

## RANDOM TREE GROWTH BY VERTEX SPLITTING

F. DAVID, W.M.B. DUKES, T. JONSSON, S.Ö. STEFÁNSSON.

ABSTRACT. We study a model of growing planar tree graphs where in each time step we separate the tree into two components by splitting a vertex and then connect the two pieces by inserting a new link between the daughter vertices. We develop a mean field theory for the vertex degree distribution, prove that the mean field theory is exact in some special cases and check that it agrees with numerical simulations in general. We calculate various correlation functions and show that the intrinsic Hausdorff dimension can vary from one to infinity, depending on the parameters of the model.

## CONTENTS

1. Introduction	2
2. A vertex splitting model	4
2.1. Description of the model	4
2.2. Distribution of vertex degrees for linear splitting weights	6
2.3. Explicit solutions	11
2.4. Discussion of the conditions on the weights	12
2.5. Mean field equation for general weights	12
2.6. The $d = \infty$ model with linear weights	13
3. Subtree structure probabilities	14
3.1. Probabilities related to subtree structure	15
3.2. Two-point functions	19
4. Hausdorff dimension	20
4.1. Definition	20
4.2. Geodesic distances and 2 point functions	21
4.3. Scaling and the Hausdorff dimension	22
4.4. An upper bound on the Hausdorff dimension	24
4.5. Explicit solutions and numerical results for $d = 3$ .	26
4.6. Hausdorff dimension for general weights	26
5. Correlations	28
5.1. Solution in the simplest case	30
5.2. Results for non-linear weights	32
6. Discussion	32
Acknowledgements	36
Appendix A. Illustration of recursion equations for correlations	36
References	42

## 1. INTRODUCTION

Random trees arise in many branches of science ranging from mathematics through computer science, physics and chemistry to biology and sociology. Trees are the simplest generalizations of random walks and are used to model various relationships like family trees [13] or phylogenetic trees [16, 19] as well as physical objects that look like trees, e.g., branched polymers. Information about more complicated random geometrical objects like random surfaces which arise in quantum gravity [4] can sometimes be encoded in labelled random trees [18]. Such trees have also been used to encode information regarding the secondary structure of macromolecules, RNA folding being one example [10].

There are two principal classes of models that have been used to model trees. The first one is equilibrium statistical mechanics where trees are assigned a certain weight factor (Boltzmann factor) which induces a probability measure on the class of trees under consideration. The second class is growing trees where one starts with a simple tree and then continues to add vertices and links according to some growth rules. After a fixed time  $t$  the growth rules induce a probability measure on the trees that can arise after time  $t$ . Time can be taken to be discrete or continuous but here we will work exclusively with discrete time. The trees themselves are also discrete and are therefore always tree graphs in the usual sense.

The first class of models is the one that appears most frequently in physics whereas the second class is the natural one to use to analyse, e.g., growing family trees or citation networks. It is well-known that in certain cases these two approaches to studying random trees are equivalent. For example, the so called *generic trees* that come up in the study of gravity are equivalent to a critical Galton-Watson process [11]. So called *causal trees* are believed to be equivalent to certain classes of growing trees [5].

In this paper we study a model which is closely related to a model of growing trees that arises in the theory of RNA secondary structures [10]. A special case of this model is also equivalent to the preferential attachment model which describes for example the growth of branched polymers out of a soup of monomers or the growth of the internet [9]. In a limiting case the model reduces to a model of phylogenetic trees [12]. Let us describe this model informally, a precise definition will be given in the next section. Suppose that at an initial time  $t_0$  we are given a tree which will always be assumed to be finite, rooted and planar (i.e. at each vertex the attached links are ordered cyclically). This assumption of planarity is convenient but not essential. In order to evolve the tree we choose a vertex at random with a relative probability  $w_i$  which only depends on the order  $i$  of the vertex. We split the chosen vertex into two vertices of order  $j$  and  $k$  which are linked to each other and attach the links of the original vertex of order  $i$  to the two new vertices so  $j+k = i+2$  and the planarity is preserved in this process, see Fig. 1. We sometimes let the relative probability of this splitting,  $w_{j,k}$ , depend

on  $j$  and  $k$  but sometimes we take the uniform distribution on the  $i(i+1)/2$  possibilities. The parameters  $w_i$  and  $w_{j,k}$  define the model. Most of our results are for the case when there is a maximal vertex degree, i.e.,  $w_{j,k} = 0$  if either  $j$  or  $k$  exceeds a cutoff  $d$ . The reader may ask why this growth process is natural. First of all the process models the arch deposition which is used to describe RNA folding [10] and secondly it becomes the preferential attachment growth model when we take  $w_{j,k} = 0$  unless  $j$  or  $k$  is equal to 1.

We are interested in the structure of the trees after a large number of steps in the growth process. The main quantities we study are the distribution of the order of vertices, the correlation between the degrees of neighbouring vertices and the intrinsic Hausdorff dimension of the trees. In order to get at the Hausdorff dimension we define and analyse certain correlation functions which describe the subtree structure of the random tree. The case when  $w_j$  depends linearly on  $j$  is much easier to analyse than the general case because then the probability  $p_j$  of splitting a vertex of degree  $j$  in a tree  $T$  is given by

$$p_j = \frac{w_j}{F(|T|)}$$

where  $F$  is a function only of the number of vertices  $|T|$  in  $T$  but does not depend on the structure of  $T$ .

For the linear case and finite  $d$  we prove that mean field theory (MFT) for the vertex degree distribution is exact. By mean field theory we mean that the density of vertices with a given degree has a well defined limit as the size of the tree goes to infinity and for a finite but large trees there are small (in a suitable sense) fluctuations about this limit. For infinite  $d$  and linear splitting probabilities we can calculate the vertex degree distribution in closed form using MFT and show that it agrees with numerical simulations. The vertex degree distribution is found to fall off factorially in this case. The vertex degree distribution always becomes independent of the initial tree as time goes to infinity. The main mathematical tool used to obtain these results is the Perron-Frobenius theorem.

Keeping track of the environment of the vertex created at a given time in the growth process we can obtain information about the subtree structure. For example we find an expression for the probability that in a tree of size  $N$  we have a link with vertices of degree  $j$  and  $k$  which are attached to trees of total size  $N_1$  and  $N_2$ , respectively, with  $N_1 + N_2 = N - 1$ . With some simple scaling assumptions this allows us to calculate the Hausdorff dimension in the linear case and find an explicit formula for  $d_H$  in the case  $d = 3$ . These results agree with simulations and we find Hausdorff dimensions ranging from 1 to  $\infty$  depending on the parameters of the model. In the nonlinear case the same method gives results which agree well with simulations.

We study the correlations between the degrees of neighbouring vertices. This amounts to studying the density  $\rho_{ij}$  of links with vertices of degrees  $i$  and  $j$ . We adopt the convention that the vertex of degree  $i$  is the one closer to the root. Then  $i \neq 1$  unless this vertex is the root. In simple cases this

correlation can be calculated explicitly and one sees that there are always nontrivial correlations, i.e.,

$$\rho_{ij} \neq \frac{\rho_i \rho_j}{1 - \rho_1}$$

where  $\rho_i$  is the density of vertices of degree  $i$ . Correlations between vertices which are further away from each other can be studied by the same method. Degree-degree correlations have been studied for a different class of random graphs in [6].

This paper is organized as follows. In the next section we study the vertex degree distribution and mean field theory. In Sections 3 and 4 we analyse the subtree structure and the intrinsic Hausdorff dimension. In Section 5 we discuss correlations of vertex degrees. In a final section we discuss open problems and the connection of our work with other models of random trees, in particular models of phylogenetic trees [2, 3, 8, 12, 15, 17]. A few technical issues are relegated to appendices.

## 2. A VERTEX SPLITTING MODEL

**2.1. Description of the model.** A rooted planar tree is a tree graph embedded into the plane  $\mathbb{R}^2$  which contains a distinguished vertex that we call the *root*. We will always assume that the root has order one. The planarity condition is for convenience only and can also be implemented by cyclically ordering the links incident on each vertex.

Let  $\mathcal{T}$  be the collection of all rooted planar trees for which every vertex has finite degree at most  $d$ . Denote the number of vertices in a tree  $T \in \mathcal{T}$  by  $|T|$  and the number of vertices of degree  $i$  in  $T$  by  $n_i(T)$ . Let  $\mathcal{T}_N$  be those trees  $T \in \mathcal{T}$  with  $|T| = N$ . Let

$$\mathbf{M} = \begin{bmatrix} 0 & w_{1,2} & w_{1,3} & \cdots & w_{1,d-1} & w_{1,d} \\ w_{2,1} & w_{2,2} & w_{2,3} & \cdots & w_{2,d-1} & w_{2,d} \\ w_{3,1} & w_{3,2} & w_{3,3} & \cdots & w_{3,d-1} & 0 \\ w_{4,1} & w_{4,2} & w_{4,3} & & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ w_{d,1} & w_{d,2} & 0 & \cdots & 0 & 0 \end{bmatrix}$$

be a symmetric matrix with nonnegative entries that we call *partitioning weights*. We define a collection of non-negative numbers called *splitting weights*,  $w_1, w_2, \dots, w_d$ , by

$$w_i = \frac{i}{2} \sum_{j=1}^{i+1} w_{j,i+2-j}. \quad (2.1)$$

We now define a growth rule for planar trees which we call *vertex-splitting*. Given a tree  $T \in \mathcal{T}_N$

- (i) Choose a vertex  $v$  of  $T$  with probability  $w_i/\mathcal{W}(T)$  where  $i$  is the order of  $v$  and

$$\mathcal{W}(T) = \sum_{j=1}^d w_j n_j(T). \quad (2.2)$$

- (ii) Partition the edges incident with  $v$  into two disjoint sets  $V$  and  $V'$  of adjacent edges with probability

$$\frac{w_{k,i+2-k}}{w_i}.$$

The set  $V$  contains  $k-1$  of the edges and  $V'$  contains  $i-(k-1)$  of these edges,  $k = 1, \dots, i$ . For a given  $k$ , all  $i$  such partitionings are taken to be equally likely.

- (iii) Move all edges in  $V'$  from  $v$  to a new vertex  $v'$  and create an edge joining  $v$  to  $v'$ . If  $v$  is the root, then the new vertex of order one is taken to be the root.

This vertex-splitting operation is illustrated in the following diagram (the root vertex is circled).

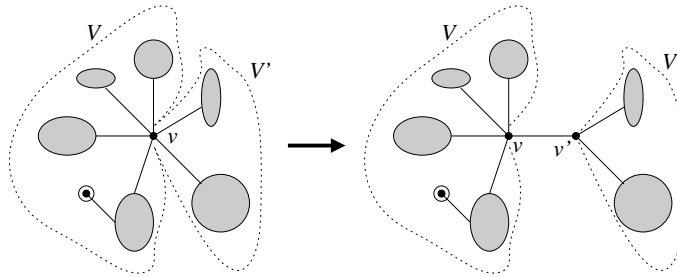


FIGURE 1. Illustration of the splitting process for  $i = 6$  and  $k = 5$ .

After the splitting operation, the degree of vertex  $v$  is  $k$  and the degree of vertex  $v'$  is  $i + 2 - k$ . Since the maximum allowed vertex degree is  $d$  we define  $w_{d+1,1} = w_{1,d+1} = 0$ , i.e. we do not allow splittings of vertices of degree  $d$  that produce vertices of degree  $d + 1$ . If the partitioning weights are chosen such that  $w_{i,j} = 0$  for  $i \neq 1$  or  $j \neq 1$ , then the vertex-splitting model is equivalent to the preferential attachment model discussed in [9].

In Sections 3 and 4 we will find it convenient to label the vertices according to their time of creation. In this case we append the following to our rules:

- (iv) Let  $a$  be the label of the vertex chosen in (i). If  $v$  is further away from the root than  $v'$  in step (iii) then we let  $v$  keep the label  $a$  and give  $v'$  the label  $N$ . Otherwise label  $v$  with  $N$  and label  $v'$  with  $a$ .

