decomposition, which later results also use. Write $t_{n,k}$ for the normalized weights in (10) below and $H_r = e\sqrt{n}\, k_*^{r-5/2} e^{s^2/2}$.

Lemma 3.2 (Three-range decomposition)**.** *Under the hypotheses of Theorem 3.1, call $k \leq 1/\eta$ the low range, $1/\eta < k < k_*/2$ the intermediate range and $k \geq k_*/2$ the high range. Uniformly for $a_n \leq s \leq b_n$:*

1. *for every fixed function g with $0 \leq g(k) \leq k$, the low-range sum of $g(k)t_{n,k}$ equals*

$$\sum_{k\geq 1} g(k)\frac{e^{1-k}k^{k-2}}{k!} + o(1);$$

2. *for $r = 0, 1$, the intermediate-range sum of $k^r t_{n,k}$ is $o(1 + H_r)$, and $\sum_{k=1}^{n} k^r t_{n,k} = O(1 + H_r)$;*

3. *for $r = 0, 1$, the high-range sum of $k^r t_{n,k}$ is $H_r(1 + o(1))$, and for every bounded continuous ϕ on $\mathbb{R}$ and every choice of $s = s_n \in [a_n, b_n]$,*

$$\frac{\sum_{k\geq k_*/2} t_{n,k}\, \phi((k-k_*)/\sqrt{n})}{\sum_{k\geq k_*/2} t_{n,k}} \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \phi(z) e^{-z^2/2}\, dz.$$

Proof of Lemma 3.2 and Theorem 3.1. Divide each size weight by n and write

$$t_{n,k} := \frac{w_{n,k}\lambda^{k-1}}{n} = \frac{c^{k-1}k^{k-2}}{k!} \prod_{j=0}^{k-1} \left(1 - \frac{j}{n}\right), \qquad 1 \leq k \leq n. \tag{10}$$

At $k = 1$ the formula means $t_{n,1} = 1$. For $r = 0, 1$ put

$$S_r = \sum_{k=1}^{n} k^r t_{n,k}.$$

We prove

$$S_0 = \frac{e}{2} + H_0 + o(1 + H_0), \qquad S_1 = e + H_1 + o(1 + H_1), \tag{11}$$

and keep track of the low and high pieces for the two other moments. All constants and limiting estimates below are uniform for $a_n \leq s \leq b_n$. In particular $\eta \to 0$, $s \to \infty$, and $k_*^3/n^2 \leq b_n^3/\sqrt{n} \to 0$.

For the low range $k \leq 1/\eta$, the factor $e^{\eta k}$ is bounded by e, and the product in (10) is at most one. Stirling bounds dominate $k^r t_{n,k}$ by $Ck^{r-5/2}$ for $r = 0, 1$. This is summable. At each fixed k the weight tends uniformly to $e^{1-k}k^{k-2}/k!$, and the cutoff tends uniformly to infinity. Dominated summation shows that the low range contributes its full critical sum, with uniform additive error $o(1)$. The same conclusion holds with any fixed factor g satisfying $0 \leq g(k) \leq k$, in particular with $k - 1$ and $h(k)$; this is part (i) of Lemma 3.2.

Next consider $1/\eta < k < k_*/2$. For every $k \le n$,

$$\log \prod_{j<k}(1-j/n) \le -\frac{k(k-1)}{2n}.$$

Stirling bounds thus give

$$k^r t_{n,k} \le C k^{r-5/2} \exp\left\{\eta k - \frac{k^2}{2n} + \frac{k}{2n}\right\}. \tag{12}$$

The last positive term is bounded by $1/2$. On $[1/\eta, 5/\eta]$ the sum is $O(\eta^{3/2-r}) = o(1)$. On $[5/\eta, k_*/2]$ the function

$$F_r(k) = k^{r-5/2} \exp\{\eta k - k^2/(2n)\}$$

is increasing once n is large. To see this, multiply its logarithmic derivative by k: the result is the concave quadratic $\eta k - k^2/n - (5/2 - r)$, positive at both endpoints for large s. The sum is consequently bounded by its number of terms times its upper-end value, giving

$$\sum_{1/\eta<k<k_*/2} k^r t_{n,k} \le o(1) + C k_*^{r-3/2} e^{3s^2/8} = o(1) + O\big(H_r s e^{-s^2/8}\big).$$

Empty intervals and rounded endpoints cause no change. This is $o(1 + H_r)$ uniformly, even when the intermediate sum is not absolutely small.

