Math. Proc. Camb. Phil. Soc. (1986), **100**, 319
Printed in Great Britain

Random trees in a graph and trees in a random graph

BY SVANTE JANSON
Department of Mathematics, Uppsala University, Sweden

(*Received* 16 *October* 1985)

0. *Introduction*

This paper treats two related sets of problems in the theory of random graphs. In Sections 2 and 3 we study random spanning subtrees of a complete graph (or, equivalently, random labelled trees). It is shown that the number of common edges of two such random trees asymptotically has a Poisson distribution with expectation 2. Similar results are obtained for the number of edges in the intersection or union of more than two random trees.

In Sections 4 and 5 we study the random graphs $G_{n,p}$ and $G_{n,N}$ (see these sections for definitions). The results of the preceding sections are applied to give asymptotic expressions for the moments of the complexity, i.e. the number of spanning subtrees, of the random graphs. The results for the two models $G_{n,p}$ and $G_{n,N}$ are quite different; if we normalize the complexity by dividing it by its expectation, then we obtain convergence to a log-normal distribution for $G_{n,p}$ and to a constant for $G_{n,N}$.

Section 6 contains results for arbitrary subtrees analogous to the results for spanning subtrees in Sections 2–5.

1. *Notation*

We use the following notations. (G, G_1, G_2 are graphs.)

$|G|$ denotes the number of vertices of G.

$\|G\|$ denotes the number of edges of G.

$G_1 \subset G$ means that G_1 is a subgraph of G, i.e. the sets of vertices and edges of G_1 are subsets of the corresponding sets for G.

$G_1 \cap G_2$ denotes the graph consisting of all edges and vertices common to G_1 and G_2.

A tree is a connected graph without cycles. If T is a tree then $\|T\| = |T| - 1$. Note that a graph G with $|G| = 1$ is a tree.

A forest is a disjoint union of trees, i.e. a graph whose connected components are trees.

T_n, T'_n etc. denote random trees (see Section 2), while t, t_1, etc. denote non-random trees.

$I(\ldots)$ denotes the indicator function which is 1 if the condition in the parenthesis is satisfied, and 0 otherwise.

$f \sim g$ means that $f(n)/g(n) \to 1$ as $n \to \infty$.

Some further notation is introduced later when it is used.

2. *Intersections of random spanning trees*

Consider the complete graph K_n on n vertices. A spanning subtree of K_n is a tree $t \subset K_n$ with $|t| = |K_n| = n$. There is an obvious isomorphism between the set of spanning subtrees of K_n and the set of labelled trees on n points. Cayley [2] proved that

https://doi.org/10.1017/S0305004100066111 Published online by Cambridge University Press

the number of such trees equals n^{n-2}. Choosing one of these trees at random (with equal probabilities n^{2-n}), we obtain a random spanning tree, denoted by T_n.

Let $G \subset K_n$ be a forest and let its connected components be the trees $t_1, \dots, t_m$. Then Moon ([7], theorem 2) has shown, generalizing Cayley's result above, that the number of spanning subtrees of K_n that contain G equals $n^{n-2} \prod_{i=1}^m |t_i| n^{1-|t_i|}$. Hence

$$P(T_n \supset G) = \prod_{i=1}^m |t_i| n^{1-|t_i|}. \tag{2·1}$$

In other words, the probability that T_n contains a given tree $t \subset K_n$ equals $|t| n^{1-|t|}$ and if $t_1, \dots, t_m$ are disjoint subtrees of K_n, then the events $\{t_1 \subset T_n\}, \dots, \{t_m \subset T_n\}$ are independent. Since obviously $P(G \subset T_n) = 0$ if G is not a forest, we may more generally say that if $G_1, \dots, G_m$ are any disjoint subgraphs of K_n then the events $\{G_i \subset T_n\}$ are independent.

Let T_n' be an independent copy of T_n. (In other words, T_n and T_n' are obtained by sampling with replacement from the set of all spanning trees in K_n.) Denote the number of common edges in T_n and T_n' by Δ_n, i.e. $\Delta_n = \|T_n \cap T_n'\|$. We will prove that the random variable Δ_n is asymptotically Poisson distributed.

THEOREM 1. $\Delta_n \xrightarrow{d} \mathrm{Po}(2)$ *as* $n \to \infty$, *i.e.*

$$P(\Delta_n = k) \to \frac{2^k}{k!} e^{-2} \quad \text{as} \quad n \to \infty$$

for each $k = 0, 1, \dots$.

Proof. Let $\mathbf{E}$ denote the set of edges of K_n. For $e \in \mathbf{E}$, let X_e be the indicator variable $I(e \in T_n \cap T_n') = I(e \in T_n) I(e \in T_n')$. Then

$$\Delta_n = \|T_n \cap T_n'\| = \sum_e X_e. \tag{2·2}$$

By (2·1), $P(e \in T_n) = 2/n$ for every e. Since T_n and T_n' are independent and identically distributed,

$$P(X_e = 1) = P(e \in T_n \quad \text{and} \quad e \in T_n') = P(e \in T_n) P(e \in T_n') = (2/n)^2. \tag{2·3}$$

