THE MINIMAL SPANNING TREE IN A COMPLETE GRAPH AND A FUNCTIONAL LIMIT THEOREM FOR TREES IN A RANDOM GRAPH.

Svante Janson

Abstract. The minimal weight of a spanning tree in a complete graph K_n with independent, uniformly distributed random weights on the edges, is shown to have an asymptotic normal distribution. The proof uses a functional limit extension of results by Barbour and Pittel on the distribution of the number of tree components of given sizes in a random graph.

1. Introduction and results

Assign random weights T_{ij}, $1 \le i < j \le n$, to the edges of the complete graph K_n with vertex set $\{1, \dots, n\}$, and let W_n be the minimum weight of a spanning tree of K_n. We assume that the weights are independent and identically distributed, with a uniform distribution on $[0, 1]$. It was proved by Frieze [5] that

$$W_n \to \zeta(3) = \sum_{k=1}^{\infty} k^{-3} = 1.202\dots$$

in probability as $n \to \infty$, see also Bollobs [3]. The main purpose of the present paper is to show that W_n has an asymptotic normal distribution.

Theorem 1. *Let W_n be the weight of the minimal spanning tree. Then*

$$n^{1/2}\bigl(W_n - \zeta(3)\bigr) \xrightarrow{\mathrm{d}} N(0, \sigma^2)$$

with

$$\sigma^2 = \frac{\pi^4}{45} - 2\sum_{i=0}^{\infty}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty} \frac{(i+k-1)!\,k^k\,(i+j)^{i-2}j}{i!\,k!\,(i+j+k)^{i+k+2}}.$$

Remark 1.1. We do not know whether there exists any simpler expression for σ^2. A numerical summation yields $\sigma^2 \approx 1.6857$.

As is well-known (and will be described in detail in Section 3), W_n is closely linked to properties of the number of components of a random graph, in particular the number of tree components. Our proof of Theorem 1 uses this, and is based on a functional limit theorem for component counts. We define a random graph process $G_n(t)$ (where $t \ge 0$ is interpreted as time) by taking the edge set of $G_n(t)$ to be $\{ij : T_{ij} \le t\}$, with T_{ij} as above. Hence T_{ij} now is interpreted as the (random) time the edge ij appears. Note that

Supported by the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine
Prepared on October 9, 1997

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

$G_n(t) = K_n$ for $t \geq 1$; hence all the evolution occurs on $[0,1]$, although we for convenience allow also $t > 1$. Define, for $k \geq 1$ and $0 \leq t < \infty$,

$$\mu_k(t) = \frac{k^{k-2}}{k!} t^{k-1} e^{-kt}$$

and, for $k, l \geq 1$ and $t, u \geq 0$,

$$\sigma_{kl}(t,u) = \begin{cases} (t-1)kl\mu_k(t)\mu_l(u), & k > l,\ t \leq u, \\ (t-1)kl\mu_k(t)\mu_l(u) + e^{-k(u-t)}\mu_k(t), & k = l,\ t \leq u, \\ (t-1)kl\mu_k(t)\mu_l(u) + \frac{k^{k-1}}{k!(l-k)!} t^{k-1}(u-t)(lu-kt)^{l-k-1}e^{-lu}, & k < l,\ t \leq u, \end{cases}$$

and

$$\sigma_{kl}(t,u) = \sigma_{lk}(u,t), \qquad t > u.$$

Theorem 2. *Let $X_{kn}(t)$ be the number of isolated trees of order k in $G_n(t)$. Then, for $k \geq 1$,*

$$\frac{X_{kn}(t/n) - \operatorname{E} X_{kn}(t/n)}{\sqrt{n}} \xrightarrow{\mathrm{d}} V_k(t) \tag{1.1}$$

and

$$\frac{X_{kn}(t/n) - n\mu_k(t)}{\sqrt{n}} \xrightarrow{\mathrm{d}} V_k(t) \tag{1.2}$$

as $n \to \infty$, in $D([0,\infty])$, jointly for any finite set of k, where $\big(V_k(t)\big)_{k=1}^{\infty}$ are continuous jointly Gaussian stochastic processes with $V_k(0) = V_k(\infty) = 0$, $\operatorname{E} V_k(t) = 0$, $0 \leq t \leq \infty$, and covariance functions $\operatorname{Cov}(V_k(t), V_l(u)) = \sigma_{kl}(t,u)$.

The convergence here is in the Skorohod topology on $D([0,\infty])$, which can be obtained from the perhaps better know space $D([0,1])$, see Billingsley [2], by a (monotone) change of parameter. For convergence to a continuous limit (the only case that really interests us), convergence in $D([0,\infty])$ is equivalent to uniform convergence. We refer to [2] for properties of the Skorohod topology and of convergence in distribution in it.

Theorem 2 implies, in particular, that (1.1) holds for every fixed t; this has been proven earlier, by different methods, by Barbour [1] (fixed k) and Pittel [10] (joint convergence for all k). Moreover, the cases $k = 1$ (isolated vertices) and $k = 2$ (isolated edges) were treated in [8, Sections 11 and 12].

Remark 1.2. Pittel [10, Theorem 1] also gives a similar (but somewhat different) limit theorem for the random graph G_{nm}, with $m = [cn/2]$ for a fixed $c > 0$. This too follows by our method, more precisely by [8, Theorem 9(iv)] using [8, Proposition 4.2] and estimates from Section 2 below. We will not consider G_{nm} further in this paper.

Remark 1.3. Pittel [10] used his results on tree counts to prove asymptotic normality of the order of the giant component in a random graph $G(n, c/n)$, $c > 1$. It should be possible to use our Theorem 2 to obtain a functional version of this, i.e. process convergence of the order of the giant component in $G(n, t/n)$, $t > 1$, to a Gaussian stochastic process on $(1, \infty)$, but some details remain to be verified and we leave this as a conjecture. (The same applies to the number of edges in the giant component of $G(n, t/n)$ or the number of components of $G(n, t/n)$.)

Remark 1.4. Results similar to Theorem 2 may be obtained by the same method for counts of specific types of trees, for example the numbers of isolated paths of different lengths in $G_n(t)$.

Remark 1.5. Donald Knuth (personal communication) has recently suggested studying the *minimum weight spanning pseudoforest* for random weights as above, where a spanning pseudoforest is a subgraph of K_n with n edges and only unicyclic components. (See Gabow and Tarjan [6] for the problem of finding the minimum weight spanning pseudoforest for given (non-random) weights.) Let W'_n denote the minimum weight of a spanning pseudoforest. We will show in Section 3 that Theorem 1 holds also with W_n replaced by W'_n.

Acknowledgement

This research was done during a visit to IMA, Minneapolis. I thank Boris Pittel and other participants for stimulating discussions.

2. Proof of Theorem 2.

We begin by studying the first two moments of $X_{kn}(t)$, but let us first introduce some notations. We let $C_1, C_2, \dots$ denote universal constants whose values are unimportant (the industrious reader may substitute suitable numbers), while $C(k)$ denotes a "constant" that depends on k but not on any other variable. Similarly, O and o denote estimates that are uniform in all relevant variables, while $f = O_k(g)$ means the same as $|f| \le C(k)g$. We let $(n)_k = \prod_1^k (n-i+1)$, the decreasing factorial, and observe for future use the following estimate.

