Acta Math. Hungar., **145** (**1**) (2015), 205–219
DOI: 10.1007/s10474-014-0463-8
First published online November 16, 2014

LIMITS OF RANDOM TREES. II*

A. DEÁK

MTA-ELTE "Numerical Analysis and Large Networks" Research Group
e-mail: deak51@cs.elte.hu

(Received May 5, 2014; revised July 29, 2014; accepted July 31, 2014)

Abstract. Local convergence of bounded degree graphs was introduced by Benjamini and Schramm. This result was extended further by Lyons to bounded average degree graphs. In this paper we study the convergence of random tree sequences with given degree distributions. Denote by $\mathcal{D}_n$ the set of possible degree sequences of a labeled tree on n nodes. Let $\mathbf{D}_n$ be a random variable on $\mathcal{D}_n$ and $\mathbf{T}(\mathbf{D}_n)$ be a uniform random labeled tree with degree sequence $\mathbf{D}_n$. We show that the sequence $\mathbf{T}(\mathbf{D}_n)$ converges in probability if and only if $\mathbf{D}_n \to \mathbf{D} = (\mathbf{D}(i))_{i=1}^{\infty}$, where $\mathbf{D}(i) \sim \mathbf{D}(j)$, $\mathbb{E}(\mathbf{D}(1)) = 2$ and $\mathbf{D}(1)$ is a random variable on $\mathbb{N}^+$.

1. Introduction

In recent years the study of the structure and behavior of real world networks has received wide attention. The degree sequence of these networks appear to have special properties (like power law degree distribution). Classical random graph models (like the Erdős–Rényi model) have very different degree sequences. An obvious solution is to study a random graph with given degree sequence. More generally generate a random graph with a degree sequence from a family of degree sequences. In [7] Chatterjee, Diaconis and Sly studied random dense graphs (graphs whose number of edges is comparable to the square of the number of vertices) with a given degree sequence.

It is not always easy to generate a truly random graph with a given degree sequence. There is a fairly large literature on the configuration model (for the exact definition of the model see [5]), where for a given degree sequence for each node i we consider d_i stubs and take a random pairing of the stubs and connect the corresponding nodes with an edge. This model creates the required degree distribution, but gives a graph with possible loops and parallel edges.

* Research was partially supported by the ERC Grant Nr.: 227701.
Key words and phrases: sparse graph limit, random tree.
Mathematics Subject Classification: 05C80.

A notion of convergence for (dense) graph sequences was developed by Borgs, Chayes, Lovász, Sós and Vesztergombi in [6]. The limit objects were described by Lovász and Szegedy in [12]. Using this limit theory, the authors in [7] described the structure of random (dense) graphs from the configuration model. They defined the convergence of degree sequences and for convergent degree sequences they gave a sufficient condition on the degree sequence, which implies the convergence of the random graph sequence (taken from the configuration model).

What can we say if the graphs we want to study are sparse (the number of edges is comparable to the number of vertices) and not dense? Is there a similar characterization for sparse graphs with given degree sequence? We establish a characterization for random trees with given (possibly random) degree sequence. There are various limit theories and convergence notions for trees introduced by Aldous [1] and by Elek and Tardos [10]. We use the notion of convergence introduced for bounded degree graphs (that is the degree of each vertex is bounded above by some uniform constant d) first introduced by Benjamini and Schramm [3]. This notion was extended by Lyons [13] to bounded average degree graphs.

In [8] the author described the behavior of a random tree sequence with a given degree distribution. In this paper we extend this result and prove a similar characterization as in [7] for random trees with given degree sequence. We define the convergence of degree sequences and give a necessary and sufficient condition on the degree sequence, which implies the convergence of the tree sequence $\mathbf{T}(\mathbf{D}_n)$ in the sense of Lyons [13]. In the case of convergence we describe the limit object.

This paper is organized as follows: In Section 2, we give the basic definitions and notations. In Section 3, we describe the basic properties and the limit of a sequence of random degree sequences. At the end of the section we state our main theorem. In Section 4, we deal with labeled homomorphisms and in Section 5, we describe the limit object.

2. Basic definitions and notations

2.1. Random weak limit of graph sequences. Let $G = G(V, E)$ be a finite simple graph on n nodes. For $S \subseteqq V(G)$ denote by $G[S]$ the subgraph of G spanned by the vertices $v \in S$. For a finite simple graph G on n nodes, let $B_G(v, R)$ be the rooted R-ball around the node v, also called as the R-neighborhood of v, that is the subgraph induced by the nodes at distance at most R from v:

$$B_G(v, R) = G\big[\big\{\, u \in V(G) : \operatorname{dist}_G(u, v) \leqq R\big\}\big].$$

Two rooted graphs G_1, G_2 are rooted isomorphic if there is an isomorphism between them which maps the root of G_1 to the root of G_2. Given a positive integer R, a finite rooted graph F and a probability distribution ρ

on rooted graphs, let $p(R,F,\rho)$ denote the probability that the graph F is rooted isomorphic to the R-ball around the root of a rooted graph chosen with distribution ρ. It is clear that $p(R,F,\rho)$ depends only on the component of the root of the graph chosen from ρ. So we will assume that ρ is concentrated on rooted connected graphs. For a finite graph G, let $U(G)$ denote the distribution on rooted graphs obtained by choosing a uniform random vertex of G as root of G. It is easy to see that for any finite graph G we have

$$p\big(R,F,U(G)\big) = \frac{\big|\big\{v \in V(G) :\ B_G(v,R) \text{ is rooted isomorhpic to } F\big\}\big|}{\big|V(G)\big|}.$$

DEFINITION 1. Let (G_n) be a sequence of finite graphs on n nodes, ρ a probability distribution on infinite rooted graphs. We say that the random weak limit of G_n is ρ if for any positive integer R and finite rooted graph F we have

$$\lim_{n\to\infty} p\big(R,F,U(G_n)\big) = p(R,F,\rho). \tag{1}$$

If G_n is a sequence of random finite graphs, then $p\big(R,F,U(G_n)\big)$ is a random variable, so by convergence we mean convergence in probability.