Furthermore, it follows from the discussion after (2·1) that if e and f are two edges with disjoint pairs of endpoints, then X_e and X_f are independent and, more generally, if $\mathbf{E}_1$ and $\mathbf{E}_2$ are two sets of edges in K_n such that no vertex is a common endpoint of an edge in $\mathbf{E}_1$ and an edge in $\mathbf{E}_2$, then $\{X_e\}_{e \in \mathbf{E}_1}$ and $\{X_e\}_{e \in \mathbf{E}_2}$ are independent. If we change the notation slightly by setting $X_{ij} = X_{e_{ij}}$ where e_{ij} is the edge joining i and j $(i \neq j)$ (we assume that the vertices of K_n are numbered $1, \dots, n$), this means that the set $\{X_{ij}\}$ is a dissociated set of random variables in the terminology of McGinley and Sibson [5], and the desired convergence of $\Delta_n = \sum_{i<j} X_{ij}$ follows by a theorem of Barbour and Eagleson ([1], theorem 2) provided that, as $n \to \infty$ and with $\lambda = 2$,

$$\sum_{i<j} EX_{ij} \to \lambda, \tag{2·4}$$

$$\sum_{\substack{i \neq j \\ i \neq k}} EX_{ij} EX_{ik} \to 0, \tag{2·5}$$

$$\sum_{\substack{i \neq j \\ i \neq k \\ j \neq k}} EX_{ij} X_{ik} \to 0. \tag{2·6}$$

https://doi.org/10.1017/S0305004100066111 Published online by Cambridge University Press

There conditions are verified as follows. By (2·3),

$$E\Delta_n = \sum_{i<j} EX_{ij} = \binom{n}{2}\frac{4}{n^2} = 2\left(1-\frac{1}{n}\right)\to 2, \tag{2·7}$$

$$\Sigma EX_{ij}EX_{ik} = \sum_{ijk}\frac{16}{n^4} \leqslant n^3\frac{16}{n^4}\to 0.$$

Furthermore, if i, j, k are distinct, then these three vertices and the two edges e_{ij}, e_{jk} form a tree, t say, with $|t| = 3$. By independence and (2·1) we obtain

$$\begin{aligned} EX_{ij}X_{ik} &= P(e_{ij}\in T_n \cap T'_n \quad \text{and} \quad e_{jk}\in T_n \cap T'_n) \\ &= P(t \subset T_n \cap T'_n) = P(t\subset T_n)P(t\subset T'_n) \\ &= P(t\subset T_n)^2 = (3n^{-2})^2 = 9n^{-4}. \end{aligned}$$

Hence

$$\sum_{i\neq j\neq k\neq i} EX_{ij}X_{ik} \leqslant n^3.9n^{-4} = 9/n\to 0. \quad \blacksquare$$

If X is any non-negative integer-valued random variable, then its probability generating function ψ_X is defined by $\psi_X(s) = Es^X$. $\psi_X(s)$ is well-defined for every $s \geqslant 0$ if we allow $+\infty$ as a possible value. Recall that a sequence $\{X_n\}$ of non-negative integer valued random variables converges in distribution iff $\psi_{X_n}(s)\to\psi_X(s)$ for every $s\in[0,1]$ (see Feller [3], section XI·6), but that, in general, convergence for $s > 1$ is not guaranteed. We will improve Theorem 1 by proving that in this case the probability generating functions do converge on the entire positive real axis. We need the following standard lemma.

LEMMA 1. *Suppose that $X, X_1, X_2, \ldots$ are non-negative integer-valued random variables such that $X_n \xrightarrow{d} X$ as $n\to\infty$ and $\sup_n \psi_{X_n}(u) < \infty$ for some $u \leqslant 0$. Then $\psi_{X_n}(s)\to\psi_X(s)$ as $n\to\infty$ for all s, $0 \leqslant s < u$.*

Proof. Since the conclusion holds for $0 \leqslant s \leqslant 1$, we may assume that $1 < s < u$. Let $p = \log u/\log s > 1$. Since $E(s^{X_n})^p = Eu^{X_n} = \psi_{X_n}(u)$, the random variables s^{X_n} have uniformly bounded pth moments, and are thus uniformly integrable. Since s^{X_n} converges in distribution to s^X, it follows that $\psi_{X_n}(s) = Es^{X_n}\to Es^X = \psi_X(s)$. $\blacksquare$

Remark 2·1. It follows similarly from the assumptions in Lemma 1 that

$$\psi_{X_n}(s)\to\psi_X(s)$$

for every complex s with $|s| < u$. Hence the generating functions in the theorems below are entire functions converging uniformly on compact subsets of the complex plane.

Recall that the probability generating function of a random variable with the Poisson distribution Po(λ) is

$$\sum_0^\infty \frac{\lambda^k}{k!}e^{-\lambda}s^k = e^{\lambda(s-1)}.$$

THEOREM 2. *For every $s \geqslant 0$, $\psi_{\Delta_n}(s)\to e^{2(s-1)}$ as $n\to\infty$.*

Proof. We derive an explicit formula for $\psi_{\Delta_n}(s)$. We start with the identity, valid for every subgraph H of K_n,

$$s^{\|H\|} = \sum_{G\subset K_n}{}' I(G\subset H)(s-1)^{\|G\|}, \tag{2·8}$$

where Σ' here and in the remainder of the proof signifies that we only sum over graphs

https://doi.org/10.1017/S0305004100066111 Published online by Cambridge University Press