Lemma 2.1. *If $1 \le k \le n$ and $n \ge 4$, then*

$$(n)_k \le C_1 (n - k/2 - 3/2)^k.$$

Proof. By the assumptions $n - k/2 - 3/2 \ge (n-3)/2 \ge n/8 > 0$. By concavity of the logarithm (or because $(n+1-i)(n-k+i) \le (n-k/2+1/2)^2$ by the conjugate rule),

$$\ln(n)_k = \sum_{i=1}^k \ln(n-i+1) \le k \ln(n - k/2 + 1/2)$$

and thus

$$\begin{aligned}\frac{(n)_k}{(n-k/2-3/2)^k} &\le \left(\frac{n-k/2+1/2}{n-k/2-3/2}\right)^k = \left(1 + \frac{2}{n-k/2-3/2}\right)^k \\ &\le \left(1+\frac{16}{n}\right)^k \le e^{16k/n} \le e^{16}.\end{aligned}$$

∎

We introduce some more notation. Let $\mathcal{T}_{kn}$ be the set of trees of order k in the complete graph K_n; by Cayley's theorem [4],

$$|\mathcal{T}_{kn}| = \binom{n}{k} k^{k-2} = \frac{k^{k-2}}{k!} n^k \big(1 + O_k(n^{-1})\big).$$

For each $\alpha \in \mathcal{T}_{kn}$, let $I_\alpha(t)$ be the indicator that is 1 if α is an isolated tree in $G_n(t)$, and 0 otherwise. Hence

$$X_{kn}(t) = \sum_{\alpha \in \mathcal{T}_{kn}} I_\alpha(t).$$

The expectation $\mathrm{E}\, I_\alpha(t) = \mathrm{P}(I_\alpha(t) = 1)$ is the same for each $\alpha \in \mathcal{T}_{kn}$. We denote it by $p_{kn}(t)$, and it is easily seen that, for $1 \le k \le n$ and $0 \le t \le 1$,

$$\mathrm{E}\, I_\alpha(t) = p_{kn}(t) = t^{k-1}(1-t)^{(n-k)k+\binom{k}{2}-k+1}, \qquad \alpha \in \mathcal{T}_{kn}.$$

Hence, for $0 \le t \le 1$,

$$\mathrm{E}\, X_{kn}(t) = |\mathcal{T}_{kn}| p_{kn}(t) = \binom{n}{k} k^{k-2} t^{k-1}(1-t)^{nk-k^2/2-3k/2+1}.$$

We state this formula together with two useful estimates derived from it as a lemma.

Lemma 2.2.

(i) *If* $0 \le t \le 1$, *then*

$$\mathrm{E}\, X_{kn}(t) = \binom{n}{k} k^{k-2} t^{k-1}(1-t)^{nk-k^2/2-3k/2+1}.$$

(ii) *For* $k \ge 1$, $n \ge 3$ *and* $0 \le t < \infty$,

$$\mathrm{E}\, X_{kn}(t) = n\mu_k(nt) + O_k(e^{-nt/2}).$$

(iii) *If* $k, n \ge 1$ *and* $0 \le t \le 1$, *then*

$$\mathrm{E}\, X_{kn}(t) \le C_2 n k^{-5/2} e^{-nt/2}.$$

For $0 \le t < 1$ *and* $n \ge 4$, *also*

$$(1-t)^{-2}\, \mathrm{E}\, X_{kn}(t) \le C_3 n k^{-5/2} e^{-nt/2}.$$

Proof. (i) is already proved.

For (ii), consider first the case $nt^2 \le 1$. Then $0 \le t \le 1$, and by (i), for $n \ge k$,

$$\begin{aligned}
\mathrm{E}\, X_{kn}(t) &= \frac{k^{k-2}}{k!}(n)_k t^{k-1} \exp\bigl(-nkt + O(k^2 t) + O(nkt^2)\bigr) \\
&= n\mu_k(nt) \exp\bigl(O(k^2 t) + O(nkt^2) + O(k^2/n)\bigr) \\
&= n\mu_k(nt)\, \bigl(1 + O_k(t + nt^2 + 1/n)\bigr) \\
&= n\mu_k(nt) + O_k\bigl(\bigl((nt)^{k-1} + (nt)^{k+1}\bigr)\, e^{-knt}\bigr) \\
&= n\mu_k(nt) + O_k(e^{-knt/2}).
\end{aligned}$$

If $nt^2 > 1$ we obtain similarly, by (i) and the trivial fact that $X_{kn}(t) = 0$ for $t \ge 1$ and $n \ge 3$,

$$\begin{aligned}
\mathrm{E}\, X_{kn}(t) &\le n^k t^{k-1} e^{-tnk+k^2} \\
&\le n^{k+1} t^{k+1} e^{-knt+k^2} \\
&= O_k(e^{-knt/2})
\end{aligned}$$

and

$$n\mu_k(nt) \le n^2 t^2 \mu_k(nt) = O_k(e^{-knt/2}).$$

Hence (ii) follows in both cases.

For (iii), we may assume that $1 \le k \le n$ and $n \ge 3$. Then, setting $x = (n - k/2 - 3/2)t \ge nt/2 - 3t/2$ and using Lemma 2.1 and Stirling's formula together with the estimate $xe^{-x} \le e^{-1}$,

$$\begin{aligned}
\mathrm{E}\, X_{kn}(t) &\le \frac{k^{k-2}}{k!}(n)_k t^{k-1}(1-t)^{k(n-k/2-3/2)} \\
&\le C_1 \frac{k^{k-2}}{k!}(n-k/2-3/2)^k t^{k-1} e^{-kt(n-k/2-3/2)} \\
&\le C_1 \frac{k^{k-2}}{k!} n x^{k-1} e^{-kx} \\
&\le C_1 n \frac{k^{k-2}}{k!} (xe^{-x})^{k-1} e^{-x} \\
&\le C_1 n \frac{k^{k-2}}{k^{k+1/2}e^{-k}} e^{-k+1} e^{-nt/2+3/2} \\
&\le C_1 e^{5/2} n k^{-5/2} e^{-nt/2}.
\end{aligned}$$

The same argument proves the final estimate. ☐

We turn to second moment calculations.

Consider $\mathrm{Cov}\,(I_\alpha(t), I_\beta(u))$, where $\alpha \in \mathcal{T}_k, \beta \in \mathcal{T}_l$ and $0 \le t \le u \le 1$ (for simplicity we here and sometimes in the sequel omit the subscript n). Recall that $\mathrm{E}\,(I_\alpha(t)I_\beta(u))$ is the probability that α is a component of $G_n(t)$ and that β is a component of $G_n(u)$ at a later time u. There are three cases:

(i) α and β are (vertex-)disjoint. If we expand $\mathrm{E}\, I_\alpha(t)I_\beta(u)$ and $\mathrm{E}\, I_\alpha(t)\, \mathrm{E}\, I_\beta(u)$ as products over the edges in K_n, the only difference is that the former contains one factor $1-u$ for each of the kl edges joining α and β, while the latter contains two factors $1-t$ and $1-u$. Hence

$$\mathrm{E}\, I_\alpha(t)I_\beta(u) = p_k(t)p_l(u)(1-t)^{-kl}. \tag{2.1}$$

(ii) $\alpha \cap \beta \neq \emptyset$, but $\alpha \not\subseteq \beta$. If α is a component of $G_n(t)$ and $u \ge t$, then any component of $G_n(u)$ that intersects α has to contain α. Hence

$$\mathrm{E}\, I_\alpha(t)I_\beta(u) = 0. \tag{2.2}$$

(iii) $\alpha \subseteq \beta$. In order for α to be a component in $G_n(t)$ that grows to a component β in $G_n(u)$, we have to have $T_e \le t$ for each edge e in α, $t < T_e \le u$ for each e in β with exactly one endpoint in α, $T_e \le u$ for each e in $\beta \setminus \alpha$, and $T_e > u$ for every other e with one or two endpoints in β. Hence, if there are j edges joining α to $\beta \setminus \alpha$,

$$\mathrm{E}\, I_\alpha(t)I_\beta(u) = t^{k-1}(u-t)^j u^{l-k-j}(1-u)^{(n-l)l+\binom{l}{2}-l+1}. \tag{2.3}$$

