

Growth of the Number of Spanning Trees of the Erdős–Rényi Giant Component

RUSSELL LYONS^{1†}, RON PELED^{2‡} and ODED SCHRAMM³

¹ Department of Mathematics, Indiana University, Bloomington, IN 47405-5701, USA
(e-mail: rdlyons@indiana.edu)

² Department of Statistics, University of California at Berkeley,
367 Evans Hall, Berkeley, CA 94720-3860, USA
(e-mail: peledron@stat.berkeley.edu)

³ Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA

Received 12 November 2007; revised 13 May 2008; first published online 30 June 2008

The number of spanning trees in the giant component of the random graph $\mathcal{G}(n, c/n)$ ($c > 1$) grows like $\exp\{m(f(c) + o(1))\}$ as $n \rightarrow \infty$, where m is the number of vertices in the giant component. The function f is not known explicitly, but we show that it is strictly increasing and infinitely differentiable. Moreover, we give an explicit lower bound on $f'(c)$. A key lemma is the following. Let $\text{PGW}(\lambda)$ denote a Galton–Watson tree having Poisson offspring distribution with parameter λ . Suppose that $\lambda^* > \lambda > 1$. We show that $\text{PGW}(\lambda^*)$ conditioned to survive forever stochastically dominates $\text{PGW}(\lambda)$ conditioned to survive forever.

1. Introduction

Methods of enumeration of spanning trees in a finite graph G and relations to various areas of mathematics and physics have been investigated for more than 150 years. The number of spanning trees is often called the *complexity* of the graph, denoted here by $\tau(G)$. The usual Erdős–Rényi model of random graphs, $\mathcal{G}(n, p)$, is a graph on n vertices, each pair of which is connected by an edge with probability p , independently of other edges. Fix $c > 1$. It is well known that with probability approaching 1 as $n \rightarrow \infty$, the largest component of $\mathcal{G}(n, c/n)$ has size proportional to n , while the second-largest component is of logarithmic size. (See, e.g., [4] or [3].) The largest component is thus called the *giant component* and will be denoted by $G_n = G_n(c)$. As an example of a general theory,

[†] Partially supported by NSF grants DMS-0406017 and DMS-0705518 and Microsoft Research.

[‡] Supported by Microsoft Research and NSF grant DMS-0605166.

Lyons [7] proved that there is a number $f(c)$ such that

$$f(c) = \lim_{n \rightarrow \infty} \frac{1}{|V(G_n)|} \log \tau(G_n)$$

in probability. In the same paper it was shown that $f(c) > 0$ for $c > 1$, that $f(1^+) = 0$, and that f is continuous on $[1, \infty)$. Lyons [7] asked whether f is strictly increasing and real analytic on $(1, \infty)$. Note that as c increases, both the number of trees $\tau(G_n)$ as well as the number of vertices $|V(G_n)|$ increase, so that it is not clear which increase dominates. Here we prove that f is strictly increasing and C^∞ ; prior to our work, it was not known even that f was non-decreasing.

Let $\text{PGW}(c)$ be the law of a rooted Galton–Watson tree (T, o) with $\text{Poisson}(c)$ offspring distribution. Write $\text{PGW}^*(c)$ for the law of $\text{PGW}(c)$ conditioned on non-extinction. Sometimes we also write this measure as PGW_c^* . The event of extinction has probability $q(c)$, which is well known to be the smallest positive solution of the equation

$$q(c) = e^{-c(1-q(c))}. \tag{1.1}$$

Let $p_k(x; G)$ denote the probability that simple random walk on a graph G started at a vertex x is back at x after k steps. Lyons [7] proved that

$$f(c) = \int \left(\log \deg_T(o) - \sum_{k \geq 1} \frac{1}{k} p_k(o; T) \right) d\text{PGW}_c^*(T, o). \tag{1.2}$$

Theorem 1.1. *The function f is strictly increasing and C^∞ on $(1, \infty)$. In fact,*

$$f'(c) > \frac{(c-1)e^{-cq(c)}}{c^2} > 0$$

for $c > 1$.

From (1.2), it is not hard to see that for any $\epsilon > 0$, we have that $(f(c + \epsilon) - f(c))/\epsilon \sim 1/c$ as $c \rightarrow \infty$. Since

$$ce^{-c} = cq(c)e^{-cq(c)} \tag{1.3}$$

and the function $x \mapsto xe^{-x}$ is unimodal in $(0, \infty)$ and vanishes at 0 and ∞ , it follows that $\lim_{c \rightarrow \infty} cq(c) = 0$. Using this, we find that our lower bound for $f'(c)$ in Theorem 1.1 has the same asymptotic, $1/c$, as $c \rightarrow \infty$. We do not have any information on $f'(1)$.

A key lemma to prove Theorem 1.1 is the following.

Theorem 1.2. *If $c' > c \geq 1$, then $\text{PGW}^*(c')$ stochastically dominates $\text{PGW}^*(c)$.*

Here, $\text{PGW}^*(1)$ denotes the weak limit as $c \downarrow 1$ of $\text{PGW}^*(c)$; see [1, Lemma 23].

We now recall what the stochastic domination referred to in the theorem means. If (T, o) and (T', o') are rooted trees, we say that (T, o) dominates (T', o') if there is an isomorphism from T' to a subtree of T that takes o' to o . A probability measure on the collection of rooted trees is said to stochastically dominate another probability measure on the collection of rooted trees if they may be coupled so that the sample from the first measure a.s. dominates the sample from the second measure.

Of course, $\text{PGW}(c')$ dominates $\text{PGW}(c)$ when $c' > c$. It is the conditioning that makes Theorem 1.2 non-trivial. Indeed, the offspring distribution that has 1 or 3 children with probability 1/2 each stochastically dominates the offspring distribution that has 0 or 3 children with probability 1/2 each, but if we condition on survival, the domination does not persist since conditioning does not change the former, but forces the latter to have 3 children of the root.