G that do not have any isolated vertices. To prove (2·8), note that graphs without isolated vertices are completely specified by their sets of edges. Thus, if $\|H\| = m$, then for every j there are exactly $\binom{m}{j}$ subgraphs of H with exactly j edges and no isolated vertices; hence the right hand side of (2·8) equals

$$\sum_{j=0}^{m} \binom{m}{j} (s-1)^j = (1+(s-1))^m = s^m.$$

Letting $H = T_n \cap T_n'$, we obtain by taking expectations in (2·8), again using the fact that T_n and T_n' are independent and identically distributed,

$$\begin{aligned} \psi_{\Delta_n}(s) = Es^{\|T_n \cap T_n'\|} &= \sum_{G \subset K_n}' P(G \subset T_n \cap T_n')\,(s-1)^{\|G\|} \\ &= \sum_{G \subset K_n}' P(G \subset T_n)^2\,(s-1)^{\|G\|}. \end{aligned} \tag{2·9}$$

If G contains any cycle, then $P(G \subset T_n) = 0$. Hence we only have to sum over forests $G \subset K_n$. Furthermore, if G is a forest consisting of disjoint trees $t_1, \dots, t_m$ with $|t_i| = k_i$, $i = 1, \dots, m$, then $P(G \subset T_n)^2 = \prod_1^m k_i^2 n^{2-2k_i}$ by (2·1). To count the number of such subforests of K_n, let us temporarily order the components of the forests and distinguish between two forests consisting of the same trees in different order. The number of ordered forests in K_n with m components consisting of $k_1, \dots, k_m$ vertices, respectively, equals

$$\binom{n}{k_1, \dots, k_m} \prod_1^m k_i^{k_i-2},$$

because the sets of vertices of the components may be chosen in $\binom{n}{k_1, \dots, k_m}$ ways and the ith of these sets may be made into a tree in $k_i^{k_i-2}$ ways by Cayley's formula.

Multiplying this number by the corresponding probability given above, summing over all $k_1, \dots, k_m \geqslant 2$ (since no isolated vertices are allowed), dividing by $m!$ since every forest is counted with $m!$ different orderings of its components, and finally summing over m, we obtain

$$\psi_{\Delta_n}(s) = \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{k_1=2}^{\infty} \cdots \sum_{k_m=2}^{\infty} \binom{n}{k_1, \dots, k_m} \prod_{i=1}^{m} k_i^{k_i-2+2} n^{2-2k_i} (s-1)^{k_i-1}. \tag{2·10}$$

Using the estimate

$$\binom{n}{k_1, \dots, k_m} \leqslant \frac{n^{k_1+\dots+k_m}}{k_1! \dots k_m!}$$

we see that, for each $s > 1$,

$$\begin{aligned} \psi_{\Delta_n}(s) &\leqslant \sum_{m=0}^{\infty} \frac{1}{m!} \sum_{k_1=2}^{\infty} \cdots \sum_{k_m=2}^{\infty} \prod_{i=1}^{m} \frac{k_i^{k_i}}{k_i!} n^{2-k_i} (s-1)^{k_i-1} \\ &= \sum_{m=0}^{\infty} \frac{1}{m!} \left(\sum_{k=2}^{\infty} \frac{k^k}{k!} n^{2-k} (s-1)^{k-1} \right)^m = \exp\left(\sum_{2}^{\infty} \frac{k^k}{k!} n^{2-k} (s-1)^{k-1} \right). \end{aligned} \tag{2·11}$$

By Stirling's formula, $k! \geqslant k^k e^{-k}$ and thus

$$\psi_{\Delta_n}(s) \leqslant \exp\left(\sum_{2}^{\infty} e^k n^{2-k} (s-1)^{k-1} \right) \quad (s > 1). \tag{2·12}$$

https://doi.org/10.1017/S0305004100066111 Published online by Cambridge University Press

Clearly, the right-hand side of (2·12) is decreasing in n (for fixed s) and finite if $n > e(s-1)$. Hence $\sup_{n>es} \psi_{\Delta_n}(s) < \infty$. Since $\psi_{\Delta_n}(s)$ is finite for all s, n, this gives

$$\sup_n \psi_{\Delta_n}(s) < \infty$$

for every $s > 1$, and the theorem follows by Theorem 1 and Lemma 1. ∎

Remark 2·2. In fact, it is easy to show the result directly, without using Lemma 1 and Theorem 1. By (2·11),

$$\limsup_{n\to\infty} \psi_{\Delta_n}(s) \leqslant \exp\left(\frac{2^2}{2!}(s-1)\right) = e^{2(s-1)} \quad \text{for} \quad s > 1.$$

Conversely, by (2·10), keeping only terms with $k_1 = \ldots = k_m = 2$,

$$\psi_{\Delta_n}(s) \geqslant \sum_{m=0}^{n/2} \frac{1}{m!} \frac{n!}{(n-2m)!\,2^m} (2^2 n^{-2}(s-1))^m \geqslant \sum_{m=0}^{n/2} \frac{1}{m!} \left(\frac{n-2m}{n}\right)^{2m} (2(s-1))^m.$$

By monotone convergence, the right hand side converges to $e^{2(s-1)}$. This furnishes an alternative proof of Theorem 1. However, this method is less convenient in the generalizations below.

Remark 2·3. The convergence of the probability generating functions for (some) $s > 1$ implies convergence of all moments.

