# Tree-valued Markov chains and Poisson-Galton-Watson distributions

David Aldous

ABSTRACT. The Poisson-Galton-Watson distribution on finite trees, and the related $\mathrm{PGW}^{\infty}(1)$ distribution on infinite trees with one end, arise in several contexts, in particular as $n \to \infty$ weak limits within various size-$n$ combinatorial models. We review this topic, introducing slick notation for describing such distributions. We then describe a family of continuous-time Markov chains whose marginal distributions are of Poisson-Galton-Watson type. Different chains in this family have connections with different parts of the extensive literature on tree and forest-valued chains (the random graph process, stochastic coalescence, spanning tree chains) and also illustrate the classical state-space-compactification theory of continuous-time countable-state chains.

## 1. Introduction

Tree- and forest-valued Markov chains have been studied in several areas.

- Evolution of tree-based data structures (Knuth [**13**], Mahmoud [**16**]).
- The *Markov chain tree theorem* (Pemantle [**18**], Lyons-Peres [**15**]), which to a finite-state Markov chain associates a chain on the set of spanning trees.
- The classical random graph process (Bollobás [**7**], Alon-Spencer [**5**]). By ignoring edges created within components, the components grow as a forest-valued process.
- Reversible models of polymerization (Whittle [**20**], Pittel et al. [**17**]).
- Irreversible models of stochastic coalescence (Aldous [**3**]).
- Spin systems on infinite trees. If the $\{0,1\}$-valued r.v.'s are assigned to edges instead of vertices, we may regard the system as a forest-valued process.

These areas have different specific motivations and have largely developed independently. Setting aside applications, consider the question

> amongst the unlimited number of tree-valued chains one could define, which seem mathematically fundamental?

The purpose of this paper is to describe (sections 3 and 4) one family of six or seven related chains which do seem fundamental, and which have connections with several of the areas above. The chains in our family have the common features
(a) their marginal (i.e. fixed time) distributions are related to Poisson-Galton-Watson distributions.

---

1991 *Mathematics Subject Classification.* Primary 60J27, 60C05, 60J80.

Research supported in part by N.S.F. Grant DMS 9622859.

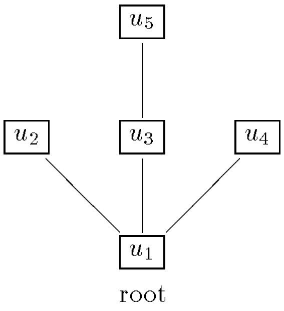


FIGURE 1. A $u$-labeled tree $\mathbf{t}$

(b) They arise within $n \to \infty$ weak limits of various Markov chains whose state spaces are sets of size-$n$ combinatorial objects.
In other words, our chains provide "time-evolving" extensions of known "static" results asserting that Poisson-Galton-Watson distributions arise as weak limits within combinatorial models, so we start by reviewing that topic in section 2.

Our presentation is somewhat informal, for reasons explained in section 3.2. The chains of sections 3.4 and 3.5 are studied rigorously in Aldous-Pitman [**4**]. The chains in section 4 are novel and may be worthy of rigorous study in future.

## 2. Poisson-Galton-Watson distributions as weak limits

For $0 \leq \lambda \leq 1$ the *Poisson-Galton-Watson* distribution PGW($\lambda$) on finite trees is the distribution of the family tree of a Galton-Watson branching process with one progenitor and Poisson($\lambda$) offspring distribution. ¿From some viewpoints these are the "canonical" distributions on finite trees, and the related distribution $\mathrm{PGW}^{\infty}(1)$ on a set of infinite trees ("trees with one end") is the canonical distribution on that set. In this section we describe how these distributions arise as weak limits of distributions on size-$n$ structures. The material is essentially known and straightforward: the main novelty is the slick[1] notation we now introduce.

**2.1. Notation for distributions on trees.** In this paper, trees (and tree-components of forests) are rooted and unordered (i.e. in the usual parent-child interpretation of edges, we do not distinguish birth order of children). There are two standard conventions to specify what is meant by the set of all $n$-vertex trees. One can regard vertices as labeled by the integers $\{1, 2, \ldots, n\}$, with Cayley's formula asserting there exist $n^{n-1}$ distinct such trees. Or one can regard trees as unlabeled, giving a smaller number of "shapes" of trees, where two trees have the same shape iff there is a root-preserving bijection between vertices which preserves edges. For discussing tree-valued chains, the second convention is unsatisfactory because, in the joint distribution of the chain at times $s_1$ and $s_2$, one cannot determine whether a particular vertex at time $s_1$ is the "same vertex" as some vertex at time $s_2$. The first convention cannot be used directly because our trees have varying size. Instead we introduce the idea of labeling vertices by distinct *real numbers* $u \in (0, 1)$. Define $\mathbf{T}$ to be the set of finite rooted trees in which vertices $i$ (including the root) are labeled with distinct $u_i \in (0, 1)$. Call an element $\mathbf{t} \in \mathbf{T}$ a *u-labeled tree.* Write $|\mathbf{t}|$ for the number of vertices of $\mathbf{t}$. Figure 1 illustrates a typical element $\mathbf{t}$ of $\mathbf{T}$.

[1](for non-native English speakers) *slick* has two meanings. (i) clever, concise (ii) slippery, untrustworthy

Given a random rooted tree $\mathcal{T}$, we may regard it as a random element of $\mathbf{T}$ by assigning independent $U(0,1)$ (i.e. uniform on $(0,1)$) random labels to the vertices. The ultimate purpose is that, when we formalize tree-valued Markov chains as $\mathbf{T}$-valued chains $(\mathcal{T}_s, 0 \le s)$, we can look at joint distributions $(\mathcal{T}_{s_1}, \mathcal{T}_{s_2})$ and tell whether a specific vertex of $\mathcal{T}_{s_2}$ is also a vertex of $\mathcal{T}_{s_1}$. But it turns out that formalizing random trees as $u$-labeled trees leads to a very useful way of describing distributions on trees, at least in settings related to Poisson-Galton-Watson distributions. Let $f : \mathbf{T} \to [0,\infty)$ be a function such that $f(\mathbf{t})$ depends only on the shape of $\mathbf{t}$, not on the values $(u_i)$. Say a random tree $\mathcal{T}$ has *density* $f(\mathbf{t})$ if the chance that $\mathcal{T}$ is a tree with the same shape as $\mathbf{t}$ and with labels in $[u_i, u_i + du_i]$ , $1 \le i \le |\mathbf{t}|$, equals $f(\mathbf{t})\ du_1 \ldots du_{|\mathbf{t}|}$. This definition is difficult to assimilate at first; we hope that repeated use throughout the paper will make it seem less mysterious. Here is a "discrete" reformulation. Given a random finite rooted tree $\mathcal{T}$, introduce an artificial parameter $N$, label vertices of $\mathcal{T}$ by sampling without replacement from $\{1, 2, \ldots, N\}$ (with some unimportant convention for the case $|\mathcal{T}| > N$) and call the resulting random tree $\mathcal{T}_{(N)}$. For given $\mathbf{t} \in \mathbf{T}$ write $\mathbf{t}_{\mathrm{int}}$ for the tree obtained from $\mathbf{t}$ by an arbitrary relabeling of vertices as $\{1, 2, \ldots, |\mathbf{t}|\}$. Then the property

$$\forall \mathbf{t} \in \mathbf{T}: \quad P(\mathcal{T}_{(N)} = \mathbf{t}_{\mathrm{int}}) \sim N^{-|\mathbf{t}|} f(\mathbf{t}) \text{ as } N \to \infty \tag{2.1}$$

is equivalent to saying $\mathcal{T}$ has density $f(\cdot)$. Let us illustrate by calculating the density for a PGW($\lambda$) distributed tree $\mathcal{T}$. Take $\mathbf{t}$ as in Figure 1 and let $\mathbf{t}_{\mathrm{int}}$ be the relabeling of $\{u_1, \ldots, u_5\}$ as $\{1, 2, \ldots, 5\}$. Then

$$P(\mathcal{T}_{(N)} = \mathbf{t}_{\mathrm{int}}) = \frac{1}{N} \times \frac{\lambda^3 e^{-\lambda}}{3!}$$

$$\times \frac{3!}{(N-1)(N-2)(N-3)} \times (e^{-\lambda} \times \lambda e^{-\lambda} \times e^{-\lambda}) \times \frac{1}{N-4} \times e^{-\lambda}$$

where successive terms represent: chance root gets label 1, chance of 3 children, chance children labeled as $2, 3, 4$ in some order, chance children have $(0, 1, 0)$ grandchildren, chance grandchild gets label 5, chance it has 0 children. Thus $P(\mathcal{T}_{(N)} = \mathbf{t}_{\mathrm{int}}) = \frac{\lambda^4 e^{-5\lambda}}{N(N-1)\ldots(N-4)}$ and so this particular $\mathbf{t}$ has $f(\mathbf{t}) = \lambda^4 e^{-5\lambda}$. Applying the same argument to a general tree $\mathbf{t}$ we see that the PGW($\lambda$) density is

$$f_\lambda(\mathbf{t}) = \lambda^{(|\mathbf{t}|-1)} e^{-\lambda |\mathbf{t}|} \tag{2.2}$$

and in particular for $\lambda = 1$

$$f_1(\mathbf{t}) = e^{-|\mathbf{t}|}. \tag{2.3}$$

The focus of this paper is that this density notation (loosely reminiscent of statistical physics) turns out to be very useful for specifying or verifying transition rates for certain tree-valued chains (look ahead to section 3.1). It is not intended to facilitate explicit calculations, though as an exercise the reader might deduce from (2.2) the well-known fact that, for $\mathcal{T}$ with PGW($\lambda$) distribution, the size $|\mathcal{T}|$ has *Borel*($\lambda$) distribution

$$P(|\mathcal{T}| = i) = \frac{(\lambda i)^{i-1}}{i!}\, e^{-\lambda i},\ i \ge 1. \tag{2.4}$$

As abuse of notation, when a random tree $\mathcal{T}$ has density $f(\cdot)$ we sometimes write $P(\mathcal{T} = \mathbf{t})$ in place of $f(\mathbf{t})$.