2. Tree domination

Let $T_n = T_n(\lambda)$ be a $\text{PGW}(\lambda)$ tree conditioned to have n vertices, where $n \in \mathbb{N}_+ \cup \{\infty\}$. We consider the values of T_n to be equivalence classes of rooted trees under isomorphisms that preserve the root. It is easy to check that the distribution of T_n does not depend on λ . It turns out that it is the same as the distribution obtained by forgetting the labels of a uniform tree on n vertices with uniform root. Also, the probability that a $\text{PGW}(\lambda)$ tree has k vertices is given by the Borel(λ) distribution, namely,

$$\frac{(\lambda e^{-\lambda})^k k^{k-1}}{\lambda k!}. \tag{2.1}$$

These facts are well known and have a variety of proofs; see Pitman [10] for some of them.

Luczak and Winkler [6] show the following.

Theorem 2.1. T_{n+1} stochastically dominates T_n for every $n \in \mathbb{N}_+$.

More precisely, in their Theorem 4.1, for each $d \geq 2$ they show such a statement for conditioned trees having offspring distribution binomial with parameters $(d, 1/d)$. Taking a limit as $d \rightarrow \infty$ gives Theorem 2.1. It is interesting to note that this is the same as saying that a uniformly rooted uniform tree on $n + 1$ vertices dominates a uniformly rooted uniform tree on n vertices.

Define $\theta(c)$ as the survival probability of $\text{PGW}(c)$, that is,

$$\theta(c) := 1 - q(c).$$

For $\mu > \lambda > 0$, set

$$\alpha(\lambda, \mu) := \log \frac{e^\mu - 1}{\mu} - \log \frac{e^\lambda - 1}{\lambda} > 0.$$

Lemma 2.2. For $\mu > \lambda > 1$, we have

$$\alpha(\lambda, \mu) < \mu - \lambda$$

and

$$\alpha(\lambda\theta(\lambda), \mu\theta(\mu)) > \lambda q(\lambda) - \mu q(\mu).$$

Proof. The first inequality states that $x \mapsto \log((e^x - 1)/x) - x$ is decreasing for $1 < x < \infty$, which in turn is a consequence of the inequality $e^x > 1 + x$.

By (1.1), we have

$$\frac{e^{\lambda\theta(\lambda)} - 1}{\lambda\theta(\lambda)} = \frac{1 - q(\lambda)}{\lambda q(\lambda)\theta(\lambda)} = \frac{1}{\lambda q(\lambda)}.$$

Therefore

$$\alpha(\lambda\theta(\lambda), \mu\theta(\mu)) = \log(\lambda q(\lambda)) - \log(\mu q(\mu)). \tag{2.2}$$

We next note that $x \mapsto xe^{-x}$ is strictly increasing on $(0, 1)$ and strictly decreasing on $(1, \infty)$. Recalling from (1.3) that $\lambda \exp(-\lambda) = \lambda q(\lambda) \exp(-\lambda q(\lambda))$ and using that $\lambda > 1$, we deduce that $\lambda q(\lambda) < 1$. We similarly deduce that $\mu q(\mu) < \lambda q(\lambda)$ since $\mu > \lambda$. The second claimed inequality therefore follows from (2.2) and the fact that $x \mapsto \log x - x$ is increasing on $(0, 1)$. \square

Let Q_λ denote a $\text{Poisson}(\lambda)$ random variable and Q_λ^* denote a random variable whose distribution is the same as that of Q_λ conditioned on $Q_\lambda > 0$.

Lemma 2.3. *Let $\mu > \lambda > 0$ and set $\alpha = \alpha(\lambda, \mu)$. Then Q_μ^* stochastically dominates the sum of mutually independent copies of Q_λ^* and Q_α . Moreover, this does not hold for any larger α .*

Proof. Consider some $\beta \in (0, \mu - \lambda)$, and let Z denote the sum of two mutually independent copies of Q_λ^* and Q_β . For $k \in \mathbb{N}_+$, set

$$a_k := \mathbf{P}[Z = k] = \frac{e^{-\lambda-\beta}}{1 - e^{-\lambda}} \sum_{j=1}^k \frac{\lambda^j \beta^{k-j}}{j!(k-j)!} = \frac{e^{-\beta}}{e^\lambda - 1} \frac{(\lambda + \beta)^k - \beta^k}{k!}, \tag{2.3}$$

and

$$b_k := \mathbf{P}[Q_\mu^* = k] = \frac{1}{e^\mu - 1} \frac{\mu^k}{k!}.$$

In order for Q_μ^* to dominate Z , it is necessary that $a_1 \geq b_1$. This translates and simplifies to $\beta \leq \alpha$. Now fix $\beta = \alpha$; note that $\alpha < \mu - \lambda$ by Lemma 2.2. Let J be the set of $k \in \mathbb{N}_+$ such that $a_k \geq b_k$. We claim that there is a $k_0 \in \mathbb{R}$ such that $J = \mathbb{N}_+ \cap [1, k_0]$. Before proving the claim, we shall demonstrate that the lemma follows from it. Indeed, to prove domination it is sufficient to show that

$$\forall k \in \mathbb{N}_+ \quad \sum_{j=k}^\infty b_j \geq \sum_{j=k}^\infty a_j. \tag{2.4}$$

This is clear if $k > k_0$, because $b_j > a_j$ for $j > k_0$. On the other hand, if $k \leq k_0$, then $a_j \geq b_j$ for $j \leq k$, which gives

$$\sum_{j=1}^k b_j \leq \sum_{j=1}^k a_j.$$

Now subtracting both sides from $\sum_{j \in \mathbb{N}_+} b_j = 1 = \sum_{j \in \mathbb{N}_+} a_j$ gives (2.4).