Let B_{kl} be the number of pairs (α, β) with $\alpha \in \mathcal{T}_k$, $\beta \in \mathcal{T}_l$ and $\alpha \cap \beta \neq \emptyset$, and let $b_{kl} = B_{kl}/|\mathcal{T}_k||\mathcal{T}_l|$. Note that, for fixed k and l,

$$b_{kl} = 1 - \frac{(n-k)_l}{(n)_l} = \frac{kl}{n} + O_{k,l}(n^{-2}). \tag{2.4}$$

Now, $\mathrm{Cov}\big(X_k(t), X_l(u)\big) = \sum_{\substack{\alpha \in \mathcal{T}_k \\ \beta \in \mathcal{T}_l}} \mathrm{E}\,(I_\alpha(t)I_\beta(u)) - \mathrm{E}\, X_k(t)\, \mathrm{E}\, X_l(u)$. There are $|\mathcal{T}_k||\mathcal{T}_l| - B_{kl} = (1-b_{kl})|\mathcal{T}_k||\mathcal{T}_l|$ terms of type (i) in the sum, and if $t \le u$ we obtain from (2.1)–(2.3),

$$\begin{aligned}
\mathrm{Cov}\big(X_k(t), X_l(u)\big) &= (1-b_{kl})|\mathcal{T}_k||\mathcal{T}_l| p_k(t)p_l(u)(1-t)^{-kl} + \sum_{\alpha \subseteq \beta} \mathrm{E}\, I_\alpha(t)I_\beta(u) \\
&\qquad - \mathrm{E}\, X_k(t)\, \mathrm{E}\, X_l(u) \\
&= \mathrm{E}\, X_k(t)\, \mathrm{E}\, X_l(u)(1-t)^{-kl}(1-b_{kl}-(1-t)^{kl}) + \sum_{\alpha \subseteq \beta} \mathrm{E}\, I_\alpha(t)I_\beta(u).
\end{aligned} \tag{2.5}$$

Again there are three cases:

(i) If $k > l$ and $t \le u \le 1$, then $\alpha \subseteq \beta$ cannot occur, and

$$\operatorname{Cov}\bigl(X_k(t), X_l(u)\bigr) = \operatorname{E} X_k(t) \operatorname{E} X_l(u)(1-t)^{-kl}\bigl(1-(1-t)^{kl} - b_{kl}\bigr). \tag{2.6}$$

(ii) If $k = l$ and $t \le u \le 1$, the only terms with $\alpha \subseteq \beta$ are those with $\alpha = \beta$. Since $\operatorname{E} I_\alpha(t) I_\alpha(u) = t^{k-1}(1-u)^{(n-k)k+\binom{k}{2}-k+1} = (t/u)^{k-1} p_k(u)$ when $\alpha \in \mathcal{T}_k$,

$$\sum_\alpha \operatorname{E}\bigl(I_\alpha(t) I_\alpha(u)\bigr) = (t/u)^{k-1} \operatorname{E} X_k(u)$$

and (2.5) yields

$$\begin{aligned}\operatorname{Cov}&\bigl(X_k(t), X_l(u)\bigr)\\ &= \operatorname{E} X_k(t) \operatorname{E} X_l(u)(1-t)^{-kl}\bigl(1-(1-t)^{kl} - b_{kl}\bigr) + (t/u)^{k-1} \operatorname{E} X_k(u),\end{aligned} \tag{2.7}$$

(iii) If $k < l$ and $t \le u \le 1$, we partition the sum over $\alpha \subset \beta$ further.

Let N_{klj} be the number of pairs (α, β) with $\alpha \in \mathcal{T}_k$, $\beta \in \mathcal{T}_l$ and $\alpha \subset \beta$, such that there are j vertices in $\beta \setminus \alpha$ adjacent to α, $1 < j \le l-k$. This number may be calculated as follows.

There are $|\mathcal{T}_k|$ choices of α. For each of these, the remaining $l-k$ vertices in β may be chosen in $\binom{n-k}{l-k}$ ways. We may then select the j vertices adjacent to α in $\binom{l-k}{j}$ ways and connect them to α in k^j ways. Finally, we complete β by adding a rooted forest on the $l-k$ new vertices, with the j selected vertices as roots; there are $j(l-k)^{l-k-j-1}$ such forests, see [4] and [11]. Hence we find

$$\begin{aligned} N_{klj} &= |\mathcal{T}_k| \binom{n-k}{l-k} \binom{l-k}{j} k^j j (l-k)^{l-k-j-1} \\ &= |\mathcal{T}_k| \binom{n-k}{l-k} \binom{l-k-1}{j-1} k^j (l-k)^{l-k-j} \end{aligned}$$

and, by (2.3), with $L = l^2/2 + 3l/2 - 1$,

$$\begin{aligned} \sum_{\alpha \subset \beta} \operatorname{E} I_\alpha(t) I_\beta(u) &= \sum_{j=1}^{l-k} N_{klj} t^{k-1} (u-t)^j u^{l-k-j} (1-u)^{nl-L} \\ &= |\mathcal{T}_k| \binom{n-k}{l-k} t^{k-1} (1-u)^{nl-L} \sum_{j=1}^{l-k} \binom{l-k-1}{j-1} k^j (u-t)^j (l-k)^{l-k-j} u^{l-k-j} \\ &= |\mathcal{T}_k| \binom{n-k}{l-k} t^{k-1} (1-u)^{nl-L} k(u-t) \bigl(k(u-t) + (l-k)u\bigr)^{l-k-1} \\ &= k |\mathcal{T}_k| \binom{n-k}{l-k} t^{k-1} (u-t)(lu-kt)^{l-k-1} (1-u)^{nl-L}, \end{aligned}$$

where we used the binomial theorem (changing the summation index to $j-1$). Hence (2.5) yields

$$\begin{aligned}\operatorname{Cov}\bigl(X_k(t), X_l(u)\bigr) &= \operatorname{E} X_k(t) \operatorname{E} X_l(u)(1-t)^{-kl}\bigl(1-(1-t)^{kl} - b_{kl}\bigr) \\ &\qquad + k|\mathcal{T}_k| \binom{n-k}{l-k} t^{k-1} (u-t)(lu-kt)^{l-k-1}(1-u)^{nl-L}.\end{aligned} \tag{2.8}$$

The formulae (2.4) and (2.6)–(2.8) yield immediately the following asymptotical result.

Lemma 2.3. *For any fixed $k, l \geq 1$ and $t, u \geq 0$,*

$$\frac{1}{n}\operatorname{Cov}\Bigl(X_{kn}\Bigl(\frac{t}{n}\Bigr), X_{ln}\Bigl(\frac{u}{n}\Bigr)\Bigr) \longrightarrow \sigma_{kl}(t,u) \qquad \text{as } n \to \infty.$$

□

Lemma 2.4. *For fixed $k \geq 1$ and $t > 0$,*

$$\operatorname{Var}\Bigl(X_{kn}\bigl(\frac{t}{n}\bigr)\Bigr) \asymp n \qquad \text{as } n \to \infty.$$

Proof. By definition,

$$\sigma_{kk}(t,t) = (t-1)k^2\mu_k(t)^2 + \mu_k(t) = \mu_k(t)\bigl(1 + (t-1)k^2\mu_k(t)\bigr).$$

If $t \geq 1$, this is obviously positive. If $0 < t < 1$ and $k \geq 2$, then

$$(1-t)k^2\mu_k(t) = (1-t)t\frac{k^k}{k!}(te^{-t})^{k-2}e^{-2t} \leq \tfrac{1}{4}\frac{e^k}{\sqrt{2\pi k}}e^{-(k-2)} \leq \frac{e^2}{4\sqrt{2\pi}} < 1.$$

Similarly, if $0 < t < 1$ and $k = 1$, then $(1-t)k^2\mu_k(t) = (1-t)e^{-t} < 1$. Hence $\sigma_{kk}(t,t) > 0$ whenever $t > 0$, and the result follows by Lemma 2.3.

