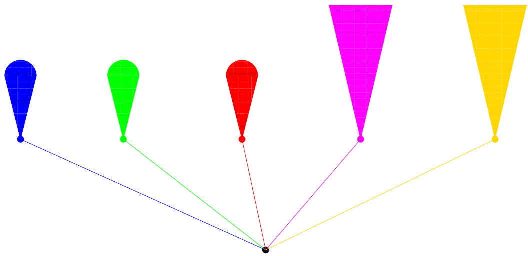
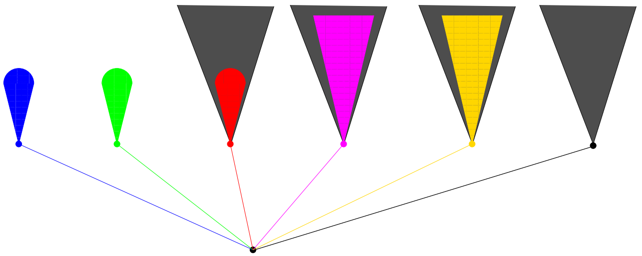
Spanning Trees, Random Graphs, and Random Walks

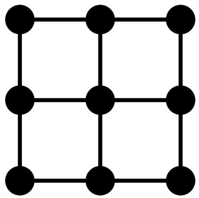
BY RUSSELL LYONS
(Indiana University, visiting Microsoft Research)

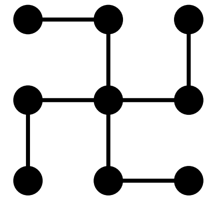
Joint work with RON PELED and ODED SCHRAMM

Combin. Probab. Comput. **17** (2008), 711–726

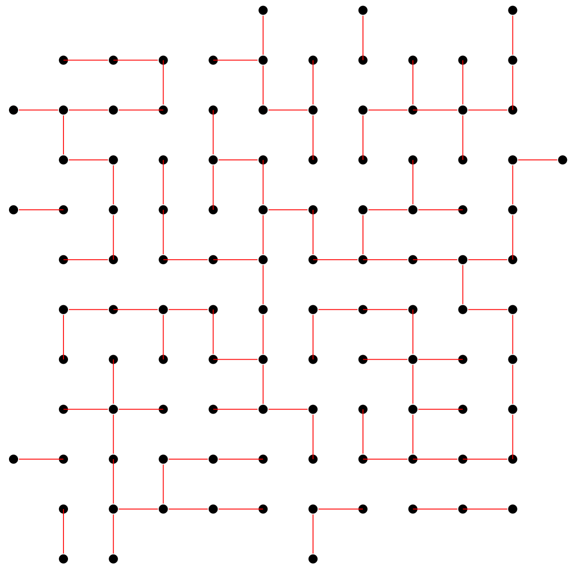








left-facing ancient religious symbol



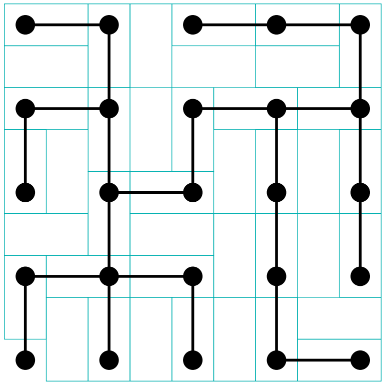
$\tau(G) :=$ number of spanning trees of G

Kirchhoff (1847): electrical networks

Example: Let a and z be two vertices of G. The effective resistance between a and z is

$$\frac{\tau(G/\{a,z\})}{\tau(G)}\,.$$

Temperley's bijection with dominos/dimers/matchings:



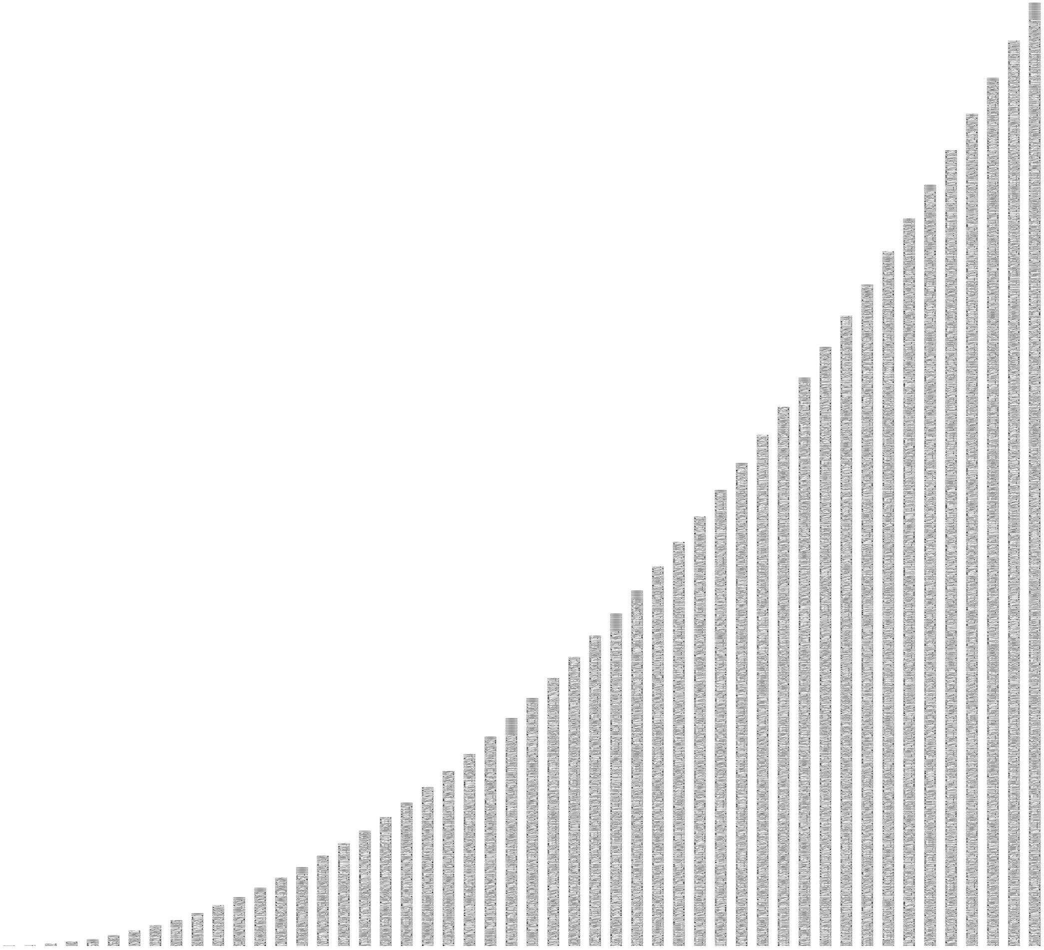
(Kenyon)

counting spanning trees is in every graph theory text

Number τ of spanning trees in $n \times n$ portion of square lattice:

n	τ(square)
1	1
2	4
3	192
4	100352
5	557568000
6	32565539635200
7	19872369301840986111

n	τ(torus)
1	1
2	32
3	11664
4	42467328
5	1562500000000
6	587312954081280000
7	2266101334892340404752383



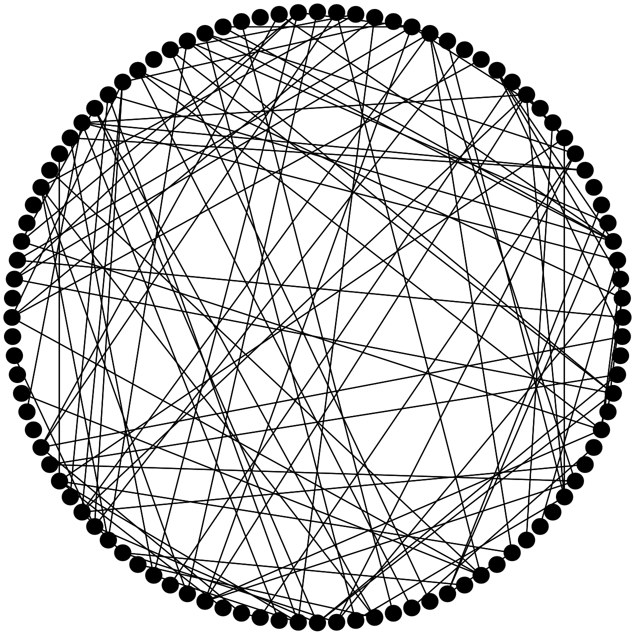
What is the growth rate? There are a certain number of choices at each vertex; in fact, pick a root and then there is precisely one edge leading toward the root from each vertex. So the natural asymptotic is $\lim_{n\to\infty} |\mathsf{V}(G_n)|^{-1} \log \tau(G_n)$ when $|\mathsf{V}(G_n)| \to \infty$ "within some family".

In each of the above two cases, squares and tori, it is well known (because of dimers) that this limit exists and both limits are the same, *viz.*, $4\mathrm{G}/\pi \approx 1.16624$, where

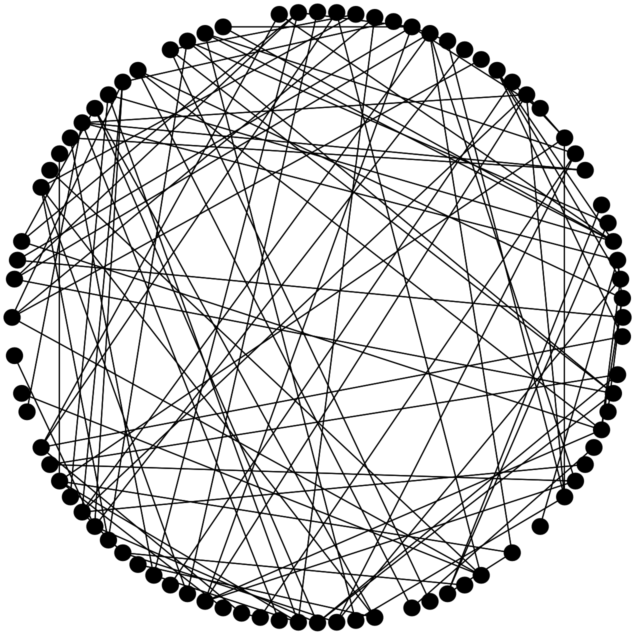
$$\mathrm{G} := \sum_{k=0}^{\infty} (-1)^k/(2k+1)^2 = 0.9160^-$$