It remains to prove that $J = \mathbb{N}_+ \cap [1, k_0]$ for some $k_0 \in \mathbb{R}$. Note that

$$\frac{a_k}{b_k} = \frac{(e^\mu - 1)e^{-\beta}}{e^\lambda - 1} \left(\left(\frac{\lambda + \beta}{\mu} \right)^k - \left(\frac{\beta}{\mu} \right)^k \right).$$

Now think of the right-hand side as a function $g(k)$ of positive real k . As such, it may be written in the form $A(B^k - C^k)$, with constants A, B, C satisfying $A > 0$ and $1 > B > C > 0$. We claim that g does not have any local minimum. Indeed, $g'(k) = A(B^k \log B - C^k \log C)$ and $g''(k) = A(B^k(\log B)^2 - C^k(\log C)^2)$. Therefore, when $g'(k) = 0$, we have $(\log B)/(\log C) = C^k B^{-k}$ and so $g''(k) = A(\log C)^2 C^k (C^k B^{-k} - 1) < 0$. This verifies that g does not have a local minimum. Hence, the set of $k \in (0, \infty)$ such that $g(k) \geq 1$ is an interval. By our choice of $\beta = \alpha$, this interval contains 1. This proves the claim, and completes the proof of the lemma. \square

Given a rooted tree T with root o and a node v in T , let $N(v)$ denote the cardinality of the set of nodes in the subtree of T corresponding to v , that is, the number of nodes in T that are not in the connected component of o in $T \setminus \{v\}$. Given a random rooted tree T , let $n_k = n_k(T)$ denote the number of children v of the root satisfying $N(v) = k$.

Let $T(\lambda)$ denote a sample from $\text{PGW}(\lambda)$. Note that the random variables $(n_k(T(\lambda)) : k \in \mathbb{N}_+ \cup \{\infty\})$ are independent Poisson random variables. It follows that the random variables $(n_k(T_\infty(\lambda)) : k \in \mathbb{N}_+ \cup \{\infty\})$ are independent, and $n_k(T_\infty(\lambda))$ has the same law as $n_k(T(\lambda))$ when $k \in \mathbb{N}_+$, while $n_\infty(T_\infty(\lambda))$ has the law of $n_\infty(T(\lambda))$ conditioned on being positive.

By (2.1), we have

$$\forall k \in \mathbb{N}_+ \quad \mathbf{E}[n_k(T(\lambda))] = \frac{(\lambda e^{-\lambda})^k k^{k-1}}{k!}.$$

Observe that this is monotone decreasing in λ in the range $\lambda \geq 1$.

Since $n_\infty(T(\lambda))$ is Poisson with parameter $\lambda \theta$, we have

$$\sum_{k \in \mathbb{N}_+} \mathbf{E}[n_k(T(\lambda))] = (1 - \theta) \lambda.$$

If T and T' are rooted trees, we write $T \leq_1 T'$ if there is an injective map i from the children of the root in T to the children of the root in T' such that $N(i(v)) \geq N(v)$ for every child v of the root in T . If T and T' are random rooted trees, we write $T \leq_1^{\mathcal{L}} T'$ if T and T' may be coupled in such a way that $T \leq_1 T'$ a.s. Observe that $\leq_1^{\mathcal{L}}$ is a partial order relation.

Theorem 1.2 will follow easily from the next lemma.

Lemma 2.4. *Let $\mu > \lambda > 1$ and let $n \in \mathbb{N}_+$. Then $T_n \leq_1^{\mathcal{L}} T_{n+1} \leq_1^{\mathcal{L}} T_\infty(\lambda) \leq_1^{\mathcal{L}} T_\infty(\mu)$.*

Proof. We start by proving $T_\infty(\lambda) \leq_1^{\mathcal{L}} T_\infty(\mu)$. Let $(Z_k : k \in \mathbb{N}_+)$ and $(Z'_k : k \in \mathbb{N}_+)$ be independent Poisson random variables with $\mathbf{E}[Z_k] = \mathbf{E}[n_k(T(\mu))]$ and $\mathbf{E}[Z'_k] = \mathbf{E}[n_k(T(\lambda))] - \mathbf{E}[n_k(T(\mu))]$. (Recall that the latter is non-negative.) Let $Z' := \sum_{k \in \mathbb{N}_+} Z'_k$, which is a

Poisson random variable. By the above,

$$\mathbf{E}[Z'] = (1 - \theta(\lambda))\lambda - (1 - \theta(\mu))\mu.$$

By Lemma 2.2, we have

$$\alpha(\lambda\theta(\lambda), \mu\theta(\mu)) \geq \lambda q(\lambda) - \mu q(\mu) = \mathbf{E}[Z']. \tag{2.5}$$

Consequently, by Lemma 2.3, $n_\infty(T_\infty(\mu))$ may be coupled to dominate $n_\infty(T_\infty(\lambda))$ plus an independent copy of Z' . Thus, we may take $n_\infty(T_\infty(\lambda))$ independent from $(Z_k : k \in \mathbb{N}_+)$ and $(Z'_k : k \in \mathbb{N}_+)$ and take $n_\infty(T_\infty(\mu))$ independent from $(Z_k : k \in \mathbb{N}_+)$ such that $n_\infty(T_\infty(\mu)) \geq Z' + n_\infty(T_\infty(\lambda))$. For $k \in \mathbb{N}_+$ we take $n_k(T_\infty(\lambda)) = Z_k + Z'_k$, and $n_k(T_\infty(\mu)) = Z_k$. Since $Z' = \sum_{k \in \mathbb{N}_+} Z'_k$, with these choices we have $T_\infty(\lambda) \leq_1 T_\infty(\mu)$. This proves that $T_\infty(\lambda) \leq_1^{\mathcal{L}} T_\infty(\mu)$.

The fact that $T_n \leq_1^{\mathcal{L}} T_{n+1}$ follows from Theorem 2.1. Recall that the limit in law of T_n as $n \rightarrow \infty$ is the same as the limit in law of $T_\infty(\lambda)$ as $\lambda \searrow 1$; see [1, Lemma 23]. Let $T_\infty(1)$ denote a random tree with this limit law. Then $T_n \leq_1^{\mathcal{L}} T_\infty(1)$. By taking the limit as $\lambda' \searrow 1$ in $T_\infty(\lambda') \leq_1^{\mathcal{L}} T_\infty(\lambda)$, we find that $T_\infty(1) \leq_1^{\mathcal{L}} T_\infty(\lambda)$. Thus, $T_n \leq_1^{\mathcal{L}} T_\infty(\lambda)$ follows. \square

Proof of Theorem 1.2. This follows by repeatedly applying Lemma 2.4 at each node of the $T_\infty(\lambda)$ tree, while keeping the corresponding couplings appropriately conditionally independent. \square

3. Return probabilities

A general result on monotonicity [8] combined with Theorem 1.2 implies the monotonicity claim in Theorem 1.1. Here, we analyse in more detail the expression (1.2) in order to gain an explicit lower bound on the derivative of $f(c)$, which that general result does not supply.