3. *Unions of random spanning trees*

It is easy to see that the intersection of three independent random spanning trees T_n, T_n', T_n'' almost surely (in the language of graph theory) contains no edges at all. In fact, it follows as in (2·7) that

$$E\|T_n \cap T_n' \cap T_n''\| = \binom{n}{2}\left(\frac{2}{n}\right)^3 < \frac{4}{n} \to 0 \quad \text{as} \quad n \to \infty.$$

Motivated by the application in the next section, we instead study unions. We generalize to m ($\geqslant 1$) spanning trees and change the notation slightly. Let

$$T_n^{(1)}, T_n^{(2)}, \ldots, T_n^{(m)}$$

be m independent copies of T_n and define

$$\Delta_{m;n} = \sum_1^m \|T_n^{(i)}\| - \left\| \bigcup_1^m T_n^{(i)} \right\| = m(n-1) - \left\| \bigcup_1^m T_n^{(i)} \right\|. \tag{3·1}$$

Note that $\Delta_{m;n} \geqslant 0$, and that $\Delta_{m;n} = 0$ iff $T_n^{(1)}, \ldots, T_n^{(m)}$ have disjoint sets of edges. Furthermore, $\Delta_{1;n} = 0$ while $\Delta_{2;n} = \|T_n^{(1)}\| + \|T_n^{(2)}\| - \|T_n^{(1)} \cup T_n^{(2)}\| = \|T_n^{(1)} \cap T_n^{(2)}\|$ is another name for Δ_n studied above.

The following theorem thus generalizes Theorems 1 and 2.

THEOREM 3. $\Delta_{m;n} \xrightarrow{d} \mathrm{Po}(m(m-1))$ *as* $n \to \infty$ *for every* $m = 1, 2, \ldots$. *The probability generating functions converge correspondingly on the entire positive real axis, i.e.*

$$\psi_{\Delta_{m;n}}(s) \to e^{m(m-1)(s-1)} \quad \text{as} \quad n \to \infty \quad (s \geqslant 0).$$

Proof. We may assume that $m \geqslant 2$. Again, let $\mathbf{E}$ be the set of edges in K_n. Let for

https://doi.org/10.1017/S0305004100066111 Published online by Cambridge University Press

each $e \in \mathbf{E}$, $Y_e = \sum_1^m I(e \in T_n^{(i)})$, i.e. the number of $T_n^{(1)}, \dots, T_n^{(m)}$ that contain e, and define $Z_e = Y_e - I(Y_e \geqslant 1) = \max(Y_e - 1, 0)$ and $X_e = I(Y_e = 2)$. Then

$$\sum_{\mathbf{E}} Z_e = \sum_{\mathbf{E}} Y_e - \sum_{\mathbf{E}} I(Y_e \geqslant 1) = \sum_1^m \sum_{\mathbf{E}} I(e \in T_n^{(i)}) - \sum_{\mathbf{E}} I\left(e \in \bigcup_1^m T_n^{(i)}\right) = \Delta_{m;n}. \tag{3·2}$$

Furthermore, Z_e and X_e differ only when $Y_e \geqslant 3$, and thus, by symmetry,

$$P(Z_e \neq X_e) \leqslant \binom{m}{3} P(e \in T_n^{(1)} \cap T_n^{(2)} \cap T_n^{(3)}) = \binom{m}{3}\left(\frac{2}{n}\right)^3.$$

Thus,

$$P(\Delta_{m;n} \neq \sum_{\mathbf{E}} X_e) = P(\sum_{\mathbf{E}} Z_e \neq \sum_{\mathbf{E}} X_e) \leqslant \sum_{\mathbf{E}} P(Z_e \neq X_e) \leqslant \frac{4}{n}\binom{m}{3} \to 0,$$

as $n \to \infty$. To prove convergence in distribution of $\Delta_{m;n}$, it therefore suffices to prove that $\sum_{\mathbf{E}} X_e \xrightarrow{d} \mathrm{Po}(m(m-1))$, which again follows by Barbour and Eagleson [1] provided (2·4)–(2·6) holds for these X_e with $\lambda = m(m-1)$. (X_{ij} is defined as in the proof of Theorem 1.)

Note that the new variables X_{ij} still are dissociated; e.g. if e and f are two edges without a common endpoint, then X_e is a function of $I(e \in T_n^{(1)}), \dots, I(e \in T_n^{(m)})$ while X_f is a function of $I(f \in T_n^{(1)}), \dots, I(f \in T_n^{(m)})$, and these variables are all independent, whence X_e and X_f are independent. To verify (2·4), we observe that

$$EX_e = \binom{m}{2}\left(\frac{2}{n}\right)^2\left(1 - \frac{2}{n}\right)^{m-2}$$

and thus

$$E\sum_{\mathbf{E}} X_e = \sum_{i<j} EX_{ij} = \binom{n}{2}\binom{m}{2}\left(\frac{2}{n}\right)^2\left(1 - \frac{2}{n}\right)^{m-2} \to m(m-1) \quad \text{as} \quad n \to \infty.$$

(2·5) is verified as before and similar arguments verify (2·6). The proof of the first assertion is complete.