DEFINITION 2. Let (G_n) be a sequence of random finite graphs on n nodes, ρ a probability distribution on infinite rooted graphs. We say that the random weak limit of G_n is ρ if $\forall\,\varepsilon > 0$, $R \in \mathbb{N}^+$ and finite rooted graph F we have

$$\lim_{n\to\infty} \mathbb{P}\Big(\big|p\big(R,F,U(G_n)\big) - p(R,F,\rho)\big| > \varepsilon\Big) = 0. \tag{2}$$

The formal meaning of this formula is, that the statistics $p\big(R,F,U(G_n)\big)$ as random variables are concentrated.

2.2. Other notations. We will denote random variables with bold characters. For a probability space $(\Omega,\mathcal{B},\mu)$ and $A \in \mathcal{B}$ denote by $\mathbf{I}(A)$ the indicator variable of the event A. Denote by $\mathcal{D}_n$ the set of possible degree sequences of a labeled tree on n nodes. Throughout the paper we consider labeled trees on n nodes unless stated otherwise. Let $\mathbf{D}_n$ be a random variable on $\mathcal{D}_n$. Denote by $\mathbf{T}(\mathbf{D}_n)$ the uniform random tree on n labeled nodes, with degree sequence $\mathbf{D}_n$. Denote the degree sequence of a tree T by $D_T = \big(D_T(i)\big)_{i=1}^n$. For a given degree sequence $D = \big(D(i)\big)_{i=1}^n$ there are

$$\binom{n-2}{D(1)-1, D(2)-1, \dots, D(n)-1}$$

labeled trees with degree sequence D. It follows that for an arbitrary tree T

$$\mathbb{P}(\mathbf{T}(\mathbf{D}_n) = T) = \frac{\mathbb{P}(\mathbf{D}_n = D_T)}{\binom{n-2}{D_T(1)-1, D_T(2)-1, \ldots, D_T(n)-1}}.$$

If it does not cause any confusion, we will use $D, \mathbf{T}_n$ instead of $D_T, \mathbf{T}(\mathbf{D}_n)$ respectively.

A finite rooted graph G with root v is said to be l-deep if the largest distance from the root is l, that is

$$l = \max_{u \in V(G)} \operatorname{dist}(v, u).$$

Denote by U^l the set of equivalence classes of finite unlabeled l-deep rooted graphs with respect to root-preserving isomorphisms. Let T_x^l be an l-deep rooted tree on k nodes with root x. Denote the vertices at distance i from the root by T_i, and let $t_i = |T_i|$ (t_0 is 1, t_1 is the degree of the root). For every finite graph G, $p\big(R, F, U(G)\big)$ induces a probability measure on U^R which we call the R-neighborhood statistics of G. If G is a tree then $p\big(R, F, U(G)\big)$ is concentrated on rooted trees.

Let $\mathcal{T}$ be the set of all countable, connected infinite rooted trees. For an infinite rooted tree $T \in \mathcal{T}$ denote by $T(R)$ the R-neighborhood of the root of T. For an R-deep rooted tree F define the set

$$\mathcal{T}(F) = \{T \in \mathcal{T} : T(R) \text{ is rooted isomorphic to } F\}.$$

Let $\mathcal{F}$ be the sigma-algebra generated by the sets $\big(\mathcal{T}(F)\big)_F$, where F is an arbitrary finite rooted tree. $(\mathcal{T}, \mathcal{F})$ is a probability field. We call a probability measure μ on $\mathcal{T}$ an infinite rooted random tree.

Every infinite random tree μ has the property that for any $F \in U^R$,

$$p(R, F, \mu) = \sum_{H \in U^{R+1},\ H(R) \cong F} p(R+1, H, \mu). \tag{3}$$

Actually every distribution on rooted infinite graphs has the above property. Note that if we want to prove the convergence of a random tree sequence to a certain limit distribution ρ, then we need to have the convergence of the neighborhood densities and also (3), the consistency of these densities, which ensures that ρ will be concentrated on infinite rooted trees. These together will imply (2).

3. Limits of degree sequences

Consider a random degree sequence $\mathbf{D}_n = \big(\mathbf{D}_n(i)\big)_{i=1}^n$ and construct a labeled tree $\mathbf{T}(\mathbf{D}_n)$ with uniform distribution given the degree sequence. We want to describe the limit of $\mathbf{T}(\mathbf{D}_n)$ as $n \to \infty$. We give a characterization of the degree sequences for which $\mathbf{T}(\mathbf{D}_n)$ has an infinite random tree as a limit. To describe the model and the limit, we need to define and understand the limit of a random degree sequence $\mathbf{D}_n$. Here we only deal with degree sequences of trees. We further assume that $\mathbf{D}_n$ is an exchangeable sequence, that is for any $\sigma \in S_n$ we have

$$\big(\mathbf{D}_n(i)\big)_{i=1}^n \sim \big(\mathbf{D}_n\big(\sigma(i)\big)\big)_{i=1}^n.$$

Exchangeability is a way to eliminate exceptional vertices and allows us to use the limit theory of exchangeable sequences. For more on exchangeable random variables we refer to [2].