This book-keeping device has no effect on the dynamics of the model. The single root vertex (which is the only tree in  $\mathcal{T}_1$ ) has label 0. We will often think of the number of vertices (or equivalently, the number of links) as time and denote it by  $t + 1$  (or  $t$  in the case of links) assuming we start with the single vertex tree at time  $t = 0$ .

If the partitioning weights are chosen such that the splitting weights are linear,

$$w_i = ai + b \quad (2.3)$$

for some  $a, b \geq 0$  then the model is easier to analyze since the weight of a tree  $T \in \mathcal{T}$  depends only on the size of the tree

$$\mathcal{W}(T) = (2a + b)|T| - 2a. \quad (2.4)$$

This is easily seen from the two constraints on the vertex degrees,

$$\sum_{i=1}^d n_i(T) = |T| \quad \text{and} \quad \sum_{i=1}^d in_i(T) = 2(|T| - 1). \quad (2.5)$$

By abuse of notation, in this case we will write  $\mathcal{W}(|T|) = \mathcal{W}(T)$ . We will also sometimes restrict to uniform partitioning weights, i.e.

$$w_{i,k+2-i} = \begin{cases} w_k / \binom{k+1}{2} & \text{for } i = 1, \dots, k+1, & \text{if } k < d, \\ w_k / \binom{k}{2} & \text{for } i = 2, \dots, k, & \text{if } k = d. \end{cases} \quad (2.6)$$

**2.2. Distribution of vertex degrees for linear splitting weights.** Start from a finite tree  $T_0$  at time  $t_0 = |T_0|$  and perform vertex splitting according to the rules described in the previous subsection  $\tau$  times. We then obtain a tree in  $\mathcal{T}_{t_0+\tau}$ . Let  $t = t_0 + \tau$ . The vertex splitting operation induces a probability measure  $\mu_t$  on  $\mathcal{T}_t$ , which of course depends on the initial tree  $T_0$ . In this section we will drop  $T_0$  from function arguments with the understanding that it is implied, unless otherwise stated.

Let  $p_t(n_1, \dots, n_d)$  be the probability that  $T \in \mathcal{T}_t$  has  $(n_1(T), \dots, n_d(T)) = (n_1, \dots, n_d)$  according to the measure  $\mu_t$ . Let  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ . Through consideration of the splitting operation, the probability generating function

$$\mathcal{H}_t(\mathbf{x}) = \sum_{n_1 + \dots + n_d = t} p_t(n_1, \dots, n_d) x_1^{n_1} \dots x_d^{n_d} \quad (2.7)$$

satisfies the recurrence

$$\mathcal{H}_{t+1}(\mathbf{x}) = \sum_{n_1 + \dots + n_d = t} \frac{p_t(n_1, \dots, n_d)}{\sum_{i=1}^d n_i w_i} \mathbf{c}(\mathbf{x}) \cdot \nabla (x_1^{n_1} \dots x_d^{n_d}) \quad (2.8)$$

for all  $t \geq t_0$ , where

$$\mathbf{c}(\mathbf{x}) = (c_1(\mathbf{x}), c_2(\mathbf{x}), \dots, c_d(\mathbf{x})) \quad (2.9)$$

with

$$c_i(\mathbf{x}) = \frac{i}{2} \sum_{j=1}^{i+1} w_{j,i+2-j} x_j x_{i+2-j} \quad (2.10)$$

and

$$\nabla = \left( \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_d} \right). \quad (2.11)$$

One may write down a master equation for these probabilities by comparing the coefficients of  $x_1^{n_1} \cdots x_d^{n_d}$  in (2.8). For linear weights (2.3), equation (2.8) becomes

$$\mathcal{H}_{t+1}(\mathbf{x}) = \frac{1}{\mathcal{W}(t)} \mathbf{c}(\mathbf{x}) \cdot \nabla \mathcal{H}_t(\mathbf{x}). \quad (2.12)$$

The remainder of this subsection concerns linear weights only. We wish to find the mean value of  $n_k(T)$  for  $T \in \mathcal{T}_t$  with respect to the measure  $\mu_t$ . Denote this value by  $\bar{n}_{t,k}$ . We have

$$\bar{n}_{t,k} = \sum_{n_1 + \dots + n_d = t} p_t(n_1, \dots, n_d) n_k = \partial_k \mathcal{H}_t(\mathbf{x})|_{\mathbf{x}=\mathbf{1}}, \quad (2.13)$$

where  $\mathbf{1} = (1, 1, \dots, 1)$ . To get a recursion equation for  $\bar{n}_{t,k}$ , differentiate both sides of (2.12) with respect to  $x_k$  and set  $\mathbf{x} = \mathbf{1}$  to find

$$\bar{n}_{t+1,k} = \frac{1}{\mathcal{W}(t)} \left( \sum_{i=k-1}^d i w_{k,i+2-k} \bar{n}_{t,i} + \sum_{i=1}^d w_i \partial_i \partial_k \mathcal{H}_t(\mathbf{x})|_{\mathbf{x}=\mathbf{1}} \right). \quad (2.14)$$

Since the weights are linear we can use the constraints in (2.5) to rewrite the last term in (2.14) as

$$\sum_{i=1}^d w_i \partial_i \partial_k \mathcal{H}_t(\mathbf{x})|_{\mathbf{x}=\mathbf{1}} = (-w_k + \mathcal{W}(t)) \bar{n}_{t,k}. \quad (2.15)$$

Inserting this into (2.14) we see that the equations close

$$\bar{n}_{t+1,k} = \frac{1}{\mathcal{W}(t)} \left( -w_k \bar{n}_{t,k} + \sum_{i=k-1}^d i w_{k,i+2-k} \bar{n}_{t,i} \right) + \bar{n}_{t,k}. \quad (2.16)$$

We can also write the recursion in terms of the density of vertices of a given degree  $\rho_{t,k} \equiv \bar{n}_{t,k}/t$  and find

$$\rho_{t+1,k} = \frac{t}{\mathcal{W}(t)} \left( -w_k \rho_{t,k} + \sum_{i=k-1}^d i w_{k,i+2-k} \rho_{t,i} \right) + t(\rho_{t,k} - \rho_{t+1,k}). \quad (2.17)$$

The above equation can be put in the matrix form

$$\boldsymbol{\rho}_{t+1} = \mathbf{A}_t \boldsymbol{\rho}_t \quad (2.18)$$

where

$$\boldsymbol{\rho}_t = (\rho_{t,1}, \rho_{t,2}, \dots, \rho_{t,d})^T, \quad \mathbf{A}_t = \frac{t}{t+1} \left( \mathbf{I} + \frac{1}{\mathcal{W}(t)} \mathbf{B} \right), \quad (2.19)$$

$$\mathbf{B} = \begin{bmatrix} w_{1,2} & 2w_{1,3} & \cdots & (d-2)w_{1,d-1} & (d-1)w_{1,d} & 0 \\ w_{2,1} & 2w_{2,2} & \cdots & (d-2)w_{2,d-2} & (d-1)w_{2,d-1} & dw_{2,d} \\ 0 & 2w_{3,1} & \cdots & (d-2)w_{3,d-3} & (d-1)w_{3,d-2} & dw_{3,d-1} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & & \ddots & (d-2)w_{d-1,1} & (d-1)w_{d-1,2} & dw_{d-1,3} \\ 0 & \cdots & 0 & 0 & (d-1)w_{d,1} & dw_{d,2} \end{bmatrix} - \text{diag}(w_i)_{1 \leq i \leq d} \quad (2.20)$$

and  $\mathbf{I}$  is the identity matrix.

If we denote the vertex degree densities of the initial tree  $T_0$  by  $\boldsymbol{\rho}_{t_0}$  we can write the densities for trees on  $t$  vertices which grow from the initial tree as

$$\boldsymbol{\rho}_t = \left( \prod_{i=t_0}^{t-1} \mathbf{A}_i \right) \boldsymbol{\rho}_{t_0} = \frac{t_0}{t} \left( \prod_{i=t_0}^{t-1} \left( \mathbf{I} + \frac{1}{\mathcal{W}(i)} \mathbf{B} \right) \right) \boldsymbol{\rho}_{t_0}. \quad (2.21)$$

We will establish convergence of the right hand side by imposing some technical restrictions on  $\mathbf{B}$ . It turns out that the limiting distribution is independent of the initial distribution  $\boldsymbol{\rho}_{t_0}$ . We begin with some necessary lemmas.

**Lemma 2.1.** *If  $\lambda$  is an eigenvalue of  $\mathbf{B}$  with corresponding eigenvector  $\mathbf{e}_\lambda = (e_{\lambda 1}, \dots, e_{\lambda d})$ , i.e.*

$$\mathbf{B} \mathbf{e}_\lambda = \lambda \mathbf{e}_\lambda, \quad (2.22)$$

then the following holds:

$$\lambda \sum_{i=1}^d e_{\lambda i} = \sum_{i=1}^d w_i e_{\lambda i} \quad \text{and} \quad (2.23)$$

$$\lambda \sum_{i=1}^d i e_{\lambda i} = 2 \sum_{i=1}^d w_i e_{\lambda i}. \quad (2.24)$$

*Proof.* We prove the second identity. The first identity is established by a similar calculation. Multiply the  $i$ -th component of the eigenvalue equation



(2.22) by  $i$  and sum over  $i$  to get

$$\begin{aligned} \lambda \sum_{i=1}^d i e_{\lambda_i} &= - \sum_{i=1}^d i w_i e_{\lambda_i} + \sum_{i=1}^d i \sum_{k=i-1}^d k w_{i,k+2-i} e_{\lambda_k} \\ &= - \sum_{i=1}^d i w_i e_{\lambda_i} + \sum_{k=1}^d k \left( \sum_{i=1}^{k+1} i w_{i,k+2-i} \right) e_{\lambda_k}. \end{aligned} \quad (2.25)$$

Using  $w_{i,j} = w_{j,i}$  we find that

$$\sum_{i=1}^{k+1} i w_{i,k+2-i} = \frac{k+2}{2} \sum_{i=1}^{k+1} w_{i,k+2-i} \quad (2.26)$$

and this together with the definition of the splitting weights (2.1) proves the identity.  $\square$

**Lemma 2.2.** *If*

- (1)  $w_{k,1} = w_{1,k} > 0$  for  $k = 1, \dots, d$  (i.e. it is possible to produce vertices of degree  $d$ ) and
- (2)  $w_{i,d+2-i} > 0$  for at least one  $i$  with  $2 \leq i \leq d-1$ ,

then  $w_2$  is a positive, simple eigenvalue of  $\mathbf{B}$ . All other eigenvalues of  $\mathbf{B}$  have a smaller real part. The corresponding eigenvector  $\mathbf{e}_{w_2}$  can be taken to have all entries positive.

*Proof.* We begin by choosing a number  $\gamma > \max_{1 \leq k \leq d} \{w_k - k w_{k,2}\}$  and define  $\mathbf{P} = \mathbf{B} + \gamma \mathbf{I}$ . The matrix  $\mathbf{P}$  has only non-negative entries and the conditions (1) and (2) on  $\mathbf{B}$  guarantee that it is primitive, i.e. there is a number  $k$  such that all entries of the matrix  $\mathbf{P}^k$  are positive. Therefore, by the Perron-Frobenius theorem [20],  $\mathbf{P}$  has a simple positive eigenvalue  $r$  and all other eigenvalues of  $\mathbf{P}$  have a smaller modulus. The corresponding eigenvector  $\mathbf{e}_r$  can be taken to have all entries positive. We normalize the eigenvector such that

$$\sum_{i=1}^d e_{r_i} = 1. \quad (2.27)$$

Shifting back to the matrix  $\mathbf{B}$  we find that  $w \equiv r - \gamma$  is a simple real eigenvalue of  $\mathbf{B}$  with the largest real part and the corresponding eigenvector is  $\mathbf{e}_w = \mathbf{e}_r$ . We see right away from (2.23) and with the chosen normalization that

$$w = \sum_{i=1}^d w_i e_{w_i}. \quad (2.28)$$

Since the weights are linear, Lemma 2.1 shows that  $w = w_2$ .  $\square$

Note that the first condition on the weights in the above lemma is natural since we have fixed a maximal degree  $d$  and therefore we want to be able to produce vertices of degree  $d$ . The second condition, however, does not seem to be necessary for the results to hold but we still require it in order to use

the Perron-Frobenius theorem for primitive matrices. This condition is not very restrictive in the case of linear weights since it holds for all  $a$  and  $b$  except when  $ad + b = 0$ .