To prove the second assertion, we use the trivial estimate

$$Z_e \leqslant m \max_{1 \leqslant k < l \leqslant m} I(e \in T_n^{(k)} \cap T_n^{(l)}).$$

Thus

$$\Delta_{m;n} = \sum_{\mathbf{E}} Z_e \leqslant m \sum_{\mathbf{E}} \sum_{k<l} I(e \in T_n^{(k)} \cap T_n^{(l)}) = m \sum_{k<l} \|T_n^{(k)} \cap T_n^{(l)}\| \leqslant m^3 \max_{k<l} \|T_n^{(k)} \cap T_n^{(l)}\|$$

and, for $s > 1$,

$$\begin{aligned}\psi_{\Delta_{m;n}}(s) = Es^{\Delta_{m;n}} &\leqslant E \max_{k<l} s^{m^3\|T_n^{(k)} \cap T_n^{(l)}\|} \leqslant E \sum_{k<l} s^{m^3\|T_n^{(k)} \cap T_n^{(l)}\|} \\ &= \binom{m}{2} Es^{m^3\|T_n^{(1)} \cap T_n^{(2)}\|} = \binom{m}{2}\psi_{\Delta_n}(s^{m^3}).\end{aligned}$$

The last term is bounded as $n \to \infty$ for every fixed $s > 1$ by Theorem 2; hence the proof is completed by Lemma 1. ∎

Remark 3·1. The probability generating functions converge uniformly on bounded subsets of R_+. Hence the result may be strengthened to:

$$s_n \to s \quad \text{implies} \quad \psi_{\Delta_{m;n}}(s_n) \to e^{m(m-1)(s-1)} \quad (n \to \infty).$$

https://doi.org/10.1017/S0305004100066111 Published online by Cambridge University Press

Remark 3·2. In the theorems above, $T_n^{(1)}, \dots, T_n^{(m)}$ are obtained by sampling with replacement from the set of all spanning trees in K_n. It follows easily that the same asymptotic results hold for sampling without replacement.

4. *Spanning trees in a random graph*

Let G be a graph. We define $c(G)$ to be the number of spanning subtrees of G, i.e. the number of trees $t \subset G$ with $|t| = |G|$. (For example, $c(K_n) = n^{n-2}$. $c(G) > 0$ iff G is connected.) $c(G)$ is known as the complexity of the graph G.

The random graph $G_{n,p}$, $0 \leqslant p \leqslant 1$, is the subgraph of K_n obtained by deleting each edge, independently of the other edges, with probability $1-p$. Hence each edge appears with probability p and, if G is a fixed subgraph of K_n, then

$$P(G \subset G_{n,p}) = p^{\|G\|}. \tag{4·1}$$

We may now compute moments of the random variable $c(G_{n,p})$.

THEOREM 4. *Let* $0 < p < 1$ *and* $m \geqslant 1$. *Then*

$$E(c(G_{n,p}))^m = (n^{n-2}p^{n-1})^m \psi_{\Delta_m;n}\left(\frac{1}{p}\right). \tag{4·2}$$

Hence, with $q = 1-p$,

$$E(c(G_{n,p}))^m \sim (n^{n-2}p^{n-1})^m e^{m(m-1)q/p} \quad \text{as} \quad n \to \infty, \tag{4·3}$$

(*with equality for* $m = 1$ *and every* n), *i.e. the moments of the normalized variable*

$$c(G_{n,p})/Ec(G_{n,p}) = (n^{n-2}p^{n-1})^{-1} c(G_{n,p})$$

converge to the moments of the log–normal distribution e^Z, *with* $Z \in N(-q/p, 2q/p)$.

Proof. Clearly, for any graph $G \subset K_n$ with $|G| = n$,

$$c(G) = \sum_t{}^* I(t \subset G),$$

where $\sum^*$ here and below denotes summation over all spanning trees in K_n. Hence

$$c(G)^m = \sum_{t_1}{}^* \dots \sum_{t_m}{}^* I(t_1 \subset G) \dots I(t_m \subset G) = \sum_{t_1}{}^* \dots \sum_{t_m}{}^* I\left(\bigcup_1^m t_i \subset G\right).$$

We apply this to $G_{n,p}$, take the expectation and obtain by (4·1)

$$Ec(G_{n,p})^m = \sum_{t_1}{}^* \dots \sum_{t_m}{}^* P\left(\bigcup_1^m t_i \subset G_{n,p}\right) = \sum_{t_1}{}^* \dots \sum_{t_m}{}^* p^{\|\bigcup_1^m t_i\|}.$$

If we divide this equation by $(n^{n-2})^m$, then the right hand side becomes the average of $p^{\|\bigcup_1^m t_i\|}$ over all possible choices of $t_1, \dots, t_m$. This equals, by definition, the expectation of this quantity when $t_1, \dots, t_m$ are chosen at random. Using our previous notation we thus obtain

$$\begin{aligned}(n^{n-2})^{-m} Ec(G_{n,p})^m &= Ep^{\|\bigcup_1^m T_n^{(i)}\|} = Ep^{m(n-1)-\Delta_{m;n}} \\ &= (p^{n-1})^m E\left(\frac{1}{p}\right)^{\Delta_{m;n}} = (p^{n-1})^m \psi_{\Delta_m;n}\left(\frac{1}{p}\right),\end{aligned} \tag{4·4}$$

which proves (4·2). Note in particular that $m = 1$ gives

$$Ec(G_{n,p}) = n^{n-2}p^{n-1}. \tag{4·5}$$

https://doi.org/10.1017/S0305004100066111 Published online by Cambridge University Press

(4·3) follows by Theorem 3, while the last assertion follows from the fact that if $Z \in N(\mu, \sigma^2)$, then e^Z has moments $E(e^Z)^m = Ee^{mZ} = e^{m\mu + m^2\sigma^2/2}$. |

Unfortunately, the moments of the log-normal distribution grow too fast for the method of moments to be applicable. In the next section, convergence in distribution will be established by a different method.