For the high range $k \ge k_*/2$, let $z = (k - k_*)/\sqrt{n}$. Uniform Stirling estimates give

$$t_{n,k} = \frac{e}{\sqrt{2\pi}}(1 + O(\eta) + O(k^{-1}))k^{-5/2} e^{\eta k} \prod_{j<k}(1 - j/n).$$

On every bounded z interval the logarithm of the product is $-k^2/(2n) + o(1)$ uniformly: the remainder is $O(k/n + k^3/n^2) = o(1)$ there. Also $k/k_* = 1 + z/s \to 1$ uniformly. Completing the square yields

$$\eta k - \frac{k^2}{2n} = \frac{s^2}{2} - \frac{z^2}{2}.$$

After dividing the summand $k^r t_{n,k}$ by $k_*^{r-5/2} e^{s^2/2}$, the global bound (12) and $k/k_* \ge 1/2$ give the majorant $Ce^{-z^2/2}$, for both $r = 0, 1$. On bounded z intervals the normalized summand tends uniformly to $(e/\sqrt{2\pi})e^{-z^2/2}$. Its mesh is $1/\sqrt{n}$; the lower boundary $-s/2$ tends uniformly to minus infinity, and the upper boundary tends uniformly to plus infinity. Gaussian mesh-tail bounds remove $|z| > M$ uniformly. Local Riemann sums, followed by $M \to \infty$, therefore prove

$$\sum_{k \ge k_*/2} k^r t_{n,k} = H_r(1 + o(1)), \qquad r = 0, 1. \tag{13}$$

This is a relative estimate even when $H_r \to 0$. Inserting the bounded factor $\phi((k - k_*)/\sqrt{n}) = \phi(z)$ in the same Riemann sums, and dividing by the case $\phi = 1$, proves part (iii) of Lemma 3.2; part (ii) is the intermediate-range bound above together with the other two ranges.

The three ranges prove (11). In the high range replacing k by $k-1$ changes the sum by its zeroth mass; replacing it by $h(k)$ changes it by at most three times that mass. Since $H_0/H_1 = 1/k_* \to 0$, these changes are $o(H_1)$. The intermediate ranges are bounded by the $r = 1$ estimate and the low ranges have the constants in (9). It follows that

$$\frac{B_n}{n} = \frac{e}{2} + H_1 + o(1 + H_1), \qquad \frac{A_n}{n} = \frac{e}{3} + H_1 + o(1 + H_1).$$

Now $H_0 = (e/2)p$ and $H_1 = (e/2)q$. The positive background constants convert these additive estimates to the relative forms (5)–(7). Substitution into (3), using $(n-1)/(n-2) \to 1$, proves (8). □

Remark 3.3. The first and second size moments S_1 and S_2 are the one-point function and the susceptibility of Liu and Slade [3]. The high-range term H_2 formally coincides with $n^{1/4}$ times the large-s form $e\, s^{-1/2} e^{s^2/2}$ of their profile, while $H_1 = H_2/k_*$ is a correction of order $n^{-1/4}$ at fixed s, below the resolution of their one-point limit e, which matches the background of S_1. Their parameter $s_{\mathrm{LS}} = \sqrt{n}(e^{s/\sqrt{n}} - 1)$ satisfies $s_{\mathrm{LS}}^2 - s^2 = O(s^3/\sqrt{n}) = o(1)$ in the present range. Their scaling theorems hold for fixed s. The increments here are uniformity as $s \to \infty$ with $s = o(n^{1/6})$, and the unrooted mass S_0, whose crossover is delayed by the factor k_*.

4 The first-moment and mass windows

Corollary 4.1 (First-moment window)**.** *If*

$$s^2 = \frac{1}{2}\log n + \frac{3}{2}\log\log n + x_n, \qquad x_n \to x, \qquad Q = 2^{7/4} e^{x/2},$$

then $q \to Q$, $p \to 0$, *and*

$$\frac{R_{n,\lambda}}{n} \longrightarrow \frac{3}{4}\frac{2/3 + Q}{(1+Q)^2}.$$

Moreover K *converges weakly to the probability distribution*

$$\pi_k = \frac{2e^{-k}k^{k-2}}{k!}, \qquad k \geq 1, \tag{14}$$

the Borel law with parameter one reweighted by $1/k$, *whose mean is* 2, *whereas* $\mathbb{E}K \to 2 + Q$.

Proof. Substitution into (4) gives $q \to Q$; the powers of $\log n$ cancel because $s \sim \sqrt{(\log n)/2}$. Also $p = q/k_* \to 0$. The ratio limit follows from Theorem 3.1. For each fixed k, divide (10) by $N_n/n \to e/2$ to obtain (14). Its probabilities sum to one by $U(1/e) = 1/2$, so pointwise convergence implies weak convergence on the positive integers. Its mean is $2T(1/e) = 2$. Finally (6) and (5) give $\mathbb{E}(K-1) \to 1 + Q$. In particular, for each fixed integer M,

$$\lim_{n\to\infty} \mathbb{E}[K\mathbf{1}_{\{K>M\}}] = 2 + Q - \sum_{k\leq M} k\pi_k.$$