**Lemma 2.3.** *Let  $\lambda \in \mathbb{C}$ . Then*

$$\frac{t_0}{t} \prod_{i=t_0}^{t-1} \left( 1 + \frac{1}{\mathcal{W}(i)} \lambda \right) \longrightarrow \begin{cases} \frac{t_0 w_2}{t_0 w_2 - 2a} & \text{if } \lambda = w_2, \\ 0 & \text{if } \operatorname{Re}(\lambda) < w_2 \end{cases} \quad (2.29)$$

as  $t \rightarrow \infty$ .

*Proof.* The result follows from the identity

$$\frac{t_0}{t} \prod_{i=t_0}^{t-1} \left( 1 + \frac{1}{\mathcal{W}(i)} \lambda \right) = \frac{t_0}{t} \frac{\Gamma\left(t - \frac{2a-\lambda}{w_2}\right) \Gamma\left(t_0 - \frac{2a}{w_2}\right)}{\Gamma\left(t - \frac{2a}{w_2}\right) \Gamma\left(t_0 - \frac{2a-\lambda}{w_2}\right)}. \quad (2.30)$$

□

**Theorem 2.4.** *With the assumptions on  $\mathbf{B}$  in Lemma 2.2 and the additional assumption that  $\mathbf{B}$  is diagonalizable, the limit as  $t \rightarrow \infty$  of the right hand side of equation (2.21) exists and is given by the eigenvector  $\mathbf{e}_{w_2}$  of  $\mathbf{B}$  normalized such that*

$$\sum_{i=1}^d e_{w_2 i} = 1. \quad (2.31)$$

*Proof.* We use the normalization in (2.31) and expand  $\boldsymbol{\rho}_{t_0}$  in the basis of eigenvectors of  $\mathbf{B}$ . Using the results of Lemmas 2.1 and 2.2 and that  $T_0$  satisfies the equations in (2.5) we see that the expansion is of the form

$$\boldsymbol{\rho}_{t_0} = \frac{w_2 t_0 - 2a}{w_2 t_0} \mathbf{e}_{w_2} + \sum_{i=1}^{d-1} a_i \mathbf{e}_{\lambda_i} \quad (2.32)$$

where  $\lambda_i$ ,  $i = 1, \dots, d-1$  are the eigenvalues of  $\mathbf{B}$  with real part less than  $w_2$ . The result now follows from Lemma 2.3. □

Theorem 2.4 shows that with the above conditions on  $\mathbf{B}$  the limit of the vertex degree densities exists, is independent of the initial tree and is given by

$$\boldsymbol{\rho} \equiv \lim_{n \rightarrow \infty} \boldsymbol{\rho}_t = \mathbf{e}_{w_2}. \quad (2.33)$$

The limiting densities are therefore the unique positive solution to equation (2.22), i.e.

$$\rho_k = -\frac{w_k}{w_2} \rho_k + \sum_{i=k-1}^d i \frac{w_{k,i+2-k}}{w_2} \rho_i. \quad (2.34)$$

**2.3. Explicit solutions.** We discuss three simple special cases.

1) When  $d = 3$  we find that

$$\mathbf{B} = \begin{bmatrix} 0 & 2w_{1,3} & 0 \\ w_{2,1} & w_{2,2} - 2w_{3,1} & 3w_{3,2} \\ 0 & 2w_{3,1} & 0 \end{bmatrix}. \quad (2.35)$$

If the weights satisfy the conditions in Lemma 2.2 it is easy to see that  $\mathbf{B}$  is diagonalizable. For linear splitting weights  $w_i = ai + b$  and uniform partitioning weights the positive solution of (2.34) is

$$\rho_1 = \rho_3 = \frac{2}{7} \quad \text{and} \quad \rho_2 = \frac{3}{7} \quad (2.36)$$

for all values of  $a$  and  $b$  as can easily be seen from the simple structure of the  $\mathbf{B}$  in this case.

2) When  $d = 4$ , the splitting weights linear and the partitioning weights uniform one can check that

$$\mathbf{B} = \begin{bmatrix} 0 & \frac{2}{3}(2a+b) & \frac{1}{2}(3a+b) & 0 \\ a+b & -\frac{1}{3}(2a+b) & \frac{1}{2}(3a+b) & \frac{2}{3}(4a+b) \\ 0 & \frac{2}{3}(2a+b) & -\frac{1}{2}(3a+b) & \frac{2}{3}(4a+b) \\ 0 & 0 & \frac{1}{2}(3a+b) & -\frac{1}{3}(4a+b) \end{bmatrix}. \quad (2.37)$$

When  $4a + b > 0$  the weights satisfy the conditions in Lemma 2.2. The eigenvalues of  $\mathbf{B}$  are  $-\frac{1}{12}(33a + 13b \pm \sqrt{a^2 - 78ab - 15b^2})$ ,  $w_2$  and 0. This shows that  $\mathbf{B}$  is diagonalizable except when  $a/b = 39 \pm 16\sqrt{6}$ . One can analyze these cases separately using a basis of generalized eigenvectors and show that the right hand side of equation (2.21) still converges to  $\mathbf{e}_{w_2}$ .

3) Fix a maximal degree  $d$ . Choose partitioning weights

$$w_{1,i} = w_{i,1} = (i-1)^{-1}, \quad i = 2, \dots, d, \\ w_{2,d} = w_{d,2} = d^{-1}$$

and all other weights equal to zero. The splitting weights are then  $w_i = 1$  for  $i = 1, \dots, d$ . Note that if we take the limit  $d \rightarrow \infty$  we get a special case of the preferential attachment model. These weights satisfy the conditions in Lemma 2.2. The nonzero matrix elements of  $\mathbf{B}$  are

$$B_{i+1,i} = B_{1,i} = -B_{i,i} = B_{2,1} = B_{2,d} = 1, \quad 1 < i < d. \quad (2.38)$$

The characteristic polynomial of  $\mathbf{B}$  is

$$p_d(\lambda) = (-1)^d (1 - \lambda) \left(1 - (1 + \lambda)^{d-1}\right) \quad (2.39)$$

which can easily be proved by induction. The roots of the characteristic

polynomial are  $\lambda = 1$  and  $\lambda = \exp\left(\frac{2\pi ik}{d-1}\right) - 1$ ,  $k = 1, \dots, d-1$  and they are all distinct which shows that  $\mathbf{B}$  is diagonalizable. The solution to (2.34) is

$$\rho_k = \frac{2^{d-k+\delta_{kd}-1} - \delta_{k1}}{2^{d-1} - 1}, \quad k = 1, \dots, d. \quad (2.40)$$

**2.4. Discussion of the conditions on the weights.** It is not obvious how restrictive the condition that  $\mathbf{B}$  must be diagonalizable is regarding the collection of weights one can consider. In the previous subsection we saw that for  $d = 3$  and  $d = 4$  the condition was not very restrictive. Also we saw that for every  $d$  there is at least one choice of weights which satisfies the conditions in Lemma 2.2 and yields a diagonalizable matrix  $\mathbf{B}$ . We will now show that this guarantees that almost all weights give a diagonalizable  $\mathbf{B}$ .

Fix a maximal degree  $d$ . Let  $\mathcal{B}_d$  be the set of matrices  $\mathbf{B}$  which correspond to partitioning weights that give linear splitting weights and satisfy the conditions in Lemma 2.2. It is clear that if  $\mathbf{B}, \mathbf{B}' \in \mathcal{B}_d$  then  $t\mathbf{B} + (1-t)\mathbf{B}' \in \mathcal{B}_d$  for all  $t \in [0, 1]$  and so  $\mathcal{B}_d$  is convex. Let

$$\mathcal{B}'_d = \{\mathbf{B} \in \mathcal{B}_d \mid \mathbf{B} \text{ is diagonalizable.}\}.$$

From the previous subsection we know that  $\mathcal{B}'_d \neq \emptyset$ . Since  $\mathcal{B}_d$  is convex and  $\mathcal{B}'_d \neq \emptyset$  then by Corollary 1 in [14],  $\mathcal{B}'_d$  is dense in  $\mathcal{B}_d$ .

We believe that it is possible to extend the result of convergence of the right hand side of (2.21) to all partitioning weights giving linear splitting weights, relaxing both the condition of diagonalizability of  $\mathbf{B}$  and condition (2) in Lemma 2.2. We also believe, in view of simulations, that equation (2.34) even describes correctly the vertex degree distribution for non-linear splitting weights and for the case  $d = \infty$ . We will look at this more closely in the next two subsections.

**2.5. Mean field equation for general weights.** To generalize equation (2.34) beyond the case of linear splitting weights we notice that Lemmas 2.1 and 2.2 do not rely on the linearity of the weights except in the conclusion of Lemma 2.2 where we show that  $w = w_2$ . We therefore conjecture that in general the limiting vertex degree densities are the unique positive solution to

$$\rho_k = -\frac{w_k}{w}\rho_k + \sum_{i=k-1}^d i \frac{w_{k,i+2-k}}{w} \rho_i, \quad (2.41)$$

subject to the constraints

$$\rho_1 + \dots + \rho_d = 1 \quad (2.42)$$

$$w_1\rho_1 + \dots + w_d\rho_d = w. \quad (2.43)$$

Recall that  $w$  is the unique simple positive eigenvalue of  $\mathbf{B}$  defined in (2.20) with the largest real part of all the eigenvalues and  $\rho_k$ ,  $k = 1, \dots, d$  are the components of the associated eigenvector with the proper normalization.

The existence and uniqueness of a positive solution to (2.41) satisfying (2.42) and (2.43) follows from the Perron-Frobenius argument in the proof of Lemma 2.2. In order to distinguish (2.41) from (2.34) we refer to it as the *mean field equation for vertex degree densities*. One can also arrive directly at this equation by assuming that for large  $t$  an equilibrium with small enough fluctuations is established, and then performing the splitting procedure on this equilibrium.

The solution to the mean field equation for the  $d = 3$  model and uniform partitioning weights is

$$\rho_3 = \frac{7\alpha - \sqrt{\alpha(\alpha + 24\beta + 24)}}{6(2\alpha - \beta - 1)} \quad (2.44)$$

where  $\alpha = \frac{w_2}{w_1}$  and  $\beta = \frac{w_3}{w_1}$ . Note that from the constraints we have  $\rho_1 = \rho_3$  and  $\rho_2 = 1 - 2\rho_3$ . This solution (and solutions in general) only depends on the ratio of the weights. In Figure 2 we compare the above solution to simulations.

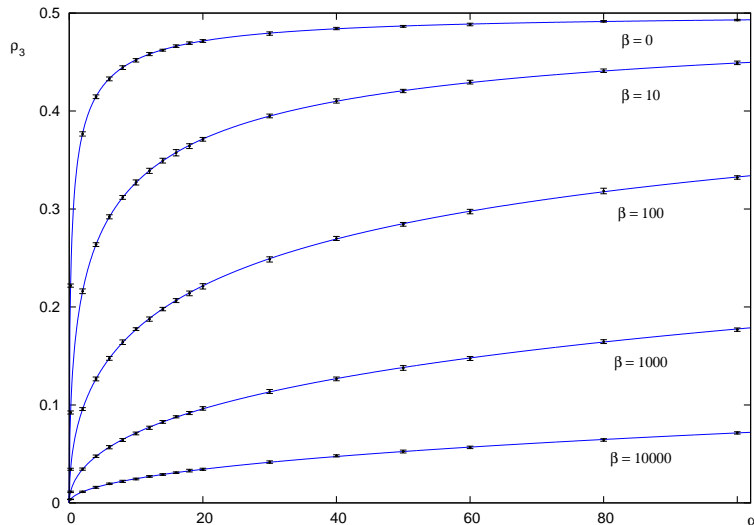


FIGURE 2. The value of  $\rho_3$  as given in (2.44) compared to results from simulations. Each point is calculated from 20 trees on 10000 vertices.