is Catalan's constant and

$$\pi/4 = \sum_{k=0}^{\infty} (-1)^k/(2k+1)\,.$$



$\mathcal{G}(100, 2/100)$

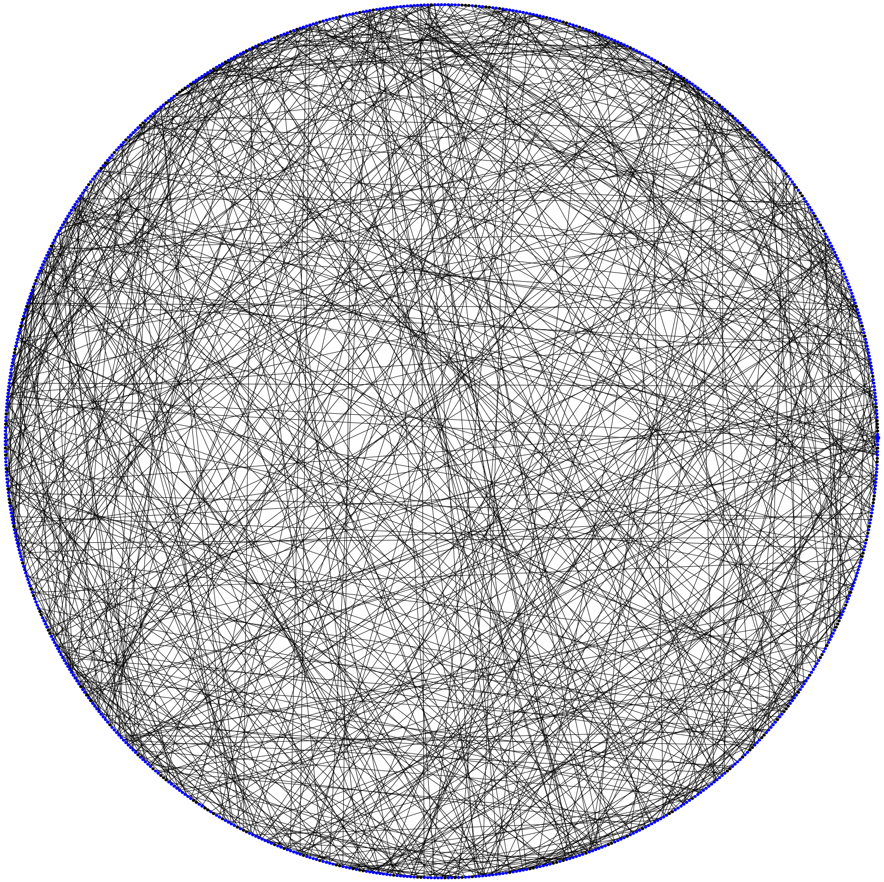


giant component

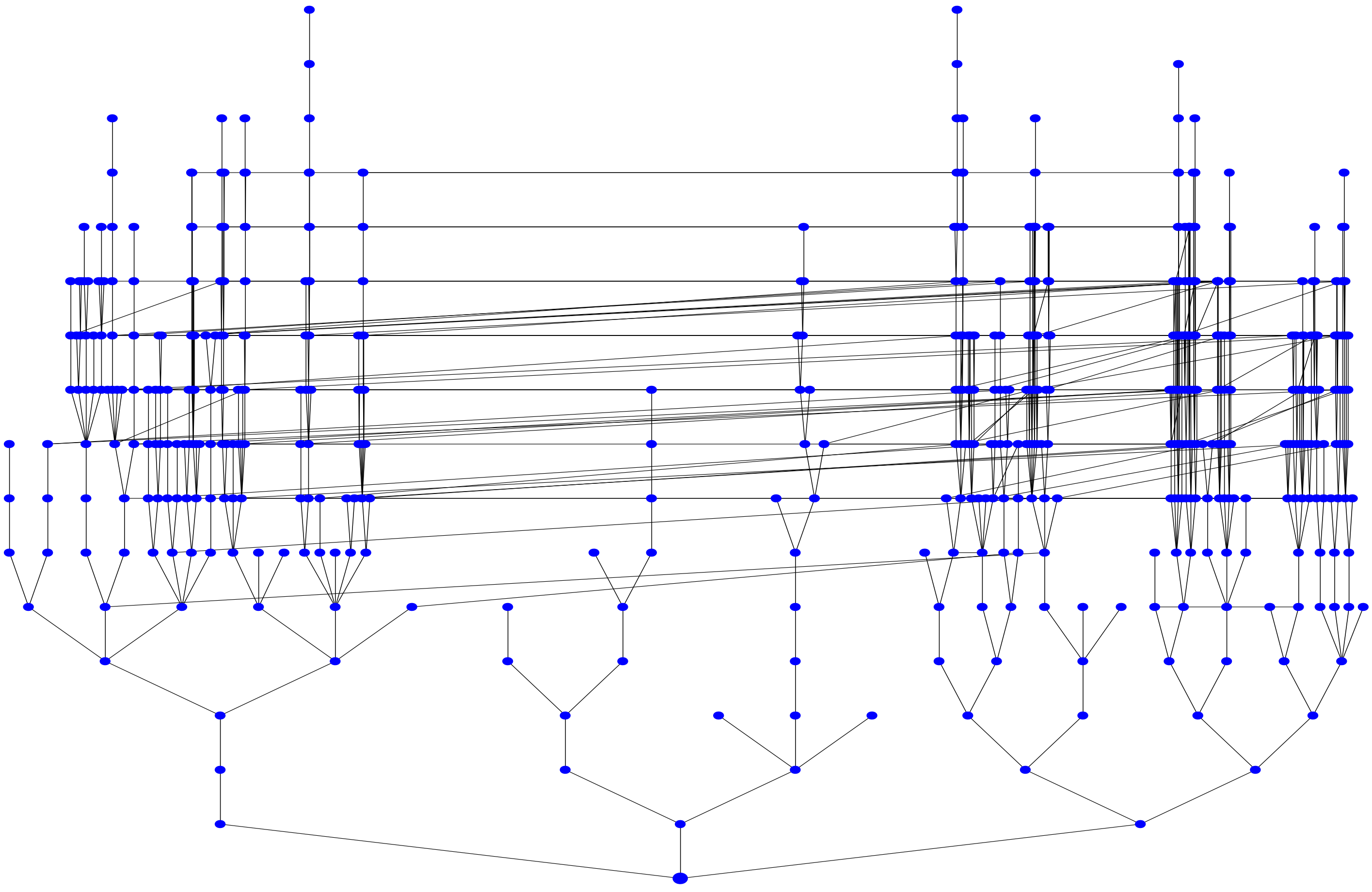
The usual Erdős-Rényi model of random graphs, $\mathcal{G}(n, p)$, is a graph on n vertices, each pair of which is connected by an edge with probability p, independently of other edges. This is a classic model that is studied in detail.