Definition 3. We say that an exchangeable sequence $\mathbf{D}_n$ is convergent and $\mathbf{D}_n \to \mathbf{D}$ where $\mathbf{D}$ is a random infinite sequence if for every $k \in \mathbb{N}$ we have

$$\big(\mathbf{D}_n(i)\big)_{i=1}^k \xrightarrow{\mathbb{P}} \big(\mathbf{D}(i)\big)_{i=1}^k.$$

It is easy to see that if $\mathbf{D}_n$ is exchangeable and $\mathbf{D}_n \to \mathbf{D}$ then $\mathbf{D}$ is also an exchangeable sequence. The following theorem of Hewitt and Savage [11], which is a generalization of de Finetti's theorem, describes the limits of exchangeable sequences.

Theorem 1. *Let $\mathbf{X}$ be a random infinite exchangeable array. Then $\mathbf{X}$ is a mixture of infinite dimensional IID distributions*

$$\mathbf{X} = \int_{IID} \lambda \, dp(\lambda),$$

where p is a distribution on infinite dimensional IID distributions λ.

As a result, the limit of an exchangeable degree sequence is an infinite exchangeable sequence and so a mixture of IID distributions. Note that if $\mathbf{D}_n$ is not exchangeable then we can take a random permutation $\sigma \in S_n$ and define the exchangeable degree sequence $\tilde{\mathbf{D}}_n(i) = \mathbf{D}_n\big(\sigma(i)\big)$.

Lemma 1. *Let $\mathbf{X}$ be an infinite exchangeable random sequence. Further assume that we have*

$$\mathbb{P}\big(\mathbf{X}(1) = i,\ \mathbf{X}(2) = i\big) = \mathbb{P}\big(\mathbf{X}(1) = i\big)\,\mathbb{P}\big(\mathbf{X}(2) = i\big).$$

Then $\mathbf{X}$ is an infinite IID distribution (p is concentrated on one distribution).

PROOF. From Jensen's inequality we have that

$$\int_{\mathrm{IID}} \lambda(i)^2 \, dp(\lambda) \geqq \left(\int_{\mathrm{IID}} \lambda(i) \, dp(\lambda) \right)^2. \tag{4}$$

Also from Theorem 1 we have

$$\int_{\mathrm{IID}} \lambda(i)^2 \, dp(\lambda) = \mathbb{P}\big(\mathbf{X}(1) = i, \mathbf{X}(2) = i\big)$$

$$= \mathbb{P}\big(\mathbf{X}(1) = i\big)\, \mathbb{P}\big(\mathbf{X}(2) = i\big) = \left(\int_{\mathrm{IID}} \lambda(i) \, dp(\lambda) \right)^2$$

It follows that in (4) equality holds which means that p is a degenerate distribution and so it proves our lemma. □

We will see that if $\mathbf{T}(\mathbf{D}_n)$ is convergent then $\mathbf{D}_n$ satisfies the assumptions in Lemma 1. So for a convergent random tree sequence $\mathbf{T}(\mathcal{D}_n)$ the limit of the degree sequence $\mathbf{D}_n$ needs to be an infinite IID distribution.

EXAMPLE 1. Let $\mathbf{X}$ be a uniform random element of $[n]$. Consider the degree sequence

$$\mathbf{D}(i) = \begin{cases} n-1, & \text{if } \ i = \mathbf{X} \\ 1, & \text{otherwise} \end{cases}$$

Let $\mathbf{St}_n = \mathbf{T}(\mathbf{D}_n)$ be the star-graph on n nodes. The limit degree sequence is just the constant 1 vector $\bullet = (1, 1, \dots)$. Obviously in the limit the expected degree of a node is 1. It is not hard to see that if F is not a single edge, then $p(R, F, \mathbf{St}_n) = 0$ for every $n > \big|V(F)\big|$. Thus there is no limit distribution ρ on infinite graphs such that $\mathbb{P}\big(\big|p(R, F, U(\mathbf{St}_n) - p(R, F, \rho)\big| > \varepsilon\big) \to 0$ for every F.

Example 1 shows that if only the average degree is bounded, too many unbounded degree vertices destroy convergence. As the average degree of a tree on n nodes is $2\frac{n-1}{n}$, one would expect that in the limit distribution the expected degree of a node is 2, that is $\mathbb{E}\big(\mathbf{D}(i)\big) = 2$ for every i.

It turns out that it is enough to have that the degree sequence converges and $\mathbb{E}\big(\mathbf{D}(i)\big) = 2$ holds for all i. Now we are ready to state our main theorem which describes the degree sequence of convergent random tree sequences.

THEOREM 2. *Let $\mathbf{D}_n$ be a sequence of random degree sequences $(\mathbf{D}_n \in \mathcal{D}_n)$. The random tree sequence $\mathbf{T}(\mathbf{D}_n)$ is convergent and converges to an infinite random tree if and only if $\mathbf{D}_n \to \mathbf{D}$, where $\mathbf{D} = (\mathbf{D}_0, \mathbf{D}_0, \dots)$ is an infinite IID sequence and $\mathbb{E}(\mathbf{D}_0) = 2$.*