**2.6. The  $d = \infty$  model with linear weights.** In this subsection we drop the assumption that there is an upper bound on the vertex degrees but we still assume that all vertex degrees are finite. If we assume that equation (2.34) holds for  $d = \infty$ , then it is possible to find an exact solution in the case of linear splitting weights,  $w_i = ai + b$ , and uniform partitioning weights.

Equation (2.34) becomes

$$\rho_k = -\frac{w_k}{w_2}\rho_k + \sum_{i=k-1}^{\infty} \frac{2}{i+1} \frac{w_i}{w_2} \rho_i. \quad (2.45)$$

Subtracting from this the same equation for  $\rho_{k+1}$  we find

$$\rho_k \left(1 + \frac{w_k}{w_2}\right) - \rho_{k+1} \left(1 + \frac{w_{k+1}}{w_2}\right) = \frac{2}{k} \frac{w_{k-1}}{w_2} \rho_{k-1}. \quad (2.46)$$

Let  $x = b/a$ . The recursion (2.46) has the solution

$$\rho_k(x) = \begin{cases} \frac{2}{C(-1)} & \text{if } x = -1 \text{ and } k = 1 \\ \frac{1}{C(x)} \frac{2^{k-1} \Gamma(k+x)}{\Gamma(k) \Gamma(k+3+2x)} (k+1+2x) & \text{otherwise,} \end{cases}$$

where

$$C(x) = \frac{e\sqrt{\pi} 2^{-\frac{3}{2}-x} I_{\frac{1}{2}+x}(1)}{2+x} \quad (2.47)$$

is a normalization constant such that  $\sum_i \rho_i = 1$ . Here,  $I_\nu$  is the modified Bessel function of the first kind. The variable  $x$  can take values from  $-1$  to  $\infty$ . The asymptotic behaviour of  $\rho_k(x)$  for large  $k$  is

$$\rho_k(x) = \frac{1}{C(x)} \frac{1}{k!} 2^{k-1} k^{-1-x} \left(1 + O\left(\frac{1}{k}\right)\right). \quad (2.48)$$

The special case  $x = \infty$  corresponds to constant weights for which the solution is

$$\rho_k(\infty) = \frac{1}{e} \frac{1}{(k-1)!}. \quad (2.49)$$

In Figure 3 we compare the above solutions to simulations for five different values of  $x$ .

### 3. SUBTREE STRUCTURE PROBABILITIES

In this section we consider the model in which vertices are labeled with their time of creation as explained in the definition of the splitting process (item (iv)). For convenience we will take time to be the number of edges in a tree, denoted by  $\ell$ , starting from the single vertex tree at time 0 and adding one link and one vertex at each time step. We consider only linear splitting weights  $w_i = ai + b$  but comment on generalizations in the next section. To simplify the notation we define

$$W(\ell) \equiv \mathcal{W}(T) - w_1 = (2a + b)\ell - a \quad (3.1)$$

where the last equality follows from the linearity of the weights.

We derive exact expressions for probabilities of particular subtree structures as seen from the vertex created at a given time. By averaging over these probabilities and assuming the existence of a scaling limit, we shall show how to extract the Hausdorff dimension of the tree and derive bounds

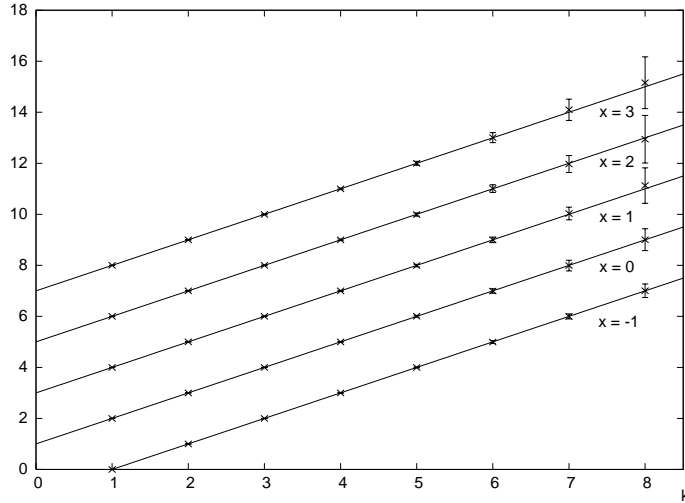


FIGURE 3. The solid lines are  $k + 1 + 2x$  plotted against  $k$  for five different values of  $x$  in comparison to results from simulations of 100 trees on  $10^6$  vertices. The datapoints on the graph are calculated, for a given  $k$  and  $x$ , from the measured density  $\rho_{k,\text{meas.}}(x)$  by

$$C(x) \frac{\Gamma(k) \Gamma(k + 3 + 2x)}{2^{k-1} \Gamma(k + x)} \rho_{k,\text{meas.}}(x)$$

with an obvious modification if  $x = -1$ .

on this dimension. In special cases we give an exact expression for the Hausdorff dimension.

**3.1. Probabilities related to subtree structure.** Consider a tree on  $\ell$  edges generated with the splitting procedure starting from the single vertex tree at time 0. Let  $p_R(\ell; s)$  be the probability that the vertex created at time  $s$  is the root. If  $s < \ell$  we find that

$$p_R(\ell; s) = \frac{1}{W(\ell - 1) + w_1} W(\ell - 1) p_R(\ell - 1; s), \quad (3.2)$$

since we can split any vertex except the root in order to get from a tree at time  $\ell - 1$  to a tree at time  $\ell$ . This contributes the factor  $p_R(\ell - 1; s)$  to  $p_R(\ell; s)$ . Similarly,

$$p_R(\ell; \ell) = \frac{1}{W(\ell - 1) + w_1} \sum_{s=0}^{\ell-1} w_1 p_R(\ell - 1; s), \quad (3.3)$$

since if we create a new root vertex at time  $\ell$  the previous root vertex, labelled  $s$  in (3.3) could have been created at any time before  $\ell$ . We depict these processes in Fig. 4.

$$\begin{aligned} \text{Diagram 1: } & \text{Vertex } s \text{ --- Circle } l = \frac{W(\ell-1)}{W(\ell-1) + w_1} \text{ Vertex } s \text{ --- Circle } \ell-1 \\ \text{Diagram 2: } & \text{Vertex } l \text{ --- Circle } l = \frac{1}{W(\ell-1) + w_1} \sum_{s=0}^{\ell-1} w_1 \text{ Vertex } s \text{ --- Circle } \ell-1 \end{aligned}$$

FIGURE 4. Diagrams representing equations (3.2) and (3.3).

If  $v$  is a vertex of order  $k$  in a tree  $T$ , then there is a unique link  $\ell_1$  incident on  $v$  leading towards the root (unless  $v$  is the root). Let  $\ell_2, \dots, \ell_k$  be the other links incident on  $v$ . The largest subtree of  $T$  which contains the root and  $\ell_1$  but none of the links  $\ell_i$  with  $i \geq 2$  will be called the *left subtree* (with respect to  $v$ ). The maximal subtrees which contain one  $\ell_j$  with  $j \neq 1$  and no other link  $\ell_i$  will be called the *right subtrees* (with respect to  $v$ ). If  $k = 1$  then there are of course no right subtrees and if  $v$  is the root then we view the left subtree as being empty. Let  $p_k(\ell_1, \dots, \ell_k; s)$  denote the probability that the vertex created at time  $s$  has a left subtree on  $\ell_1$  edges and right subtrees on  $\ell_2, \dots, \ell_k$  edges, where  $\ell_1 + \dots + \ell_k = \ell$ . By the nature of the splitting operation,  $p_k(\ell_1, \ell_2, \dots, \ell_k; s)$  is symmetric under permutations of  $(\ell_2, \dots, \ell_k)$ . We will sometimes refer to the vertex created at time  $s$  as the  $s$ -vertex.

By the definition of the relabeling when we split we have

$$p_1(\ell; \ell) = 0, \quad (3.4)$$

because the vertex closer to the root gets a new label and therefore no leaf except the root can have the maximal label. In the case  $s < \ell$  we find the recursion

$$\begin{aligned} p_1(\ell; s) &= \frac{1}{W(\ell-1) + w_1} \left[ W(\ell-1)p_1(\ell-1; s) \right. \\ &\quad \left. + \sum_{i=1}^{d-1} i w_{i+1,1} \sum_{\ell'_1 + \dots + \ell'_i = \ell-1} p_i(\ell'_1, \dots, \ell'_i; s) + \delta_{\ell 1} w_1 \right]. \end{aligned} \quad (3.5)$$

The first term in the square bracket corresponds to the case when we do not split the vertex with label  $s$ . The second term corresponds to splitting the  $s$ -vertex which can have any order up to  $d-1$ . Finally the last term corresponds to the special case when we have  $\ell = 1$  so the  $s$ -vertex is the root of the trivial tree, see Fig. 5.



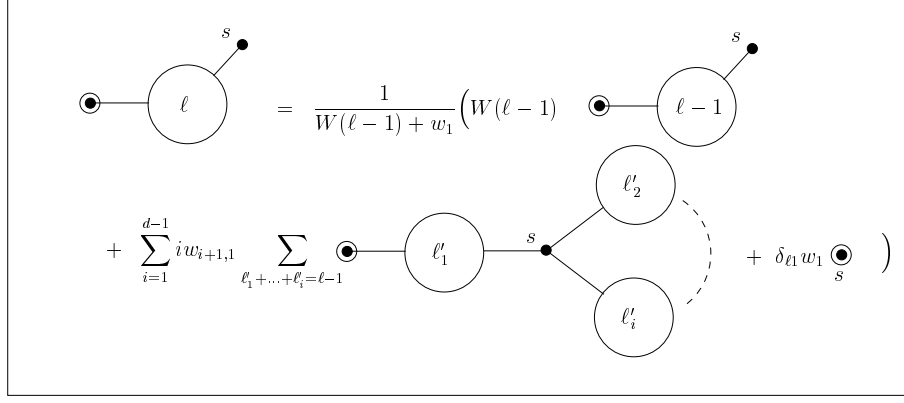


FIGURE 5. A diagram representing equation (3.5).

For a general  $k \geq 2$  and  $s < \ell$  the recursion can be written

$$\begin{aligned}
 p_k(\ell_1, \dots, \ell_k; s) &= \frac{1}{W(\ell-1) + w_1} \times \\
 &\left[ \delta_{k2} \delta_{\ell_1 1} w_1 p_R(\ell-1; s) + \sum_{i=1}^k W(\ell_i - 1) p_k(\ell_1, \dots, \ell_i - 1, \dots, \ell_k; s) \right. \\
 &\left. + \sum_{i=k}^d (i+1-k) w_{k, i-k+2} \sum_{\ell'_1 + \dots + \ell'_{i+1-k} = \ell_1 - 1} p_i(\ell'_1, \dots, \ell'_{i+1-k}, \ell_2, \dots, \ell_k; s) \right],
 \end{aligned} \tag{3.6}$$

see Fig.6. The first term corresponds to the case when the  $s$ -vertex is the root before the splitting in which case we have  $\ell_1 = 1$  and  $k = 2$ . The second term corresponds to the case when we split a vertex different from the  $s$ -vertex and the last term arises when we split the  $s$ -vertex in the step from time  $\ell - 1$  to time  $\ell$ . Finally we have

$$\begin{aligned}
 p_k(\ell_1, \dots, \ell_k; \ell) &= \frac{1}{W(\ell-1) + w_1} \times \\
 &\sum_{s=0}^{\ell-1} \sum_{j=2}^k \sum_{i=k-1}^{d-1} \sum_{\substack{\ell'_1 + \dots + \ell'_{i+1-k} \\ = \ell_j - 1}} w_{k, i-k+2} p_i(\ell_1, \dots, \ell_{j-1}, \ell'_1, \dots, \ell'_{i+1-k}, \ell_{j+1}, \dots, \ell_k; s),
 \end{aligned} \tag{3.7}$$

where  $\ell_1 + \dots + \ell_k = \ell$ , see Fig.7. Here  $s$  is the label of the vertex that is split in the step from time  $\ell - 1$  to time  $\ell$  and we sum over all possible degrees of the  $s$ -vertex and all ways of splitting it.