Fix $c > 1$. With probability approaching 1 as $n \to \infty$, there is a unique connected component, called the **giant component**, of $\mathcal{G}(n, c/n)$, that has $\Omega(n)$ vertices. If we fix a root vertex o, then the component of o has a limit of its neighborhood laws as $n \to \infty$ under $\mathcal{G}(n, c/n)$, which is $\mathsf{PGW}(c)$, where $\mathsf{PGW}(c)$ is the law of a rooted Galton-Watson tree with Poisson(c) offspring distribution. This means the root has a Poisson(c) number of offspring; each of the offspring, should there be any, independently also have a Poisson(c) number of offspring; etc. We draw an edge between a vertex and each of its offspring.

If we condition o to be in the giant component, then the component of o has a weak limit $\mathsf{PGW}^*(c)$, which is $\mathsf{PGW}(c)$ conditioned on nonextinction.



$\mathcal{G}(800, 2/800)$



A theorem of Lyons (2005) shows that for any graphs G_n, the existence of a weak limit in an appropriate sense implies the existence of

$$\lim_{n\to\infty} \frac{\log \tau(G_n)}{|\mathsf{V}(G_n)|}\,.$$

If the weak limit of G_n is a probability measure ρ on rooted graphs (G, o), then the above asymptotic equals the "tree entropy" of ρ,

$$\mathbf{E}_\rho\Big[\log \deg_G(o) - \sum_{k\geq 1} \frac{1}{k} p_k(o; G)\Big]\,,$$

where $p_k(o; G)$ is the probability that simple random walk on G started at o is again at o after k steps.

Let $f(c)$ be the tree entropy of $\mathsf{PGW}^*(c)$. Intuitively $f(c)$ increases for large c and for small c. Is it increasing for all c? Is $\mathsf{PGW}^*(c)$ increasing in c?

If X and Y are real-valued random variables, we say that X is **stochastically dominated** by Y if there are copies X' and Y' such that $X' \leq Y'$ a.s. We then write $X \preccurlyeq Y$.

For example, $\mathsf{Pois}(c) \preccurlyeq \mathsf{Pois}(c')$ for $c < c'$. One way to couple them is

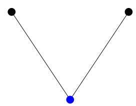
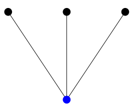
$$\mathsf{Pois}(c) \oplus \mathsf{Pois}(c' - c) = \mathsf{Pois}(c') \,.$$

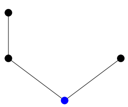
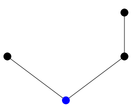
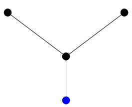
There is a similar notion for random rooted trees, with $T_1 \leq T_2$ meaning $T_1 \subseteq T_2$ and they have the same root. Thus, $\mathsf{PGW}(c) \preccurlyeq \mathsf{PGW}(c')$ for $c < c'$.

What's the distribution of a $\mathsf{PGW}(c)$ tree given to have k vertices in total? If the numbers of children of the vertices of a tree of total size k are $d_1, \ldots, d_k$ in depth-first order, then the probability of this tree is

$$\prod_{i=1}^{k} e^{-c} \frac{c^{d_i}}{d_i!} = e^{-ck} c^{k-1} \prod_{i=1}^{k} \frac{1}{d_i!} \,.$$

When we condition that the size is k, then the terms involving c disappear.