Our main aim in this section is to prove the following result.

Theorem 3.1. *The expression*

$$\int \sum_{k \geq 1} \frac{1}{k} p_k(o; T) d\text{PGW}_c^*(T, o)$$

is monotonic decreasing in $c > 1$.

In light of (1.2) and Theorem 1.2, this implies the monotonicity claim in Theorem 1.1 and will lead to an explicit lower bound on the derivative in Section 5. It also implies the following lower bound for $f(c)$ itself:

$$f(c) \geq \sum_{k \geq 1} \frac{e^{-c} c^k (1 - q(c)^k) \log k}{\theta(c) k!} - \sum_{k \geq 0} \frac{e^{-1} \log(1 + k)}{k!} \geq 0.$$

To see this, note first that by, say, (2.3), we have that

$$\text{PGW}_c^*[\text{deg}_T(o) = k] = \frac{e^{-c} c^k (1 - q(c)^k)}{\theta(c) k!}.$$

Second, recall that $\lim_{c \downarrow 1} f(c) = 0$. Therefore, Theorem 3.1 and (1.2) imply that

$$f(c) \geq \int \log \deg_T(o) d\text{PGW}_c^*(T, o) - \int \log \deg_T(o) d\text{PGW}_1^*(T, o) \geq 0,$$

and this equals the above expression by the well-known form of $\text{PGW}^*(1)$ [1, Corollary 3]. This lower bound should be compared to the trivial upper bound

$$f(c) \leq \sum_{k \geq 1} \frac{e^{-c} c^k (1 - q(c)^k) \log k}{\theta(c) k!}.$$

To prove Theorem 3.1, let $V(s, T, o) := \sum_{k \geq 0} p_k(o; T) s^k$. Since

$$\int_0^1 \frac{\mathbf{E}[V(s, T, o)] - 1}{s} ds = \int \sum_{k \geq 1} \frac{1}{k} p_k(o; T) d\text{PGW}_c^*(T, o),$$

Theorem 3.1 will be a consequence of the following result.

Theorem 3.2. *For all $s \in (0, 1)$, the expectation $\int V(s, T, o) d\text{PGW}_c^*(T, o)$ is decreasing in $c > 1$.*

Fix $\mu > \lambda > 1$ and let T and T' have the distributions $\text{PGW}^*(\lambda)$ and $\text{PGW}^*(\mu)$, respectively. Let X count the number of visits to the root in a random walk on the tree T started from the root in which at each step the walker has probability $1 - s$ of dying, independent of the other steps (note that $X \geq 1$ since we start from the root). Let X' be the same for a walk on T' . Because $V(s, T, o) = \mathbf{E}[X]$, Theorem 3.2 follows from the next result.

Theorem 3.3. *X stochastically dominates X' .*

In words, larger trees have fewer returns of simple random walk for this model. We shall need a technical lemma for the proof.

Lemma 3.4. *Fix integers $a \geq 1$ and $b \geq 0$. Let F be a convex increasing function on $[0, \infty)$. Let $X_1, \dots, X_a, Y_1, \dots, Y_b$ be independent non-negative random variables. If each X_i stochastically dominates each Y_j , then*

$$\mathbf{E}F\left(\frac{1}{a+b} \left(\sum_{i=1}^a X_i + \sum_{i=1}^b Y_i\right)\right) \leq \mathbf{E}F\left(\frac{1}{a} \left(\sum_{i=1}^a X_i\right)\right).$$

Proof. Define an auxiliary random vector A , with $a + b$ coordinates to be uniformly chosen among the $N := \binom{a+b}{a}$ vectors containing exactly a values equal to $1/a$ and b zeros. Condition on the X s and Y s and consider the random variable

$$R := F\left(\mathbf{E}_A \left(\sum_{i=1}^a A_i X_i + \sum_{i=1}^b A_{i+a} Y_i\right)\right),$$

where \mathbf{E}_A denotes expectation over A . On the one hand, we have

$$R = F\left(\frac{1}{a+b}\left(\sum_{i=1}^a X_i + \sum_{i=1}^b Y_i\right)\right). \tag{3.1}$$

On the other hand, by Jensen’s inequality (since the X s and Y s are non-negative),

$$R \leq \mathbf{E}_A F\left(\sum_{i=1}^a A_i X_i + \sum_{i=1}^b A_{i+a} Y_i\right) = N^{-1} \sum_{i=1}^N M_i, \tag{3.2}$$

where each M_i is a random variable of the form

$$F\left(\frac{1}{a} \sum_{k=1}^{a_1} X_{i_k} + \frac{1}{a} \sum_{k=1}^{b_1} Y_{i_k}\right);$$

here, $(X_{i_k})_{k=1}^{a_1}$ is a subset of a_1 of the X s, $(Y_{i_k})_{k=1}^{b_1}$ is a subset of b_1 of the Y s, $a_1 \leq a, b_1 \leq b$ and $a_1 + b_1 = a$. Now taking expectation over the X s and Y s and using that each X_i stochastically dominates each Y_j , we get

$$\mathbf{E}(M_i) \leq \mathbf{E}F\left(\frac{1}{a} \sum_{k=1}^a X_k\right), \tag{3.3}$$

since the X_i s and Y_j s are independent, non-negative and each M_i is increasing in each of the random variables. Putting (3.1), (3.2) and (3.3) together, we get

$$\mathbf{E}F\left(\frac{1}{a+b}\left(\sum_{i=1}^a X_i + \sum_{i=1}^b Y_i\right)\right) \leq \mathbf{E}F\left(\frac{1}{a} \sum_{i=1}^a X_i\right),$$

proving the lemma. □

Proof of Theorem 3.3. It is enough to show that, for each integer $M \geq 2$, we have

$$\mathbf{P}(X \geq M) \geq \mathbf{P}(X' \geq M). \tag{3.4}$$

We couple the two trees according to the coupling given in the preceding section, in the proof of Lemma 2.4 and Theorem 1.2, for $T_\infty(\lambda)$ and $T_\infty(\mu)$. It is then enough to show the inequality (3.4) conditioned on the number of subtrees of each size that the roots of T and T' have (the variables $n_k(T)$ and $n_k(T')$). Henceforth we always condition on these values.