5. *More on spanning trees in random graphs*

A different model of random graphs is $G_{n,N}$, $0 \leqslant N \leqslant \binom{n}{2}$, which has n vertices and N edges; $G_{n,N}$ is obtained by choosing at random (with equal probabilities) one of the $\binom{n(n-1)/2}{N}$ sets of N edges in K_n. For most properties of random graphs, $G_{n,N}$ and $G_{n,p}$ with $p = N/\binom{n}{2}$ behave almost identically, at least asymptotically as $n \to \infty$. We will see that this is not the case for the number of spanning subtrees; we obtain different asymptotics for the moments and for the distributions of $c(G_{n,N})$ and $c(G_{n,p})$. We begin by computing the moments (asymptotically). (The expectation $Ec(G_{n,N})$ is given by Ford and Uhlenbeck [4]; the second moment is given by Moon [6].)

THEOREM 5. *Suppose that* $n \to \infty$ *and* $N\Big/\binom{n}{2} \to p$ *with* $0 < p \leqslant 1$. *Then for each* $m \geqslant 1$, *with* $q = 1-p$,

$$E(c(G_{n,N}))^m \sim \left(n^{n-2}\left(N\Big/\binom{n}{2}\right)^{n-1} e^{-q/p}\right)^m \quad \text{as} \quad n \to \infty. \tag{5·1}$$

Hence all moments of the normalized variable $c(G_{n,N})/Ec(G_{n,N})$ *or*

$$\left(n^{n-2}\left(N\Big/\binom{n}{2}\right)^{n-1} e^{-q/p}\right)^{-1} c(G_{n,N})$$

converge to 1, *whence these variables converge to* 1 *in probability.*

Remark 5·1. The above formulation with a p is chosen to facilitate comparisons with Theorem 4. In fact, as is easily shown, if q/p is replaced by $\binom{n}{2}\Big/N - 1$, then the results are true provided only $n \to \infty$ and $\liminf N\Big/\binom{n}{2} > 0$.

Proof. Let us write $M = \binom{n}{2}$. There are $\binom{M}{N}$ possible choices of $G_{n,N}$, and if G is a fixed subgraph of K_n then exactly $\binom{M - \|G\|}{N - \|G\|}$ of these contain G. Consequently (4·1) is replaced by

$$P(G \subset G_{n,N}) = \binom{M-\|G\|}{N-\|G\|}\Big/\binom{M}{N} = \frac{(N)_{\|G\|}}{(M)_{\|G\|}}. \tag{5·2}$$

Proceeding as in the proof of Theorem 4 we obtain, cf. (4·4), (with $\Delta = \Delta_{m;n}$)

$$(n^{n-2})^{-m} Ec(G_{n,N})^m = E\frac{(N)_{m(n-1)-\Delta}}{(M)_{m(n-1)-\Delta}} = E\frac{(N)_{m(n-1)}}{(M)_{m(n-1)}} \prod_1^{\Delta_{m;n}} \left(\frac{N - m(n-1) + i}{M - m(n-1) + i}\right)^{-1}. \tag{5·3}$$

https://doi.org/10.1017/S0305004100066111 Published online by Cambridge University Press

Put $$p_1 = \frac{N}{M} \quad \text{and} \quad p_2 = \frac{N-m(n-1)}{M-m(n-1)}.$$

Then, for $1 \leqslant i \leqslant m(n-1)$,

$$p_2 \leqslant \frac{N-m(n-1)+i}{M-m(n-1)+i} \leqslant p_1.$$

Hence $$\psi_{\Delta_{m;n}}(p_1^{-1}) = Ep_1^{-\Delta_{m;n}} \leqslant \left(\frac{(N)_{m(n-1)}}{(M)_{m(n-1)}}(n^{n-2})^m\right)^{-1} Ec(G_{n,N})^m$$
$$\leqslant Ep_2^{-\Delta_{m;n}} = \psi_{\Delta_{m;n}}(p_2^{-1}).$$

Since $p_1 \to p$ and $p_2 \to p$, Theorem 3 and Remark 3·1 show that $\psi_{\Delta_{m;n}}(p_1^{-1})$ and $\psi_{\Delta_{m;n}}(p_2^{-1})$ both converge to $e^{m(m-1)(1/p-1)}$. Furthermore, by a standard calculation based on Stirling's formula,

$$\frac{(N)_{m(n-1)}}{(M)_{m(n-1)}} \sim \left(\frac{N}{M}\right)^{m(n-1)} e^{-m^2(1/p-1)} \quad \text{as} \quad n \to \infty. \tag{5·4}$$

Combining these estimates, we obtain (5·1). The last assertion follows by the method of moments or by Chebyshev's inequality. ∎

We may now prove the announced complement to Theorem 4.