Thus, $\mathbf{P}\Big[$  $\Big] = 1/3$ and $\mathbf{P}\Big[$ $\Big] = 2/3$. Likewise, $\mathbf{P}\Big[$  $\Big] = 1/16$,

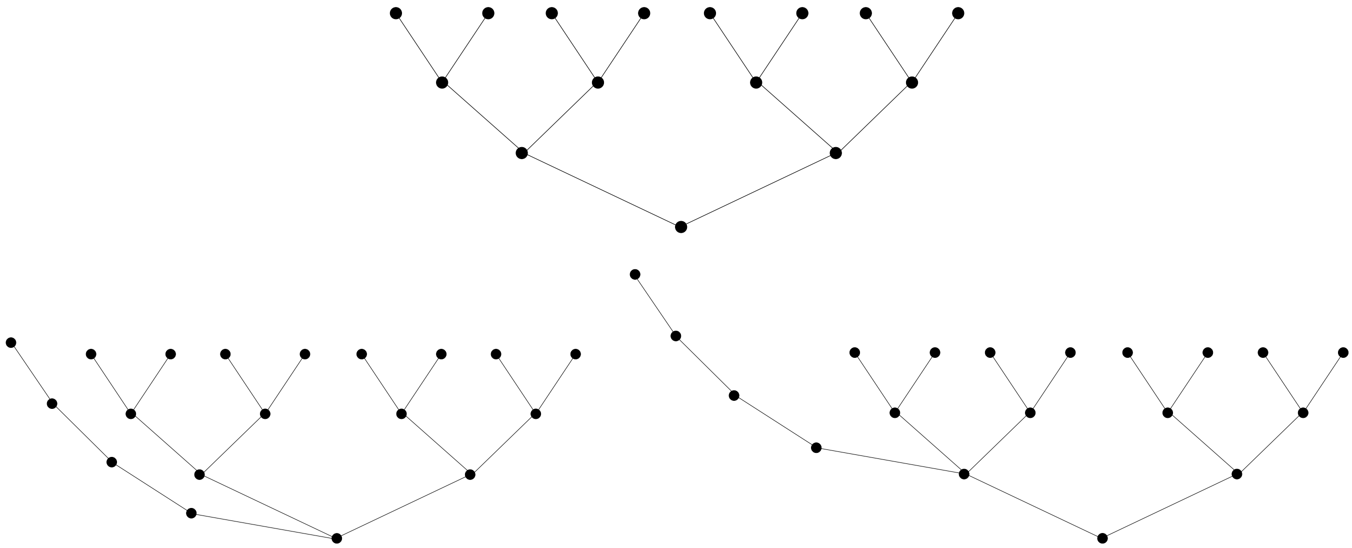
$\mathbf{P}\Big[$  $\Big] = \mathbf{P}\Big[$  $\Big] = \mathbf{P}\Big[$  $\Big] = 3/16$, and $\mathbf{P}\Big[$ $\Big] = 3/8$.

Let T_k be a PGW tree on k vertices. From the above, we see that $\mathsf{T}_1 \preccurlyeq \mathsf{T}_2 \preccurlyeq \mathsf{T}_3 \preccurlyeq \mathsf{T}_4$. In fact, Luczak and Winkler (2004) proved (essentially) that $\mathsf{T}_k \preccurlyeq \mathsf{T}_{k+1}$ for all $k \geq 1$.

If ρ increases stochastically, then does its tree entropy increase?

$$\text{tree entropy}(\rho) = \mathbf{E}_\rho\Big[\log \deg_G(o) - \sum_{k\geq 1} \frac{1}{k} p_k(o;G)\Big]$$

The first term in the tree entropy would then clearly increase, but what about the second term? Not necessarily!



However, stochastic increase of ρ (among limits of finite graphs) *does* imply increase of tree entropy (L., 2007).

Recall that $\mathsf{PGW}(c)$ increases stochastically since $\mathsf{Pois}(c)$ does. But we are interested in $\mathsf{PGW}^*(c)$, the tree conditioned on survival. Does $\mathsf{PGW}^*(c)$ increase stochastically? What does conditioning do? It is *not* true that if one offspring distribution dominates another, then the corresponding Galton-Watson trees conditioned on survival have the same relationship. Example: $p_0 = p_3 = 1/2$ compared to $p_1 = p_3 = 1/2$.

Our new work shows that $\mathsf{PGW}^*(c)$ *is* stochastically increasing: therefore $f(c)$ is strictly increasing in c. Furthermore, the second part of the tree entropy,