□

We will also use the following uniform estimates.

Lemma 2.5. *For all $k, n \geq 1$ and $t \geq 0$,*

$$\operatorname{Var} X_{kn}(t) \leq C_4 n k^{-5/2} e^{-nt/2}.$$

If $t < 1$ and $n \geq 4$, then also

$$(1-t)^{-2}\operatorname{Var} X_{kn}(t) \leq C_5 n k^{-5/2} e^{-nt/2}.$$

Proof. We may assume $0 \leq t \leq 1$ and $k \leq n$. Then, by (2.7),

$$\begin{aligned}
\operatorname{Var}\bigl(X_{kn}(t)\bigr) &= \bigl(\operatorname{E} X_{kn}(t)\bigr)^2(1-t)^{-k^2}\bigl(1-(1-t)^{k^2}-b_{kk}\bigr) + \operatorname{E} X_{kn}(t)\\
&\leq \bigl(\operatorname{E} X_{kn}(t)\bigr)^2(1-t)^{-k^2}(1-b_{kk})\bigl(1-(1-t)^{k^2}\bigr) + \operatorname{E} X_{kn}(t)\\
&= \operatorname{E} X_{kn}(t)(1-b_{kk})|\mathcal{T}_{kn}|(1-t)^{-k^2}p_{kn}(t)\bigl(1-(1-t)^{k^2}\bigr) + \operatorname{E} X_{kn}(t)\\
&= \operatorname{E} X_{kn}(t)|\mathcal{T}_{k,n-k}|p_{k,n-k}(t)\bigl(1-(1-t)^{k^2}\bigr) + \operatorname{E} X_{kn}(t)\\
&= \operatorname{E} X_{kn}(t)\bigl(\operatorname{E} X_{k,n-k}(t)(1-(1-t)^{k^2}) + 1\bigr)
\end{aligned}$$

Since $1-(1-t)^{k^2} \leq k^2 t$ and, by Lemma 2.2 (iii),

$$k^2 t \operatorname{E} X_{k,n-k}(t) \leq C_2 t(n-k)e^{-(n-k)t/2} \leq C_2,$$

this yields

$$\operatorname{Var} X_{kn}(t) \leq C_6 \operatorname{E} X_{kn}(t),$$

and the result follows by Lemma 2.2(iii) again. □

We also have to study the drift $\xi_{kn}(t)$ of $X_{kn}(t)$, i.e. the expected rate of change of $X_{kn}(t)$ in an infinitesimal interval $(t, t+dt)$. (More precisely, $\xi_{kn}(t) = \lim_{\triangle t\to 0+} \mathrm{E}\big(X_{kn}(t+\triangle t) - X_{kn}(t) \mid \mathcal{F}_t\big)/\triangle t$, where $\mathcal{F}_t$ is the σ-field generated by all events up to time t.) Isolated trees of order k may be both created and destroyed as new edges are added (except for $k = 1$), and we thus write

$$\xi_{kn}(t) = \xi^+_{kn}(t) - \xi^-_{kn}(t),$$

separating creations and destructions.

An isolated tree in $G_n(t)$ is destroyed when any of $nk - k^2/2 - 3k/2 + 1$ edges is added, and since each edge is added with intensity $(1-t)^{-1}$, conditionally on not already being present, this yields

$$\xi^-_{kn}(t) = (nk - k^2/2 - 3k/2 + 1)(1-t)^{-1} X_{kn}(t), \qquad 0 \le t < 1. \tag{2.9}$$

Similarly, an isolated tree of order k is created when two isolated trees, of orders j and $k-j$, say, are joined by an edge. For each such pair of (disjoint) isolated trees, there are $j(k-j)$ possible edges that will join the trees, and thus (the factor $\frac{1}{2}$ is needed because we have considered unordered pairs of trees),

$$\xi^+_{kn}(t) = \tfrac{1}{2}\sum_{i=1}^{k-1}\sum_{\alpha\in\mathcal{T}_i}\sum_{\substack{\beta\in\mathcal{T}_{k-i}\\ \beta\cap\alpha=\emptyset}} i(k-i)(1-t)^{-1} I_\alpha(t) I_\beta(t), \qquad 0 \le t < 1. \tag{2.10}$$

Lemma 2.6. *For each $k \ge 1$ and all $n \ge 1$ and $t \ge 0$,*

$$\operatorname{Var}\big(\xi_{kn}(t)\big) \le C_7(k) n^3 e^{-nt/2}.$$

Proof. Clearly $\xi_{kn}(t) = 0$ for $t \ge 1$; hence we may assume $t < 1$. We then have $\operatorname{Var}\big(\xi_{kn}(t)\big) \le 2\operatorname{Var}\big(\xi^+_{kn}(t)\big) + 2\operatorname{Var}\big(\xi^-_{kn}(t)\big)$, and, by (2.9) and Lemma 2.5,

$$\operatorname{Var}\big(\xi^-_{kn}(t)\big) \le n^2 k^2 (1-t)^{-2} \operatorname{Var}\big(X_{kn}(t)\big) \le C_5 n^3 e^{-nt/2}.$$

Hence it remains only to consider $\operatorname{Var}\big(\xi^+_{kn}(t)\big)$. We obtain from (2.10)

$$\operatorname{Var}\big(\xi^+_{kn}(t)\big) = \frac{1}{4}\sum_{i,j=1}^{k-1} i(k-i)j(k-j) \sum_{\alpha,\beta,\gamma,\delta} (1-t)^{-2} \operatorname{Cov}\big(I_\alpha(t)I_\beta(t), I_\gamma(t)I_\delta(t)\big), \tag{2.11}$$

where we sum over all $\alpha \in \mathcal{T}_i$, $\beta \in \mathcal{T}_{k-i}$, $\gamma \in \mathcal{T}_j$, $\delta \in \mathcal{T}_{k-j}$ such that $\alpha\cap\beta = \gamma\cap\delta = \emptyset$. Let us first consider the terms where $\alpha,\beta,\gamma,\delta$ are pairwise disjoint. Then, similarly to (2.1),

$$\mathrm{E}\, I_\alpha(t)I_\beta(t)\, \mathrm{E}\, I_\gamma(t)I_\delta(t) = (1-t)^{k^2}\, \mathrm{E}\, I_\alpha(t)I_\beta(t)I_\gamma(t)I_\delta(t),$$

and thus

$$\begin{aligned}\operatorname{Cov}\big(I_\alpha(t)I_\beta(t), I_\gamma(t)I_\delta(t)\big) &= \big(1-(1-t)^{k^2}\big)\, \mathrm{E}\, I_\alpha(t)I_\beta(t)I_\gamma(t)I_\delta(t)\\ &= \big(1-(1-t)^{k^2}\big) p_{i,n}(t) p_{k-i,n-i}(t) p_{j,n-k}(t) p_{k-j,n-k-j}(t).\end{aligned}$$