Denote $N_F := \sum_{k=1}^\infty n_k(T)$, $N_I := n_\infty(T)$, $N'_F := \sum_{k=1}^\infty n_k(T')$ and $N'_I := n_\infty(T')$. According to the coupling, we have $N_F \geq N'_F$ and $d := N_F + N_I \leq N'_F + N'_I =: d'$. We construct our coupling of T and T' to have the following properties: T is a rooted subtree of T' ; the children of the root are ordered; the first N'_F children of the root lie in T and have finite subtrees, all pairwise equal in T and T' ; the next $N_F - N'_F$ children lie in T and have finite subtrees; and the next N_I children lie in T . Given the sizes of the subtrees, recall that the pairs of coupled subtrees of the children of the root are independent, even including the $d' - d$ left-over subtrees of T' , and that all N'_I of the infinite subtrees of T' are i.i.d.

For each $1 \leq j \leq N_F$, suppose that the random walk enters in its first step the j th finite subtree of T . Let P_j^F be the probability that, continuing this random walk, we ever return to the root. Similarly, define for $1 \leq j \leq N_I$ the probability P_j^I of returning from the j th infinite subtree of T . Define analogously Q_j^F and Q_j^I on T' . Thus, $P_j^F = Q_j^F$ for $j \leq N'_F$ because of the coupling.

The inequality (3.4) can now be written as follows:

$$\mathbf{E} \left(\frac{s}{d} \left(\sum_{j=1}^{N_F} P_j^F + \sum_{j=1}^{N_I} P_j^I \right) \right)^{M-1} \geq \mathbf{E} \left(\frac{s}{d'} \left(\sum_{j=1}^{N'_F} Q_j^F + \sum_{j=1}^{N'_I} Q_j^I \right) \right)^{M-1}.$$

We prove this inequality in two steps. First we observe that

$$\mathbf{E} \left(\frac{s}{d} \left(\sum_{j=1}^{N_F} P_j^F + \sum_{j=1}^{N_I} P_j^I \right) \right)^{M-1} \geq \mathbf{E} \left(\frac{s}{d} \left(\sum_{j=1}^{N'_F} Q_j^F + \sum_{j=1}^{d-N'_F} Q_j^I \right) \right)^{M-1};$$

this is because if the walk entered a branch of T' that contains a branch of T , then certainly its probability of ever returning to the root is smaller in T' than it is in T (by coupling the walks). Now we may use Lemma 3.4 to get that

$$\mathbf{E} \left(\frac{s}{d} \left(\sum_{j=1}^{N'_F} Q_j^F + \sum_{j=1}^{d-N'_F} Q_j^I \right) \right)^{M-1} \geq \mathbf{E} \left(\frac{s}{d'} \left(\sum_{j=1}^{N'_F} Q_j^F + \sum_{j=1}^{N'_I} Q_j^I \right) \right)^{M-1}$$

since each Q_j^I is stochastically dominated by each Q_j^F (again, by coupling the walks on the coupled subtrees). This proves the theorem. □

4. Smoothness

Let $\bar{p}_k = \bar{p}_k(c) := \int p_k(o; T) d\text{PGW}_c^*(T, o)$. We shall prove the following.

Theorem 4.1. *For each $k \geq 1$, $\bar{p}_k(c)$ is real analytic in $c > 1$ and there exists $\beta > 0$ such that for $c > 1$, $k \geq 1$, and $n \geq 1$, we have*

$$\left| \frac{\partial^n \bar{p}_k}{\partial c^n} \right| \leq A^n n! k^{\beta n} e^{-ak^{1/6}},$$

where the constants $A, a > 0$ depend only on c and are bounded from 0 and infinity for c in every compact sub-interval of $(1, \infty)$.

Remark. We obtain $\beta = 1$ in the proof, but this could be reduced further.

An immediate corollary is as follows.

Corollary 4.2. *$f(c)$ is C^∞ for $c \in (1, \infty)$.*

To prove Theorem 4.1, we shall prove the following result.

Theorem 4.3. For each $k \geq 1$, \bar{p}_k can be analytically continued to the domain $\Omega_k := \{x + iy \mid x \in (1, \infty), |y| \leq ak^{-\beta}\}$ for some $\beta > 0$ and satisfies for $c \in \Omega_k$

$$|\bar{p}_k(c)| \leq Ae^{-ak^{1/6}},$$

where $A, a > 0$ depend only on $x = \operatorname{Re}(c)$ and are bounded from 0 and infinity for x in every compact sub-interval of $(1, \infty)$.

Theorem 4.1 is an immediate corollary by Cauchy estimates. To see this, for each $c > 1$ take a circle C of radius $r := \min(ak^{-\beta}, \frac{c-1}{2})$ around c . Then

$$\left| \frac{\partial^n \bar{p}_k}{\partial c^n} \right| = \left| \frac{n!}{2\pi i} \oint_C \frac{\bar{p}_k(z)}{(z-c)^{n+1}} dz \right| \leq \frac{An!e^{-ak^{1/6}}}{r^n} \leq AA'n!k^{\beta n}e^{-ak^{1/6}},$$

where $A' := \max(a^{-n}, (\frac{c-1}{2})^{-n})$.

In the rest of the section, we prove Theorem 4.3. We start by quoting a known result concerning *a priori* bounds on \bar{p}_k ; see [9, Theorem 2].

Theorem 4.4. For each $k \geq 1$, we have $\bar{p}_k(c) \leq Ae^{-ak^{1/6}}$ for $c > 1$, where $A, a > 0$ depend only on c and are bounded from 0 and infinity for c in every compact sub-interval of $(1, \infty)$.

We remark that in [9, Theorem 2] the boundedness of the constants is not claimed, just that constants exist for every $c > 1$, but this implies the theorem since \bar{p}_k is a continuous function of c .