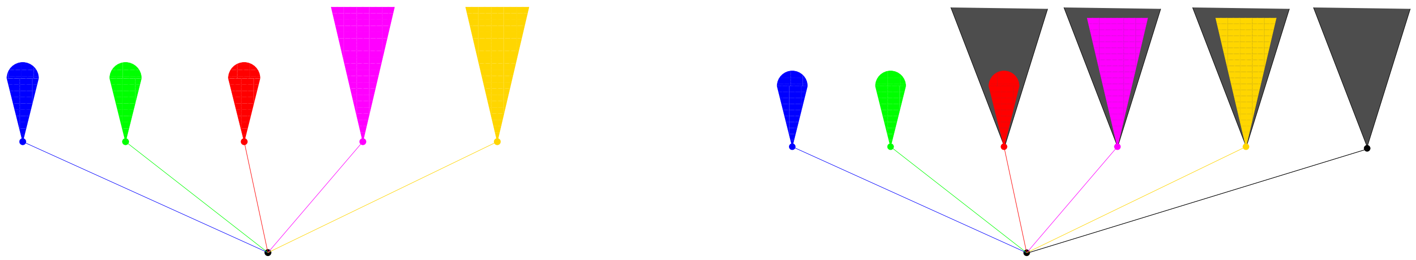
$$\mathbf{E}_{\mathsf{PGW}^*(c)}\Big[\sum_{k\geq 1}\frac{1}{k}p_k(o;G)\Big]\,,$$

is decreasing in c, which leads to an explicit lower bound on the derivative of f:

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

where $q = e^{c(q-1)}$ and $0 < q < 1$. We also show that $f \in C^\infty(1,\infty)$.

We will couple $\mathsf{PGW}^*(c)$ and $\mathsf{PGW}^*(c')$ as follows for $c < c'$:



This requires the following to hold:

- for each $k < \infty$, the number of children of the root with descendant subtree of size k decreases; in fact, the same holds simultaneously in k

- simultaneously the number of children of the root with infinite descendant subtree increases enough to make up the difference in number of finite subtrees

- finite subtrees can be embedded in infinite ones

To prove that $\mathsf{PGW}^*(c)$ stochastically increases, classify children of the root by their total progeny size $k = 1, 2, \ldots, \infty$ (where the progeny includes the individual and all its descendants).

board picture explaining children and progeny

Let N_k be the number where the progeny size is k. These are independent Poisson random variables thinned from $\mathsf{Pois}(c)$, except that N_∞ is conditioned to be at least 1.

How do they behave as c increases?

What is the probability prg_k that a $\mathsf{PGW}(c)$ tree has total size k?

Recall that the component of o in $\mathcal{G}(n, c/n)$ tends weakly to $\mathsf{PGW}(c)$. Thus, the probability prg_k is the same as the limiting probability that the component of o in $\mathcal{G}(n, c/n)$ is a tree of size k. Cayley proved that there are k^{k-2} trees on k (distinguishable) vertices. Thus, the expected number of trees of size k in $\mathcal{G}(n, c/n)$ is

$$\binom{n}{k} k^{k-2} \left(\frac{c}{n}\right)^{k-1} \left(1 - \frac{c}{n}\right)^{k(n-k)+\binom{k}{2}-k+1} \sim n \frac{k^{k-2}}{k!} c^{k-1} \left(1 - \frac{c}{n}\right)^{kn} .$$

Since all vertices have the same probability $\mathsf{prg}_k^{(n)}$ to be in a tree of size k, we have

$$n \cdot \mathsf{prg}_k^{(n)} = k \cdot \mathbf{E}[\text{number of trees of size } k] \, .$$

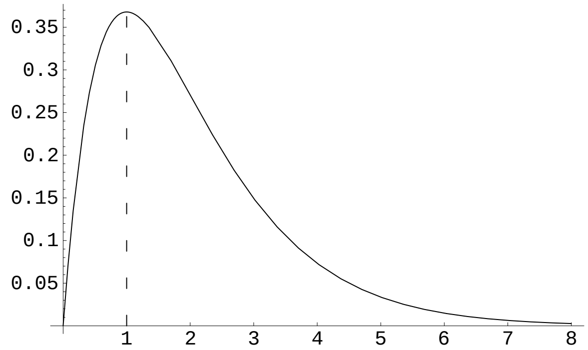
Thus,

$$\mathsf{prg}_k = \lim_{n\to\infty} \mathsf{prg}_k^{(n)} = \lim_{n\to\infty} \frac{k^{k-1}}{k!} c^{k-1} \left(1 - \frac{c}{n}\right)^{kn} = \frac{(ce^{-c})^k k^{k-1}}{ck!} \, .$$

(This is the Borel(c) distribution.) Thus,

$$N_k \sim \mathsf{Pois}\left(\frac{(ce^{-c})^k k^{k-1}}{k!}\right) \qquad \text{for } k < \infty \, .$$

$$N_k \sim \text{Pois}\left(\frac{(ce^{-c})^k k^{k-1}}{k!}\right) \qquad \text{for } k < \infty\,.$$



Since $c \mapsto ce^{-c}$ has its only local maximum at $c = 1$, the distribution of N_k is stochastically *decreasing* in c for $k < \infty$.

Let $\mathsf{Pois}^*(\lambda)$ denote $\mathsf{Pois}(\lambda)$ conditioned to be at least 1. Then $N_\infty \sim \mathsf{Pois}^*\big(c(1-q)\big)$, where $q = q(c)$ is the extinction probability. The p.g.f. of $\mathsf{Pois}(c)$ is $s \mapsto e^{c(s-1)}$, so the extinction probability of $\mathsf{PGW}(c)$ is the smallest positive solution to

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

Note that

$$ce^{-c} = cqe^{-cq} .$$

Thus, cq decreases in c, whence $c(1-q) = c - cq$ increases in c.