The number of such quadruples $(\alpha,\beta,\gamma,\delta)$ with given i and j equals $|\mathcal{T}_{i,n}||\mathcal{T}_{k-i,n-i}||\mathcal{T}_{j,n-k}|$ $|\mathcal{T}_{k-j,n-k-j}|$ and their total contribution to the inner sum in (2.11) is thus, using Lemma 2.2(iii),

$$\begin{aligned}(1-t)^{-2}\bigl(1-(1-t)^{k^2}\bigr)\operatorname{E}X_{i,n}(t)\operatorname{E}X_{k-i,n-i}(t)\operatorname{E}X_{j,n-k}(t)\operatorname{E}X_{k-j,n-k-j}(t)\\ \le C_3C_2^3k^2tn(n-i)(n-k)(n-k-j)e^{-nt/2-(n-i)t/2}\le C_8k^2n^3e^{-nt/2}.\end{aligned}$$

Summing over i and j we obtain an estimate of the required order. There still remains the terms where two (or more) of $\alpha,\beta,\gamma,\delta$ intersect. In this case, if the intersecting trees do not coincide, $I_\alpha(t)I_\beta(t)I_\gamma(t)I_\delta(t)=0$ and the covariance in (2.11) is negative. Hence, for an upper bound we only have to consider the cases $\beta=\delta$ with α,β,γ disjoint, and $\alpha=\gamma$, $\beta=\delta$ (up to relabelling). For the first case,

$$\operatorname{Cov}\bigl(I_\alpha(t)I_\beta(t),I_\gamma(t)I_\beta(t)\bigr)\le\operatorname{E}I_\alpha(t)I_\beta(t)I_\gamma(t)=p_{i,n}(t)p_{k-i,n-i}(t)p_{i,n-k}(t)$$

and the total contribution to the inner sum of (2.11) is, for fixed i, using Lemma 2.2(iii) again, bounded by

$$(1-t)^{-2}\operatorname{E}X_{i,n}(t)\operatorname{E}X_{k-i,n-i}(t)\operatorname{E}X_{i,n-k}(t)\le C_3C_2^2n^3e^{-nt/2}.$$

For the second case we similarly get, for each i, a contribution $\le C_9n^2e^{-nt/2}$. The lemma follows. □

We next use the orthogonal decomposition in [8], which can be briefly described as follows. Let, for $0\le t\le 1$, $I'_e(t)=I_e(t)-t$; note that $\operatorname{E}I'_e(t)=0$, which is the central point in the theory. Each $I_\alpha(t)$ equals a product

$$\prod_{e\in A(\alpha)}I_e(t)\prod_{e\in B(\alpha)}\bigl(1-I_e(t)\bigr)=\prod_{e\in A(\alpha)}\bigl(I'_e(t)+t\bigr)\prod_{e\in B(\alpha)}\bigl(1-t-I'_e(t)\bigr),\tag{2.12}$$

for certain edge sets $A(\alpha)$ and $B(\alpha)$; $A(\alpha)$ is the set of edges in α, and $B(\alpha)$ is the set of all other edges in K_n that have at least one endpoint in α. Expanding the product in (2.12), we see that $I_\alpha(t)$ equals a linear combination of products $\prod_{e\in C}I'_e(t)$, where C ranges over the subsets of $A(\alpha)\cup B(\alpha)$; summing over $\alpha\in\mathcal{T}_{kn}$ and collecting terms according to the isomorphism type of the graph spanned by the edges in the set C, we finally arrive at the decomposition

$$X_{kn}(t)=\sum_H\hat\psi_{kn}(H;t)S_n(H;t),$$

where H ranges over unlabelled graphs (without isolated vertices), $\hat\psi_{kn}(H;t)$ are some deterministic functions and

$$S_n(H;t)=\sum_F\prod_{e\in F}I'_e(t),$$

where F ranges over all subgraphs of K_n that are isomorphic to H (counted with multiplicity $\operatorname{aut}(H)$, for convenience). See [8] for further details.

We do not have to calculate the coefficients $\hat\psi_{kn}(H;t)$ exactly, but it will be important to know that the terms when H is not a tree are sufficiently small.

Lemma 2.7. *Suppose that H is not a tree. If H has v (non-isolated) vertices and e edges, and $t=\tau/n$ with $\tau>0$ fixed, then (for $n\geq\tau$)*

$$n^{v/2}t^{e/2}|\hat{\psi}_{kn}(H;t)|\leq C_{10}(k,H,\tau).$$

Proof. Let us consider one of the terms in the construction above. Fix α and C, and suppose that C consist of $k-1-j$ of the $k-1$ edges in $A(\alpha)$ and m of the edges in $B(\alpha)$, and that the graph F spanned by the edges in C is isomorphic to H. Further, suppose that H has c components, and that $w\geq 0$ of the vertices in α do not appear among the vertices of F (because they are not endpoints of any edge in C). We study the construction of F in three steps. First we start with α and remove j edges; this leaves a forest with k vertices, $k-1-j$ edges and $j+1$ components (including isolated vertices). Secondly, we add m edges from B, which may only reduce the number of components. Finally, the w remaining isolated vertices are removed. This yields the relations

$$\begin{aligned}e&=k-1-j+m\\ c&\leq j+1-w.\end{aligned}$$

Furthermore, the resulting coefficient of $\prod_{e\in F}I'_e(t)$, $b(t)$ say, equals $t^j(1-t)^{|B(\alpha)|-m}(-1)^m$. Hence

$$n^{e/2+c+w}t^{e/2}|b(t)|\leq n^{e/2+j+1}t^{e/2+j}=n(nt)^{e/2+j}$$

or

$$n^w|b(t)|\leq n^{1-c-e/2}t^{-e/2}(nt)^{e/2+j}. \tag{2.13}$$

When summing over α and collecting the terms that contribute to $\hat{\psi}_{kn}(H;t)$, we note that several different choices of α and $C\subset A(\alpha)\cup B(\alpha)$ may give the same F. In particular, there are $\binom{n-v}{w}\leq n^w$ choices of the w vertices in α that do not appear in F, but apart from that, the number of possibilities is finite and independent of n (for n large enough). Hence (2.13) implies

$$|\hat{\psi}_{kn}(H;t)|\leq C_{11}(k,H)n^{1-c-e/2}t^{-e/2}\sum_{j=0}^{k-1}(nt)^{e/2+j}$$

and, with $t=\tau/n$,

$$n^{v/2}t^{e/2}|\hat{\psi}_{kn}(H;t)|\leq C_{11}(k,H)n^{1-(2c+e-v)/2}\sum_{j=0}^{k-1}\tau^{e/2+j}.$$

Now, $v\leq c+e$, with strict inequality except when H is a forest. Hence $2c+e-v=c+(c+e-v)\geq 2$ unless H is a connected forest, i.e. a tree, which completes the proof. ☐

We are now prepared to apply the results of [8]. We first consider a fixed $k\geq 1$.