Letting $M \to \infty$ leaves $Q > 0$, proving failure of uniform integrability. A vanishing-probability high-size contribution carries the extra mean. □

Corollary 4.2 (Mass window)**.** *If*

$$s^2 = \frac{3}{2}\log n + \frac{5}{2}\log\log n + y_n, \qquad y_n \to y, \qquad P = 2(3/2)^{-5/4} e^{y/2},$$

then $p \to P$, $q \to \infty$, *and*

$$\frac{k_*}{n} R_{n,\lambda} \longrightarrow \frac{3}{4}\frac{1+P}{P}, \qquad \frac{K}{k_*} \;\Rightarrow\; \frac{1}{1+P}\delta_0 + \frac{P}{1+P}\delta_1.$$

Proof. Direct substitution gives the stated value of P; here $s \sim \sqrt{(3\log n)/2}$ and $q = k_* p \to \infty$. Multiplying (8) by k_* gives the ratio limit. By Lemma 3.2 the low zeroth mass tends to $e/2$, the intermediate mass is $o(1) + o(H_0) = o(1)$, and the high mass tends to $eP/2$. For a bounded continuous function f on $[0, \infty)$, the low range contributes $(e/2)f(0) + o(1)$ to the unnormalized sum with test factor $f(k/k_*)$: dominated summation reduces it to fixed k, where $k/k_* \to 0$; on the low range $k/k_* \le 1/(\eta k_*) = 1/s^2$. In the high range, on bounded z intervals $k/k_* = 1 + z/s \to 1$ uniformly. By part (iii) of Lemma 3.2, z is tight under the high-range weights, and $f(1+z/s) \to f(1)$ uniformly on compact z-sets, so the high range contributes $(eP/2)f(1) + o(1)$. The intermediate part vanishes by boundedness of f. Dividing by total mass $e(1 + P)/2 + o(1)$ proves the mixture law. □

Corollary 4.3 (Conditional Gaussian fluctuations)**.** *For every choice of $s = s_n \in [a_n, b_n]$ in the range of Theorem 3.1, conditional on $K \ge k_*/2$,*

$$\frac{K - k_*}{\sqrt{n}} \;\Rightarrow\; \mathcal{N}(0, 1).$$

This holds whether or not the conditioning event has positive limiting probability.

Proof. The conditional law of K given $K \ge k_*/2$ has weights proportional to $t_{n,k}$ on the high range, so this is part (iii) of Lemma 3.2, whose estimate is relative. □

Both windows lie in $s \to \infty$, $s = o(n^{1/6})$. Their distinct locations reflect the factor k_* between the saddle's first moment and its total mass.

5 Higher moments and fixed forest markings

Lemma 5.1 (Every fixed higher size moment)**.** *Under the uniform hypotheses of Theorem 3.1, for each fixed integer $d \ge 2$,*

$$S_d := \sum_{k=1}^{n} k^d t_{n,k} = H_d(1 + o(1)), \qquad H_d = e\sqrt{n}\, k_*^{d-5/2} e^{s^2/2} = k_*^d H_0.$$

The error is uniform for $a_n \le s \le b_n$, for each fixed d.

Proof. The bound (12) holds with r replaced by d. For $k \le 5/\eta$, summation gives $O(\eta^{3/2-d})$, since $d - 5/2 > -1$. Relative to H_d this is $O(s^{4-2d} e^{-s^2/2}) = o(1)$ uniformly. On $[5/\eta, k_*/2]$ the function $k^{d-5/2} \exp\{\eta k - k^2/(2n)\}$ is increasing. For $d = 2$, multiplying its logarithmic derivative by k gives a concave quadratic positive at both endpoints for large s; for $d \ge 3$, its logarithmic derivative is at least $\eta/2 > 0$ throughout this interval. The intermediate sum is bounded by $C k_*^{d-3/2} e^{3s^2/8}$, whose ratio to H_d is $O(s e^{-s^2/8}) = o(1)$.

In the high range the bounded-z expansion in the proof of Theorem 3.1 is unchanged. The normalized summand tends to $(e/\sqrt{2\pi}) e^{-z^2/2}$ on bounded z intervals. For positive z, $k/k_* = 1 + z/s \le 1 + z$ eventually; for negative z in this range, $1/2 \le k/k_* \le 1$. Hence a uniform integrable majorant is $C(1 + |z|^d) e^{-z^2/2}$. Polynomial Gaussian mesh-tail bounds and local Riemann sums prove that the high sum is $H_d(1 + o(1))$. Combining the three ranges proves the lemma. □

For a fixed forest F with no isolated vertices, write r for its number of incident vertices, m for its number of edges, and $d = r - m$ for its number of components. Let their orders be $r_1, \dots, r_d$

and put $c_F = \prod_i r_i$. The locations of its vertices in K_n are prescribed. The same classical extension count as in (2) gives the exact conditional identity

$$\mathbb{P}(F \subseteq T \mid K = k) = \frac{c_F}{(n)_r} \frac{(k)_r}{k^m}, \tag{15}$$

interpreted as zero for $k < r$, where $(x)_r = x(x-1)\cdots(x-r+1)$. Indeed the selected set contains the r prescribed vertices with probability $(k)_r/(n)_r$, and the conditional tree contains their forest with probability c_F/k^m.