4. Labeled subgraph densities

To prove convergence we need to understand the neighborhood statistics of the random tree $\mathbf{T}(\mathbf{D}_n)$. First we will count subgraph densities and then relate them to neighborhood statistics. For fixed unlabeled graphs F and G denote by

$$\mathrm{inj}(F,G) = \frac{\big|\{\phi : \phi \text{ is an injective homomorphism from } F \text{ to } G\}\big|}{\big|V(G)\big|}$$

the normalized number of copies of F in G. We call F the test graph. We call $\mathrm{inj}(F,G)$ the injective density of F in G. For bounded degree graphs the convergence of injective densities for every F is equivalent to the convergence of neighborhood densities for every H rooted finite graph. For bounded average degree graphs, subgraph statistics may be unbounded. For the random star tree $\mathbf{St}_n$ we have

$$\mathrm{inj}(\text{}, \mathbf{St}_n) = \frac{(n-1)(n-2)}{n}.$$

To avoid unbounded subgraph statistics we add a further structure to the test graph F. We call a pair (F,r) a numbered graph, where $r = (r_i)_{i=1}^{V(F)}$ and $r_i \in \mathbb{N}$. We call r_i the remainder degree of the node $i \in V(F)$. Let (F,r) be a numbered graph and ϕ be a homomorphism from F to a graph G. We say that ϕ is a labeled homomorphism if ϕ is a homomorphism and

$$D_G\big(\phi(v)\big) = D_F(v) + r_v, \quad \forall\, v \in V(F).$$

Let

$$\mathrm{inj}_{\mathrm{lab}}\big((F,r),G\big)$$

$$= \frac{\big|\{\phi : \phi \text{ is an injective labeled homomorphism from } F \text{ to } G\}\big|}{\big|V(G)\big|}$$

be the normalized number of numbered copies of F in G. First we want to derive properties of degree sequences $\mathbf{D}_n$ for which $\mathrm{inj}_{\mathrm{lab}}\big((F,r),\mathbf{T}(\mathbf{D}_n)\big)$ is convergent for every finite graph F and remainder degrees r. Then in Section 5 we will turn to the convergence of neighborhood statistics.

Remark 1. The convergence of $\mathrm{inj}_{\mathrm{lab}}(.,G_n)$ for every (F,r) does not imply the random weak convergence of G_n in general. $\mathrm{inj}_{\mathrm{lab}}\big((F,R),\mathbf{St}_n\big)$ is convergent for every (F,r), but as we saw earlier $\mathbf{St}_n$ is not a convergent tree sequence.

REMARK 2. Let (F,r) be an arbitrary numbered graph on k nodes. One can easily see that $\mathrm{inj}_{\mathrm{lab}}\big((F,r),G\big)$ is uniformly bounded for every G.

PROOF. To see this we will bound the number of ways we can construct an injective labeled homomorphism ψ from (F,r) to G. Let $R=\max\{r_i\}$. If we define $\psi(1)=v\in V(G)$, then $D_G\big(\psi(1)\big)=r_1$. There are at most $D_G\big(\psi(1)\big)^{D_F(1)}=r_1^{D_F(1)}\leqq R^k$ possibilities for $\psi(u)$'s $\big(u\in N_F(1)\big)$, where $N_F(1)$ is the set of neighbors of 1 in F. Following this idea we get that for every v there are at most $\left(R^k\right)^k$ possible ways to extend ψ, given $\psi(1)=v$. Hence there are at most nR^{k^2} injective labeled homomorphisms from F to G and the remark follows. □

For an arbitrary numbered tree (T,r), and $\phi: V(T)\mapsto[n]$ let

$$\mathbf{I}_n\big((T,r),\phi\big)=\mathbf{I}\big(\{\phi \text{ is an injective labeled homomorphism of } T \text{ to } \mathbf{T}_n\}\big)$$

$$\mathbf{X}_n^{(T,r)}=\sum_{\phi:\,V(T)\mapsto[n]}\mathbf{I}_n\big((T,r),\phi\big)=n\cdot\mathrm{inj}_{\mathrm{lab}}\big((T,r),\mathbf{T}(\mathbf{D}_n)\big). \tag{5}$$

We define $\mathbf{I}_n\big((F,r),\phi\big)$, $\mathbf{X}_n^{(F,r)}$ similarly for a numbered forest (F,r). If it does not cause any confusion, we will omit r from the formulas above and use $\mathbf{I}_n(T,\phi)$, X_n^T, $\mathbf{I}_n(F,\phi)$ and $\mathbf{X}_n^F$ instead to simplify notation. For random graph sequences $\mathbf{G}_n$ by the convergence of $\mathrm{inj}_{\mathrm{lab}}(.,\mathbf{G}_n)$ we mean convergence in probability.

Let $\mathbf{D}_n$ be a random degree sequence and $\mathbf{T}_n=\mathbf{T}(\mathbf{D}_n)$ be the associated random tree sequence. Let (T,r) be a numbered tree. As $\mathrm{inj}_{\mathrm{lab}}\big((T,r),\mathbf{T}_n\big)$ is bounded, we have that $\mathrm{inj}_{\mathrm{lab}}\big((T,r),\mathbf{T}_n\big)$ is convergent for every (T,r) if and only if we have that

$$\mathbb{D}^2\left(\frac{X_n^T}{n}\right)=\mathbb{D}^2\big(\mathrm{inj}_{\mathrm{lab}}\big((T,r),\mathbf{T}_n\big)\big)\to 0. \tag{6}$$

We will use this formula to prove properties of the degree sequence. We can expand the above formula using (5):

$$\mathbb{D}^2\left(\frac{X_n^T}{n}\right)=\frac{1}{n^2}\Bigg(\sum_{\psi,\phi:\,V(T)\mapsto[n]}\mathbb{E}\big(\mathbf{I}_n(T,\psi)\mathbf{I}_n(T,\phi)\big) - \sum_{\psi,\phi:\,V(T)\mapsto[n]}\mathbb{E}\big(\mathbf{I}_n(T,\psi)\big)\,\mathbb{E}\big(\mathbf{I}_n(T,\phi)\big)\Bigg)\to 0. \tag{7}$$

The following two lemmas will establish a connection between the degree sequence and the probabilities $\mathbb{P}\big(\mathbf{I}_n(T,\phi)=1\big)$. Then we will use (7) to prove that the degree sequence satisfies the conditions in Lemma 1.