THEOREM 6. *Suppose that* $0 < p < 1$. *Then, as* $n \to \infty$,

$$c(G_{n,p})/Ec(G_{n,p}) \xrightarrow{d} e^Z \quad \textit{with} \quad Z \in N(-q/p, 2q/p).$$

Proof. Again, let $M = \binom{n}{2}$, and let X denote the number of edges in $G_{n,p}$. Note that X has the binomial distribution $\mathrm{Bi}(M,p)$ and that the conditional distribution of $G_{n,p}$ given $X = N$ equals the distribution of $G_{n,N}$. Let

$$Y_1 = c(G_{n,p})/(n^{n-2}(X/M)^{n-1}e^{-q/p}) \quad \text{and} \quad Y_2 = (X/Mp)^{n-1}e^{-q/p}.$$

Thus, cf. (4·5),

$$c(G_{n,p})/Ec(G_{n,p}) = Y_1 Y_2. \tag{5·5}$$

By Theorem 5, if $n \to \infty$ and $N = N(n)$ is any sequence of positive integers with $N/M \to p$, then $P(|Y_1 - 1| > \epsilon \,|\, X = N) \to 0$ for every $\epsilon > 0$. It follows that, for every $\epsilon > 0$,

$$\sup_{N:\,|N-Mp|<M^{\frac{2}{3}}} P(|Y_1 - 1| > \epsilon \,|\, X = N) \to 0 \quad \text{as} \quad n \to \infty.$$

Hence

$$P(|Y_1 - 1| > \epsilon) \leqslant P(|X - Mp| \geqslant M^{\frac{2}{3}}) + \sup_{|N-Mp|<M^{\frac{2}{3}}} P(|Y_1 - 1| > \epsilon \,|\, X = N) \to 0,$$

as $n \to \infty$ for every $\epsilon > 0$, i.e.

$$Y_1 \xrightarrow{p} 1 \quad \text{as} \quad n \to \infty. \tag{5·6}$$

Next, $(X - Mp)/\sqrt{M} \xrightarrow{d} N(0, pq)$ as $n \to \infty$ by the Central Limit Theorem. Hence, as $n \to \infty$,

$$(n-1)\frac{X-Mp}{Mp} \xrightarrow{d} N(0, 2q/p)$$

and

$$(n-1)\left(\log\left(1 + \frac{X-Mp}{Mp}\right) - \frac{X-Mp}{Mp}\right) \xrightarrow{p} 0.$$

https://doi.org/10.1017/S0305004100066111 Published online by Cambridge University Press

Consequently,

$$(n-1)\log\frac{X}{Mp} = (n-1)\log\left(1+\frac{X-Mp}{Mp}\right) \xrightarrow{d} N(0, 2q/p)$$

and

$$\log Y_2 = (n-1)\log\frac{X}{Mp} - q/p \xrightarrow{d} N(-q/p,\ 2q/p).$$

Thus, if $Z \in N(-q/p, 2q/p)$, $Y_2 \xrightarrow{d} e^Z$. The proof is completed by (5·6) and (5·5). ∎

An 'explanation' of the difference between $G_{n,p}$ and $G_{n,N}$ is that, as is seen in the proof above, the number of spanning subtrees in $G_{n,p}$ is so strongly dependent on the number of edges in the graph, that the random fluctuations in the number of edges cause a large variation in $c(G_{n,p})$. Note also that our method does not give the asymptotic variance of $c(G_{n,N})$ (only that it is $o((Ec(G_{n,N}))^2)$).

6. *Arbitrary subtrees*

In this section we will give results corresponding to Theorems 1–6 for arbitrary subtrees (not only spanning ones). Many details will be omitted.

The number of subtrees of K_n that contain $n-k$ vertices equals, by Cayley's formula,

$$\binom{n}{k}(n-k)^{n-k-2} \sim \frac{n^k}{k!}n^{n-k-2}\left(1-\frac{k}{n}\right)^{n-k-2} \sim n^{n-2}\frac{1}{k!}e^{-k} \tag{6·1}$$

and it follows, by dominated convergence, that the total number of subtrees of K_n equals

$$\sum_0^{n-1}\binom{n}{k}(n-k)^{n-k-2} \sim n^{n-2}\sum_0^{\infty}\frac{1}{k!}e^{-k} = n^{n-2}e^{e^{-1}}. \tag{6·2}$$

Let $\tilde{T}_n$ denote a random subtree of K_n, i.e. a random, uniform choice of one of the subtrees. Thus, by (6·1) and (6·2), if k is fixed and $n\to\infty$,

$$P(|\tilde{T}_n| = n-k) \to \frac{1}{k!}e^{-k-1/e}$$

(in particular, about $e^{-1/e} \sim 69\%$ of the subtrees are spanning). In other words,

$$n - |\tilde{T}_n| \xrightarrow{d} \mathrm{Po}\,(1/e) \quad \text{as} \quad n\to\infty \tag{6·3}$$

and it is easy to see (e.g. using Lemma 1) that the corresponding probability generating functions converge on the positive real axis.

Next we study the intersection of $\tilde{T}_n$ and an independent copy $\tilde{T}_n'$ of it. The proof of Theorem 1 is easily adapted to show that, for any integers $k_1, k_2, k_3 \geqslant 0$ with

$$\max(k_1, k_2) \leqslant k_3 \leqslant k_1 + k_2,$$

the conditional distribution of $\|\tilde{T}_n \cap \tilde{T}_n'\|$ given $|\tilde{T}_n| = n-k$, $|\tilde{T}_n'| = n-k_2$ and $|\tilde{T}_n \cap \tilde{T}_n'| = n-k_3$ converges to the Poisson distribution $\mathrm{Po}\,(2)$ as $n\to\infty$. Hence $\|\tilde{T}_n \cap \tilde{T}_n'\| \to \mathrm{Po}\,(2)$ as $n\to\infty$, and $\|\tilde{T}_n \cap \tilde{T}_n'\|$ is asymptotically independent of $n-|\tilde{T}_n|$ and $n-|\tilde{T}_n'|$. Consequently, we also obtain

$$2(n-1) - \|\tilde{T}_n \cup \tilde{T}_n'\| = n-|\tilde{T}_n| + n-|\tilde{T}_n'| + \|\tilde{T}_n \cap \tilde{T}_n'\| \xrightarrow{d} \mathrm{Po}\left(\frac{1}{e}+\frac{1}{e}+2\right).$$

https://doi.org/10.1017/S0305004100066111 Published online by Cambridge University Press

Furthermore, using the above conditioning and the argument in the proof of Theorem 2, it is not difficult to show that $\psi_{\|\tilde{T}_n \cap \tilde{T}'_n\|}(s)$ is bounded, and hence converges, as $n \to \infty$ for every $s \geqslant 0$.