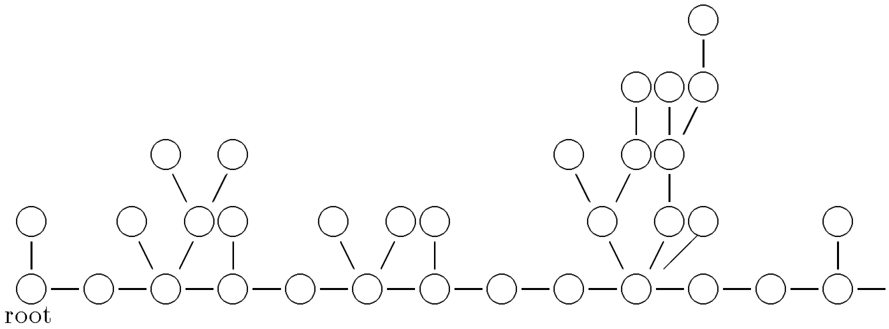


FIGURE 2. A sin-tree $\mathbf{t}$

**2.2. $\mathbf{PGW}^\infty(1)$.** Consider a locally finite rooted $u$-labeled tree $\mathbf{t}$ with an infinite number of vertices, which has a unique path (root $= u_0, u_1, u_2, \dots$) from the root to infinity. Call such a $\mathbf{t}$ a *sin*-tree (for **s**ingle **in**finite path; alternatively, a "tree with one end"). The set $\mathbf{T}^\infty$ of sin-trees may be identified as $\mathbf{T}\times\mathbf{T}\times\dots$, where $\mathbf{T}$ is the set of finite rooted trees, via the bijection $\mathbf{t}\leftrightarrow(\mathbf{t}_0,\mathbf{t}_1,\mathbf{t}_2,\dots)$ where the $(\mathbf{t}_i)$ are the trees rooted at $u_i$ obtained by deleting edges of the path $(u_0,u_1,u_2,\dots)$ in $\mathbf{t}$ (Figure 2). We can use this bijection to define a probability distribution PGW$^\infty(1)$ on $\mathbf{T}^\infty$ as the product distribution PGW(1)$\times$PGW(1)$\times\dots$.

The PGW$^\infty(1)$ distribution is in many ways the "canonical" probability distribution on $\mathbf{T}^\infty$. It arises in several $n\to\infty$ settings (see Lemmas 2.3 and 2.2 and section 2.7). We first give a technically useful equivalent description. The *height* of a vertex in a tree is its distance from the root; the height of a tree is the maximum vertex height. Let $\mathbf{T}^h$ be the set of finite rooted trees with height at most $h$. There is a natural *restriction map* $r_h:\mathbf{T}\to\mathbf{T}^h$, taking $\mathbf{t}$ to the subtree $\mathbf{t}^h$ consisting of vertices at height $\le h$, and there is a corresponding map $r_h:\mathbf{T}^\infty\to\mathbf{T}^h$, where in $r_h(\mathbf{t})$ the original "path towards infinity" is no longer distinguished.

LEMMA 2.1. *Let $\lambda\le 1$. Write $f^h_\lambda$ and $f^h_*$ for the densities of the probability distributions on $\mathbf{T}^h$ induced by $r_h$ from PGW($\lambda$) and from PGW$^\infty(1)$. Let $n(h,\mathbf{t})$ denote the number of vertices of $\mathbf{t}$ at height $h$. Then for $\mathbf{t}\in\mathbf{T}^h$*

$$f^h_\lambda(\mathbf{t}) = e^{\lambda n(h,\mathbf{t})}f_\lambda(\mathbf{t}) = \lambda^{|\mathbf{t}|-1}e^{\lambda(n(h,\mathbf{t})-|\mathbf{t}|)} \tag{2.5}$$

$$f^h_1(\mathbf{t}) = e^{n(h,\mathbf{t})-|\mathbf{t}|} \tag{2.6}$$

$$f^h_*(\mathbf{t}) = n(h,\mathbf{t})f^h_1(\mathbf{t}). \tag{2.7}$$

PROOF. Let $(u_1,\dots,u_m)$ be a subset of leaves of a tree $\mathbf{t}$, and write $A(\mathbf{t};u_1,\dots,u_m)$ for the set of trees which coincide with $\mathbf{t}$ except that vertices $u_i$ may have adjoined subtrees. Then

$$\frac{f_\lambda(\mathbf{t})}{\sum_{\mathbf{t}'\in A(\mathbf{t};u_1,\dots,u_m)}f_\lambda(\mathbf{t}')} = e^{-\lambda m}$$

because at each $u_i$ the PGW($\lambda$) process has chance $e^{-\lambda}$ to be a leaf (and the random trees are finite). By considering $\mathbf{t} \in \mathbf{T}^h$ and the set of all its vertices at height $h$,

$$\frac{f_\lambda(\mathbf{t})}{f_\lambda^h(\mathbf{t})} = e^{-\lambda n(h,\mathbf{t})}$$

which is (2.5). And (2.6) is just the $\lambda = 1$ case of (2.5). To verify (2.7), fix a height-$h$ vertex $u_h$ of $\mathbf{t} \in \mathbf{T}^h$, and write $\mathbf{t}^{(u_h)}$ for the tree with $u_h$ as distinguished vertex. Consider the path (root $= u_0, u_1, u_2, \ldots, u_h$), and write $(\mathbf{t}_0, \mathbf{t}_1, \ldots, \mathbf{t}_h)$ for the trees obtained by breaking edges of that path. ¿From the definition of PGW$^\infty(1)$ and (2.6), the chance that the restriction of PGW$^\infty(1)$, with the height-$h$ vertex on the original path to infinity distinguished, is $\mathbf{t}^{(u_h)}$, equals

$$\prod_{i=0}^{h} e^{n(h-i,\mathbf{t}_i)-|\mathbf{t}_i|} = e^{n(h,\mathbf{t})-|\mathbf{t}|}.$$

So (2.7) follows by undistinguishing $u_h$. □

In words, (2.6,2.7) say that the restriction of PGW$^\infty(1)$ to the first $h$ generations can be obtained from the restriction of PGW(1) by size-biasing according to the number of generation-$h$ individuals. Lemma 2.1 and extensions for general Galton-Watson trees go back to Kesten [**12**]. More general such size-biasing relations have recently been emphasized by Lyons et al [**14**] as systematic ways of deriving classical branching process asymptotics. We shall soon see how, in our setting, (2.5 - 2.7) are similarly effective in yielding simple proofs of asymptotics.

**2.3. Weak convergence.** On the countable set $\mathcal{T}_{\mathrm{unlabeled}}$ of all rooted finite unlabeled trees, convergence in distribution $\mathcal{T}_n \xrightarrow{d} \mathcal{T}$ is just pointwise convergence

$$P(\mathcal{T}_n = \mathbf{t}_{\mathrm{unlabeled}}) \to P(\mathcal{T} = \mathbf{t}_{\mathrm{unlabeled}})\ \forall \mathbf{t}_{\mathrm{unlabeled}} \in \mathcal{T}_{\mathrm{unlabeled}}. \tag{2.8}$$

Our convention of creating $u$-labeled trees by assigning i.i.d. $U(0,1)$ labels to vertices makes no difference to weak convergence: we can interpret $\mathcal{T}_n \xrightarrow{d} \mathcal{T}$ for $u$-labeled trees as (2.8) for unlabeled trees or equivalently in terms of densities as

$$P(\mathcal{T}_n = \mathbf{t}) \to P(\mathcal{T} = \mathbf{t})\ \forall \mathbf{t} \in \mathbf{T}.$$

To handle sin-trees, recall the restriction map $r_h : \mathbf{T} \cup \mathbf{T}^\infty \to \mathbf{T}^h$, and define $\mathcal{T}_n \xrightarrow{d} \mathcal{T}$ for finite or sin-trees via

$$r_h(\mathcal{T}_n) \xrightarrow{d} r_h(\mathcal{T}) \text{ for all } h < \infty. \tag{2.9}$$

Informally, this is "local" convergence.

The next three sections describe three fundamental examples. We write out proofs to illustrate our "density" notation.

**2.4. PGW($\lambda$) as a limit within $G(n, \lambda/n)$.** In the random graph $G(n, \lambda/n)$ there are $n$ vertices, and each of the $\binom{n}{2}$ possible edges is present independently with probability $\lambda/n$. The next result is implicit in much of classical random graph theory [**7**] and explicit (in different language) in [**1**].