REMARK 3. As the degree sequence is exchangeable, for any $\psi,\phi: V(T)\mapsto[n]$ we have

$$\mathbb{P}\big(\mathbf{I}_n(T,\phi)=1\big)=\mathbb{P}\big(\mathbf{I}_n(T,\psi)=1\big).$$

Let T be an arbitrary tree on k nodes. For a random degree sequence $\mathbf{D}_n$ and $\phi: V(T)\mapsto[n]$ let $\mathbf{D}_\phi=\big(\mathbf{D}_n\big(\phi(i)\big)\big)_{i=1}^k$.

LEMMA 2. *Let $\mathbf{D}_n\in\mathcal{D}_n$ be a random degree sequence and $\mathbf{T}_n=\mathbf{T}(\mathbf{D}_n)$. Let F be an arbitrary forest on m $(m\leqq n)$ nodes with remainder degrees $r=(r_1,\dots,r_m)$. Let $R=\sum_i r_i$ and denote by $C_1,C_2,\dots,C_c$ the connected components of F. The probability that an arbitrary $\phi: V(T)\mapsto[n]$ is an injective labeled homomorphism is*

$$\mathbb{P}\big(\mathbf{I}_n(F,\phi)=1\big)=\frac{(n-m+c-2)!}{(n-2)!}H(r,F)\mathbb{P}(\mathbf{D}_\phi=D_T),$$

where $H(r,F)=\prod_{i=1}^c\Big[\big(\sum_{j\in C_i}r_j\big)\prod_{j\in C_i}\frac{(D_F(j)+r_j-1)!}{(r_j!)}\Big]$ is a constant depending only on F and the remainder degrees r.

PROOF. We may assume that $\phi(i)=i$ for all $i\in V(F)$. Let $R_i=\sum_{j\in C_i}r_j$. Fix a degree sequence $D=\big(D(i)\big)_{i=1}^n$. It follows from the Prüfer sequence that the number of trees realizing this degree sequence is $\binom{n-2}{D(1)-1,\dots,D(n)-1}$. We need to count the trees with degree sequence D which have F spanned by the first m nodes and the remainder degree condition holds. Contract every connected component C_i of F to a single vertex u_i. Also contract the images of these components in $\mathbf{T}(\mathbf{D}_n)$. We get a tree on $n-m+c$ nodes with degree sequence

$$D'=\big(R_1,R_2,\dots,R_c,D(m+1),\dots,D(n)\big).$$

There are

$$\binom{n-m+c}{R_1-1,R_2-1,\dots,R_c-1,D(m+1)-1,\dots,D(n)-1}$$

trees realizing the degree sequence D'. For each connected component C_i we can connect the R_i edges to the vertices in

$$\frac{R_i!}{\prod_{j\in C_i}r_j!}$$

ways. It follows that the number of labeled trees realizing the degree sequence D and having F on the first m vertices is

$$\binom{n-m+c-2}{R_1-1,\dots,R_c-1,D(m+1)-1,\dots,D(n)-1}\prod_{i=1}^{c}\left[\frac{R_i!}{\prod_{j\in C_i}r_j!}\right].$$

From this it follows that

$$\mathbb{P}\big(\mathbf{I}_n(F,\phi)=1 \mid \mathbf{D}_n=D\big) \tag{8}$$

$$=\frac{\binom{n-m+c-2}{R_1-1,\dots,R_c-1,D(m+1)-1,\dots,D(n)-1}}{\binom{n-2}{D(1)-1,\dots,D(n)-1}}\prod_{i=1}^{c}\frac{R_i!}{\prod_{j\in C_i}r_j!}.$$

Note that the degree sequence D should be such that $D(i)=D_F(i)+r_i$, $i=1,\dots,m$, holds for the first m degrees. We need to sum this probability for all possible degree sequences. In our case we sum over degree sequences for which $D(i)=D_F(i)+r_i$, $i=1,\dots,m$, holds. As in equation (8) the right hand side does not depend on $D(i)$, $i>m$, we have

$$\mathbb{P}\big(\mathbf{I}_n(F,\phi)=1\big)$$

$$=\frac{(n-m+c-2)!}{(n-2)!}\prod_{i=1}^{c}\left[R_i\prod_{j\in C_i}\frac{(D_F(j)+r_j-1)!}{(r_j!)}\right]\mathbb{P}(\mathbf{D}_\phi=D_T).$$

If we take $H(r,F)=\prod_{i=1}^{c}\big[R_i\prod_{j\in C_i}\frac{(D_F(j)+r_j-1)!}{(r_j!)}\big]$, we get the desired equation. □

Let $(F_1,r_1),(F_2,r_2)$ be two labeled graphs, $\phi: V(F_1)\mapsto[n]$ and $\psi: V(F_2)\mapsto[n]$. We denote by $F_{1,2}$ the graph obtained by identifying nodes $i\in V(F_1)$, $j\in V(F_2)$ if and only if $\phi(i)=\psi(j)$. We can define remainder degrees $r_{1,2}$ on $F_{1,2}$ in a straightforward way if $\phi(i)=\psi(j)\Rightarrow r_1(i)=r_2(j)$.