We define the following mean probabilities by averaging over the vertex labels in (3.2–3.7)

$$p_R(\ell) = \frac{1}{\ell+1} \sum_{s=0}^{\ell} p_R(\ell; s) \tag{3.8}$$

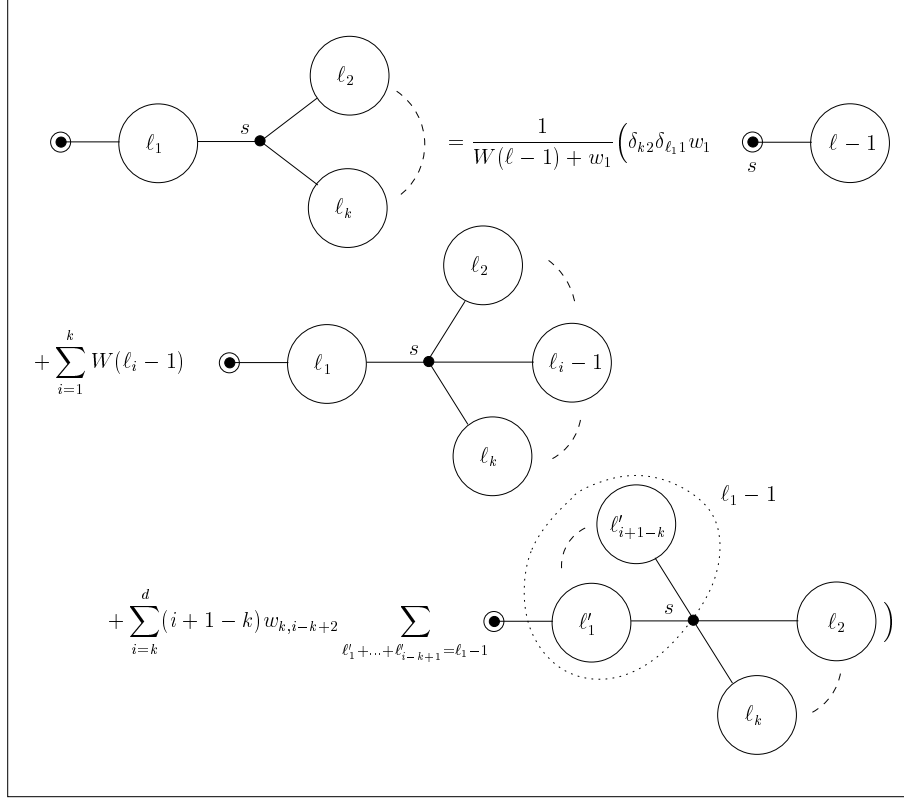


FIGURE 6. A diagram representing equation (3.6).

and

$$p_k(\ell_1, \dots, \ell_k) = \frac{1}{\ell + 1} \sum_{s=0}^{\ell} p_k(\ell_1, \dots, \ell_k; s), \quad (3.9)$$

where  $\ell_1 + \dots + \ell_k = \ell$ . From (3.8) we get a recursion for the mean probabilities, going from time  $\ell$  to  $\ell + 1$

$$p_R(\ell + 1) = \frac{\ell + 1}{\ell + 2} p_R(\ell). \quad (3.10)$$

For  $k = 1$  we obtain from (3.4), (3.5) and (3.9)

$$\begin{aligned} p_1(\ell + 1) & \quad (3.11) \\ &= \frac{\ell + 1}{\ell + 2} \frac{1}{W(\ell) + w_1} \left[ W(\ell) p_1(\ell) + \sum_{i=1}^{d-1} i w_{i+1,1} \sum_{\substack{\ell'_1 + \dots + \ell'_i \\ = \ell}} p_i(\ell'_1, \dots, \ell'_i) + 2\delta_{\ell 0} w_1 \right]. \end{aligned}$$

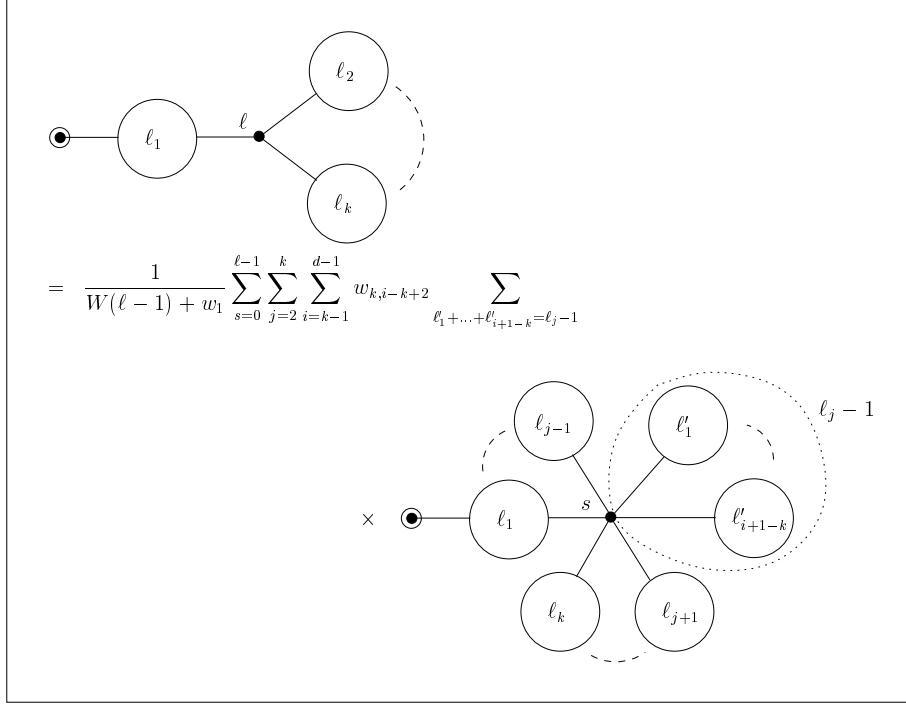


FIGURE 7. A diagram representing equation (3.7).

Finally, the general case for  $k \geq 2$  is

$$\begin{aligned}
& p_k(\ell_1, \dots, \ell_k) \\
&= \frac{\ell + 1}{\ell + 2} \frac{1}{W(\ell) + w_1} \left[ \delta_{k2} \delta_{\ell_1 1} w_1 p_R(\ell) + \sum_{i=1}^k W(\ell_i - 1) p_k(\ell_1, \dots, \ell_i - 1, \dots, \ell_k) \right. \\
&\quad \left. + \sum_{i=k}^d (i - k + 1) w_{k,i-k+2} \sum_{\substack{\ell'_1 + \dots + \ell'_{i+1-k} \\ = \ell_1 - 1}} p_i(\ell'_1, \dots, \ell'_{i+1-k}, \ell_2, \dots, \ell_k) \right] \quad (3.12) \\
&\quad \left. + \sum_{j=2}^k \sum_{i=k-1}^d w_{k,i-k+2} \sum_{\substack{\ell'_1 + \dots + \ell'_{i+1-k} \\ = \ell_j - 1}} p_i(\ell_1, \dots, \ell_{j-1}, \ell'_1, \dots, \ell'_{i+1-k}, \ell_{j+1}, \dots, \ell_k) \right]
\end{aligned}$$

where  $\ell_1 + \dots + \ell_k = \ell + 1$  and we have made use of (3.6), (3.7) and (3.9).

**3.2. Two-point functions.** One can reduce the above recursion formulas for the mean probabilities to simpler recursion formulas which suffice for the determination of the Hausdorff dimension. Define the *two-point functions*

$$q_{ki}(\ell_1, \ell_2) = \sum_{\ell'_1 + \dots + \ell'_{k-i} = \ell_1} \sum_{\ell''_1 + \dots + \ell''_i = \ell_2} p_k(\ell'_1, \dots, \ell'_{k-i}, \ell''_1, \dots, \ell''_i), \quad (3.13)$$

where  $k = 2, \dots, d$  and  $i = 1, \dots, k-1$ . In total there are  $d(d-1)/2$  of these functions. If we define

$$q_{1,0}(\ell_1, \ell_2) = \delta_{\ell_2,0} \delta_{\ell_1,\ell} p_1(\ell_1 + \ell_2)$$

then  $q_{ki}(\ell_1, \ell_2)$  is the probability that  $i$  right trees of total volume  $\ell_2$  are attached to a vertex of degree  $k$  in a tree of total volume  $\ell_1 + \ell_2$ . By summing over the equations in the previous section we get

$$\begin{aligned} q_{ki}(\ell_1, \ell_2) &= \frac{\ell+1}{\ell+2} \frac{1}{W(\ell) + w_1} \left[ \right. \\ &\quad \sum_{j=k-1}^d w_{k,j+2-k} \left( (j-i)q_{ji}(\ell_1-1, \ell_2) + iq_{j,j-(k-i)}(\ell_1, \ell_2-1) \right) \\ &\quad + \left( W(\ell_1-1) + (k-i-1)(w_2-w_3) \right) q_{ki}(\ell_1-1, \ell_2) \\ &\quad + \left( W(\ell_2-1) + (i-1)(w_2-w_3) \right) q_{ki}(\ell_1, \ell_2-1) \\ &\quad \left. + \delta_{k,2} \delta_{\ell_1,1} w_1 p_R(\ell_2) + \delta_{i,1} \delta_{\ell_2,1} w_{k,1} \sum_{\ell'_1+\dots+\ell'_{k-1}=\ell_1} p_{k-1}(\ell'_1, \dots, \ell'_{k-1}) \right] \end{aligned} \tag{3.14}$$

with  $\ell_1 + \ell_2 = \ell + 1$ . We see that the two-point functions satisfy an essentially closed system of equations. The last two terms in (3.14) do not contribute to the scaling limit which will be discussed in the next section.

#### 4. HAUSDORFF DIMENSION

In this section we relate the two-point functions defined in the previous section to the size of trees, defined in a suitable way. With the help of some scaling assumptions this relation allows us to calculate the Hausdorff dimension of the trees as a function of the partitioning weights in simple cases. As in Section 3 we assume that the weights  $w_i$  are linear in  $i$  and we shall comment on the general case at the end of the section.

**4.1. Definition.** Let  $T$  be a tree with  $\ell$  edges and  $v$  and  $w$  two vertices of  $T$ . The (intrinsic geodesic) distance  $d_T(v, w)$  between  $v$  and  $w$  is the number of edges that separate  $v$  from  $w$ . We define the radius of  $T$  with respect to the vertex  $v$  as

$$R_T(v) = \frac{1}{2\ell} \sum_w d_T(v, w) k(w), \tag{4.1}$$

where  $k(w)$  is the degree of the vertex  $w$ . Notice that  $2\ell = \sum_w k(w)$ . The global radius of  $T$  is

$$R_T = \frac{1}{2\ell} \sum_v R_T(v) k(v). \tag{4.2}$$

We define the Hausdorff dimension of the tree,  $d_H$ , from the scaling law for large trees

$$\langle R_T \rangle \sim \ell^{1/d_H} \quad \ell \rightarrow \infty. \quad (4.3)$$

The more usual definition of the local fractal dimension of a vertex  $v$  of the tree,  $d_f(v)$ , is defined by the growth of the volume of a ball of radius  $r$  around  $v$ ,  $\mathcal{B}_v(r) = \{w : d(v, w) \leq r\}$

$$\langle \text{Card}(\mathcal{B}_v(r)) \rangle \sim r^{d_f(v)} \quad 1 \ll r \ll \ell. \quad (4.4)$$

These two definitions are expected to coincide provided that the tree is a homogeneous fractal (no multifractal behaviour). We expect that the scaling behaviour (4.4) is valid independently of the point  $v$  and also that the scaling (4.3) holds for  $R_T(v)$  defined by (4.1) irrespective of the point  $v$  chosen.