LEMMA 2.2. *Let $\mathcal{G}_\lambda^n$ be the component of $G(n, \lambda/n)$ containing vertex 1, considered as a random rooted graph. Then $\mathcal{G}_\lambda^n \xrightarrow{d}$ PGW($\lambda$) as $n \to \infty$.*

Proof. We claim that $\mathcal{G}^n_\lambda$ has density (on $\mathbf{T}$)

$$\frac{n(n-1)\ldots(n-|\mathbf{t}|+1)}{n^{|\mathbf{t}|}}\ \lambda^{|\mathbf{t}|-1}\ \left(1-\frac{\lambda}{n}\right)^{Q(n,|\mathbf{t}|)}$$

where

$$Q(n,\mathbf{t}) = |\mathbf{t}|(n-|\mathbf{t}|)\ +\ \tfrac{1}{2}|\mathbf{t}|(|\mathbf{t}|-1)\ -\ (|\mathbf{t}|-1).$$

This implies that as $n\to\infty$ for fixed $\mathbf{t}$,

$$P(\mathcal{G}^n_\lambda = \mathbf{t}) \to \lambda^{|\mathbf{t}|-1}e^{-\lambda|\mathbf{t}|} = f_\lambda(\mathbf{t})$$

establishing the lemma. To derive the formula, consider the chance that the $u$-labeling of $\mathcal{G}^n_\lambda$ gives a graph which contains the tree $\mathbf{t}$ of Figure 1, with labels in $[u_i, u_i+du_i]$. The chance is

$$du_1 \times (n-1)(n-2)(n-3)(\lambda/n)^3 du_2du_3du_4 \times (n-4)(\lambda/n)du_5$$

where the first term is the chance the root gets label in $[u_1, u_1+du_1]$, the second is the chance of 3 edges to vertices having labels in $[u_i, u_i+du_i]$, $i=2,3,4$, and the third is the chance that $u_3$ has an edge to some vertex given label in $[u_5, u_5+du_5]$. Applying this argument for general $\mathbf{t}$, the chance $\mathcal{G}^n_\lambda$ contains $\mathbf{t}$ is

$$\frac{n(n-1)\ldots(n-|\mathbf{t}|+1)}{n^{|\mathbf{t}|}}\ \lambda^{|\mathbf{t}|-1}\ du_1du_2\ldots du_{|\mathbf{t}|}.$$

In order for $\mathcal{G}^n_\lambda$ be equal $\mathbf{t}$, there must be no other edges, and $Q(n,\mathbf{t})$ counts the number of forbidden edges. □

**2.5. PGW$^\infty(1)$ as a limit within the Wright-Fisher model.** In the Wright-Fisher model (Ewens [**9**]) each generation of a population has $n$ female individuals, and each individual in generation $g$ (for $-\infty < g < \infty$) is the daughter of a uniform random mother in generation $g-1$. Thus if we pick an individual in generation 0 as the root, then this individual and her descendants (indicated by $\bullet$ in the Figure 3) form a random tree $\mathcal{T}_n$; the root individual also has an infinite line of descent through generations $-1,-2,\ldots$, and the family tree linking all individuals (in generations $-\infty$ through $n$, say) defines a sin-tree $\mathcal{T}^*_n$. Near-relatives of the root are indicated by $\otimes$ in Figure 3.

Because each individual has Binomial$(n, 1/n)$ children, the following lemma is intuitively clear.

Lemma 2.3. *$\mathcal{T}_n \xrightarrow{d} PGW(1)$ and $\mathcal{T}^*_n \xrightarrow{d} PGW^\infty(1)$.*

Arguing as in Lemma 2.2, it is straightforward to state the density of $\mathcal{T}_n$ after $u$-labeling as

$$P(\mathcal{T}_n = \mathbf{t}) = \prod_{i\geq 1} b(n, n(i,\mathbf{t}))\left(1-\tfrac{n(i-1,\mathbf{t})}{n}\right)^{n-n(i,\mathbf{t})}$$

where

$$b(n,m) = \tfrac{n(n-1)\ldots(n-m+1)}{n^m}, \qquad b(n,0)=1.$$

Here the term $b(n, n(i,\mathbf{t}))$ reflects existence of edges in $\mathbf{t}$ from generation $i$ to generation $i-1$, and the term $(1-\frac{n(i-1,\mathbf{t})}{n})^{n-n(i,\mathbf{t})}$ reflects non-existence of other edges. Then as $n\to\infty$

$$P(\mathcal{T}_n = \mathbf{t}) \to \exp(-\sum_{i\geq 1} n(i-1,\mathbf{t})) = e^{-|\mathbf{t}|} = f_1(\mathbf{t}).$$

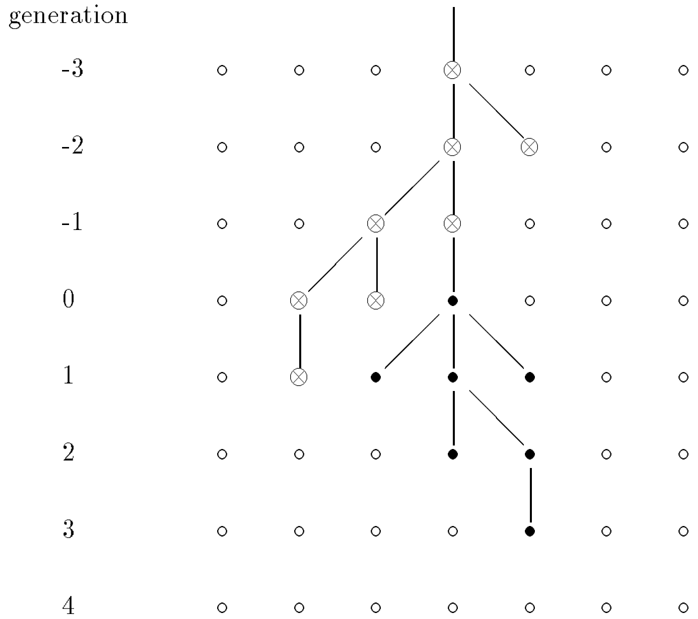


FIGURE 3. The Wright-Fisher model

The argument for $\mathcal{T}_n^* \xrightarrow{d} \mathrm{PGW}^\infty(1)$ is similar.

**2.6. $\mathbf{PGW}^\infty(1)$ as a limit within the uniform random $n$-tree.** While the previous weak convergence results (Lemmas 2.2 and 2.3) were intuitively obvious to probabilists, the next may not be. It is due to Grimmett [**10**], who used it to explain earlier combinatorial results. Cayley's formula says there are $n^{n-2}$ trees on vertex set $\{1, 2, \ldots, n\}$ rooted at vertex 1. Let $\mathcal{S}^n$ be a uniform random such tree.

LEMMA 2.4. *$\mathcal{S}^n \xrightarrow{d} PGW^\infty(1)$.*

PROOF. As in previous cases, the proof rests on an explicit formula. Fix $h$ and $\mathbf{t} \in \mathbf{T}^h$ with $k = n(h, \mathbf{t}) \geq 1$. We claim that, regarding $r_h(\mathcal{S}^n)$ as a $u$-labeled tree, it has density

$$\frac{(n-1)(n-2)\ldots(n-|\mathbf{t}|+1)k(n-|\mathbf{t}|+k)^{n-|\mathbf{t}|-1}}{n^{n-2}}. \tag{2.10}$$

Then as $n \to \infty$, this density tends to $n(h, \mathbf{t})\exp(n(h, \mathbf{t}) - |\mathbf{t}|)$, and this limit is the density $f_h^*(\mathbf{t})$ of the restriction of $\mathrm{PGW}^\infty(1)$ by (2.7).

To derive (2.10), recall another form of Cayley's formula. The number of forests on vertices $\{1, 2, \ldots, m\}$ with exactly $k$ tree-components rooted at $k$ specified vertices equals $c(m, k) = km^{m-k-1}$. Write $A_n(\mathbf{t})$ for the set of trees obtainable by giving the vertices of $\mathbf{t}$ labels from $\{1, 2, \ldots, n\}$. Then the number of trees on $\{1, 2, \ldots, n\}$ rooted at 1 whose restriction to height $h$ is in $A_n(\mathbf{t})$ equals

$$(n-1)(n-2)\ldots(n-|\mathbf{t}|+1) \times c(n-|\mathbf{t}|+k, k)$$

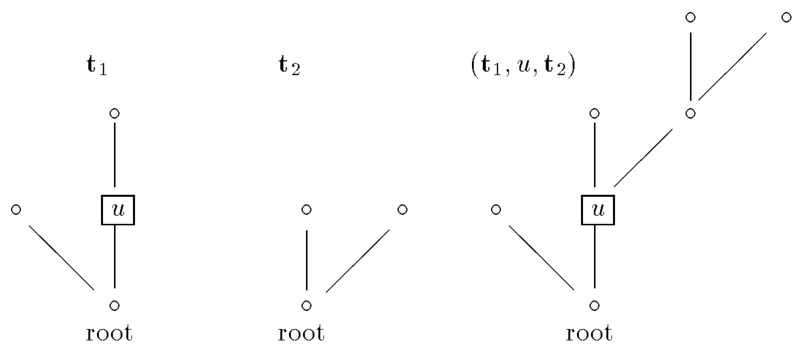


FIGURE 4. Joining

where the first term counts labelings of $\mathbf{t}$ and the second counts extensions above height $h$. So, before $u$-labeling,

$$P(r_h(\mathcal{S}_n) \in A_n(\mathbf{t})) = \text{ formula in (2.10)}.$$

But $u$-labeling a tree in $A_n(\mathbf{t})$ has chance $du_1 \ldots du_{|\mathbf{t}|}$ to assign labels in $[u_i, u_i{+}du_i]$ for all $i$, and so we get the same formula for the density after $u$-labeling. □

**2.7. Other limit representations of $\mathbf{PGW}^\infty(1)$.** Lemma 2.4 seems less mysterious if one knows that the uniform labeled $n$-tree is distributed as the PGW(1) tree conditioned on having size equal to $n$. It is well known (see [**4**] Lemma 23) that several different formalizations of "PGW(1) conditioned to be large" lead to the $\mathrm{PGW}^\infty(1)$ limit.