First, by Lemmas 2.4, 2.5 and 2.6 and [8, Proposition 4.10], with $p_n=\tau/n$ and $\beta_n=n^{1/2}$ ($\tau>0$ fixed), $X_{kn}(\tau/n)$ is "almost finitely dominated" by e.g. the family of all graphs. (This means that the variable can be expressed as a function of small subgraph counts, up to a small remainder term.) Moreover, by Lemma 2.7 and [8, Proposition 4.6], $X_{kn}(\tau/n)$ is almost finitely dominated by the family of all trees. We now apply [8, Theorem 14], with $\mathcal{H}$ the family of all trees, $p_n = 1/n$, $p = 0$, $\beta_n = n^{1/2}$, and $\varphi(t) = C_7(k)^{1/2}e^{-t/4}$. Apart from the facts established above and in Lemma 2.6, this theorem also requires that

$$n^{v/2-e/2}\hat{\psi}_{kn}(H;\tau/n)/\sqrt{n} = \hat{\psi}_{kn}(H;\tau/n) \to a_k(H;\tau)$$

for every tree H and $\tau \geq 0$ and some functions $a_k(H;\tau)$. This can be shown by the argument in the proof of Lemma 2.7, but it can also be avoided completely by the subsequence argument in the proofs of [8, Theorems 3 and 13] (it is unfortunate that this version of the theorem is not stated explicitly in [8]). Consequently, we obtain from [8, Theorem 14], using Lemma 2.3,

$$\frac{X_{kn}(t/n) - \mathrm{E}\, X_{kn}(t/n)}{\sqrt{n}} \xrightarrow{\mathrm{d}} V_k(t) \qquad \text{as } n \to \infty,$$

in $D([0,\infty])$, where $V_k(t)$ is a continuous Gaussian process with mean $\mathrm{E}\, V_k(t) = 0$ and covariance function

$$\mathrm{E}\big(V_k(t)V_k(u)\big) = \lim_{n\to\infty} \mathrm{Cov}\big(X_{kn}(t/n), X_{kn}(u/n)\big) = \sigma_{kk}(t,u).$$

This proves the first assertion in Theorem 2 for a fixed k, and the second follows because

$$n^{-1/2}\big(\mathrm{E}\, X_{kn}(t/n) - n\mu_k(t)\big) \to 0, \qquad \text{uniformly in } t \geq 0,$$

by Lemma 2.2(ii). Joint convergence for several different k follows by the argument above and the Cramr-Wold device, cf. [8, Remark 8.2]. $\square$

3. Proof of Theorem 1.

The minimal spanning tree may be found by the greedy algorithm, see Kruskal [9]; consider the edges in order of increasing weights and select all edges that do not connect two vertices that already are connected by some sequence of selected edges. In other words, we select the edges in $G_n(t)$ that decrease the number of components when they appear. If $N(t)$ is the number of selected edges up to time t, and $K(t)$ is the number of components of $G_n(t)$, then $K(t) = n - N(t)$. Hence, letting M be the set of edges in the minimal spanning tree,

$$\begin{aligned} W_n &= \sum_{e\in M} T_e = \sum_{e\in M} \int_0^1 I(t < T_e)\, dt = \int_0^1 \sum_{e\in M} \big(1 - I(t \geq T_e)\big)dt \\ &= \int_0^1 \big(n - 1 - N(t)\big)dt = \int_0^1 \big(K(t) - 1\big)\, dt. \end{aligned} \tag{3.1}$$

In order to analyze this further, we study different types of components of $G_n(t)$ separately. We already have defined $X_{kn}(t)$ as the number of components of $G_n(t)$ that are trees of order k. We further let $Y_{kn}(t)$ be the number of unicyclic components of order k, and $Z_n(t)$ the number of components with more than one cycle. Clearly,

$$K(t) = \sum_{k=1}^{n} X_{kn}(t) + \sum_{k=3}^{n} Y_{kn}(t) + Z_n(t)$$

and the theorem follows from (3.1) and the following five lemmas.

Lemma 3.1.

$$n^{1/2}\Bigl(\sum_{k=1}^{n}\int_0^1 X_{kn}(t)\,dt-\sum_{k=1}^{n}\int_0^1 \operatorname{E}X_{kn}(t)\,dt\Bigr)\ \xrightarrow{\ \mathrm{d}\ }\ N(0,\sigma^2)$$

with

$$\sigma^2=\sum_{k,l=1}^{\infty}\int_0^\infty\int_0^\infty \sigma_{kl}(t,u)\,dt\,du.$$

Lemma 3.2.

$$\sum_{k=1}^{n}\int_0^1 \operatorname{E}X_{kn}(t)\,dt=\zeta(3)+O(1/n).$$

Lemma 3.3.

$$\operatorname{E}\sum_{k=3}^{n}\int_0^1 Y_{kn}(t)\,dt=O(1/n)$$

and thus

$$n^{1/2}\int_0^\infty\sum_{k=3}^{n}Y_{kn}(t)\,dt\ \xrightarrow{\ \mathrm{p}\ }\ 0.$$

Lemma 3.4.

$$\operatorname{E}\int_0^1|Z_n(t)-1|\,dt=O(\log n/n)$$

and thus

$$n^{1/2}\int_0^1\bigl(Z_n(t)-1\bigr)\,dt\ \xrightarrow{\ \mathrm{p}\ }\ 0.$$

Lemma 3.5.

$$\sigma^2=\sum_{k,l=1}^{\infty}\int_0^\infty\int_0^\infty \sigma_{kl}(t,u)\,dt\,du=\frac{\pi^2}{45}-2\sum_{i=0}^{\infty}\sum_{j=1}^{\infty}\sum_{k=1}^{\infty}\frac{(i+k-1)!\,k^k\,(i+j)^{i-2}j}{i!\,k!\,(i+j+k)^{i+k+2}}.$$

Proof of Lemma 3.1. For convenience, we introduce the normalized variables

$$X^*_{kn}(t)=\frac{X_{kn}(t)-\operatorname{E}X_{kn}(t)}{\sqrt{n}};$$

note that $X^*_{kn}(t)=0$ when $k>n$ or $t>1$. By Theorem 2,

$$X^*_{kn}(t/n)\ \xrightarrow{\ \mathrm{d}\ }\ V_k(t)$$

in $D([0,\infty])$ as $n\to\infty$, for every k; since the integral over a finite interval is a continuous functional on $D([0,\infty])$, this implies

$$\int_0^A X^*_{kn}(t/n)\,dt\ \xrightarrow{\ \mathrm{d}\ }\ \int_0^A V_k(t)\,dt$$

for any $A < \infty$. Moreover, by the joint convergence for any finite set of k, also

$$\sum_1^N \int_0^A X^*_{kn}(t/n)\,dt \xrightarrow{\mathrm{d}} \sum_1^N \int_0^A V_k(t)\,dt \tag{3.2}$$

for every $A, N < \infty$. We want to extend this to $A = N = \infty$, and therefore observe that by Lemma 2.5,

$$\|X^*_{kn}(t/n)\|_2 = \operatorname{Var}\bigl(X^*_{kn}(t/n)\bigr)^{1/2} \le C_{12} k^{-5/4} e^{-t/4}. \tag{3.3}$$

Since $\|X^*_{kn}(t/n)\|_2 \to \sigma_{kk}(t,t)^{1/2} = \|V_k(t)\|_2$ as $n \to \infty$ by Lemma 2.3, also

$$\|V_k(t)\|_2 \le C_{12} k^{-5/4} e^{-t/4}. \tag{3.4}$$

Since $\sum_{k=1}^\infty \int_0^\infty k^{-5/4} e^{-t/4}\,dt < \infty$, (3.4) implies that the variable

$$U = \sum_{k=1}^\infty \int_0^\infty V_k(t)\,dt$$

is well-defined in L^2. Moreover, if

$$\begin{aligned} U_N &= \sum_1^N \int_0^N V_k(t)\,dt, \\ U^*_{Nn} &= \sum_{k=1}^N \int_0^N X^*_{kn}(t/n)\,dt, \\ U^*_n &= \sum_{k=1}^\infty \int_0^\infty X^*_{kn}(t/n)\,dt = \sum_{k=1}^n \int_0^n X^*_{kn}(t/n)\,dt, \end{aligned}$$

then $U_N \to U$ in L^2, and thus in distribution, as $N \to \infty$, while $U^*_{Nn} \to U^*_n$ in L^2 (and thus in probability) uniformly in n as $N \to \infty$, because by (3.3),

$$\begin{aligned} \|U^*_{Nn} - U^*_n\|_2 &\le \Bigl(\sum_1^N \int_N^\infty + \sum_{N+1}^\infty \int_0^\infty\Bigr) \|X^*_{kn}(t/n)\|_2\,dt \\ &\le \Bigl(\sum_1^N \int_N^\infty + \sum_{N+1}^\infty \int_0^\infty\Bigr) C_{12} k^{-5/4} e^{-t/4}\,dt \to 0. \end{aligned}$$