Corollary 5.2 (Every fixed disconnected forest)**.** *For every fixed nonempty forest F with $d \geq 2$ components and no isolated vertices, uniformly in the range of Theorem 3.1,*

$$\mathbb{P}(F \subseteq T) = \frac{c_F}{n^r} \frac{H_d}{e/2 + H_0}(1 + o(1)) = \frac{c_F k_*^d}{n^r} \frac{p}{1+p}(1 + o(1)). \tag{16}$$

The forest is fixed before taking the limit.

Proof. Set $g_F(k) = (k)_r/k^m$ for $k \geq r$ and $g_F(k) = 0$ otherwise. For fixed F, polynomial expansion, with the finitely many smaller sizes absorbed into the constant, gives

$$|g_F(k) - k^d| \leq C_F k^{d-1} \qquad (k \geq 1).$$

If $d \geq 3$, Lemma 5.1 makes the error in its weighted sum $O(H_{d-1}) = o(H_d)$. If $d = 2$, part (ii) of Lemma 3.2 gives error $O(1+H_1) = o(H_2)$: here $H_1/H_2 = 1/k_* \to 0$ and $H_2 = en^{1/4}s^{-1/2}e^{s^2/2} \to \infty$ uniformly. Thus $\sum_k g_F(k) t_{n,k} = H_d(1 + o(1))$. Average (15) and use $(n)_r = n^r(1 + o(1))$ and (5). Finally $H_d = k_*^d H_0$ and $H_0 = (e/2)p$ give the second expression. □

A connected tree marking has $d = 1$, and its finite critical background depends on its number of vertices r. It is therefore not covered by the higher-moment simplification in Corollary 5.2. The edge and wedge backgrounds used in (3) were evaluated explicitly above.

Corollary 5.3 (Every fixed connected tree marking)**.** *Let F be a fixed nonempty tree on $r \geq 2$ prescribed vertices, and set*

$$g_r(k) = \begin{cases} (k)_r/k^{r-1}, & k \geq r, \\ 0, & k < r, \end{cases} \qquad D_r = \sum_{k \geq r} \frac{e^{1-k} k^{k-2}}{k!} g_r(k).$$

Then $D_r = e/r$, and uniformly in the range of Theorem 3.1,

$$\mathbb{P}(F \subseteq T) = \frac{1}{n^r} \frac{2 + rq}{1+p}(1 + o(1)). \tag{17}$$

In particular $D_2 = e/2$ and $D_3 = e/3$. Together with Corollary 5.2, this gives a profile for every fixed nonempty forest with no isolated vertices.

Proof. The inequality $0 \leq g_r(k) \leq k$ makes the series defining D_r convergent. Its value is explicit. Put $k = j + r$ and cancel the falling factorial to obtain

$$D_r = e^{1-r} \sum_{j \geq 0} \frac{(j+r)^{j-1}}{j!} e^{-j}.$$

Lagrange inversion applied to $T = ze^T$ gives

$$e^{rT(z)} = \sum_{j\geq 0} \frac{r(j+r)^{j-1}}{j!} z^j.$$

For $j \geq 1$, its coefficient is $(r/j)[u^{j-1}]e^{(j+r)u} = r(j+r)^{j-1}/j!$; the $j = 0$ term is one as well. Letting $z \uparrow 1/e$ and using nonnegative coefficients and $T(1/e) = 1$ gives $D_r = e^{1-r}e^r/r = e/r$. Part (i) of Lemma 3.2, with factor g_r, gives $D_r + o(1)$ for the low range, and part (ii) bounds the intermediate range by $o(1 + H_1)$. For all positive integers k, $|g_r(k) - k| \leq C_r$, after enlarging the constant for $k < r$. Therefore the high sum differs from the high first moment by $O(H_0) = o(H_1)$. Its total is consequently $D_r + H_1 + o(1 + H_1)$. Apply (15) with $m = r - 1$ and $c_F = r$, and divide by (5). Substituting $D_r = e/r$ and $H_1 = (e/2)q$ gives the stated formula, with relative error because $D_r > 0$. The identities $g_2(k) = k - 1$ and $g_3(k) = h(k)$, including their small-size values, give the two explicit constants from (9). □

6 Connected graphs and marked observables

We now give every nonempty connected simple labelled subgraph G of K_n weight $\lambda^{|E(G)|}$, retaining the singleton graphs. Write N^{c} for its partition function and $\mathbb{P}_{\mathrm{c}}$ for the resulting probability measure. Superscript t denotes the tree model. The size variable is still $K = |V(G)|$. In this section

$$\eta = \frac{s}{\sqrt{n}}, \qquad \lambda = \frac{e^{-1+\eta}}{n}, \qquad k_* = s\sqrt{n}, \qquad a_n \leq s \leq b_n, \quad a_n \longrightarrow \infty, \quad b_n = o(n^{1/6}).$$

Every asymptotic statement below is uniform over this interval. Its endpoints may depend on n.