In a different direction, the uniform random tree $\mathcal{S}^n$ in Lemma 2.4 may be viewed as a special case (the complete graph) of the uniform random spanning tree of an given size-$n$ graph. Lemma 2.4 can be extended to spanning trees of certain other graphs [**2, 18**].

## 3. Markov chains with Poisson-Galton-Watson marginals

**3.1. Notation for tree-valued chains.** Our tree-processes evolve by "joining" and "pruning" operations. If $u$ is a vertex-label of $\mathbf{t}_1$, then let $(\mathbf{t}_1, u, \mathbf{t}_2)$ denote the tree obtained by joining $\mathbf{t}_1$ and $\mathbf{t}_2$ via a new edge from $u$ to the root of $\mathbf{t}_2$, and let the root of the new tree be the root of $\mathbf{t}_1$ (Figure 4). A transition $\mathbf{t}_1 \to (\mathbf{t}_1, u, \mathbf{t}_2)$ is a "joining" operation; the "pruning" operation of cutting an edge and retaining the subtree containing the root is a transition of the reverse form $(\mathbf{t}_1, u, \mathbf{t}_2) \to \mathbf{t}_1$.

We illustrate this notation with a simple example (somewhat different in spirit from our subsequent examples). Here is a Metropolis scheme for constructing a tree-valued discrete-time chain whose stationary distribution is (by design) PGW($\lambda$), for fixed $\lambda \le 1$, and where the only transitions allowed are to add or delete a leaf. In words, a step of the chain is

> Toss a fair coin. If Heads, pick a uniform random vertex $v$, and with chance $\lambda e^{-\lambda}$ append a leaf to $v$. If Tails, pick (if possible) a uniform random vertex $v$ other than the root, and if $v$ is a leaf, delete it.

Symbolically, writing $\bullet$ for the single-vertex tree, the transition probabilities are

$$\mathbf{t} \to (\mathbf{t}, u, \bullet) : \text{ probability } \frac{\lambda e^{-\lambda}}{2|\mathbf{t}|}$$

$$(\mathbf{t}, u, \bullet) \to \mathbf{t} : \text{ probability } \frac{1}{2|\mathbf{t}|}$$

for each vertex $u$ of $\mathbf{t}$. To check that the stationary distribution is indeed PGW($\lambda$) we need to confirm the detailed balance equation

$$f_\lambda(\mathbf{t}) \times \frac{\lambda e^{-\lambda}}{2|\mathbf{t}|} = f_\lambda(\mathbf{t}, u, \bullet) \times \frac{1}{2|\mathbf{t}|},$$

and indeed by (2.2) both sides equal $\lambda^{|\mathbf{t}|} e^{-\lambda|\mathbf{t}|-\lambda}/(2|\mathbf{t}|)$.

**3.2. Caveat.** In following sections we describe continuous-time tree-valued Markov chains by specifying their transition rates. Conceptually, the rates can be described in words, not depending on any choice of notation for trees; but our density notation is useful for discussing marginal distributions.

There is a technical issue – not pursued in this paper – of showing that rates do indeed specify a unique chain. In Aldous-Pitman [**4**] we give a rigorous constructive treatment of two of the chains (those in sections 3.4 and 3.5) without explicitly using rates or the density notation. It is not hard to make explicit constructions of the other chains, except for the "exotic" chain in section 4.3. But the present paper is intended as a more informal and intuitive presentation.

The definitions of the chains are mostly motivated as limits of finite-state chains. Writing out rigorous proofs of these weak limit results would be technically complicated, so we shall present them as "weak limit assertions" with only brief explanations. In only two places (sections 3.7 and 4.3) do there arise issues of substance rather than technique, so we write those as explicit conjectures.

We think of our chains as a "family" because their transition rates involve qualitatively similar pruning and joining rules. Note that some of the chains are (time)-homogeneous while some are non-homogeneous; and some are (time)-reversible while others are non-reversible.

**3.3. The stationary PGW($\lambda$) chain.** Fix $0 < \lambda \leq 1$. Consider the homogeneous $\mathbf{T}$-valued chain $(\mathcal{H}_s, s \geq 0)$ with transition rates

$$\mathbf{t}_1 \to (\mathbf{t}_1, u, \mathbf{t}_2) \quad : \quad \text{rate } \lambda f_\lambda(\mathbf{t}_2) \tag{3.1}$$

$$(\mathbf{t}_1, u, \mathbf{t}_2) \to \mathbf{t}_1 \quad : \quad \text{rate } 1 \tag{3.2}$$

(these rates apply to each possible transition, i.e. each possible combination of $\mathbf{t}_1$, $u$ and $\mathbf{t}_2$). In words,

> Each edge is pruned away at rate 1; at each vertex, at rate $\lambda$ there is adjoined a PGW($\lambda$)-distributed tree.

We claim this chain is reversible with stationary distribution PGW($\lambda$). We need only verify the detailed balance equation

$$f_\lambda(\mathbf{t}_1) \times \lambda f_\lambda(\mathbf{t}_2) = f_\lambda(\mathbf{t}_1, u, \mathbf{t}_2) \times 1 \tag{3.3}$$

and indeed, by (2.2) both sides equal $\lambda^{|\mathbf{t}_1|+|\mathbf{t}_2|-1} e^{-\lambda(|\mathbf{t}_1|+|\mathbf{t}_2|)}$.

As motivation for the definition, regard the random graph $G(n, \lambda/n)$ as the stationary distribution of the chain whose state space is the set of graphs on

$\{1, 2, \ldots, n\}$, in which each of the $\binom{n}{2}$ possible edges is born at rate $\lambda/n$ and dies at rate $1 - \lambda/n$. Restrict attention to the component containing vertex 1, and call this process $(\mathcal{H}^n_s, 0 \le s)$. Lemma 2.2 showed that $P(\mathcal{H}^n_s = \mathbf{t}) \to f_\lambda(\mathbf{t})$ as $n \to \infty$ for fixed $s$.

Weak Limit Assertion 3.1. $(\mathcal{H}^n_s, 0 \le s) \xrightarrow{d} (\mathcal{H}_s, 0 \le s)$, where the limit is the stationary chain with transition rates (3.1,3.2).

In brief: each edge of $\mathcal{H}^n_s$ dies at rate $1 - \lambda/n$, which is asymptotic to (3.2). Since there are $(1 - o(1))n$ vertices outside $\mathcal{H}^n_s$, at each vertex of $\mathcal{H}^n_s$ a new edge to some vertex $v$ is created at rate $\frac{\lambda}{n}(1 - o(1))n \to \lambda$, and the distribution of the joined component containing $v$ will be approximately the distribution of $\mathcal{H}^n_s$, i.e. with density $f_\lambda$.

Curiously, this chain $(\mathcal{H}_s)$ has not been studied, though (regarding the random graph as a mean-field model of percolation) it is analogous to the recently-studied topic of *flickering percolation* [**11**]. For $\lambda < 1$, using the fact that the distribution (2.4) of $|\mathcal{H}_s|$ has geometrically decreasing tail it is not hard to show that

$$T_a := \inf\{s : |\mathcal{H}_s| \ge a\} \xrightarrow{p} \infty \text{ as } a \to \infty$$

so that the chain stays in the state space $\mathbf{T}$ at all times. For $\lambda = 1$ this is more delicate. In the first draft, and at the Workshop, I asserted that $T_a < \infty$ a.s. when $\lambda = 1$. But this is false. More careful analysis, which may be presented elsewhere, shows that in fact $T_a$ goes to infinity at rate $\log a$.

**3.4. The PGW growth chain.** Define $(\mathcal{G}_\lambda, 0 \le \lambda \le 1)$ to be the non-homogeneous $\mathbf{T}$-valued Markov chain with transition rate at time $\lambda$

$$\mathbf{t}_1 \to (\mathbf{t}_1, u, \mathbf{t}_2) : \text{ rate } f_\lambda(\mathbf{t}_2) \tag{3.4}$$

where $f_\lambda(\cdot)$ is the PGW($\lambda$) distribution (2.2) and where $\mathcal{G}_0 = \bullet$ is the single-vertex tree. In words,

> At each vertex, at rate 1 there is adjoined a PGW($\lambda$)-distributed tree.