How much does $N_\infty \sim \mathsf{Pois}^*\big(c(1-q)\big)$ increase in c? If $1 < \lambda < \mu$, then

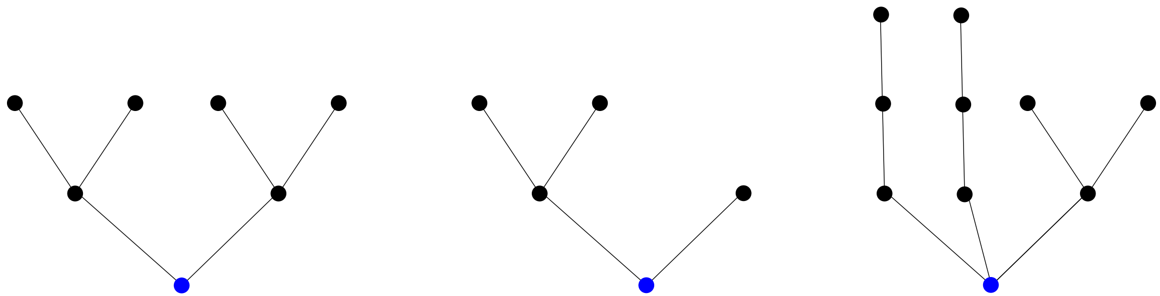
$$\mathsf{Pois}^*(\lambda) \oplus \mathsf{Pois}(\alpha) \preccurlyeq \mathsf{Pois}^*(\mu)$$

when

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

(This is best possible: look at the probability of 1.)

If T and T' are rooted trees, 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 for every child x of the root in T, the number of progeny of x is at most that of $i(x)$.



If T and T' are random rooted trees, write $T \preccurlyeq_1 T'$ if T and T' may be coupled in such a way that $T \leq_1 T'$ a.s.

Recall that T_k is a PGW tree on k vertices and that $\mathsf{T}_k \preccurlyeq \mathsf{T}_{k+1}$.
This implies $\mathsf{T}_k \preccurlyeq_1 \mathsf{T}_{k+1}$ and conversely.

Claim: If $1 < c < c'$, then $\mathsf{PGW}^*(c) \preccurlyeq_1 \mathsf{PGW}^*(c')$.

Proof: It suffices that $N_\infty(c) \oplus \mathsf{Pois}\big(cq - c'q'\big) \preccurlyeq N_\infty(c')$. This follows from

$$cq - c'q' \leq \alpha\big(c(1-q), c'(1-q')\big) \,.$$

Claim: If $c > 1$ and $k < \infty$, then $\mathsf{T}_k \preccurlyeq_1 \mathsf{PGW}^*(c)$.

Proof: Let $\mathsf{PGW}^*(1)$ denote the weak limit of $\mathsf{PGW}^*(c)$ as $c \downarrow 1$. This is the same as the (increasing) weak limit of T_k as $k \to \infty$. Since $\mathsf{T}_k \preccurlyeq_1 \mathsf{T}_{k+1}$ and $\mathsf{PGW}^*(1) \preccurlyeq_1 \mathsf{PGW}^*(c)$ by the preceding claim, the result follows.

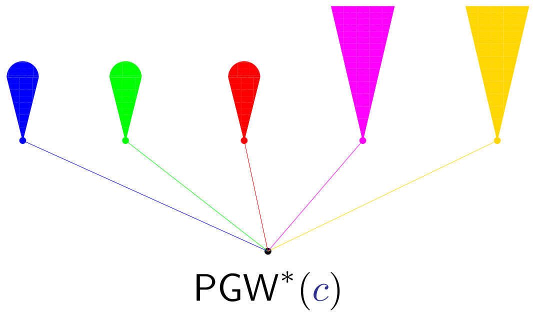
So

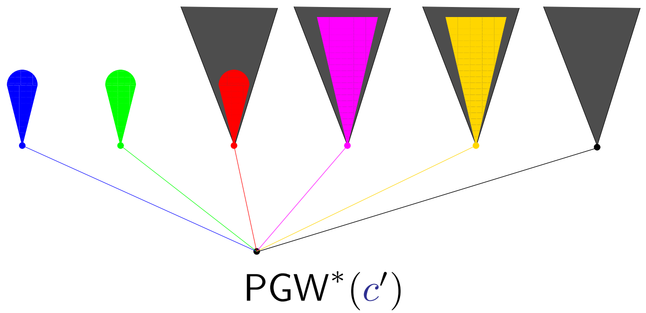
$$1 < c < c' \implies \mathsf{PGW}^*(c) \preccurlyeq_1 \mathsf{PGW}^*(c') \tag{1}$$

and

$$\mathsf{T}_k \preccurlyeq_1 \mathsf{PGW}^*(c)\,. \tag{2}$$

To complete the proof that $\mathsf{PGW}^*(c) \preccurlyeq \mathsf{PGW}^*(c')$, we apply (1) and (2) to each vertex of $\mathsf{PGW}^*(c)$.





Open Questions

Does one have similar stochastic domination for supercritical percolation on regular trees? This is almost the same as supercritical Galton-Watson trees with binomial offspring distribution.

Does one have similar stochastic domination for supercritical percolation on $\mathbb{Z}^d$? If so, one would prove the existence of the "incipient infinite cluster".

Does one have similar stochastic domination for the rooted giant component before taking the limit?