Similar modifications of the proof of Theorem 3 yield the following result.

THEOREM 7. *Let $\tilde{T}_n^{(1)}, \dots, \tilde{T}_n^{(m)}$ $(m \geqslant 1)$ be independent random subtrees of K_n and set $\tilde{\Delta}_{m;n} = m(n-1) - \|\bigcup_1^m \tilde{T}_n^{(i)}\|$. Then, as $n \to \infty$*

$$\tilde{\Delta}_{m;n} \xrightarrow{d} \mathrm{Po}\,(m(m-1) + m/e)$$

and

$$\psi_{\Delta_{m;n}}(s) \to e^{(m(m-1)+m/e)(s-1)} \quad (s \geqslant 0). \quad \blacksquare$$

Let $\tilde{c}(G)$ denote the total number of subtrees of the graph G. Theorem 7 and the proofs in Sections 4 and 5, with trivial modifications, yield an analogue of Theorems 4, 5 and 6.

THEOREM 8. *Let $0 < p \leqslant 1$. Then, as $n \to \infty$,*

$$E(\tilde{c}(G_{n,p}))^m \sim (e^{1/pe}\, n^{n-2} p^{n-1})^m\, e^{m(m-1)q/p}$$

and, if $N \Big/ \binom{n}{2} \to p$,

$$E(\tilde{c}(G_{n,N}))^m \sim \left(e^{1/pe}\, n^{n-2} \left(N \Big/ \binom{n}{2}\right)^{n-1} e^{-q/p}\right)^m.$$

The normalized variables $\tilde{c}(G_{n,p})/E\tilde{c}(G_{n,p})$ and $\tilde{c}(G_{n,N})/E\tilde{c}(G_{n,N})$ converge in distribution to the same limits as $c(G_{n,p})/Ec(G_{n,p})$ (a log–normal distribution ($p < 1$)) and

$$c(G_{n,N})/Ec(G_{n,N}) \quad (\textit{the constant } 1). \quad \blacksquare$$

It follows that $c(G_{n,N})/\tilde{c}(G_{n,N})$ converge in probability to the limit of

$$Ec(G_{n,N})/E\tilde{c}(G_{n,N}),$$

viz. $e^{-1/pe}$. By the method used in the proof of Theorem 6, the same holds for $G_{n,p}$. Thus, the proportion of spanning trees among all subtrees of the random graph $G_{n,N}$ $\left(\text{with } N \approx p\binom{n}{2}\right)$ or $G_{n,p}$ is about $e^{-1/pe}$. It may be shown by a similar argument that the proportion of the subtrees that miss exactly k vertices converges in probability to $(ep)^{-k}\, 1/k!\, e^{-1/pe}$. To give this result a more fancy formulation, let $P_G(k)$, for any graph G, denote the probability that a random subtree of G has exactly $|G| - k$ vertices. (If $G \subset K_n$, $P_G(k) = P(|\tilde{T}_n| = |G| - k | \tilde{T}_n \subset G)$.)

THEOREM 9. *If $0 < p \leqslant 1$, then the random probability measure $P_{G_{n,p}}$ on the non-negative integers converges in probability (in the weak topology and in total variation) to the Poisson measure* Po $(1/pe)$ *as $n \to \infty$. The same holds for $P_{G_{n,N}}$ provided $N \Big/ \binom{n}{2} \to p$.* $\blacksquare$

This research was mainly done during the Second Seminar on Random Graphs in Poznań, August 1985, in response to a question by Joel Cohen. I thank the organizers for providing the atmosphere that made this work possible, and several participants, in particular Joel Cohen, Colin McDiarmid and John Moon, for helpful discussions.

https://doi.org/10.1017/S0305004100066111 Published online by Cambridge University Press

REFERENCES

[1] A. D. BARBOUR and G. K. EAGLESON. Poisson convergence for dissociated statistics. *J. Roy. Statist. Soc. Ser. B* **46** (1984), 397–402.

[2] A. CAYLEY. A theorem on trees. *Quart. J. Math.* **23** (1889), 376–378.

[3] W. FELLER. *An Introduction to Probability Theory and its Applications I*, 3rd ed. (Wiley, 1968).

[4] G. W. FORD and G. E. UHLENBECK. The theory of linear graphs with applications to the virial development of the properties of gases. *Studies in Statistical Mechanics I* (North-Holland, 1962), 119–211.

[5] W. G. MCGINLEY and R. SIBSON. Dissociated random variables. *Math. Proc. Cambridge Philos. Soc.* **77** (1975), 185–188.

[6] J. W. MOON. The second moment of the complexity of a graph. *Mathematika* **11** (1964), 95–98.

[7] J. W. MOON. Enumerating labelled trees. *Graph theory and theoretical physics* (ed. F. Harary) (Academic Press, 1967), 261–272.

https://doi.org/10.1017/S0305004100066111 Published online by Cambridge University Press