It is intuitively clear (and proved in Aldous [**1**] in different language) that this process is the weak limit of the component containing vertex 1 in the usual random graph process $(G(n, \lambda/n), \lambda \ge 0)$. In particular, the marginal distribution at time $\lambda$ is PGW($\lambda$). We may verify this intrinsically by verifying the forwards equation

$$\frac{d}{d\lambda} f_\lambda(\mathbf{t}) = \sum_{(\mathbf{t}_1, u, \mathbf{t}_2) = \mathbf{t}} f_\lambda(\mathbf{t}_1) f_\lambda(\mathbf{t}_2) - f_\lambda(\mathbf{t}) \sum_{u \in \mathbf{t}, \mathbf{t}_2} f_\lambda(\mathbf{t}_2).$$

The right side becomes

$$\lambda^{-1} f_\lambda(\mathbf{t})(|\mathbf{t}| - 1) \; - \; f_\lambda(\mathbf{t})|\mathbf{t}|$$

and this is indeed the derivative of (2.2).

This process $(\mathcal{G}_\lambda)$ is studied in detail in Aldous-Pitman [**4**]. In particular, one can extend the time-interval from $0 \le \lambda \le 1$ to $0 \le \lambda < A$, where $A > 1$ is the *ascension time*, indicating (in the random graph pre-limit) the time at which the component under study joins the giant component. Qualitatively, $\mathcal{G}_{A-}$ is a finite tree; at time $A$ it acquires an edge linking it to an infinite tree. The distribution of $A$ is specified by (3.9) below.

**3.5. The process $(\mathcal{G}^*_s, 0 \le s \le 1)$.** This process, and its extension to general Galton-Watson trees, is also studied rigorously in Aldous-Pitman [**4**], which contains elaborations of the results stated here (except for the "rate" result in Lemma 3.2(b)).

Let $\mathcal{G}^*_1$ have PGW$^\infty(1)$ distribution. Attach independent $U(0,1)$ random variables $\xi_e$ to its edges; consider the subgraph consisting of edges $e$ with $\xi_e \le s$, and let $\mathcal{G}^*_s$ be the component of that subgraph containing the root. So $\mathcal{G}^*_0 = \bullet$, the single-vertex tree. Clearly the process $(\mathcal{G}^*_s, 0 \le s \le 1)$ is Markov when $s$ is decreasing, so it is automatically (non-homogeneous) Markov when $s$ is increasing. The first assertion of the lemma is that for fixed $s$ the distribution of $\mathcal{G}^*_s$ is the PGW$(s)$ distribution, size-biased by total population size.

LEMMA 3.2. *(a) For $0 \le s < 1$, the distribution $g^*_s$ of $\mathcal{G}^*_s$ is*

$$g^*_s(\mathbf{t}) = (1-s)|\mathbf{t}|f_s(\mathbf{t}).$$

*(b) $(\mathcal{G}^*_s)$ is the Markov chain with transition rates at time $s$*

$$\mathbf{t}_1 \to (\mathbf{t}_1, u, \mathbf{t}_2) : \text{ rate } \frac{|\mathbf{t}_1| + |\mathbf{t}_2|}{|\mathbf{t}_1|} f_s(\mathbf{t}_2) \tag{3.5}$$

*for each vertex $u$ of $\mathbf{t}_1$.*

PROOF. (a) In constructing $\mathcal{G}^*_s$ from $\mathcal{G}^*_1$ we may distinguish a vertex $u_h$, the first vertex on the infinite path for which $\xi_{(u_h, u_{h+1})} > s$. Fix $\mathbf{t}$ and a height-$h$ vertex $u_h$ of $\mathbf{t}$. Consider the path (root $= u_0, u_1, u_2, \ldots, u_h$), and write $(\mathbf{t}_0, \mathbf{t}_1, \ldots, \mathbf{t}_h)$ for the trees obtained by breaking edges of that path. The chance that $\mathcal{G}^*_s = \mathbf{t}$ and $u_h$ is the distinguished vertex equals $s^h(1-s)f_s(\mathbf{t}_0)f_s(\mathbf{t}_1)\ldots f_s(\mathbf{t}_h)$. But this equals

$$s^h(1-s)\prod_{i=0}^{h} s^{|\mathbf{t}_i|-1}e^{-s|\mathbf{t}_i|} = (1-s)s^{|\mathbf{t}|-1}e^{-|\mathbf{t}|} = (1-s)f_s(\mathbf{t}).$$

So (a) follows by undistinguishing $u_h$.

For (b), as $s$ decreases the transition rate at time $s$ is (from the "pruning at uniform random times" construction)

$$(\mathbf{t}_1, u, \mathbf{t}_2) \to \mathbf{t}_1 : \text{ rate } 1/s. \tag{3.6}$$

So the forwards transition rate is $1/s \times g^*_s(\mathbf{t}_1, u, \mathbf{t}_2)/g^*_s(\mathbf{t}_1)$, and applying the formula in (a) this rate becomes

$$\frac{|\mathbf{t}_1| + |\mathbf{t}_2|}{|\mathbf{t}_1|} s^{-1} \frac{f_s(\mathbf{t}_1, u, \mathbf{t}_2)}{f_s(\mathbf{t}_1)}$$

which reduces to the formula in (b). □

Several relationships between $(\mathcal{G}^*_t)$ and $(\mathcal{G}_t)$ are explored in [**4**]. The first is that in reversed time they evolve in the same way. At ([**4**] equation (85)) this is formalized as: for fixed $\mathbf{t} \in \mathbf{T}$, $0 < \lambda < \infty$ and $0 < s < 1$,

$$\text{dist}(\mathcal{G}_{t\lambda}, 0 \le t \le 1 | \mathcal{G}_\lambda = \mathbf{t}) = \text{dist}(\mathcal{G}^*_{ts}, 0 \le t \le 1 | \mathcal{G}^*_s = \mathbf{t}). \tag{3.7}$$

Here is a reformulation in terms of transition rates.

LEMMA 3.3. *Each of the process $(\mathcal{G}_s, 0 \le s \le 1)$ and $(\mathcal{G}^*_s, 0 \le s \le 1)$ evolves, as $s$ decreases, as the non-homogeneous Markov chain with transition rates at time $s$*

$$(\mathbf{t}_1, u, \mathbf{t}_2) \to \mathbf{t}_1 : \text{ rate } 1/s. \tag{3.8}$$

PROOF. For $(\mathcal{G}^*_s)$ this is (3.6). For $(\mathcal{G}_s)$ it is intuitively clear from the random graph asymptotics: here is a formal verification. Since the forwards transition rates for $\mathcal{G}_s$ are (3.4)

$$\mathbf{t}_1 \to (\mathbf{t}_1, u, \mathbf{t}_2) : \text{ rate } f_s(\mathbf{t}_2)$$

and the marginal distribution is $f_s$, the reversed transition rate is

$$\frac{f_s(\mathbf{t}_1)f_s(\mathbf{t}_2)}{f_s(\mathbf{t}_1, u, \mathbf{t}_2)} = 1/s.$$

$\square$

A second relationship is that we can regard $(\mathcal{G}^*_s, 0 \le s \le 1)$ as having the conditional distribution of $(\mathcal{G}_s, 0 \le s \le 1)$ given $|\mathcal{G}_1| = \infty$. See [**4**] sections 3.3 and 4.4 for this and the analogous result for general Galton-Watson processes. Here is a third, more surprising relationship.

PROPOSITION 3.4 ([**4**] Proposition 26).

$$(\mathcal{G}_\lambda, 0 \le \lambda < A) \stackrel{d}{=} \left(\mathcal{G}^*_{\lambda U}, 0 \le \lambda < \frac{-\log U}{1-U}\right)$$

*where $U$ is $U(0,1)$, independent of $(\mathcal{G}^*_s, 0 \le s \le 1)$.*

So in particular the distribution of $A$ is

$$A \stackrel{d}{=} \frac{-\log U}{1-U}. \tag{3.9}$$

**3.6. Stochastic models of coalescence.** We digress to give an interpretation of two previous models and motivation for the next model. Given $n$ and a kernel $K(i,j) \ge 0$, the *Marcus-Lushnikov* process takes values in the set of forests on $\{1, 2, \dots, n\}$ and evolves according to the rule: trees $\mathbf{t}_1$ and $\mathbf{t}_2$ merge at rate $K(|\mathbf{t}_1|, |\mathbf{t}_2|)/n$, by adding a linking edge between uniform random vertices of $\mathbf{t}_1$ and $\mathbf{t}_2$. Initially there are 0 edges. If we suppose existence of the deterministic limit as $n \to \infty$

$$n^{-1}(\text{number of size-}i \text{ trees at time } s) \stackrel{p}{\to} c(i,s)$$

then these limits should satisfy the *Smoluchowski coagulation equation*

$$\frac{d}{ds}c(i,t) = \tfrac{1}{2}\sum_{j=1}^{i-1} K(j, i-j)c(j,s)c(i-j,s) \;-\; \sum_{j\ge 1} K(i,j)c(i,s)c(j,s). \tag{3.10}$$

See [**3**] for a survey of such models of stochastic coalescence. Now consider the process describing the tree component containing vertex 1 in the Marcus-Lushnikov process, considered as a rooted tree. As $n \to \infty$ we expect a limit tree-valued process $(\mathcal{T}_s, s \ge 0)$ such that

$$P(|\mathcal{T}_s| = i) = ic(i,s) \tag{3.11}$$

at least for $s$ smaller than some "gelation time". Now change viewpoints, and consider seeking to define the candidate limit process directly. Combining the argument for (3.10) with the size-biasing relationship (3.11), one can see that the appropriate formula is as follows. Define $(\mathcal{T}_s)$ to be the $\mathbf{T}$-valued chain with $\mathcal{T}_0 = \bullet$ and with transition rate at time $s$

$$\mathbf{t}_1 \to (\mathbf{t}_1, u, \mathbf{t}_2) : \quad \text{rate } \frac{K(|\mathbf{t}_1|, |\mathbf{t}_2|)}{|\mathbf{t}_1|\;|\mathbf{t}_2|}\, g_s(\mathbf{t}_2) \tag{3.12}$$

where $g_s$ is the density of $\mathcal{T}_s$. Note (3.12) has the unusual feature that the formula for transition rates involves the marginal density of the process itself. Solving (3.12) for $g_s$ is essentially equivalent to solving the Smoluchowski coagulation equation (3.10) for $c(i,s)$. But (3.4) says that the densities $(f_\lambda, 0 \leq \lambda \leq 1)$ of the PGWgrowth process $(\mathcal{G}_\lambda, 0 \leq \lambda \leq 1)$ solve (3.12) for $K(i,j) = ij$. And (3.5) says that the densities $(g^*_s, 0 \leq s \leq 1)$ of the process $(\mathcal{G}^*_s, 0 \leq s \leq 1)$ solve (3.12) for $K(i,j) = i+j$, except with an extra factor $(1-s)^{-1}$ on the right side. So after making the deterministic time-change to $(g^*_{1-e^{-s}},\ 0 \leq s < \infty)$ we do get a solution of (3.12) for $K(i,j) = i+j$.