6.1 Surplus enumeration and size tails

Let $C_k(\lambda)$ be the sum of the weights of connected graphs on one fixed labelled k-set, and put

$$t_k(\lambda) = k^{k-2}\lambda^{k-1}, \qquad t_1(\lambda) = 1, \qquad D_k(\lambda) = \frac{C_k(\lambda)}{t_k(\lambda)} - 1.$$

For $k = 1, 2$, $D_k = 0$. We use the following published enumeration input.

Lemma 6.1 (Surplus bound)**.** *There is a constant C such that, for $k \geq 3$, $z > 0$, $zk^{3/2} \leq 1$ and $kz \leq 3/e$,*

$$0 \leq D_k(z) \leq Czk^{3/2} + Cz^2.$$

Let $C(k, m)$ count connected labelled graphs on k vertices with m edges, and put $S(k, z) = \sum_{\ell\geq 1} C(k, k-1+\ell)z^\ell$, so that $D_k(z) = k^{-(k-2)}S(k, z)$. Liu and Slade split $S(k, z) \leq k^{k-2}(A(k, z) + B(k, z))$ into sparse and dense parts [3, Section 3.2, (3.15)–(3.16)], with their vertex count n in the role of our k. Their Lemma 3.3(i) with $b = 1$ gives $A(k, z) \leq Czk^{3/2}$ when $zk^{3/2} \leq 1$, and their Lemma 3.4 gives $B(k, z) \leq Cz^2$ when $kz \leq 3/e$. Together these give Lemma 6.1.

Let $t_{n,k} = n^{-1}\binom{n}{k}t_k(\lambda)$ be the normalized tree size weight, and define the connected size weight

$$a_{n,k} = \frac{1}{n}\binom{n}{k} C_k(\lambda) = t_{n,k}(1 + D_k(\lambda)). \tag{18}$$

Both singleton weights equal one. The following bounds will also control observables whose probabilities vanish with n.

Lemma 6.2 (Local surplus and global tails)**.** *For each fixed $L > 0$,*

$$\sup_{k \le Lk_*} D_k(\lambda) = o(1).$$

For $r = 0, 1$, with H_r as in Lemma 3.2,

$$\sum_{k > 8k_*} k^r a_{n,k} = o(H_r). \tag{19}$$

Moreover, given fixed $D > 0$ and integer $J \ge 0$, take $C_0 > 8(D + J/2 + 2)$ and

$$M = \left\lceil \max\{8k_*, \sqrt{C_0 n \log n}\} \right\rceil. \tag{20}$$

Then

$$\sup_{k \le M} D_k(\lambda) = o(1), \qquad \lambda M = o(1), \qquad \sum_{k > M} k^J a_{n,k} = o(n^{-D}). \tag{21}$$

Proof. For $k \le Lk_*$,

$$\lambda k^{3/2} \le C n^{-1/4} b_n^{3/2} = o(1), \qquad k\lambda = o(1),$$

so Lemma 6.1 applies uniformly. The same argument applies up to M: the additional cutoff contributes at most $Cn^{-1/4}(\log n)^{3/4} = o(1)$ to $\lambda M^{3/2}$. It also gives $\lambda M = o(1)$.

To bound sizes outside this local regime, sum over a spanning tree and then all choices of further edges. Every connected graph is counted at least once, so, for $k \ge 2$,

$$1 + D_k(\lambda) \le (1 + \lambda)^{\binom{k}{2} - k + 1} \le \exp(\lambda k^2/2). \tag{22}$$

Stirling's upper bound and $\prod_{j<k}(1 - j/n) \le \exp[-k(k-1)/(2n)]$ give

$$k^r a_{n,k} \le C k^{r-5/2} \exp\left\{ \eta k - \frac{(1-c)k^2}{2n} + \frac{k}{2n} \right\}, \qquad c = e^{-1+\eta}.$$

Here $c \le 1/2$ eventually, uniformly. For $k \ge 8k_*$, $\eta k \le k^2/(8n)$ and $\exp(k/(2n)) \le e^{1/2}$, whence

$$k^r a_{n,k} \le C k^{r-5/2} e^{-k^2/(8n)}. \tag{23}$$

For $r = 0, 1$ the tail sum is bounded by

$$C n^{r/2 - 3/4} \int_{7s}^{\infty} u^{r-5/2} e^{-u^2/8}\, du.$$

The replacement of $8s$ by $7s$ absorbs the lattice endpoint uniformly. Dividing this bound by

$$H_r = e n^{r/2-3/4} s^{r-5/2} e^{s^2/2}$$

proves (19), since $s \ge a_n \to \infty$.