LEMMA 3. *Let $\mathbf{D}_n\in\mathcal{D}_n$ be a random degree sequence and $\mathbf{T}_n=\mathbf{T}(\mathbf{D}_n)$. Let F_1, F_2 be two forests on m_1 and m_2 nodes $(m_1,m_2\leqq n)$ with remainder degrees r_1, r_2. Let $\phi: V(F_1)\mapsto[n]$ and $\psi: V(F_2)\mapsto[n]$. If $F_{1,2}$ is a forest and we can define $r_{1,2}$, then let $m_{1,2}=\big|V(F_{1,2})\big|$, $c_{1,2}=\{$the number of components of $F_{1,2}\}$ and $R_{1,2}=\sum_{V(F_{1,2})}r_{1,2}(i)$. We have*

$$\mathbb{P}\big(\mathbf{I}_n(F_1,\phi)=1 \mid \mathbf{I}_n(F_2,\psi)=1\big) \tag{9}$$

$$=\frac{(n-m_{1,2}+c_{1,2}-2)!}{(n-m_2+c_2-2)!}\frac{H(r_{1,2},F_{1,2})}{H(r_2,F_2)}\mathbb{P}(\mathbf{D}_\phi=D_{F_1}\mid\mathbf{D}_\psi=D_{F_2}).$$

PROOF. The proof follows immediately from the definition of conditional probability. □

Let $\mathbf{D}_n$ be a degree sequence and $\mathbf{T}_n = \mathbf{T}(\mathbf{D}_n)$ be the associated random tree. Assume that $\mathrm{inj}_{\mathrm{lab}}\big((T,r),\mathbf{T}_n\big)$ is convergent. Then by (6), $\mathbb{D}^2(\mathbf{X}_n^T/n) \to 0$. For any tree T on k nodes we have

$$\mathbb{D}^2\left(\frac{\mathbf{X}_n^T}{n}\right) = \frac{1}{n^2} \sum_{\phi,\psi\colon V(T)\mapsto[n]} \Big(\mathbb{E}\big(\mathbf{I}_n(T,\phi)\mathbf{I}_n(T,\psi)\big) - \mathbb{E}\big(\mathbf{I}_n(T,\phi)\big)\,\mathbb{E}\big(\mathbf{I}_n(T,\psi)\big)\Big). \tag{10}$$

Now if we split the sum by the size of the intersection of $\phi\big(V(T)\big)$ and $\psi\big(V(T)\big)$ and use Remark 3, we have

$$\mathbb{D}^2\left(\frac{\mathbf{X}_n^T}{n}\right) = \frac{1}{n^2} \sum_{i=0}^{k} \sum_{|\phi(V(T))\cap\psi(V(T))|=i} n(n-1)\cdot\ldots\cdot(n-2k+i+1) \cdot \Big(\mathbb{E}\big(\mathbf{I}_n(T,\phi)\mathbf{I}_n(T,\psi)\big) - \mathbb{E}\big(\mathbf{I}_n(T,\phi)\big)\,\mathbb{E}\big(\mathbf{I}_n(T,\psi)\big)\Big) \tag{11}$$

From Lemmas 2 and 3 we can easily derive that the order of the terms corresponding to $i \neq 0$ is $\mathcal{O}\big(\frac{1}{n}\big)$. It follows that the condition $\mathbb{D}^2(\mathbf{X}_n^T) \to 0$ is equivalent to

$$\frac{(n-1)\cdot\ldots\cdot(n-2k+1)}{n}\big(\mathbb{P}\big(\mathbf{I}_n(T,\phi)\mathbf{I}_n(T,\psi)\big) - \mathbb{P}\big(\mathbf{I}_n(T,\phi)\big)\,\mathbb{P}\big(\mathbf{I}_n(T,\psi)\big)\big) \to 0.$$

Using again Lemmas 2 and 3 we can easily derive the following:

$$\forall\, T,\ \mathbb{D}^2\left(\frac{\mathbf{X}_n^T}{n}\right) \to 0 \Leftrightarrow \forall\,\phi,\psi:\ V(T)\mapsto[n],\ \phi\big(V(T)\big)\cap\psi\big(V(T)\big) = \emptyset \tag{12}$$

$$\mathbb{P}(\mathbf{D}_\phi = D_T, \mathbf{D}_\psi = D_T) \to \mathbb{P}(\mathbf{D}_\phi = D_T)\mathbb{P}(\mathbf{D}_\psi = D_T).$$

The following corollary is an easy application of Lemma 1 and (12).

COROLLARY 1. *The labeled subgraph densities of a random tree sequence converge in probability if and only if the corresponding degree sequence converges to an infinite IID sequence.*

REMARK 4. The formula in Lemma 2 yields an easy result on the probability that two vertices i, j with degrees d_i, d_j are connected:

$$\mathbb{P}\big(ij \in E\big(\mathbf{T}(\mathbf{D}_n)\big) \mid \mathbf{D}_n(i) = d_i,\ \mathbf{D}_n(j) = d_j\big) = \frac{d_i + d_j - 2}{n-2}.$$

Similarly for a given edge $ij \in E\big(\mathbf{T}(\mathbf{D}_n)\big)$ the degree distribution of the vertices i and j can be expressed as

$$\mathbb{P}\big(\mathbf{D}_n(i) = d_i, \mathbf{D}_n(j) = d_j \mid ij \in E\big(\mathbf{T}(\mathbf{D}_n)\big)\big)$$

$$= \frac{n}{n-2}\frac{d_i + d_j - 2}{2}\mathbb{P}\big(\mathbf{D}_n(i) = d_i,\ \mathbf{D}_n(j) = d_j\big).$$

5. The limit of $\mathbf{T}(\mathbf{D}_n)$

In the last section we discussed tree sequences $\mathbf{T}_n$ for which $\mathrm{inj}_{\mathrm{lab}}\big((T,r), \mathbf{T}_n\big)$ was convergent for every (T,r). We now turn to neighborhood statistics. First we want to relate them to labeled subgraph densities. We will express the neighborhood statistics as functions of the labeled subgraph densities.