Returning to the case $K(i,j) = ij$, for $\lambda > 1$ the PGW($\lambda$) distribution has a sub-probability density $f_\lambda$ on the set $\mathbf{T}$ of finite trees. But for $\lambda > 1$ , $(f_\lambda)$ is no longer a solution of (3.12) (with $\mathbf{t}_i$ finite trees), because in the random graph pre-limit there is coalescence between finite trees and the giant component which has no counterpart in (3.12). In the next section we construct a process $(\mathcal{G}^0_s)$ whose densities $(h_s)$ do solve (3.12) on $\mathbf{T}$: see (3.15). For the weak limit interpretation see Conjecture 3.6.

**3.7. The process $(\mathcal{G}^0_s, 0 \leq s < Z)$.** For $0 \leq s < \infty$ let $h_s$ be the density on $\mathbf{T}$ given by

$$\begin{aligned} h_s &= f_s,\ 0 \leq s \leq 1 \\ &= s^{-1} f_1,\ s > 1 \end{aligned}$$

where $f_s$ is the PGW($s$) density, so that $h_s$ is a sub-probability density for $s > 1$.

Proposition 3.5. *Let $V$ have $U(0,1)$ distribution, independent of $(\mathcal{G}^*_s, 0 \leq s \leq 1)$. Define a $\mathbf{T}$-valued process $(\mathcal{G}^0_s, 0 \leq s < Z)$ by*

$$\mathcal{G}^0_s = \mathcal{G}^*_{sV},\ 0 \leq s < Z = 1/V. \tag{3.13}$$

*Then*

$$P(\mathcal{G}^0_s = \mathbf{t}, Z > s) = h_s(\mathbf{t}). \tag{3.14}$$

*And $(\mathcal{G}^0_s)$ is the non-homogeneous Markov chain with transition rates at time $s$*

$$\mathbf{t}_1 \to (\mathbf{t}_1, u, \mathbf{t}_2) : \ \textit{rate } h_s(\mathbf{t}_2)\ ,\ \textit{for } \mathbf{t}_1, \mathbf{t}_2 \in \mathbf{T} \tag{3.15}$$

*stopped at $Z = \inf\{s : |\mathcal{G}^0_s| = \infty\}$.*

Note that the limit $\mathcal{G}^0_{Z-} \stackrel{d}{=} \mathcal{G}^*_1$ has PGW$^\infty(1)$ distribution. And note that by Proposition 3.4

$$\mathcal{G}^*_V \stackrel{d}{=} \mathcal{G}_1 \text{ has density } f_1 \tag{3.16}$$

and that by definition (3.13)

$$P(Z > z) = 1/z,\ z \geq 1.$$

Proof. For fixed $s \leq 1$,

$$P(\mathcal{G}^0_s = \mathbf{t}, Z > s) = P(\mathcal{G}^*_{sV} = \mathbf{t}) = f_s(\mathbf{t})$$

by Proposition 3.4. For fixed $s \geq 1$,

$$\begin{aligned} P(\mathcal{G}^0_s = \mathbf{t}, Z > s) &= P(\mathcal{G}^*_{sV} = \mathbf{t}, V < 1/s) = \\ &= s^{-1} P(\mathcal{G}^*_V = \mathbf{t}) \\ &\quad \text{because } P(V < 1/s) = 1/s \text{ and } \mathrm{dist}(sV | V < 1/s) = \mathrm{dist}(V) \\ &= s^{-1} f_1(\mathbf{t}) \text{ by (3.16).} \end{aligned}$$

¿From (3.7,3.8) the reversed transition rates are

$$(\mathbf{t}_1, u, \mathbf{t}_2) \to \mathbf{t}_1 : \text{ rate } 1/s$$

and so the forwards transition rate $\mathbf{t}_1 \to (\mathbf{t}_1, u, \mathbf{t}_2)$ is

$$s^{-1} h_s(\mathbf{t}_1, u, \mathbf{t}_2)/h_s(\mathbf{t}_1).$$

For $s \leq 1$ this is

$$s^{-1} f_s(\mathbf{t}_1, u, \mathbf{t}_2)/f_s(\mathbf{t}_1) = f_s(\mathbf{t}_2) = h_s(\mathbf{t}_2),$$

while for $s > 1$ it is

$$s^{-1} f_1(\mathbf{t}_1, u, \mathbf{t}_2)/f_1(\mathbf{t}_1) = s^{-1} f_1(\mathbf{t}_2) = h_s(\mathbf{t}_2)$$

establishing (3.15). □

Comparing Propositions 3.4 and 3.5 leads to an interesting paradox. Both processes $(\mathcal{G}_s)$ and $(\mathcal{G}^0_s)$ can be constructed as $(\mathcal{G}^*_{sV})$, stopped at times $\frac{-\log V}{1-V}$ and $1/V$ respectively. So we can say

> the process $(\mathcal{G}_s, 0 \leq s < A)$ is distributed as the process $(\mathcal{G}^0_s, 0 \leq s \leq Z)$ stopped at the random time $A = \frac{\log Z}{1-Z^{-1}}$.

But the transition rates (for $s > 1$) for the two processes are different. The resolution of the paradox is that, for a "stopped" process to have the same transition rates (before stopping, of course), the stopping time must be a (randomized) *stopping time* in the technical sense, and $A$ here is not such a stopping time.

Motivation for study of $(\mathcal{G}^0_s)$ comes from the following limit considerations. Modify the random graph process $(G(n, s/n), s \geq 0)$ by specifying that components whose size exceeds a threshold $\phi(n)$ are forbidden to coalesce with other components (imagine a physical process of aggregation in a liquid, where sufficiently large aggregates precipitate out). Thus when the process is run to time infinity, instead of getting one component of size $n$ we get a set of components, all of sizes between $\phi(n)$ and $2\phi(n)$ (except perhaps one smaller component).

Conjecture 3.6. *Let $\phi(n)$ be a function such that $\phi(n) \to \infty$ and $\phi(n)/n \to 0$ as $n \to \infty$. Consider the Marcus-Lushnikov process with*

$$K_n(x, y) = xy 1_{(\max(x,y) \leq \phi(n))}.$$

*Let $\mathcal{Q}_n(t)$ be the component containing vertex 1 at time $t$, let $Q_n(t) = |\mathcal{Q}_n(t)|$ be its size, and let $Z_n = \min\{t : Q_n(t) > \phi(n)\}$. Then as $n \to \infty$,*

$$(\mathcal{Q}_n(t), 0 \leq t < Z_n) \xrightarrow{d} (\mathcal{G}_t, 0 \leq t < Z)$$

*and so in particular*

$$P(Z_n > t) \to t^{-1}, \ t > 1$$

$$P(Q_n(t) = q, Z_n > t) \to \frac{q^{q-1} e^{-q}}{q!} t^{-1}, \ t > 1, q \geq 1.$$

$$Q_n(Z_n-) \xrightarrow{d} \infty.$$

This seems genuinely harder than our previous weak limit assertions, because we cannot write down useful exact expressions for the distribution of $\mathcal{Q}_n(t)$.

## 4. Chains with boundary behavior

The chains in section 3 either stayed in the set $\mathbf{T}$ of finite trees for all time, or were stopped upon first becoming infinite. In this section we deal with chains which run for time $0 \leq s < \infty$ but which sometimes momentarily become infinite (but are a.s. finite at fixed times).

Let us briefly mention theoretical background. Suppose we regard a continuous-time countable-state Markov chain as being specified by its finite-dimensional distributions. Then what can we say in general about its sample path behavior? The path must spend almost all (w.r.t. Lebesgue measure) time in the state space, but there are examples where in a natural topology the chain occasionally goes outside the original state space. An extensive general *compactification* theory was developed to handle such phenomena (see e.g. [**19**] Chapter III.6). But naturally-arising examples are rather scarce.

The specific chains we shall describe are novel, and indeed the notion of chains with such boundary behavior arising as weak limits within concrete discrete settings seems novel. The chain in sections 4.1 and 4.2 exhibits the qualitatively simplest boundary behavior, being infinite only at isolated times. A more exotic chain is outlined in section 4.3.