We now fix a compact sub-interval $I \subseteq (1, \infty)$ and shall work only with $c = x + iy$ such that $x \in I$. All the constants $A, a > 0$ appearing below may depend on I and it is understood that their value may change from line to line: A may increase, while a may decrease.

We record for later use the well-known structure of the $\text{PGW}^*(c)$ distribution, as was also discussed in Section 2.

Lemma 4.5. The $\text{PGW}^*(c)$ distribution is a 2-type Galton–Watson distribution, with the types called *I* and *F* (for ‘infinite’ and ‘finite’). For vertices of type *I* the number of type *I* children is distributed as $Q_{c(1-q(c))}^*$ and of type *F* children as $Q_{cq(c)}$. Vertices of type *F* have only type *F* children, the number of which is distributed as $Q_{cq(c)}$.

We next introduce the notion of the *trace of a random walk path*. This is all the information about a path on a tree that starts at its root. The trace includes the following information per step:

- (1) whether the step is up or down (up is away from the root);
- (2) if the step is up, whether it is to a type-I child or to a type-F child and which such child is it (e.g., the first type-I child, the second type-F child, etc.)

We denote by \mathcal{W}_k the set of all traces that have exactly k steps and end at the root. Given $W \in \mathcal{W}_k$, let $l_I(W), l_F(W)$ be the number of distinct vertices of types *I* and *F*, respectively, that the trace visits, so that $l_I(W) + l_F(W) \leq k$. Given a tree T , the trace W may be

feasible on T or not. It is feasible if and only if all the vertices that W visits exist in T (e.g., if on the first step, W moves to the third type-F child of the root, then the root of T must have at least 3 type-F children). Let $d_I(W) := (d_1^F, d_1^I, \dots, d_{l_I(W)}^F, d_{l_I(W)}^I)$ be the minimum required number of children of type F and type I from each of the vertices of type I that W visits in order for the walk to be feasible. Here, the subscript i indicates the i th distinct vertex of type I visited by W . Similarly, let $d_F(W) := (\tilde{d}_1^F, \dots, \tilde{d}_{l_F(W)}^F)$ be the minimum required number of children of type F from the vertices of type F that W visits. Given two vectors e and d of the same length, we write $e \geq d$ if each coordinate of e is greater than or equal to the corresponding coordinate of d . Finally, denote by $p^c(W)$ the probability under $\text{PGW}^*(c)$ of sampling a feasible tree for W and then of sampling W as a simple random walk path of length k on that tree. From all the above discussion, we have

$$\bar{p}_k(c) = \sum_{W \in \mathcal{W}_k} p^c(W) = \sum_{W \in \mathcal{W}_k} \sum_{\substack{e_I \geq d_I(W) \\ e_F \geq d_F(W)}} p^c(e_I, e_F) p(W, e_I, e_F), \tag{4.1}$$

where $p^c(e_I, e_F)$ is the probability of sampling a tree in which the vertices that W passes through have exactly the prescribed number of children e_I, e_F of each type, and $p(W, e_I, e_F)$ is the conditional probability, given e_I and e_F , of sampling W as a simple random walk path on the tree. We emphasize that $p(W, e_I, e_F)$ does not depend on c , while $p^c(e_I, e_F)$ is the same for all W that satisfy $e_I \geq d_I(W)$ and $e_F \geq d_F(W)$.

Since c and q are analytically related for $c \in (1, \infty)$ by (1.1) (which can be rewritten as $c = -\frac{\log(q)}{1-q}$), there is a unique extension of $q(c)$ to an analytic function of c for $\text{Re } c > 1$ and $|\text{Im } c| \leq \kappa(\text{Re } c)$ for some continuous function $\kappa : (1, \infty) \rightarrow (0, \infty)$. (In fact, one can extend it much further, but we shall not need that.) Hence, the same holds for $p^c(e_I, e_F)$. We shall use the same notation for the original functions as for these extensions, and likewise for similar functions below.

Note that to prove Theorem 4.3, it is enough to show that for $c = x + iy$ with $x \in I$ and $|y| \leq \kappa(x)k^{-\beta}$, the sum (4.1) converges uniformly and is bounded by $A \exp(-ak^{1/6})$. Denote by $\max(e)$ the maximal element of e .

Lemma 4.6. *If $c = x + iy$ with $x \in I$ and $|y| \leq \kappa(x)$, then*

$$|p^c(e_I, e_F)| \leq p^x(e_I, e_F) e^{Ak(\max(e_I) + \max(e_F) + 1)|y|}.$$

Proof. Using the structure Lemma 4.5, we know that

$$p^c(e_I, e_F) = \prod_{i=1}^{l_I(W)} p^c(e_I, i) \prod_{i=1}^{l_F(W)} p^c(e_F, i)$$

(the values $l_I(W)$ and $l_F(W)$ are implicit in e_I and e_F as their lengths), where

$$p^c(e_I, i) = \mathbf{P}(Q_{c(1-q(c))}^* = e_i^I) \mathbf{P}(Q_{cq(c)} = e_i^F),$$

$$p^c(e_F, i) = \mathbf{P}(Q_{cq(c)} = \tilde{e}_i^F).$$

More explicitly, denoting $j := e_I^1, m := e_F^1$ and $n := \tilde{e}_I^F$ and abbreviating $q := q(c)$, we have

$$p^c(e_I, i) = e^{-c(1-q)} \frac{(c(1-q))^j}{j!(1-e^{-c(1-q)})} e^{-cq} \frac{(cq)^m}{m!},$$

$$p^c(e_F, i) = e^{-cq} \frac{(cq)^n}{n!}.$$

We have in the first case

$$|p^c(e_I, i)| \leq \frac{e^{-x}}{j!m!} |c|^{j+m} |1-q|^j |q|^m \frac{1}{|1-e^{-c(1-q)}|}. \tag{4.2}$$