The same Gaussian bound holds for any fixed power $r = J$. For $J \ge 3$ the integrand $u^{J-5/2} e^{-u^2/8}$ is decreasing for $u \ge 2\sqrt{J}$, which holds beyond the cutoff for large n, so the sum-to-integral comparison still applies. Beyond M, the integral starts at least at $\sqrt{C_0 \log n} - n^{-1/2}$. Its bound is $n^{J/2 - 3/4 - C_0/8}$ times a fixed power of $\log n$. The strict choice of C_0 makes this $o(n^{-D})$. The endpoint error in the exponent is $O(M/n) = o(1)$. These estimates are uniform, and complete the proof. □

Proposition 6.3 (Transfer of size laws)**.** *The connected model has the same relative zeroth and first size-moment asymptotics as the tree model in the stated moderate range. Its size law and the tree size law have total variation distance $o(1)$. Their conditional size laws given $K \geq k_*/2$ also have total variation distance $o(1)$, even if this event has probability tending to zero. Consequently the two size-window limits and their conditional Gaussian limits proved for trees hold for connected graphs, with the same constants. The first-moment-window limit of $\mathbb{E}K$, and its escaped-mean identity, hold as well.*

Proof. For $r = 0, 1$, the excess normalized moment is

$$\Delta_r = \sum_k k^r t_{n,k} D_k(\lambda) \geq 0.$$

Up to $8k_*$, the local surplus bound and the estimate $\sum k^r t_{n,k} = O(1 + H_r)$ of Lemma 3.2 give $o(1 + H_r)$. Beyond $8k_*$, Lemma 6.2 and $t_{n,k} \leq a_{n,k}$ give $o(H_r)$. Thus

$$\Delta_r = o(1 + H_r), \qquad r = 0, 1. \tag{24}$$

For the functions $k - 1$ and $h(k) = k - 3 + 2/k$, their nonnegative excesses are bounded by Δ_1. Their tree moment expansions have strictly positive critical backgrounds, so the error is relative $o(1)$. The same holds for the zeroth moment.

The excess partition function is the total weight of cyclic graphs. Equation (24) for $r = 0$ shows that its fraction of the total is $o(1)$. Conditional on being a tree, the connected law is exactly the original tree law. This proves the unconditional total variation comparison. On $k_*/2 \leq k \leq 8k_*$ the excess zeroth weight is $o(1)$ times the tree weight there, which is $H_0(1+o(1))$. The far tail is $o(H_0)$ by Lemma 6.2. Thus the excess on the whole high-size event is $o(H_0)$, proving the conditional comparison.

Total variation transfers bounded-test weak limits, including the mixture and conditional Gaussian. The moment comparison for $r = 1$ separately transfers the mean in the first-moment window. Convergence of each fixed truncated mean, together with that full mean limit, then transfers the escaped-mean identity. □

6.2 Deleting a fixed marked forest

Size-law comparison alone does not compare probabilities of prescribed edges. We prove a relative comparison for every fixed forest. Let F have $m \geq 1$ edges and r incident labelled vertices, with no isolated vertices in its specified pattern, and let c_F be the product of its component sizes. The pattern F is fixed as n varies; its ambient labels may be chosen arbitrarily and distinctly. On a fixed k-set containing those labels, let $M_F(k, \lambda)$ denote the weight of connected graphs containing its edges. Its tree counterpart is

$$T_F(k, \lambda) = c_F k^{k-m-2} \lambda^{k-1}, \qquad k \geq r. \tag{25}$$

To see this classical prescribed-forest Cayley identity, contract all components of F, including the other $k - r$ singleton vertices. If their sizes are $\sigma_1, \ldots, \sigma_{k-m}$, a tree on the components contributes $\prod_i \sigma_i^{d_i}$ choices of its endpoint labels. The weighted Cayley identity, obtained from the Prüfer code, sums these choices to $(\sum_i \sigma_i)^{k-m-2} \prod_i \sigma_i = k^{k-m-2} c_F$. When there is only one component the forest is already a spanning tree; the same displayed expression equals one.

Lemma 6.4 (Fixed-forest comparison)**.** *Suppose $C_j(\lambda) \le (1+\epsilon)t_j(\lambda)$ for $1 \le j \le \kappa$. For $r \le k \le \kappa$,*

$$1 \le \frac{M_F(k,\lambda)}{T_F(k,\lambda)} \le (1+\epsilon)^{m+1} + \sum_{j=1}^{m} \frac{\binom{r}{j} j}{c_F} (1+\epsilon)^j (\lambda k)^{m+1-j}. \tag{26}$$

In particular this ratio tends uniformly to one if $\epsilon = o(1)$ and $\lambda\kappa = o(1)$.