**4.1. The backwards chain.** This chain is motivated as a weak limit of a chain on spanning trees (weak limit assertion 4.3). It will also provide an alternative explanation (Corollary 4.6) of the striking "quasistationarity" property (3.14) of $(\mathcal{G}^0_s)$: that the conditional distribution of $\mathcal{G}^0_s$ given $Z > s$ is PGW(1) for all $s \geq 1$.

We study the $\mathbf{T}$-valued chain $(\overleftarrow{\mathcal{T}}_s)$ specified in words as

> Each edge is pruned away at rate 1; at rate 1 a PGW$^\infty$(1) infinite tree is appended at the root

and symbolically by transition rates

$$\mathbf{t}_1 \to (\mathbf{t}_1, \mathrm{root}, \cdot) : \ \text{rate } f_*(\cdot)$$

$$(\mathbf{t}_1, u, \mathbf{t}_2) \to \mathbf{t}_1 : \ \text{rate } 1 \tag{4.1}$$

where $f_*$ is the density of the PGW$^\infty$(1) infinite tree. Here $(\mathbf{t}_1, \mathrm{root}, \mathbf{t}_2)$ indicates joining the root of $\mathbf{t}_2$ to the root of $\mathbf{t}_1$ by a new edge. The point is that an infinite branch is adjoined at the times at a Poisson (rate 1) process, thus taking the chain outside the set $\mathbf{T}$ of finite trees, but this new branch is immediately cut back to finite size, so the chain only spends isolated time-instants outside $\mathbf{T}$.

PROPOSITION 4.1. *The chain $(\overleftarrow{\mathcal{T}}_s)$ has unique stationary distribution PGW(1).*

PROOF. We consider restrictions to height $h$. Recall (2.6,2.7)

$$f^h(\mathbf{t}) = e^{n(h,\mathbf{t})-|\mathbf{t}|}, \ \mathbf{t} \in \mathbf{T}^h$$

$$f^h_*(\mathbf{t}) = n(h,\mathbf{t})e^{n(h,\mathbf{t})-|\mathbf{t}|}, \ \mathbf{t} \in \mathbf{T}^h$$

are the distributions of the restrictions of PGW(1) and of PGW$^\infty$(1). So we need to show that the chain on $\mathbf{T}^h$ with transition rates

$$\mathbf{t}_1 \rightarrow (\mathbf{t}_1, \text{root}, \mathbf{t}_2) : \text{ rate } f_*^{h-1}(\mathbf{t}_2) \text{ ( for } \mathbf{t}_1 \in \mathbf{T}^h, \mathbf{t}_2 \in \mathbf{T}^{h-1})$$

$$(\mathbf{t}_1, u, \mathbf{t}_2) \rightarrow \mathbf{t}_1 : \text{ rate } 1 \text{( for } (\mathbf{t}_1, u, \mathbf{t}_2) \in \mathbf{T}^h)$$

has stationary distribution $f^h$. So fix $\mathbf{t} \in \mathbf{T}^h$ and verify the balance equations for probability flow rates into and out of $\mathbf{t}$. We assert these flow rates are

$$\begin{array}{rc} \text{out, by pruning} & (|\mathbf{t}|-1)f^h(\mathbf{t}) \\ \text{out, by joining} & f^h(\mathbf{t}) \\ \text{in, by pruning} & (|\mathbf{t}|-n(\mathbf{t},h))f^h(\mathbf{t}) \\ \text{in, by joining} & n(\mathbf{t},h)f^h(\mathbf{t}) \end{array}$$

(these rates plainly do balance). The "out" rates are clear. To verify the "in, by pruning" rate, we need to show that for each vertex $u$ of $\mathbf{t}$ at height less than $h$,

$$\sum_{\mathbf{t}_2 : (\mathbf{t}, u, \mathbf{t}_2) \in \mathbf{T}^h} \frac{f^h(\mathbf{t}, u, \mathbf{t}_2)}{f^h(\mathbf{t})} = 1.$$

But, if $u$ is at distance $j$ from the root, this reduces to

$$\sum_{\mathbf{t}_2 \in \mathbf{T}^{h-j-1}} f^{h-j-1}(\mathbf{t}_2) = 1$$

which holds because $f^{h-j-1}$ is a probability distribution. To verify the "in, by joining" rate, for each edge $e$ of $\mathbf{t}$ incident at the root, we may write $\mathbf{t} = (\mathbf{t}_1, \text{root}, \mathbf{t}_{h-1})$ where $\mathbf{t}_{h-1}$ is the subtree of height at most $h-1$ which is pruned away by cutting $e$. The flow rate from $\mathbf{t}_1$ to $\mathbf{t}$ equals

$$f^h(\mathbf{t}_1) f_*^{h-1}(\mathbf{t}_{h-1}) = m(e) f^h(\mathbf{t})$$

where $m(e)$ is the number of vertices at height $h$ in $\mathbf{t}$ which lie in the branch through $e$. Summing over edges $e$ gives the stated rate. □

The special case $\lambda = 1$ of Proposition 3.4 is the assertion that $\mathcal{G}_V^*$ has PGW(1) distribution. Here is an interesting independent proof of that fact, restated in words.

COROLLARY 4.2. *Take a realization from $PGW^\infty(1)$, and independently let $U$ have uniform distribution on $(0,1)$. Cut each edge of the $PGW^\infty(1)$-tree with probability $U$. Then the component containing the original root has distribution PGW(1).*

PROOF. By the PASTA (Poisson arrivals see time-averages) principle (e.g. [**6**]), observing the stationary process $(\overleftarrow{\mathcal{T}}_s)$ only at the instants immediately before an infinite branch is attached defines a discrete-time Markov chain, with PGW(1) stationary distribution, whose steps are specified by

> join a realization of PGW$^\infty$(1) at the root; prune each edge with probability $1 - e^{-V}$

where $V$ (representing the time until the next infinite branch is attached) has exponential(1) distribution. But (from the definitions) joining a PGW$^\infty$(1) tree to a PGW(1) tree gives a PGW$^\infty$(1) tree, and $1 - e^{-V} \stackrel{d}{=} U$. □

Our motivation for the definition of $(\overleftarrow{\mathcal{T}}_s)$ came from the following weak limit assertion. Let $(Q_i^n)$ be the discrete-time Markov chain taking values in the set of all $n^{n-1}$ rooted trees on labeled vertices $\{1, \ldots, n\}$, which in one step goes from $\mathbf{t}$ to a tree $\mathbf{t}'$ as follows.

> Pick a uniform random vertex $I$. If $I = \mathrm{root}(\mathbf{t})$ then let $\mathbf{t}' = \mathbf{t}$. Otherwise, draw an edge from $\mathrm{root}(\mathbf{t})$ to $I$, creating a cycle, and delete the other edge in the cycle incident at $\mathrm{root}(\mathbf{t})$: let $\mathbf{t}'$ be the resulting tree, rooted at $I$.

It is easy to check that the stationary distribution is uniform. This chain may be regarded as the specialization (to the complete graph) of the now well-known *random walk algorithm* for constructing a uniform random spanning tree of a general finite undirected graph [**2, 8**], which is part of the circle of ideas surrounding the Markov chain tree theorem [**18, 15**]. Now define $F_i^n$ to be the subtree of $Q_i^n$ rooted at vertex 1 obtained by cutting the edge at vertex 1 which leads toward $\mathrm{root}(Q_i^n)$ (if $\mathrm{root}(Q_i^n) = 1$ let $F_i^n = Q_i^n$). Regard $(F_i^n, i \geq 0)$ as a stationary process taking values in the set $\mathbf{T}$ of $u$-labeled trees. Lemma 2.4 implies that $F_0^n \xrightarrow{d} \mathrm{PGW}(1)$.

Weak Limit Assertion 4.3. As $n \to \infty$,

$$(F^n_{\lfloor ns \rfloor}, s \geq 0) \xrightarrow{d} (\overleftarrow{\mathcal{T}}_s, s \geq 0).$$

*Outline.* If the vertex $I$ chosen at step $i+1$ is not a vertex of $F_i^n$ then $F_{i+1}^n = F_i^n$. If $I$ is a non-root vertex $u$ of $F_i^n$ then $F_{i+1}^n$ is obtained from $F_i^n$ by pruning the edge from $u$ to its parent. After time-rescaling, in $F^n_{\lfloor ns \rfloor}$ each edge is being pruned at rate asymptotic to 1. When $I$ is vertex 1 (which happens at rescaled rate 1) the new tree $F_{i+1}^n$ is obtained from $F_i^n$ by attaching to the root an edge to a vertex $v$ which is the root of a random tree on the remaining $n - |F_i^n|$ vertices. By Lemma 2.4 this attached tree has distribution asymptotic to $\mathrm{PGW}^\infty(1)$.

**4.2. The forwards chain.** The time-reversal $(\overrightarrow{\mathcal{T}}_s)$ of the stationary chain $(\overleftarrow{\mathcal{T}}_s)$ is automatically a Markov chain with the same PGW(1) stationary distribution.