We know that when $x \in I$ and $|y| \leq \kappa(x)$, we have $|q(x+iy) - q(x)| \leq A|y|$ since q is an analytic function of c . Hence

$$|c|^{j+m} = x^{j+m} \left| 1 + \frac{y^2}{x^2} \right|^{\frac{j+m}{2}} \leq x^{j+m} e^{A y^2 (j+m)},$$

$$|1 - q(x+iy)|^j \leq (1 - q(x))^j \left(1 + \frac{A|y|}{1 - q(x)} \right)^j \leq (1 - q(x))^j e^{A|y|},$$

$$|q(x+iy)|^m \leq q(x)^m \left(1 + \frac{A|y|}{q(x)} \right)^m \leq q(x)^m e^{Am|y|},$$

and

$$\frac{1}{|1 - e^{-(x+iy)(1-q(x+iy))}|} \leq \frac{1}{1 - e^{-x(1-q(x))}} + A|y| \leq \frac{1}{1 - e^{-x(1-q(x))}} (1 + A|y|)$$

$$\leq \frac{e^{A|y|}}{1 - e^{-x(1-q(x))}}.$$

Substituting back into (4.2), we get

$$|p^c(e_I, i)| \leq p^x(e_I, i) e^{A(j+m+1)|y|}.$$

This bound was for $p^c(e_I, i)$, but we obtain analogously that $|p^c(e_F, i)| \leq p^x(e_F, i) e^{A(n+1)|y|}$. Hence

$$|p^c(e_I, e_F)| \leq p^x(e_I, e_F) e^{A(l_I(W) + l_F(W))(\max(e_I) + \max(e_F) + 1)|y|}$$

$$\leq p^x(e_I, e_F) e^{Ak(\max(e_I) + \max(e_F) + 1)|y|}. \quad \square$$

To continue, say that a vertex of a tree is L -big if it has either exactly L type-I children or exactly L type-F children or both. Let $E_{k,L}$ be the event that if we sample a tree and do a simple random walk on it (from the root), then the walk returns to the root after exactly k steps and visits an L -big vertex along the way but does not visit an M -big vertex along the way for any $M > L$. We observe that

$$\sum_{W \in \mathcal{W}_k} \sum_{\substack{e_I \geq d_I(W) \\ e_F \geq d_F(W) \\ \max(\max(e_I), \max(e_F)) = L}} p^x(e_I, e_F) p(W, e_I, e_F) = \text{PGW}_x^*(E_{k,L}) \leq Ak e^{-aL \log L}, \tag{4.3}$$

where the last inequality follows since there are no more than k vertices along any path W and since the tails of a Poisson(c) random variable decay as $Ae^{-aL \log L}$, even when conditioned to be at least 1.

Thus, we find that if $c = x + iy$ with $x \in I$ and $|y| \leq \kappa(x)/k$, then from Lemma 4.6 (for k large enough as a function of I), we have

$$\begin{aligned}
 |\bar{p}_k(c)| &\leq \sum_{W \in \mathcal{W}_k} \sum_{\substack{e_I \geq d_I(W) \\ e_F \geq d_F(W)}} |p^c(e_I, e_F)| p(W, e_I, e_F) \\
 &\leq \sum_{W \in \mathcal{W}_k} \sum_{L=1}^{\infty} e^{A(L+1)} \sum_{\substack{e_I \geq d_I(W) \\ e_F \geq d_F(W) \\ \max(\max(e_I), \max(e_F)) = L}} p^x(e_I, e_F) p(W, e_I, e_F) \\
 &\leq \underbrace{e^{A\delta k^{1/6}} \sum_{W \in \mathcal{W}_k} \sum_{L \leq \delta k^{1/6}} (\dots)}_{(C)} + \underbrace{\sum_{W \in \mathcal{W}_k} \sum_{L > \delta k^{1/6}} e^{AL} (\dots)}_{(D)}.
 \end{aligned}$$

By (4.3), we have

$$(D) \leq Ak \sum_{L > \delta k^{1/6}} e^{AL - aL \log L} \leq Ae^{-a\delta k^{1/6} \log k},$$

and by Theorem 4.4, we have

$$(C) \leq e^{A\delta k^{1/6}} \bar{p}_k(x) \leq Ae^{(A\delta - a)k^{1/6}} \leq Ae^{-ak^{1/6}},$$

where the last inequality follows by taking δ small enough (as a function of I). Putting everything together, we get

$$|\bar{p}_k(c)| \leq Ae^{-ak^{1/6}}.$$

The calculation was made for k large enough as a function of I , but the inequality will be true for smaller k as well by taking A large enough. This completes the proof of Theorem 4.3.

5. Derivative

By Theorem 3.1 and (1.2), we have

$$f'(c) \geq \frac{d}{dc} \int \log \deg_T(o) d\text{PGW}_c^*(T, o) = \frac{d}{dc} \sum_{k \geq 1} \frac{e^{-c} c^k (1 - q(c)^k) \log k}{\theta(c) k!}.$$

Although this lower bound appears to be a fairly simple expression, the presence of the logarithm makes it hard to evaluate. For that reason, it seems desirable to have a more explicit lower bound.

Write $r_k(c)$ for the probability that the root has degree k under the $\text{PGW}^*(c)$ distribution. We seek a lower bound for

$$\begin{aligned}
 \sum_{k \geq 1} r'_k(c) \log k &= \sum_{k \geq 0} r'_k(c) \log^+ k = \sum_{k \geq 0} s'_k(c) [\log^+(k+1) - \log^+ k] \\
 &= \sum_{k \geq 1} s'_k(c) \log \frac{k+1}{k} > \sum_{k \geq 1} s'_k(c) \frac{1}{k+1},
 \end{aligned}$$

where $s_k(c) := \sum_{j>k} r_j(c)$ and we have used Lemma 2.4 for the fact that $s'_k(c) \geq 0$ (i.e., the degree distribution of the root under $\text{PGW}^*(c)$ is stochastically increasing in c). Now the degree of the root has the same law as $X_c := Q_{c\theta(c)}^* + Q_{cq(c)}$. Let $c > 1$ and $\delta > 0$. Define

$$g(c, \delta) := \alpha(c\theta(c), (c + \delta)\theta(c + \delta)) - [cq(c) - (c + \delta)q(c + \delta)].$$