Finally, $U^*_{Nn} \xrightarrow{\mathrm{d}} U_N$ as $n \to \infty$ for each N by (3.2). This implies, by [2, Theorem 4.2], that

$$U^*_n \xrightarrow{\mathrm{d}} U \qquad \text{as } n \to \infty. \tag{3.5}$$

Since U is the limit of (Riemann) sums of the jointly normal random variables $V_k(t)$, U has a normal distribution; the parameters are given by

$$\mathrm{E}\,U = \sum_1^\infty \int_0^\infty \mathrm{E}\,V_k(t)\,dt = 0$$

and

$$\begin{aligned}\operatorname{Var} U = \mathrm{E}\, U^2 &= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\int_0^{\infty}\int_0^{\infty} \mathrm{E}\big(V_k(t)V_l(u)\big)\,dt\,du\\ &= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty}\int_0^{\infty}\int_0^{\infty} \sigma_{kl}(t,u)\,dt\,du.\end{aligned}$$

Thus $U \sim N(0,\sigma^2)$. The result now follows from (3.5), because by our definitions,

$$\begin{aligned}U_n^* &= n\sum_{k=1}^{n}\int_0^1 X_{kn}^*(t)\,dt\\ &= \sqrt{n}\sum_{k=1}^{n}\int_0^1 \big(X_{kn}(t) - \mathrm{E}\,X_{kn}(t)\big)\,dt.\end{aligned}$$

□

Proof of Lemma 3.2. Let $\mu_{kn} = \int_0^1 \mathrm{E}\,X_{kn}(t)\,dt$. By Lemma 2.2(i) and evaluation of a beta integral

$$\begin{aligned}\mu_{kn} &= \binom{n}{k} k^{k-2}\frac{\Gamma(k)\Gamma(nk-k^2/2-3k/2+2)}{\Gamma(nk-k^2/2-k/2+2)}\\ &= \frac{(n)_k}{k!}k^{k-2}\frac{(k-1)!}{(nk-k^2/2-k/2+1)_k}\\ &= k^{k-3}\frac{(n)_k}{(nk-k^2/2-k/2+1)_k}.\end{aligned}$$

A Taylor expansion of the logarithm yields, for $k \le n/2$ say,

$$\log(n)_k = k\log n + \sum_1^{k-1}\log(1-\frac{i}{n}) = k\log n - \frac{k(k-1)}{2n} + O\Big(\frac{k^3}{n^2}\Big),$$

and similarly

$$\begin{aligned}\log(nk-k^2/2-k/2+1)_k &= k\log(nk-k^2/2-k/2+1) + O\Big(\frac{k^2}{nk}\Big)\\ &= k\log(nk) + k\log(1-\frac{k}{2n}-\frac{1}{2n}+\frac{1}{nk}) + O\Big(\frac{k^2}{nk}\Big)\\ &= k\log(nk) - \frac{k^2}{2n} + O\Big(\frac{k}{n}+\frac{k^3}{n^2}\Big).\end{aligned}$$

Hence, if $1 \le k \le \sqrt{n}$,

$$\mu_{kn} = k^{k-3}\frac{n^k}{(nk)^k}\exp\Big(O(\frac{k}{n})\Big) = k^{-3}\Big(1+O(\frac{k}{n})\Big) = k^{-3} + O(n^{-1}k^{-2}).$$

For $\sqrt{n} < k \le n$ we instead use Lemma 2.1 to conclude (assuming as we may $n \ge 4$),

$$\mu_{kn} \le C_1 k^{k-3}\frac{(n-k/2-3/2)^k}{(nk-k^2/2-3k/2)^k} = C_1 k^{-3}.$$

For $k > n$, we trivially have $\mu_{kn} = 0$.

Consequently,

$$\begin{aligned}\Big|\sum_{1}^{n} \mu_{kn} - \sum_{1}^{\infty} k^{-3}\Big| &\le \sum_{1}^{\sqrt{n}} |\mu_{kn} - k^{-3}| + \sum_{\sqrt{n}}^{\infty} (1 + C_1) k^{-3} \\ &\le \frac{C_{13}}{n} \sum_{1}^{\infty} \frac{1}{k^2} + C_{14} \sum_{\sqrt{n}}^{\infty} k^{-3} \\ &\le \frac{C_{15}}{n}.\end{aligned}$$

□

Proof of Lemma 3.3. Let $m(k)$ be the number of connected graphs with k edges on k (labelled) vertices. Then

$$\operatorname{E} Y_{kn}(t) = \binom{n}{k} m(k) t^k (1-t)^{nk - k^2 + \binom{k}{2} - k}$$

and thus, evaluating a beta integral as in the proof of Lemma 3.2,

$$\begin{aligned}\int_0^1 \operatorname{E} Y_{kn}(t)\, dt &= \binom{n}{k} m(k) \frac{\Gamma(k+1)}{(nk - k^2/2 - k/2 + 1)_{k+1}} \\ &= m(k) \frac{(n)_k}{(nk - k^2/2 - k/2 + 1)_{k+1}}.\end{aligned}$$

Since $m(k) = \frac{1}{2}(k-1)! \sum_{j=0}^{k-3} k^j/j! < (k-1)! e^k \le C_{16} k^{k-1/2}$, see e.g. [3, Theorem 18], we obtain using Lemma 2.1 for $3 \le k \le n$ and $n \ge 4$,

$$\begin{aligned}\int_0^1 \operatorname{E} Y_{kn}(t)\, dt &\le C_{17} k^{k-1/2} \frac{(n - k/2 - 3/2)^k}{(nk - k^2/2 - 3k/2)^{k+1}} \\ &= C_{17} \frac{k^{-3/2}}{n - k/2 - 3/2} \\ &\le C_{18} k^{-3/2} n^{-1}.\end{aligned}$$

The result follows because $\sum_1^\infty k^{-3/2} < \infty$. □

Proof of Lemma 3.4. We consider the integrals over $[0, 3\log n/n]$ and $[3\log n/n, 1]$ separately. Let $\widetilde{Z}_n$ be the number of times during the evolution of $G_n(t)$ that a component with at least two cycles is formed (either by the creation of an edge joining two unicyclic components or by the creation of an edge joining two vertices in the same unicyclic component). Clearly $Z_n(t) \le \widetilde{Z}_n$ for every t, and thus

$$\int_0^{3\log n/n} \operatorname{E}|Z_n(t) - 1|\, dt \le \int_0^{3\log n/n} \operatorname{E}(\widetilde{Z}_n + 1)\, dt = \frac{3\log n}{n}(\operatorname{E}\widetilde{Z}_n + 1).$$

It was proved in [7] that $\operatorname{E}\widetilde{Z}_n = O(1)$ (in fact, $E\widetilde{Z}_n \to 5\pi/8\sqrt{3}$ as $n \to \infty$); hence

$$\int_0^{3\log n/n} \operatorname{E}|Z_n(t) - 1|\, dt = O\Big(\frac{\log n}{n}\Big).$$

On the other hand, let $\mathcal{E}$ be the event that $G_n(3\log n/n)$ is connected and has more than n edges. If $\mathcal{E}$ happens, then $Z_n(t) = 1$ for all $t \geq 3\log n/n$; thus we conclude from the trivial estimate $|Z_n(t)-1| \leq n$ that

$$\int_{3\log n/n}^{1} \mathrm{E}\,|Z_n(t)-1|\,dt \leq \int_{3\log n/n}^{1} n\,\mathrm{P}(\mathcal{E}^c)\,dt \leq n\,\mathrm{P}(\mathcal{E}^c) = O(n^{-1}),$$

since $\mathrm{P}(\mathcal{E}^c) = O(n^{-2})$ as is easily shown, cf. [3, Exercise II.15]. The lemma follows. □

Proof of Lemma 3.5. We first observe that

$$|\sigma_{kl}(t,u)| \leq \|V_k(t)\|_2 \|V_l(u)\|_2 \leq C_{12}^2 k^{-5/4} e^{-t/4} l^{-5/4} e^{-u/4}$$

by Cauchy-Schwarz' inequality and (3.4), and thus $\sum\sum\int\int |\sigma_{kl}(t,u)| < \infty$. This and the gamma integrals $\int \mu_k(t)\,dt = \int t\mu_k(t)\,dt = k^{-3}$ imply easily that all integrals and sums below are finite.