Proposition 4.4. *$(\overrightarrow{\mathcal{T}}_s)$ is the homogeneous chain with transition rates*

$$\mathbf{t}_1 \to (\mathbf{t}_1, u, \mathbf{t}_2) : \ \text{rate } f_1(\mathbf{t}_2)$$

*together with the rule:*

*at the instant the tree becomes infinite, the infinite branch is pruned away at the root.*

Proof. To check that the transition rate is as asserted is to check

$$f_1(\mathbf{t}_1) \times f_1(\mathbf{t}_2) =^? f_1(\mathbf{t}_1, u, \mathbf{t}_2) \times 1$$

where the 1 is the reversed flow rate (4.1). But this equality is the $\lambda = 1$ case of (3.3). The final rule is just a sample path property inherited from the corresponding property of $(\overleftarrow{\mathcal{T}}_s)$. ◻

*Remark.* Our construction of $(\overrightarrow{\mathcal{T}}_s)$ via time-reversal finesses various difficulties which would arise if we tried to argue directly that the properties in Proposition 4.4 specify a positive-recurrent chain. First, it is not obvious (though in fact true, by the corresponding fact for $(\overleftarrow{\mathcal{T}}_s)$) that the restrictions to height $h$, i.e. the process

$(r_h(\overrightarrow{\mathcal{T}}_s), s \geq 0)$, is Markov. (The difficulty being that by only looking at the restriction we cannot see the infinite branch being grown and pruned). Thus we would have to deal directly with a compactification of $\mathbf{T}$ as state-space, and it is not elementary to write down the equations one must verify in order to check that PGW(1) is a stationary distribution.

Write $T_\infty > 0$ for the first time that $|\overrightarrow{\mathcal{T}}_s| = \infty$.

COROLLARY 4.5. *(a) For the stationary chain, $T_\infty$ has exponential(1) distribution. Moreover, for each fixed $s > 0$ the distribution of $\overrightarrow{\mathcal{T}}_s$ is PGW(1) and is independent of the event $\{T_\infty > s\}$.*

*(b)*

$$P(T_\infty > s | \overrightarrow{\mathcal{T}}_0 = \mathbf{t}) = (H(s))^{|\mathbf{t}|},$$

*where*

$$H(s) = \exp(1 - s - e^{-s}).$$

PROOF. In the stationary $(\overleftarrow{\mathcal{T}}_s)$ chain, let $C$ be the first time that an infinite branch is joined. Clearly $P(C > s) = e^{-s}$ and, for fixed $s$, the event $\{C > s\}$ is independent of $\overleftarrow{\mathcal{T}}_0$. The assertions of (a) follow immediately by time-reversal. For (b), from each vertex of $\mathbf{t}$ there grow independent and identically distributed processes of subtrees, so the first assertion holds for some $H(s)$. Then by (a)

$$e^{-s} = \Phi(H(s))$$

where $\Phi(z)$ is the generating function of the size of PGW(1), i.e. of the Borel(1) distribution (2.4). But it is elementary and well-known that $\Phi(z) = z\exp(\Phi(z) - 1)$, and hence that $\Phi^{-1}(u) = ue^{1-u}$. □

We now relate $(\overrightarrow{\mathcal{T}}_s)$ to $(\mathcal{G}^0_s)$.

COROLLARY 4.6.

$$(\mathcal{G}^0_s, 1 \leq s \leq Z) \stackrel{d}{=} (\overrightarrow{\mathcal{T}}_{\log s} : 1 \leq s \leq e^{T_\infty}).$$

PROOF. Corollary 4.5(a) shows that

$$P(\overrightarrow{\mathcal{T}}_{\log s} = \mathbf{t}, s < e^{T_\infty}) = s^{-1} f_1(\mathbf{t}) = P(\mathcal{G}^0_s = \mathbf{t}, s < Z).$$

Proposition 4.4 implies that the transition rates $\mathbf{t}_1 \to (\mathbf{t}_1, u, \mathbf{t}_2)$ for $(\overrightarrow{\mathcal{T}}_{\log s})$ are

$$\frac{d \log s}{ds} \times f_1(\mathbf{t}_2) = h_s(\mathbf{t}_2),$$

agreeing with the transition rates (3.15) for $(\mathcal{G}^0_s)$. □

Note that we could take Corollary 4.6 as a construction of $\mathcal{G}^0_s$, and then the quasistationarity property of $\mathcal{G}^0_s$ would be a consequence of the independence assertion for the stationary $\overrightarrow{\mathcal{T}}$-chain in Corollary 4.5(a).

Note also that Corollary 4.5(b) now translates to give distributional information for $(\mathcal{G}^0_s)$.

COROLLARY 4.7.

$$P(Z > z | \mathcal{G}_1^0 = \mathbf{t}) = (H_1(z))^{|\mathbf{t}|}, z \geq 1$$

*where*

$$H_1(z) = z^{-1} \exp(1 - z^{-1}).$$

**4.3. An exotic process.** Let us end by indicating a process with more exotic boundary behavior. Recall from section 4.1 the non-reversible discrete-time chain $(Q_i^n, i \geq 0)$ on the set of all rooted trees on $\{1, 2, \ldots, n\}$, with uniform stationary distribution. Here is a reversible chain with the same property. In one step, go from $\mathbf{t}$ to a tree $\mathbf{t}'$ as follows.

> Pick two distinct vertices $v_1, v_2$ uniformly at random. If $(v_1, v_2)$ is an edge of $\mathbf{t}$, let $\mathbf{t}' = \mathbf{t}$. Otherwise, add an edge $(v_1, v_2)$ thereby creating a cycle, and delete a uniformly-chosen other edge from that cycle to get $\mathbf{t}'$.

Call this chain $(R_i^n)$, and regard the tree as rooted at vertex $n$. As in section 4.1, define $F_i^n$ to be the subtree of $R_i^n$ rooted at vertex 1 obtained by cutting the edge at vertex 1 which leads toward the root. Lemma 2.4 says that $F_0^n \xrightarrow{d}$ PGW(1). The arguments in the outline proof of weak limit assertion 4.3 suggest

$$(F_{\lfloor ns \rfloor}^n, s \geq 0) \xrightarrow{d} (\mathcal{Z}_s, s \geq 0),$$

where $(\mathcal{Z}_s)$ is a certain stationary reversible chain with stationary distribution PGW(1). For finite $\mathbf{t}$, the transition rates of $(\mathcal{Z}_s)$ are as follows.

> (i) At each vertex of $\mathbf{t}$ an infinite PGW$^\infty$(1)-distributed branch is attached at rate 1/2;
> (ii) at each vertex $v$ of $\mathbf{t}$, at rate 1/2 the edge at the root leading toward $v$ is cut, and simultaneously an infinite PGW$^\infty$(1)-distributed branch is attached to the root.

To complete a verbal specification of $(\mathcal{Z}_s)$ we need to define transitions which are qualitatively reverse to (i) when $\mathbf{t}$ is infinite. Obviously it is a challenge to formalize such a definition and show that it does specify a unique process.

*Acknowledgement.* This work has benefited from an ongoing collaboration with Jim Pitman. I also thank Vlada Limic for carefully reading a draft.

## References

[1] D.J. Aldous. A random tree model associated with random graphs. *Random Structures Algorithms*, 1:383–402, 1990.

[2] D.J. Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Discrete Math.*, 3:450–465, 1990.

[3] D.J. Aldous. Deterministic and stochastic models for coalescence: a review of the mean-field theory for probabilists. Unpublished. Available via homepage http://www.stat.berkeley.edu/users/aldous, 1996.

[4] D.J. Aldous and J. Pitman. Tree-valued Markov chains derived from Galton-Watson processes. Unpublished, 1997.

[5] N. Alon and J. H. Spencer. *The Probabilistic Method.* Wiley, 1992.

[6] S. Asmussen. *Applied Probability and Queues.* Wiley, 1987.

[7] B. Bollobás. *Random Graphs.* Academic Press, London, 1985.

[8] A. Broder. Generating random spanning trees. In *Proc. 30'th IEEE Symp. Found. Comp. Sci.*, pages 442–447, 1989.
[9] W. J. Ewens. *Mathematical Population Genetics.* Springer-Verlag, Berlin, 1979.
[10] G. R. Grimmett. Random labelled trees and their branching networks. *J. Austral. Math. Soc. (Ser. A)*, 30:229–237, 1980.
[11] O. Haggstrom. Dynamical percolation: early results and open problems. In this volume.
[12] H. Kesten. Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Probab. Statist.*, 22:425–487, 1987.
[13] D.E. Knuth. *The Art of Computer Programming*, volume 3. Addison-Wesley, 1973.
[14] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proof of $L \log L$ criteria for mean behavior of branching processes. *Ann. Probab.*, 23:1125–1138, 1995.
[15] R. Lyons and Y. Peres. Probability and trees. Book in preparation, 1997.
[16] H. M. Mahmoud. *Evolution of Random Search Trees.* Wiley, 1992.
[17] J.A. Mann, B. Pittel, and W. A. Woycznski. Random tree-type partitions as a model for acyclic polymerization: Holtsmark (3/2 stable) distribution of the supercritical gel. *Ann. Probab.*, 18:319–341, 1990.
[18] R. Pemantle. Uniform random spanning trees. In J. Laurie Snell, editor, *Topics in Contemporary Probability*, pages 1–54, Boca Raton, FL, 1995. CRC Press.
[19] L.C.G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales*, volume 1,Foundations. Wiley, 1994. 2nd. edition.
[20] P. Whittle. *Systems in Stochastic Equilibrium.* Wiley, 1986.

Department of Statistics, University of California, 367 Evans Hall # 3860, Berkeley, CA 94720-3860