**4.2. Geodesic distances and 2 point functions.** The radius  $R_T(v)$  can be extracted from the two point functions calculated in the previous section. Let  $T$  be a tree and  $v$  a vertex of  $T$ . Let  $i$  be an edge of  $T$ . If we cut this edge then the tree is split into two connected components, a tree  $T_1$  which contains  $v$  and a tree  $T_2$  that does not contain  $v$  (see Figure 8). Let  $\ell_2(v; i)$  be the number of edges of  $T_2$ . We have the simple remarkable result

$$\sum_w d_T(v, w) k(w) = \sum_i (2\ell_2(v; i) + 1) \quad (4.5)$$

which we will now prove. For the tree  $T$  with  $\ell$  edges, we may assign two

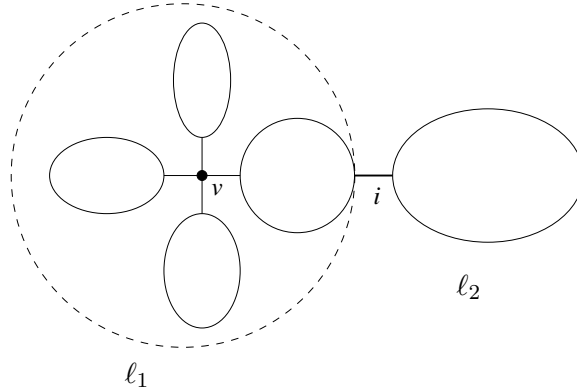


FIGURE 8. Cutting a tree along the edge  $i$ .

labels to every edge in the following way. Starting from  $v$ , we walk around the tree while always keeping the tree to the left. Drop the labels 1 to  $2\ell$  on the sides of edges as we pass them.

An example of such a walk and labelling is shown in Figure 9. Let us mention that the initial direction from  $v$  is unimportant. In what follows we will denote these new labels by greek letters.

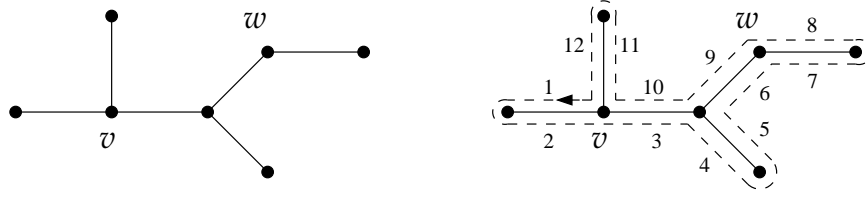


FIGURE 9. A tree and its labels.

Given  $1 \leq \alpha < \beta \leq 2\ell$ , define  $\phi_v(\alpha, \beta)$  to be 1 if  $\alpha$  and  $\beta$  are labels of the same edge, and zero otherwise. In the above example we have  $\phi_v(6, 9) = 1$  whereas  $\phi_v(6, 12) = 0$ . For any vertex  $w \in T$ , let us define  $\omega(w)$  to be the smallest label of the edges adjacent to  $w$ . In the example above  $\omega(w) = 6$  and  $\omega(v) = 1$ .

We now have for any  $w \in T$

$$d_T(v, w) = \sum_{\alpha, \beta: \alpha \leq \omega(w) < \beta} \phi_v(\alpha, \beta) \quad (4.6)$$

and it is easy to see that

$$\sum_{w \in T} d_T(v, w) k(w) = \sum_{\alpha, \beta, \gamma: \alpha \leq \gamma < \beta} \phi_v(\alpha, \beta) = \sum_{\alpha, \beta: \alpha < \beta} \phi_v(\alpha, \beta) (\beta - \alpha). \quad (4.7)$$

If  $\phi(\alpha, \beta) = 1$ , i.e. if  $\alpha$  and  $\beta$  correspond to the two faces of the edge  $a$ , then

$$\beta - \alpha = 2\ell_2(v, a) + 1 \quad (4.8)$$

and equation (4.5) follows.

We now apply (4.5) by choosing for  $v$  the root  $r$  of the tree and averaging over all trees obtained by the splitting process. We notice that the average number of links giving the volume  $\ell_2(r, i)$  is simply the number of links  $\ell$  times the proportion of vertices  $w$  which have a left tree (containing the root  $r$ ) on  $\ell_1 = \ell - \ell_2$  edges and an arbitrary number of right trees (on a total number of  $\ell_2$  edges). We obtain

$$\langle R_T(r) \rangle = \frac{\ell + 1}{2\ell} \sum_{\ell_2=0}^{\ell} (2\ell_2 + 1) \sum_{k=1}^d q_{k, k-1}(\ell - \ell_2, \ell_2), \quad (4.9)$$

see Fig. 10 for a simple diagrammatic representation of this identity. Thus, if we know how the two point functions  $q_{k,i}(\ell_1, \ell_2)$  scale for large  $\ell$ , we know how the radius of the tree scales with  $\ell$  and we can compute the Hausdorff dimension  $d_H$ .

**4.3. Scaling and the Hausdorff dimension.** We assume that the following scaling holds for the two-point functions  $q_{ki}$  with  $\ell$  large

$$q_{ki}(\ell_1, \ell_2) = \ell^{-\rho} (\bar{\omega}_{ki}(x) + \bar{\gamma}_{ki}(x)\ell^{-1} + O(\ell^{-2})) \quad (4.10)$$

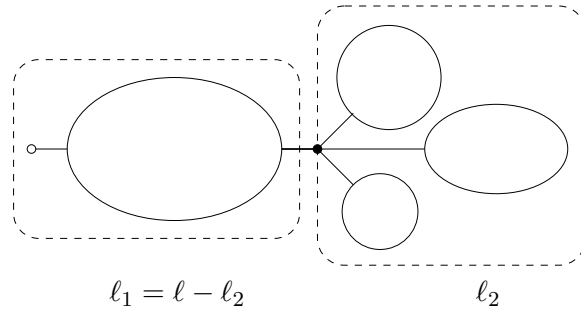


FIGURE 10. Illustration of eq. 4.9.

where  $\ell_1 + \ell_2 = \ell$ ,  $x = \ell_1/\ell \in ]0, 1[$  and where  $\bar{\omega}_{ki}, \bar{\gamma}_{ki}$  are some functions. It must hold that  $\bar{\omega}_{ki} > 0$  and we assume that the scaling exponent  $\rho$  satisfies

$$1 < \rho \leq 2. \quad (4.11)$$

Note that for  $\ell$  finite, the probabilities  $q_{k,i}(\ell_1, \ell_2)$  are of order  $\ell^{-1}$  when  $\ell_1$  is of order 1 and are of order 1 when  $\ell_2$  is of order 1. This implies that the scaling functions  $\bar{\omega}_{ki}(x)$  should scale when  $x \rightarrow 0$  or  $x \rightarrow 1$ , respectively, as

$$\bar{\omega}_{ki}(x) \underset{x \rightarrow 0}{\sim} x^{1-\rho} \quad \text{and} \quad \bar{\omega}_{ki}(x) \underset{x \rightarrow 1}{\sim} (1-x)^{-\rho}. \quad (4.12)$$

Using this ansatz and (4.9) the mean radius scales as

$$\langle R_T(r) \rangle \simeq \ell^{2-\rho} C, \quad C = \int_0^1 dx (1-x) \bar{\omega}(x), \quad \bar{\omega}(x) = \sum_k \bar{\omega}_{k,k-1}(x). \quad (4.13)$$

Equations (4.12) and (4.11) ensure that the integral  $C$  is convergent when  $\rho < 2$ . Equation (4.3) then implies that the Hausdorff dimension of the tree is given by

$$2 - \rho = \frac{1}{d_H}. \quad (4.14)$$

For  $\rho = 2$  we see that  $C$  is logarithmically divergent and this corresponds to an infinite Hausdorff dimension.

Inserting (4.10) into the recursion equation (3.14) for the two point functions and expanding in  $\ell^{-1}$  gives (dropping the function argument  $x$  in an

obvious way)

$$\begin{aligned}
& \bar{\omega}_{ki} - \rho \bar{\omega}_{ki} \ell^{-1} + \bar{\gamma}_{ki} \ell^{-1} + O(\ell^{-2}) \\
&= \frac{1}{w_2} \ell^{-1} \left( 1 - \frac{w_1 + 2w_2 - w_3}{w_2} \ell^{-1} + O(\ell^{-2}) \right) \\
&\quad \times \left[ \sum_{j=k-1}^d w_{k,j+2-k} \left( (j-i) \bar{\omega}_{ji} + i \bar{\omega}_{j,j-(k-i)} + O(\ell^{-1}) \right) \right. \\
&\quad + \ell \left( w_2 x + (-w_3 + (k-i-1)(w_2 - w_3)) \ell^{-1} \right) \left( \bar{\omega}_{ki} + \bar{\gamma}_{ki} \ell^{-1} + O(\ell^{-2}) \right) \\
&\quad \left. + \ell \left( w_2(1-x) + i(w_2 - w_3) \ell^{-1} \right) \left( \bar{\omega}_{ki} + \bar{\gamma}_{ki} \ell^{-1} + O(\ell^{-2}) \right) \right]. \quad (4.15)
\end{aligned}$$

The equation is trivially satisfied in zeroth order of  $\ell^{-1}$ . When we go to the next order we see that the following must hold

$$(2 - \rho) \bar{\omega}_{ki} = \frac{1}{w_2} \sum_{j=k-1}^d w_{k,j+2-k} \left( (j-i) \bar{\omega}_{ji} + i \bar{\omega}_{j,j-(k-i)} \right) - \frac{w_k}{w_2} \bar{\omega}_{ki}. \quad (4.16)$$

This eigenvalue equation may be rewritten as

$$\mathbf{C} \boldsymbol{\omega} = w_2(2 - \rho) \boldsymbol{\omega} \quad (4.17)$$

where  $\mathbf{C}$  is a  $\binom{d}{2} \times \binom{d}{2}$  matrix indexed by a pair of two indices  $ki$  with  $k > i$ ,  $k = 2, \dots, d$  and  $\boldsymbol{\omega}$  is a vector with two such indices. The matrix elements of  $\mathbf{C}$  are

$$C_{ki,jn} = w_{k,j+2-k} \left( (j-i) \delta_{in} + i \delta_{n,j-(k-i)} \right) - w_k \delta_{kj} \delta_{in}. \quad (4.18)$$

We use the convention that  $w_{i,j} = 0$  if  $i$  or  $j$  is less than 1 or greater than  $d$ . Thus,  $w_2(2 - \rho)$  is an eigenvalue of the matrix  $\mathbf{C}$  and the associated eigenvector must have components  $\geq 0$ . We now show that there is in general a unique solution to this eigenvalue problem.

Since the only possibly negative elements of  $\mathbf{C}$  are on the diagonal we can make the matrix non-negative by adding a positive multiple  $\gamma$  of the identity to both sides of (4.17) and choosing  $\gamma$  large enough.

If enough of the weights  $w_{i,j}$  are non-zero ( $w_{1,i} > 0$  for  $2 \leq i \leq d$  and  $w_{j,3} > 0$  for  $2 \leq j \leq d-1$  is for example sufficient) then one can check that the matrix  $\mathbf{C} + \gamma \mathbf{I}$  is primitive. Then, by the Perron-Frobenius theorem, it has a simple positive eigenvalue of largest modulus and its corresponding eigenvector can be taken to have all entries positive cf. Lemma 2.2. Therefore this largest positive eigenvalue gives the  $\rho$  we are after.

**4.4. An upper bound on the Hausdorff dimension.** We can get an upper bound on  $\rho$  by a straight forward estimate from (4.16). The off-diagonal terms in the sum are all non-negative so we disregard them and get the inequality



$$\rho \leq 2 - \left( k \frac{w_{k,2}}{w_2} - \frac{w_k}{w_2} \right), \quad k = 2, \dots, d. \quad (4.19)$$

Since  $1 < \rho \leq 2$  and for  $k \geq 3$

$$w_k = kw_{k,2} + \frac{k}{2} \sum_{\substack{i=1 \\ i \neq 2, i \neq k}}^{k+1} w_{i,k+2-i} > kw_{k,2} \quad (4.20)$$

the best we can get from this upper bound is when  $k = 2$  which yields

$$\rho \leq 2 - \frac{w_{2,2} - 2w_{1,3}}{w_{2,2} + 2w_{1,3}}. \quad (4.21)$$

Now,  $2 - \rho = \frac{1}{d_H}$  and therefore, if  $w_{2,2} > 2w_{1,3}$ , we obtain the upper bound

$$d_H \leq \frac{w_{2,2} + 2w_{1,3}}{w_{2,2} - 2w_{1,3}}. \quad (4.22)$$

If  $w_{2,2} \leq 2w_{1,3}$  the upper bound in (4.21) gives no information about the Hausdorff dimension.