How does this give a rate of increase? Recall that

$$f(c) = \mathbf{E}_{\mathsf{PGW}^*(c)}\Big[\log \deg_T(o) - \sum_{k\geq 1} \frac{1}{k} p_k(o;T)\Big] .$$

The expectation of the first term is clearly increasing in c. We claim the expectation of the second term (without the negative sign) is decreasing in c. Recall this is not true for $T_1 \subset T_2$, unless their roots have the same number of children.

Let $X(T,o) := \sum_{k\geq 0} p_k(o;T) s^k$, the expected number of visits V to the root for a random walk on T with geometric killing. Since

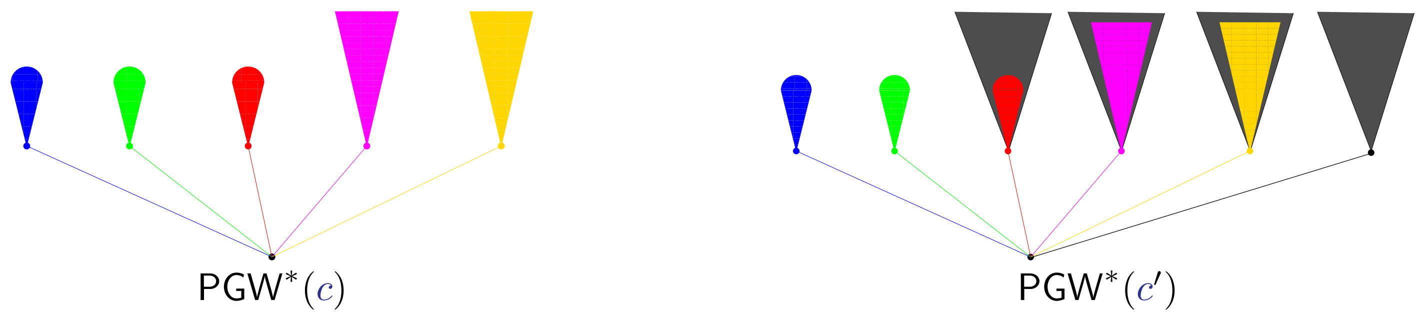
$$\int_0^1 \frac{\mathbf{E}[V]-1}{s}\, ds = \int_0^1 \frac{\mathbf{E}[X]-1}{s}\, ds = \mathbf{E}\left[\sum_{k\geq 1} \frac{1}{k} p_k(o;T)\right] ,$$

it suffices to show that the $\mathsf{PGW}^*(c)$-law of the random variable V is stochastically decreasing in $c > 1$. Again, this is true for fixed trees $T_1 \subset T_2$ with the same root degree.

Let $Q = Q(T, o)$ be the probability of at least one return to the root. If the root has d children and (T, o) is obtained by identifying the roots of the d subtrees (T_i, o), then

$$Q = \frac{1}{d} \sum_{i=1}^{d} Q_i \,.$$

For $1 < c < c'$, we can couple the trees T and T' so that $d \le d'$, $T_i \subseteq T'_i$ for $1 \le i \le d$, all T_i are independent, all T'_i are independent, and the distribution of T'_j for $j > d$ dominates that of T_i for $i \le d$. This means that $Q_i \ge Q'_i$ for $i \le d$ and the distribution of Q_i for each $i \le d$ dominates that of Q'_j for all $j > d$.



We want to show that for each $k \geq 1$,

$$\mathbf{P}[V(T, o) > k] \geq \mathbf{P}[V(T', o) > k] .$$

Now

$$\mathbf{P}[V(T, o) > k] = \mathbf{E}[Q^k] = \mathbf{E}\left[\left(\frac{1}{d}\sum_{i=1}^{d} Q_i\right)^k\right] .$$

We want to show this is at least

$$\mathbf{E}\left[\left(\frac{1}{d'}\Big[\sum_{i=1}^{d} Q'_i + \sum_{i=d+1}^{d'} Q'_i\Big]\right)^k\right] .$$

We write the average of d' terms as the expectation of a random average of d terms:

$$\mathbf{E}\left[\left(\mathbf{E}_\chi \frac{1}{d}\Big[\sum_{i=1}^{d} \chi_i Q'_i + \sum_{i=d+1}^{d'} \chi_i Q'_i\Big]\right)^k\right] .$$

$$\mathbf{E}\left[\left(\mathbf{E}_\chi \frac{1}{d}\Big[\sum_{i=1}^{d} \chi_i Q_i' + \sum_{i=d+1}^{d'} \chi_i Q_i'\Big]\right)^k\right].$$

By convexity, this is at most

$$\mathbf{E}\left[\mathbf{E}_\chi\left(\frac{1}{d}\Big[\sum_{i=1}^{d} \chi_i Q_i' + \sum_{i=d+1}^{d'} \chi_i Q_i'\Big]\right)^k\right].$$

By monotonicity, this is at most

$$\mathbf{E}_\chi \mathbf{E}\left[\left(\frac{1}{d}\sum_{i=1}^{d} Q_i\right)^k\right] = \mathbf{E}\left[\left(\frac{1}{d}\sum_{i=1}^{d} Q_i\right)^k\right].$$

This completes the proof.