As before let U^l denote the set of all finite l-deep rooted tree. Consider an l-deep rooted tree with root x: $T_x^l \in U^l$, with $|T_x^l| = k$. Let us denote the nodes at distance i from the root by T_i, and $|T_i| = t_i$ (t_0 is just 1, t_1 is the degree of the root). $B_G(v,l)$ is the rooted l-ball around v in G and $\mathbf{T}_n = \mathbf{T}(\mathbf{D}_n)$ is a random labeled tree with degree distribution $\mathbf{D}_n$.

Let $\sigma, \rho \in \mathrm{Aut}(T_x^l)$ be two rooted automorphisms of the rooted tree T_x^l. We say that $\sigma \sim \rho$ if and only if there exists $\tau \in \mathrm{Aut}(T_x^l)$, such that τ fixes every vertex not in T_l and $\sigma \circ \tau = \rho$. $\sim$ is an equivalence relation. The equivalence classes have $\prod_{i \in T_{l-1}}(D(i)-1)!$ elements, hence it follows

$$\big|\mathrm{Aut}(T_x^l)\big| = \big|\mathrm{Aut}(T_x^l)/\sim\big| \prod_{i \in T_l}\big(D(i)-1\big)!. \tag{13}$$

It is easy to see that

$$p(l, T_x^l, \mathbf{T}_n) = \frac{1}{n}\frac{\mathbf{X}_n^{\left(T', (r_i')_{i=1}^{|T'|}\right)}}{\big|\mathrm{Aut}(T_x^l)/\sim\big|}, \tag{14}$$

where $T' = T_x^l \setminus T_l$, and

$$r_i' = \begin{cases} 0 & i \notin T_l \cup T_{l-1} \\ D_{T_x^l}(i) - 1 & i \in T_{l-1}. \end{cases} \tag{15}$$

If $\mathbf{T}_n$ is a convergent random tree sequence then from (2) and (14) we have that for any T' defined above

$$\mathbb{D}^2\left(\frac{X_n^{T'}}{n}\right) \to 0. \tag{16}$$

For bounded degree graphs the convergence of the neighborhood densities implies the convergence of the graph sequence in the sense of Benjamini and Schramm. We saw earlier in Example 1 that for bounded average degree graphs this is not the case. The convergence of the neighborhood densities alone is not enough. We need also (3) to hold.

The reason is that for fixed k the k-neighborhood of the large degree nodes is large ($\mathcal{O}(n)$). In Example 1 even if $k=1$, every node "sees" the center node (e.g. every neighborhood with radius 1 contains the center node) and so every 2 radius neighborhood contains $\mathcal{O}(n)$ vertices, which is unbounded.

Assign remainder degrees r ($r_i = 0,\ \forall\, i \notin T_l$) to the rooted tree T_x^l and forget the root, then using Lemma 2

$$\mathbb{E}(\mathbf{X}_n^T) = \mathbb{E}\Bigg(\sum_{\phi\colon V(T)\mapsto[n]} \mathbf{I}_n(T,\phi)\Bigg) = \frac{n!}{(n-k)!}\mathbb{P}\big(\mathbf{I}_n(T,\phi)=1\big) \tag{17}$$

$$= n\frac{n-1}{n-k}\mathbb{P}\big(\mathbf{D}_n\big(\{1,2,\dots,k\}\big) = D_T\big)H(r,T).$$

From (15) we have that

$$p(l,T_x^l,\mathbf{T}_n) = \frac{1}{n}\frac{\mathbf{X}_n^{T'}}{\big|\operatorname{Aut}(T_x^l)/\sim\big|}.$$

We want to define an infinite random rooted tree which is the limit of $\mathbf{T}_n$. Let

$$\mu_n(T_x^l) = \frac{1}{n}\frac{\mathbb{E}(\mathbf{X}_n^{T'})}{\big|\operatorname{Aut}(T_x^l)/\sim\big|}.$$

Assume we have a convergent sequence of random trees $\mathbf{T}_n$ with degree sequence $\mathbf{D}_n$. Further assume that $\mathbf{D}_n \to \mathbf{D} = (\mathbf{D}_0, \mathbf{D}_0, \dots)$ and let $\gamma = \mathbb{E}(D_0) - 1$. Define

$$p(T_x^l) = \lim_{n\to\infty} \mu_n(T_x^l) \tag{18}$$

$$= \lim_{n\to\infty} \frac{1}{\big|\operatorname{Aut}(T_x^l)/\sim\big|}\frac{n-1}{n-k}\mathbb{P}\big(\mathbf{D}_n\big(\{1,2,\dots,k\}\big) = D_{T'}\big)H(r,T')$$

$$= \frac{\prod_{i \notin T_l} \mathbb{P}(D_0 = d_i)(d_i - 1)!}{\big|\operatorname{Aut}(T_x^l)\big|} t_l.$$

We can expand the formula

$$H(r, T') = \sum_{i \in V(T')} r_i \prod_{i \in V(T')} \frac{(d_i + r_i - 1)!}{r_i!}$$

$$= \sum_{i \in T_{l-1}} (d_i - 1) \prod_{i \notin T_{l-1} \cup T_l} (d_i - 1)! = t_l \prod_{i \notin T_{l-1} \cup T_l} (d_i - 1)!.$$

Then the last equation in (18) follows using equation (13) and the expansion of $H(r, T')$.

Define $\mu\big(\mathcal{T}(F)\big) = p(F)$. As the sets $\mathcal{T}(F)$ generate the σ-algebra, we can extend μ to $\mathcal{T}$ if μ satisfies (3). If this is the case then μ is a random infinite rooted tree.