By Lemma 2.2, we have $g(c, \delta) > 0$. By Lemma 2.3, we have that $X_{c+\delta}$ stochastically dominates

$$Q_{c\theta(c)}^* + Q_{\alpha(c\theta(c), (c+\delta)\theta(c+\delta))} + Q_{(c+\delta)q(c+\delta)},$$

which has the same distribution as $X_c + Y_{c,\delta}$, where $Y_{c,\delta} := Q_{g(c,\delta)}$ is independent of X_c . Therefore,

$$s_k(c + \delta) - s_k(c) = \mathbf{P}[X_{c+\delta} > k] - \mathbf{P}[X_c > k] \geq \mathbf{P}[X_c + Y_{c,\delta} > k] - \mathbf{P}[X_c > k].$$

It follows that

$$s'_k(c) \geq r_k(c)\beta(c),$$

where

$$\beta(c) := \lim_{\delta \rightarrow 0} g(c, \delta)/\delta.$$

By (2.2) and (1.3), we have that

$$\begin{aligned} g(c, \delta) &= \log(cq(c)e^{-cq(c)}) - \log((c + \delta)q(c + \delta)e^{-(c+\delta)q(c+\delta)}) \\ &= \log(ce^{-c}) - \log((c + \delta)e^{-(c+\delta)}) = \log c - c - \log(c + \delta) + c + \delta. \end{aligned}$$

Therefore,

$$\beta(c) = 1 - \frac{1}{c}.$$

Thus, we obtain

$$\begin{aligned} f'(c) &> \sum_{k \geq 1} \frac{r_k(c)\beta(c)}{k+1} = \frac{e^{-c}\beta(c)}{\theta(c)} \int_0^1 (e^{cs} - e^{csq(c)}) ds \\ &= \left(1 - \frac{1}{c}\right) \left(\frac{1 - e^{-c}}{c\theta(c)} - \frac{e^{-c\theta(c)} - e^{-c}}{cq(c)\theta(c)}\right) \\ &= \frac{(c-1)e^{-cq(c)}}{c^2} > 0. \end{aligned}$$

This completes the proof of Theorem 1.1.

6. Open problems

A number of questions suggest themselves in light of our results, some of which arose in conversation with Itai Benjamini.

(1) Given two finite graphs H and G , say that $H \preceq G$ if there is a coupling of uniform vertices Y of H and X of G such that there is an isomorphism φ of the component of Y

in H to a subgraph of G such that $\varphi(Y) = X$. Let $\mathcal{G}(n, M)$ denote the random graph on n vertices with M edges. Write $\mathcal{G}^*(n, M)$ for the union of all components of $\mathcal{G}(n, M)$ that have the maximum number of edges (the maximum being taken over all the components of $\mathcal{G}(n, M)$; for large M , there is likely to be only one such component). One very strong finitary version of Theorem 1.2 would say that $\mathcal{G}^*(n, M) \preceq \mathcal{G}^*(n, M + 1)$ for $M < \binom{n}{2}$. Does this hold?

(2) Consider a $(d + 1)$ -regular tree and $p_2 > p_1 > 1/d$. Let $T(p)$ denote the component of the root under Bernoulli(p) percolation conditioned on the event that this component is infinite. Does $T(p_2)$ stochastically dominate $T(p_1)$? Unpublished work of Erik Broman and the first author here shows that for the (slightly different) case of d -ary trees, this holds for $d = 2, 3$.

(3) More generally, let G be a transitive graph, especially such as \mathbb{Z}^d , and $p_2 > p_1 > p_c(G)$, where $p_c(G)$ is the critical probability for Bernoulli (bond or site) percolation on G . Fix $o \in G$ and let $G(p)$ denote the component of o given that it is infinite. Does $G(p_2)$ stochastically dominate $G(p_1)$? If this holds, then there is a weak limit of $G(p)$ as $p \downarrow p_c(G)$, which could be called the incipient infinite cluster. (It is conjectured that there is no infinite component at $p_c(G)$; see [2].) Such a limit is not known to exist in \mathbb{Z}^d for $d \geq 3$, although another incipient infinite cluster has been constructed for $d \geq 19$ by van der Hofstad and Járai [5].

(4) Again, if G is a transitive graph, $o \in G$, and $n \geq 1$, let T_n denote a uniformly chosen random subtree of G rooted at o and with n vertices. Is $T_n \preceq T_{n+1}$?

(5) Let ρ_1 and ρ_2 be two Galton–Watson measures on rooted trees. If $k \geq 2$ and $\rho_1 \preceq \rho_2$, then is it necessarily the case that $\int p_k(o; T) d\rho_1(T, o) \geq \int p_k(o; T) d\rho_2(T, o)$?

Acknowledgement

We thank Yuval Peres for several conversations.

References

- [1] Aldous, D. and Pitman, J. (1998) Tree-valued Markov chains derived from Galton–Watson processes. *Ann. Inst. H. Poincaré Probab. Statist.* **34** 637–686.
- [2] Benjamini, I. and Schramm, O. (1996) Percolation beyond \mathbb{Z}^d , many questions and a few answers. *Electron. Comm. Probab.* **1** 71–82 (electronic).
- [3] Bollobás, B. (2001) *Random Graphs*, Vol. 73 of *Cambridge Studies in Advanced Mathematics*, 2nd edn, Cambridge University Press.
- [4] Erdős, P. and Rényi, A. (1960) On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** 17–61.
- [5] van der Hofstad, R. and Járai, A. A. (2004) The incipient infinite cluster for high-dimensional unoriented percolation. *J. Statist. Phys.* **114** 625–663.
- [6] Luczak, M. and Winkler, P. (2004) Building uniformly random subtrees. *Random Struct. Alg.* **24** 420–443.

- [7] Lyons, R. (2005) Asymptotic enumeration of spanning trees. *Combin. Probab. Comput.* **14** 491–522.
- [8] Lyons, R. (2007) Identities and inequalities for tree entropy. Preprint, arXiv:0712.3035.
- [9] Piau, D. (1998) Théorème central limite fonctionnel pour une marche au hasard en environnement aléatoire. *Ann. Probab.* **26** 1016–1040.
- [10] Pitman, J. (1998) Enumerations of trees and forests related to branching processes and random walks. In *Microsurveys in Discrete Probability* (D. Aldous and J. Propp, eds), Vol. 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, AMS, Providence, RI, pp. 163–180.