It is easy to verify that (4.22) is an equality if we choose the weights such that  $w_{i,j} = 0$  if  $i \neq 1$  or  $j \neq 1$  with the exception that  $w_{2,2} > 0$ . This condition means that we only allow vertices to evolve by link attachment, except that we can split vertices of degree 2. With this choice the matrix  $\mathbf{C}$  is lower triangular and we can simply read the eigenvalues from the diagonal. Note that  $\mathbf{C}$  is not primitive in this case and therefore we cannot use the Perron-Frobenius theorem to determine which eigenvalue gives the scaling exponent. However, with these simple weights one can show explicitly that there is precisely one eigenvector with strictly positive components and the corresponding eigenvalue is the one that saturates the inequality in (4.22).

Also note that with this choice we have set  $w_d = 0$  and since the weights are linear,  $w_i = ai + b$ , we have fixed  $a$  and  $b$  so that  $w_i = 1 - \frac{i}{d}$ . Therefore there is only one free parameter which we can choose to be  $w_{2,2}$ . Then we can write the Hausdorff dimension as

$$d_H = \frac{1 - 2/d}{2w_{2,2} - (1 - 2/d)} \quad (4.23)$$

with

$$\frac{1}{2}(1 - 2/d) < w_{2,2} < 1 - 2/d. \quad (4.24)$$

We see that for any  $d$  the Hausdorff dimension can vary continuously from 1 to infinity.

**4.5. Explicit solutions and numerical results for  $d = 3$ .** When the maximal degree is  $d = 3$ , the splitting weights are taken to be linear  $w_i = ai + b$  and the partitioning weights uniform, it is easy to solve equation (4.16) for the Hausdorff dimension. Since the solution only depends on the ratio of the weights there is only one independent variable and we choose it to be  $y := w_3/w_2$  where  $0 \leq y \leq 2$ . The solution is

$$d_H = \frac{3(1 + \sqrt{1 + 16y})}{8y} \quad (4.25)$$

In Figure 11 we compare this equation to results from simulations. The agreement of the simulations with the formula is excellent.

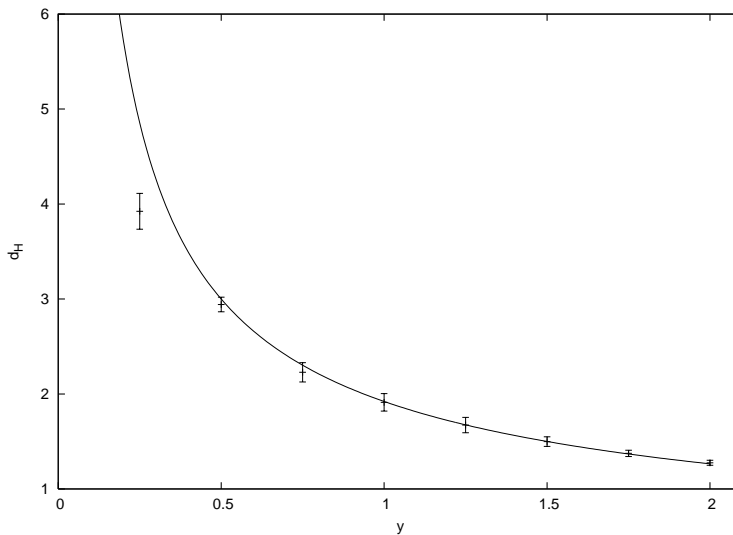


FIGURE 11. Equation (4.25) compared to simulations. The Hausdorff dimension,  $d_H$ , is plotted against  $y = w_3/w_2$ . The two leftmost datapoints are calculated from 50 trees on 50000 vertices and the others are calculated from 50 trees on 10000 vertices. Finite size effects are responsible for the poor fit of the leftmost datapoint.

#### 4.6. Hausdorff dimension for general weights.

**4.6.1. General mean field argument.** Our argument to compute the Hausdorff dimension relies on the recursion relations for the substructure probabilities, studied in Section 3, which are valid only when the splitting weights  $w_i$  are linear functions of the vertex degree  $i$  ( $w_i = ai + b$ ). In this case the total probability weight  $\mathcal{W}(T)$  for a given tree  $T$  depends only on its size  $\ell$  (number of edges) and mean field arguments can be made exact.

In the general case where the  $w_{i,j}$  are arbitrary and the  $w_i$  are not linear with  $i$ , these recursion relations are no more exact. We can use a

mean field argument and assume that they are still valid for large “typical” trees, provided that we replace in these recursion relations the exact weights  $W(\ell) + w_1 = \mathcal{W}(T)$  by their mean field value for large trees

$$W(\ell) + w_1 \rightarrow \overline{W}(\ell) = \sum_j w_j \overline{n}_{\ell+1,j} \quad (4.26)$$

where  $\overline{n}_{\ell+1,j}$  is the average number of  $i$ -vertices in a tree with  $\ell$  edges, studied in Section 2. From the mean field analysis of Section 2.5, we expect that these  $\overline{n}_{\ell+1,j}$  scale with  $\ell$  as

$$\overline{n}_{\ell+1,j} \simeq \ell \rho_j \quad (4.27)$$

with the vertex densities  $\rho_j$  given by the mean field equations (2.41,2.42,2.43) as the components of the eigenvector  $\boldsymbol{\rho}$  associated to the largest eigenvalue  $w$  of the matrix  $\mathbf{B}$ . Thus the mean field approximation amounts to replacing

$$W(\ell) + w_1 \rightarrow \overline{W}(\ell) = w \ell + \dots \quad (4.28)$$

in the recursion relations of Section 3, in particular in the recursion relation (3.14) for the two point function  $q_{ki}$ .

With this assumption, we can repeat the scaling argument of Section 4.3, and we end up with equation (4.16), with the normalisation factor  $\frac{1}{w_2}$  in the r.h.s. replaced by the mean field normalisation factor  $\frac{1}{w}$

$$(2 - \rho) \overline{\omega}_{ki} = \frac{1}{w} \sum_{j=k-1}^d w_{k,j+2-k} \left( (j-i) \overline{\omega}_{ji} + i \overline{\omega}_{j,j-(k-i)} \right) - \frac{w_k}{w} \overline{\omega}_{ki}. \quad (4.29)$$

This equation is still an eigenvalue equation of the form

$$\mathbf{C} \boldsymbol{\omega} = w(2 - \rho) \boldsymbol{\omega} \quad (4.30)$$

where  $\mathbf{C}$  is the  $\binom{d}{2} \times \binom{d}{2}$  matrix with coefficients given in (4.18).

If we denote by  $\chi$  the largest eigenvalue of this matrix  $\mathbf{C}$  and if  $w$  is the largest eigenvalue of  $\mathbf{B}$  then the Perron-Frobenius argument can be applied to show that  $\chi$  is non-negative and that the eigenvector  $\boldsymbol{\omega}$  has non-negative components, which is a consistency requirement for the argument, since the  $\omega_{ki}$  are rescaled probabilities. We end up with a mean field prediction for the Hausdorff dimension of the simple form

$$d_H = \frac{1}{2 - \rho} = \frac{w}{\chi}. \quad (4.31)$$

4.6.2. *General solution for  $d = 3$ .* In the  $d = 3$  case (trees with only  $k = 1, 2$  and 3 vertices), the  $\mathbf{B}$  and  $\mathbf{C}$  matrices are

$$\mathbf{B} = \begin{bmatrix} 0 & 2w_{3,1} & 0 \\ w_{2,1} & w_{2,2} - 2w_{3,1} & 3w_{3,2} \\ 0 & 2w_{3,1} & 0 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} w_{2,2} - 2w_{3,1} & 2w_{3,2} & w_{3,2} \\ w_{3,1} & 0 & 0 \\ 2w_{3,1} & 0 & 0 \end{bmatrix} \quad (4.32)$$

and we find

$$d_H = \frac{(w_{2,2} - 2w_{3,1}) + \sqrt{(w_{2,2} - 2w_{3,1})^2 + 8w_{3,1}(w_{2,1} + 3w_{3,2})}}{(w_{2,2} - 2w_{3,1}) + \sqrt{(w_{2,2} - 2w_{3,1})^2 + 16w_{3,1}w_{3,2}}}. \quad (4.33)$$

4.6.3. *Numerical test of the mean field prediction.* We have tested our prediction when  $d = 3$  and the partitioning weights  $w_i$  are uniform. In this case

$$d_H = \frac{\alpha - \sqrt{\alpha(\alpha + 24 + 24\beta)}}{\alpha - \sqrt{\alpha(\alpha + 16\beta)}} \quad (4.34)$$

where  $\alpha = \frac{w_2}{w_1}$  and  $\beta = \frac{w_3}{w_1}$ . In Figure 12 we compare this equation to results from simulations. There is a good agreement for small values of  $\alpha$  and  $\beta$ , but the precision of the numerics becomes poor for large values of  $\alpha$  and  $\beta$ . This is expected since in this case, the trees will have a large Hausdorff dimension and one must go to very large trees to see the scaling.

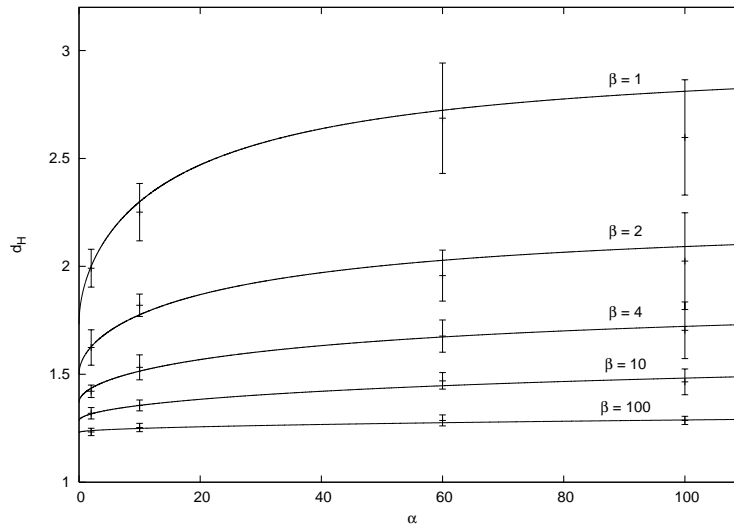


FIGURE 12. Equation (4.34) compared to simulations. Each datapoint is calculated from 50 trees on 10000 vertices.

## 5. CORRELATIONS

Consider a tree on  $\ell$  edges generated by the splitting procedure starting from the single vertex tree at time 0. We are interested in determining the density of edges which have endpoints of degrees  $j$  and  $k$  in the limit when  $\ell \rightarrow \infty$ . Denote this density by  $\rho_{jk}$ , assuming it exists, and for convenience

we let the vertex of degree  $j$  be the one closer to the root. Therefore  $\rho_{1k} = 0$  for all  $k$  and in general  $\rho_{jk} \neq \rho_{kj}$ .