Proof. We use the classical rooted spanning-forest identity: for a specified j-set of roots in a k-set, the number of forests with one root per component is jk^{k-j-1}. Indeed the rooted matrix-tree theorem gives the determinant of the complete-graph Laplacian minor obtained by deleting the root rows and columns. This minor is $kI - J$ of size $k-j$, with eigenvalues k of multiplicity $k-j-1$ and j of multiplicity one. Its determinant is the asserted count. If $j = k$, the empty minor equals one, consistently with the formula.

Delete the m prescribed edges of F from a connected graph containing it. Every remaining component contains a marked vertex, since a component without any marked vertex could not have been attached by a deleted edge. There are $j \le m+1$ components. Reinserting the edges of F gives a connected quotient multigraph on those components.

If $j = m+1$, that quotient has j vertices and $j-1$ edges, hence is a tree. In particular it has no loops and no parallel-edge cycles. All prescribed edges run between components, so the component weights are unrestricted connected weights. Replacing each by its tree weight gives precisely the trees containing F: deletion and reinsertion are inverse. This leading contribution is bounded by $(1+\epsilon)^{m+1}T_F$.

For $j \le m$, we may ignore both quotient connectivity and any forbidden internal occurrence of a deleted edge, since we need an upper bound. After replacing each connected component weight by at most $(1+\epsilon)$ times its tree weight, count tree forests whose components each contain at least one of the r marks. Choose one marked root in each component and take a union over all $\binom{r}{j}$ root sets. The rooted forest identity bounds their number by $\binom{r}{j}jk^{k-j-1}$. Multiple counting only enlarges this bound. Multiplication by λ^{k-j} for the forest edges and λ^m for reinsertion, followed by division by (25), gives the jth term of (26). The lower bound follows by including all trees. □

Theorem 6.5 (Fixed-forest inclusion probabilities)**.** *For every fixed nonempty labelled forest F as above,*

$$\frac{\mathbb{P}_{\mathrm{c}}(F \subseteq G)}{\mathbb{P}_{\mathrm{t}}(F \subseteq T)} = 1 + o(1), \tag{27}$$

uniformly throughout the stated moderate range. Here inclusion prescribes edges, not an induced subgraph, and the forest may be disconnected.

Proof. Use Lemma 6.2 with $J = 0$, $D = r+2$ and $C_0 = 16(r+4)$. At its cutoff M, $\epsilon = \sup_{j \le M} D_j(\lambda) = o(1)$ and $\lambda M = o(1)$. Lemma 6.4 compares all marked weights below M with relative error $o(1)$. Multiplication by $\binom{n-r}{k-r}$ and summation preserve that comparison for the ambient marked numerators below M.

The raw connected partition tail is

$$T_{>M} = \sum_{k>M} \binom{n}{k} C_k(\lambda) = n \sum_{k>M} a_{n,k} = o(n^{-r-1}).$$

Each fixed-forest marked tail is at most this quantity, because the inclusion event is a subset of the tail outcomes. Tree tails are no larger than connected tails. If F is disconnected, F itself is not an admissible connected outcome. Instead join its components by fixed edges on the same r

marked vertices to obtain a tree containing F. This tree contributes λ^{r-1} to the full marked tree numerator. Since $\lambda \sim 1/(en)$ uniformly, both tail numerators are negligible relative to that lower bound. Hence the full connected marked numerator is its tree counterpart times $1 + o(1)$.

Finally, Proposition 6.3 gives a partition-function ratio of $1 + o(1)$. Dividing proves the probability comparison. The empty forest, if included, has inclusion probability one in both models and requires no argument. The comparison concerns forests, since a cyclic prescribed pattern has zero tree probability. □

Corollary 6.6 (Edge and adjacent-edge profiles)**.** *For prescribed distinct adjacent edges e, f,*

$$\frac{\mathbb{P}_{\mathrm{c}}(e)}{\mathbb{P}_{\mathrm{t}}(e)} = 1 + o(1), \qquad \frac{\mathbb{P}_{\mathrm{c}}(e,f)}{\mathbb{P}_{\mathrm{t}}(e,f)} = 1 + o(1), \qquad \frac{R^{\mathrm{c}}_{n,\lambda}}{R^{\mathrm{t}}_{n,\lambda}} = 1 + o(1), \tag{28}$$

where $R^{\mathrm{c}}_{n,\lambda} = \mathbb{P}_{\mathrm{c}}(e,f)/\mathbb{P}_{\mathrm{c}}(e)^2$; by label symmetry $\mathbb{P}_{\mathrm{c}}(e) = \mathbb{P}_{\mathrm{c}}(f)$. With p, q as in Theorem 3.1,

$$\frac{R^{\mathrm{c}}_{n,\lambda}}{n} = \frac{3}{4}\frac{(1+p)(2/3+q)}{(1+q)^2}(1 + o(1)). \tag{29}$$