By the definitions,

$$\begin{aligned}
\sigma^2 &= 2\sum_{k,l=1}^{\infty} \iint_{t\leq u} \sigma_{kl}(t,u)\,dt\,du = 2\sum_{k,l=1}^{\infty} \int_0^{\infty}\int_0^{\infty} \sigma_{kl}(t,t+s)\,ds\,dt \\
&= 2\Big(\sum_{k,l=1}^{\infty} I_1(k,l) + \sum_{k=1}^{\infty} I_2(k) + \sum_{k=1}^{\infty}\sum_{l=k+1}^{\infty} I_3(k,l)\Big),
\end{aligned}$$

where

$$\begin{aligned}
I_1(k,l) &= \int_0^{\infty}\int_0^{\infty} kl(t-1)\mu_k(t)\mu_l(t+s)\,ds\,dt, \\
I_2(k) &= \int_0^{\infty}\int_0^{\infty} e^{-ks}\mu_k(t)\,ds\,dt, \\
I_3(k,l) &= \int_0^{\infty}\int_0^{\infty} \frac{k^{k-1}}{k!\,(l-k)!} t^{k-1} s\big(ls+(l-k)t\big)^{l-k-1} e^{-ls-lt}\,ds\,dt.
\end{aligned}$$

Binomial expansions and standard integrations yield

$$\begin{aligned}
I_1(k,l) &= \frac{k^{k-1}}{k!}\frac{l^{l-1}}{l!}\int_0^{\infty}\int_0^{\infty} (t-1)t^{k-1}(t+s)^{l-1}e^{-(k+l)t-ls}\,dt\,ds \\
&= \frac{k^{k-1}}{k!}\frac{l^{l-1}}{l!}\sum_{i=0}^{l-1}\int_0^{\infty}\int_0^{\infty}\binom{l-1}{i}(t^{k+i}-t^{k+i-1})s^{l-1-i}e^{-(k+l)t-ls}\,dt\,ds \\
&= \sum_{i=0}^{l-1}\frac{k^{k-1}}{k!}\frac{l^{l-1}}{l!}\frac{(l-1)!}{i!\,(l-1-i)!}\frac{(k+i-1)!\,\big(k+i-(k+l)\big)}{(k+l)^{k+i+1}}\frac{(l-i-1)!}{l^{l-i}} \\
&= -\sum_{i=0}^{l-1}\frac{(k+i-1)!\,k^{k-1}l^{i-2}(l-i)}{k!\,i!\,(k+l)^{k+i+1}}, \\
I_2(k) &= \int_0^{\infty} e^{-ks}\,ds\int_0^{\infty}\frac{k^{k-2}}{k!}t^{k-1}e^{-kt}\,dt = k^{-1}\frac{k^{k-2}}{k!}\frac{(k-1)!}{k^k} = k^{-4},
\end{aligned}$$

and, for $l > k$,

$$I_3(k,l) = \frac{k^{k-1}}{k!\,(l-k)!} \sum_{i=0}^{l-k-1} \int_0^\infty \int_0^\infty \binom{l-k-1}{i} (l-k)^i t^{k-1+i} l^{l-k-1-i}$$
$$s^{l-k-i} e^{-ls-lt}\,ds\,dt$$
$$= \sum_{i=0}^{l-k-1} \frac{k^{k-1}}{k!\,(l-k)!} \frac{(l-k-1)!}{i!\,(l-k-1-i)!} (l-k)^i l^{l-k-1-i} \frac{(k+i-1)!}{l^{k+i}} \frac{(l-k-i)!}{l^{l-k-i+1}}$$
$$= \sum_{i=0}^{l-k-1} \frac{(k+i-1)!\,k^{k-1}(l-k)^{i-1}}{k!\,i!\,l^{k+i+2}} (l-k-i).$$

Summing and substituting $l = i + j$ in I_1 and $l = k + i + j$ in I_3, we obtain

$$\sigma^2 = 2\sum_{k=1}^\infty k^{-4} + 2\sum_{k=1}^\infty \sum_{i=0}^\infty \sum_{j=1}^\infty \frac{(k+i-1)!\,k^{k-1}(i+j)^{i-2}j}{k!\,i!\,(k+i+j)^{k+i+2}} \big(-(k+i+j)+i+j\big)$$

and the result follows, using the well-known formula $\sum_1^\infty k^{-4} = \pi^4/90$. $\square$

We have completed the proof of Theorem 1, and it remains only to indicate the small modifications needed to show the corresponding result for spanning pseudoforests stated in Remark 1.5.

It is not difficult to see that the minimum weight pseudoforest consists of the edges in $G_n(t)$ that have at least one endpoint in a tree component when they appear (cf. Gabow and Tarjan [6, Lemma 2.1]). Each such edge decreases the number of tree components by one; thus, if $N'(t)$ denotes the number of selected edges that has appeared up to time t, then $N'(t) = n - X(t)$, where $X(t) = \sum_{k=1}^n X_{kn}(t)$ is the number of tree components of $G_n(t)$. An argument as above then shows

$$W_n' = \int_0^1 \big(n - N'(t)\big)\,dt = \int_0^1 X(t)\,dt = \sum_{k=1}^n \int_0^1 X_{kn}(t)\,dt$$

and the result follows by Lemmas 3.1, 3.2 and 3.5. (Thus Lemmas 3.4 and 3.5 are not needed for spanning pseudoforests.)

References

1. A.D. Barbour, Poisson convergence and random graphs, *Math. Proc. Camb. Phil. Soc.* **92** (1982), 349–359.
2. P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.
3. B. Bollobás, *Random Graphs*, Academic Press, London, 1985.
4. A. Cayley, A theorem on trees, *Q. J. Pure Appl. Math.* **23** (1889), 376–378.
5. A.M. Frieze, On the value of a minimal spanning tree problem, *Discrete Appl. Math.* **10** (1985), 47–56.
6. H.N. Gabow and R.E. Tarjan, A linear-time algorithm for finding a minimum spanning pseudoforest, *Information Processing Letters* **27** (1988), 259–263.
7. S. Janson, Multicyclic components in a random graph process, *Random Struct. Alg.* **4** (1993), 71–84.
8. S. Janson, Orthogonal decompositions and functional limit theorems for random graph statistics, *Memoirs Amer. Math. Soc.* **534** (1994).
9. J.B. Kruskal, The expected number of components under random mapping functions, *Amer. Math. Monthly* **61** (1954), 392–397.
10. B. Pittel, On tree census and the giant component in sparse random graphs, *Random Struct. Alg.* **1** (1990), 311–342.
11. A. Rényi, Some remarks on the theory of trees, *Publ. Math. Inst. Hungar. Acad. Sci.* **4** (1959), 73–85.

Department of Mathematics, Uppsala University, PO Box 480, S-751 06 Uppsala, Sweden
E-mail address: svante.janson@math.uu.se