LEMMA 4. *Let $\mathbf{D}_n$ be an exchangeable random degree sequence and assume that $\mathbf{D}_n \to \mathbf{D}$, where $\mathbf{D}$ is an infinite IID random sequence of the variable $\mathbf{D}_0$. Let μ be the associated measure defined above. μ extends to a probability measure on $\mathcal{G}$ if and only if $\mathbb{E}(\mathbf{D}_0) = 2$ (or equivalently $\gamma = 1$).*

PROOF. We only need to show that μ satisfies (3):

$$p(T_x^{l-1}) = \sum_{T_x^l \,:\, B_{T_x^l}(x, l-1) \cong T_x^{l-1}} p(T_x^l) \Leftrightarrow \gamma = 1 \tag{19}$$

We have

$$p(T_x^l) = \frac{\prod_{i \notin T_l} \mathbb{P}(\mathbf{D}_0 = d_i)(d_i - 1)!}{\big|\operatorname{Aut}(T_x^l)\big|} t_l.$$

Now rearranging the sum by the degrees of the leaves of T_x^{l-1} in T_x^l we have

$$\sum_{T_x^l \,:\, B_{T_x^l}(x, l-1) \cong T_x^{l-1}} p(T_x^l) = \sum_{D_{T_x^l}(i) = 1,\, i \in T_{l-1}}^{\infty} p(T_x^{l-1} \cup (d_i)_{i \in T_{l-1}})$$

$$= \prod_{j \notin T_{l-1} \cup T_l} \mathbb{P}(\mathbf{D}_0(j) = d_j)(d_j - 1)!$$

$$\times \sum_{D_{T_x^l}(i) = 1,\, i \in T_{l-1}}^{\infty} \frac{\prod_{i \in T_{l-1}} \mathbb{P}(\mathbf{D}_0(j) = d_j)(d_i - 1)!}{\big|\operatorname{Aut}(T_x^l)\big|} t_l$$

$$= \frac{\prod_{j \notin T_l \cup T_{l-1}} \mathbb{P}(\mathbf{D}_0(j) = d_j)(d_j - 1)!}{\big|\operatorname{Aut}(T_x^l \setminus T_l)\big|} t_{l-1}\gamma,$$

where the last equation follows from the fact that for fixed d_i, $i \in T_{l-1}$ every $\sigma \in \operatorname{Aut}(T_x^{l-1})$ has only one extension in $\operatorname{Aut}(T_l)/\sim$. Now (19) will hold only if $\gamma = 1$. It follows that (19) holds if and only if $\mathbb{E}(D_0) = 2$ ($\gamma = 1$). □

Proof of Theorem 2. Let $\mathbf{D}_n$ be a degree sequence and $\mathbf{T}(\mathbf{D}_n) = \mathbf{T}_n$ be the associated random tree sequence. First assume, that the degree sequence converges to the distribution $\mathbf{D} = (\mathbf{D}_0, \mathbf{D}_0, \dots)$ and $\mathbb{E}(\mathbf{D}_0) = 2$. From equation (12) we get that for an arbitrary tree T, $\mathbb{D}^2\big(\frac{\mathbf{X}_n^T}{n}\big) \to 0$. Then by equation (14) we have that for every T_x^l l-deep rooted tree, the neighborhood statistics converge in probability to a limiting distribution $p(T_x^l)$. As the assumptions of Lemma 4 hold, $p(T_x^l)$ defines a measure μ on infinite rooted trees and so $\mathbf{T}_n \to \mu$.

On the other hand assume that $\mathbf{T}_n$ converges to a random infinite rooted tree μ. Then by equation (14) we get that the number of degree d vertices is concentrated. Using that our degree distribution is exchangeable we get that $\mathbf{D}_n \to \mathbf{D} = (\mathbf{D}_0, \mathbf{D}_0, \dots)$ and $\mathbb{E}(\mathbf{D}_0) = 2$. □

References

[1] D. Aldous, The continuum random tree III, *Ann. Probab.,* **21** (1993), 248–289.

[2] D. Aldous, *Exchangeability and Related Topics,* Ecole d'Ete St Flour 1983. Springer Lecture Notes in Math 1117 (1985).

[3] I. Benjamini and O. Schramm, Recurrence of distributional limits of finite planar graphs, *Electronic J. Probab.,* **6** (2001), 1–13.

[4] I. Benjamini, O. Schramm and A. Shapira, Every minor-closed property of sparse graphs is testable, in: *40th Ann. ACM Symp. on Th. Comp.* (2008), pp. 393–402.

[5] B. Bollobás, *Random Graphs,* 2nd edition, Academic Press (1998).

[6] C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós and K. Vesztergombi, Counting graph homomorphisms, in: *Topics in Discrete Mathematics* (ed. M. Klazar, J. Kratochvil, M. Loebl, J. Matousek, R. Thomas, P. Valtr), Springer (2006), pp. 315–371.

[7] S. Chatterjee, P. Diaconis and A. Sly, Random graphs with a given degree sequence, *Ann. of Appl. Probab.,* **21** (2011), 1400–1435.

[8] A. Deák, Limits of random trees, *Acta Math. Hungar.,* **141** (2013), 185–201.

[9] G. Elek, On limits of finite graphs, *Combinatorica,* **27** (2007), 503–507.

[10] G. Elek and G. Tardos, *Limits of Trees,* Oberwolfach Report No. 11/2010, 566–568.

[11] E. Hewitt and L. J. Savage, Symmetric measures on Cartesian products, *Trans. Amer. Math. Soc.,* **80** (1955), 470–501.

[12] L. Lovász and B. Szegedy, Limits of dense graph sequences, *J. Combin. Theory Ser. B,* **96** (2004), 933–957.

[13] R. Lyons, Asymptotic enumeration of spanning trees, *Combinatorics, Probability and Computing,* **14** (2005), 491–522.