The individual probabilities have the relative forms

$$\mathbb{P}_{\mathrm{c}}(e) = \frac{2(1+q)}{n(n-1)(1+p)}(1 + o(1)), \tag{30}$$

$$\mathbb{P}_{\mathrm{c}}(e,f) = \frac{3(2/3+q)}{n(n-1)(n-2)(1+p)}(1 + o(1)). \tag{31}$$

Proof. Apply Theorem 6.5 to a single edge and to a three-vertex path, and substitute the tree formulas from Theorem 3.1 and the exact identities (2)–(3). □

At the first-moment window of Corollary 4.1, $q \to Q = 2^{7/4}e^{x/2}$ and $p \to 0$, so

$$\frac{R^{\mathrm{c}}_{n,\lambda}}{n} \longrightarrow \frac{3}{4}\frac{2/3+Q}{(1+Q)^2}.$$

At the mass window of Corollary 4.2, $p \to P = 2(3/2)^{-5/4}e^{y/2}$ and $q = k_* p \to \infty$, so

$$\frac{k_*}{n}R^{\mathrm{c}}_{n,\lambda} \longrightarrow \frac{3}{4}\frac{1+P}{P}.$$

Remark 6.7 (Two disjoint edges). The fixed-forest comparison and Corollary 5.2 also give a different profile for prescribed disjoint edges e_1, e_2. Their forest has $r = 4$, $d = 2$ and $c_F = 4$, so in either model

$$R^{\mathrm{disj}}_{n,\lambda} := \frac{\mathbb{P}(e_1,e_2)}{\mathbb{P}(e_1)\mathbb{P}(e_2)} = \frac{k_* q(1+p)}{(1+q)^2}(1 + o(1)).$$

At the first-moment window, $R^{\mathrm{disj}}_{n,\lambda}/k_* \to Q/(1+Q)^2$; at the mass window, $R^{\mathrm{disj}}_{n,\lambda} \to (1+P)/P$. Thus adjacent and disjoint edge pairs probe different size moments and have different normalizations in the same two windows.

7 Regimes of the adjacent-edge ratio

In the tree model the adjacent-edge ratio takes the following forms. For fixed $\lambda > 0$ it is $3/4 + 3/(4e\lambda n) + o(1/n)$ [6]. For $\lambda = c/n$ with fixed $c \le 1/e$ the normalized weights are dominated by $Ck^{-5/2}$, so by dominated convergence the size moments in (3) converge to positive limits and R is of order n; at $c = 1/e$, (9) gives $R/n \to 1/2$. In the critical window of bounded s the profile of Liu and Slade [3] applies. For $\lambda = e^{-1+s/\sqrt{n}}/n$ with $s \to \infty$ and $s = o(n^{1/6})$, Corollary 4.1 gives the limit of R/n near the first-moment scale; between the two scales, where $q \to \infty$ and $p \to 0$, Theorem 3.1 gives $R \sim \frac{3}{4}n/q$, and beyond the mass scale, where $p \to \infty$, it gives $R \sim \frac{3}{4}n/k_*$. The cutoff $s = o(n^{1/6})$ keeps $k_*^3/n^2 \to 0$, so that the product in (10) is Gaussian to leading order at the saddle.

Assistance and data availability

AI-assisted tools were used in mathematical exploration and manuscript preparation. The arguments supporting the conclusions are contained in this article, with the external enumeration input explicitly identified. The proof uses no numerical experiment or external dataset.

References

[1] A. Chin, G. Gordon, K. MacPhee and C. Vincent, *Random subtrees of complete graphs*, arXiv:1308.4613.

[2] L. Fredes and J.-F. Marckert, *Models of random subtrees of a graph*, Probability Surveys **20** (2023), 722–801. doi:10.1214/23-PS22.

[3] Y. Liu and G. Slade, *Critical scaling profile for trees and connected subgraphs on the complete graph*, Canadian Mathematical Bulletin **69** (2026), no. 2, 550–565. doi:10.4153/S0008439525100763; arXiv:2412.05503.

[4] Z. Luo, K. Xu and J. Tian, *Random subtrees and unimodal sequences in graphs*, Discrete Mathematics **347** (2024), no. 1, article 113654. doi:10.1016/j.disc.2023.113654.

[5] S. Wagner, *On the probability that a random subtree is spanning*, arXiv:1910.07349.

[6] S. Schubr, *A sharp lower bound for adjacent-edge correlations in random subtrees*, Luminary Research Archive, edition 4, 1 November 2026. `https://research.luminaryoddities.org/sharp-lower-bound-adjacent-edge-correlations-random-subtrees/v/4`.