To arrive at these densities we use the same labelling techniques as in Section 3. We start by defining

$$p_{jk}(\ell'_1, \dots, \ell'_{j-1}; \ell''_1, \dots, \ell''_{k-1}; s)$$

as the probability that a vertex created at time  $s$  is of degree  $k$  and has  $\ell''_1, \dots, \ell''_{k-1}$  right trees and that the vertex left to it is of degree  $j$  with an  $\ell'_1$  left tree and  $\ell'_2, \dots, \ell'_{j-1}$  right trees (excluding the right tree which contains  $s$ ). Note that it is symmetric under permutations of both  $(\ell'_2, \dots, \ell'_{j-1})$  and  $(\ell''_1, \dots, \ell''_{k-1})$  and

$$\ell'_1 + \dots + \ell'_{j-1} + \ell''_1 + \dots + \ell''_{k-1} = \ell - 1.$$

We use the methods of Section 3 to derive recursion equations for  $p_{jk}(\ell'_1, \dots, \ell'_{j-1}; \ell''_1, \dots, \ell''_{k-1}; s)$  for linear splitting weights. We then average over the label  $s$  as before and get recursions for the average probabilities  $p_{jk}(\ell'_1, \dots, \ell'_{j-1}; \ell''_1, \dots, \ell''_{k-1})$ . All equations and diagrams are given in Appendix A. Finally we define the densities  $\rho_{jk}(\ell)$  by averaging out the volume dependence of the average probabilities

$$\rho_{jk}(\ell) = \sum_{\ell'_1 + \dots + \ell'_{j-1} + \ell''_1 + \dots + \ell''_{k-1} = \ell - 1} p_{jk}(\ell'_1, \dots, \ell'_{j-1}; \ell''_1, \dots, \ell''_{k-1})$$

and similarly we denote the vertex degree density by

$$\rho_j(\ell) \equiv \rho_{\ell,j} = \sum_{\ell_1 + \dots + \ell_j = \ell} p_j(\ell_1, \dots, \ell_j),$$

cf. Section 2.2. We have the following recursion for the densities

$$\begin{aligned} \rho_{jk}(\ell + 1) = & \frac{\ell + 1}{\ell + 2} \frac{1}{W(\ell) + w_1} \left\{ (\ell w_2 - w_j - w_k + 2w_1 - w_2) \rho_{jk}(\ell) + (j - 1) w_{j,k} \rho_{j+k-2}(\ell) \right. \\ & + (j - 1) \sum_{i=j-1}^d w_{j,i+2-j} \rho_{ik}(\ell) + (k - 1) \sum_{i=k-1}^d w_{k,i+2-k} \rho_{ji}(\ell) \\ & \left. + \delta_{j1} \delta_{\ell'_1, \ell-1} w_1 p_R(\ell) + \delta_{j1} \delta_{\ell 0} w_1 \right\} \end{aligned}$$

for  $i, j \geq 1$ , see Appendix A. Now assume that  $\rho_{jk}(\ell) = \rho_{jk} + r_{jk} \ell^{-1} + O(\ell^{-2})$  and that a similar expansion holds for  $\rho_j(\ell)$ . Expanding the above recursion

in  $\ell^{-1}$  gives

$$\begin{aligned} \rho_{jk} + r_{jk}\ell^{-1} + O(\ell^{-2}) &= \left(1 - \frac{w_1 + 2w_2 - w_3}{w_2}\ell^{-1} + O(\ell^{-2})\right) \left\{ \right. \\ &\quad \left(1 + \frac{-w_j - w_k + 2w_1 - w_2}{w_2}\ell^{-1}\right) (\rho_{jk} + r_{jk}\ell^{-1} + O(\ell^{-2})) \\ &\quad + \frac{\ell^{-1}}{w_2} \left[ (j-1)w_{j,k}(\rho_{j+k-2} + O(\ell^{-1})) + (j-1) \sum_{i=j-1}^d w_{j,i+2-j} (\rho_{ik} + O(\ell^{-1})) \right. \\ &\quad \left. \left. + (k-1) \sum_{i=k-1}^d w_{k,i+2-k} (\rho_{ji} + O(\ell^{-1})) \right] \right\}. \end{aligned}$$

This equation is trivially satisfied in zeroth order of  $\ell^{-1}$ . When we go to the next order we get the following equation for the limits of the densities

$$\begin{aligned} \rho_{jk} &= -\frac{w_j + w_k}{w_2}\rho_{jk} + (j-1)\frac{w_{j,k}}{w_2}\rho_{j+k-2} + (j-1) \sum_{i=j-1}^d \frac{w_{j,i+2-j}}{w_2}\rho_{ik} \\ &\quad + (k-1) \sum_{i=k-1}^d \frac{w_{k,i+2-k}}{w_2}\rho_{ji}. \end{aligned} \tag{5.1}$$

We can also arrive directly at this equation by assuming that for large  $\ell$  an equilibrium with small enough fluctuations is established, and then perform the splitting procedure on this equilibrium. With the same methods, it is possible to derive an equation like (5.1) for the density  $\rho_{j_1, j_2, \dots, j_R}$  of linear paths of length  $R-1$  directed towards the root containing vertices of degrees  $j_1, \dots, j_R$ .

Existence of solutions to equation (5.1) can be established by the Perron-Frobenius argument like in the previous sections. In the following subsections we will find an explicit solution for linear weights and discuss generalizations for non-linear weights. In both cases we compare the results with simulations.

**5.1. Solution in the simplest case.** When  $d = 3$ , the splitting weights are linear and the partitioning weights uniform, we know that  $\rho_1 = \rho_3 = 2/7$  and  $\rho_2 = 3/7$ , see Section 2.3. Let  $y = w_3/w_2$ . Then the solutions to equation (5.1) are

$$\begin{aligned}
\rho_{21} &= \frac{4(3-y)}{7(11-2y)}, & \rho_{31} &= \frac{10}{7(11-2y)}, \\
\rho_{22} &= \frac{4y^2 - 12y + 105}{7(2y+7)(11-2y)}, & \rho_{32} &= \frac{2(-8y^2 + 18y + 63)}{7(2y+7)(11-2y)}, \\
\rho_{23} &= \frac{2(-4y^2 + 20y + 21)}{7(2y+7)(11-2y)}, & \rho_{33} &= \frac{8(3y-14)}{7(2y+7)(2y-11)}.
\end{aligned} \tag{5.2}$$

Note that the following sum rules hold for the solutions

$$\begin{aligned}
\rho_{21} + \rho_{31} &= \rho_1 = 2/7 \\
\rho_{22} + \rho_{32} &= \rho_2 = 3/7 \\
\rho_{23} + \rho_{33} &= \rho_3 = 2/7, \\
\rho_{21} + \rho_{22} + \rho_{23} &= \rho_2 = 3/7 \\
\rho_{31} + \rho_{32} + \rho_{33} &= 2\rho_3 = 4/7.
\end{aligned} \tag{5.3}$$

These relations show that there are only two independent link densities. We plot  $\rho_{21}$  and  $\rho_{22}$  in Figure 13 and compare to simulations.

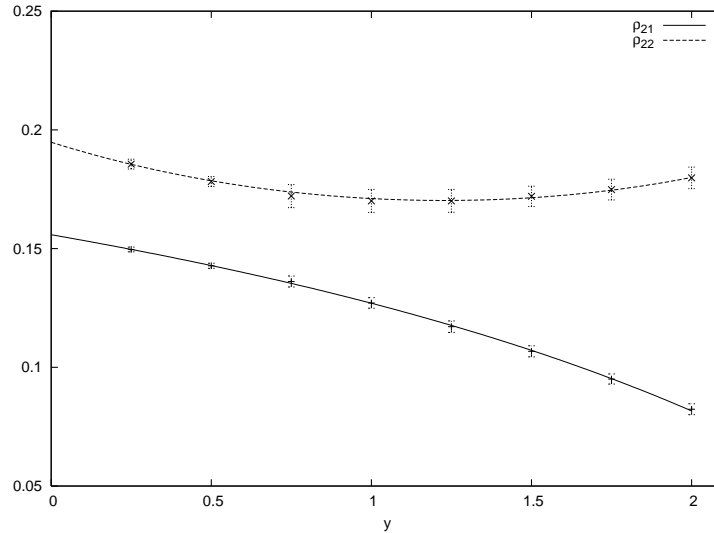


FIGURE 13. Two independent solutions given in (5.2) plotted against  $y = w_3/w_2$  and compared to simulations. The two leftmost datapoints on each line come from simulations of 50 trees on 50000 vertices. The other datapoints come from simulations of 50 trees on 10000 vertices.

We can compare the above results to the case when no correlations are present. Denote the uncorrelated densities by  $\tilde{\rho}_{ij}$ . Then we simply have

$$\tilde{\rho}_{ij} = \frac{\rho_i \rho_j}{1 - \rho_1}.$$

The denominator comes from the fact that the vertex closer to the root is of degree one with probability zero. Inserting the values from (5.3) into this equation gives

$$\begin{aligned}\tilde{\rho}_{21} &= 6/35, & \tilde{\rho}_{31} &= 4/35 \\ \tilde{\rho}_{22} &= 9/35, & \tilde{\rho}_{32} &= 6/35 \\ \tilde{\rho}_{23} &= 6/35, & \tilde{\rho}_{33} &= 4/35\end{aligned}$$

showing that in general  $\rho_{ij} \neq \tilde{\rho}_{ij}$  and so correlations are present between degrees of vertices.

**5.2. Results for non-linear weights.** We can generalize equation (5.1) to a mean field equation, valid for arbitrary weights, by replacing  $w_2$ , where it appears in a denominator, with  $w$  as we did with the equation for vertex degree densities in Section 2. For  $d = 3$  and uniform partitioning weights the two independent densities  $\rho_{21}$  and  $\rho_{22}$  are given by

$$\rho_{21} = \frac{1}{3} \frac{(3 + \beta)(7\alpha - \gamma)}{(2\alpha - \beta - 1)(3\alpha + 2\beta + \gamma + 6)} \quad (5.4)$$

$$\begin{aligned}\rho_{22} &= \frac{16}{3} \left( 284\alpha^2\beta^4\gamma - 177\alpha^5\beta\gamma + 3564\alpha^3 + 18\alpha^6\gamma + 161\alpha\beta^5\gamma - 873\gamma + 11979\alpha^2\beta^3 \right. \\ &\quad - 2259\alpha^5 - 39\alpha^6\beta - 207\alpha^5\gamma + 6516\alpha^2\beta^4 - 5205\alpha^5\beta - 1419\alpha^4\beta\gamma + 996\alpha\beta^5 \\ &\quad - 5994\alpha^4 - 892\alpha^4\beta^2\gamma + 1543\alpha^2\beta^5 - 18\alpha^7 - 668\alpha^3\beta^4 + 324\alpha^2\gamma + 909\alpha\beta^3\gamma \\ &\quad - 2600\alpha^5\beta^2 - 975\alpha^3\beta^3 + 222\alpha\beta^6 - 1533\alpha^3\beta^2\gamma + 10206\alpha^2\beta^2 - 11799\alpha^4\beta \\ &\quad - 5300\alpha^4\beta^3 - 1521\alpha^3\beta\gamma + 1899\alpha^2\beta^2\gamma + 1059\alpha^2\beta^3\gamma + 1269\alpha^3\beta^2 + 3240\alpha^2\beta \\ &\quad + 756\alpha\beta^3 + 4860\alpha^3\beta + 6\beta^6\gamma - 11703\alpha^4\beta^2 + 1728\alpha^2\beta\gamma - 162\alpha^3\gamma + 486\alpha\beta^2\gamma \\ &\quad \left. + 18\beta^4\gamma + 1530\alpha\beta^4 + 624\alpha\beta^4\gamma - 772\alpha^3\beta^3\gamma - 9\alpha^6 + 24\beta^5\gamma \right) / \left( (3\alpha + 2\beta + \gamma + 6) \right. \\ &\quad \left. \times (11\alpha^2 + 25\alpha\beta + 5\alpha\gamma + 3\beta\gamma + 12\alpha + 4\beta^2)(-\alpha + \gamma)(1 - 2\alpha + \beta)(7\alpha + 2\beta + \gamma)^2 \right)\end{aligned}$$

where  $\alpha = \frac{w_2}{w_1}$ ,  $\beta = \frac{w_3}{w_1}$  and  $\gamma = \sqrt{\alpha(\alpha + 24\beta + 24)}$ . These solutions are compared to simulations in Figures 14 and 15. The other densities are obtained by using the sum rules (5.3).

## 6. DISCUSSION

In this paper we introduced a simple but general model of random tree growth by vertex splitting. The probability weights  $w_{i,j}$  for each splitting process can be chosen independently and are the parameters of the model. We studied the large time/large tree size asymptotics of this model. A mean field theory was presented for the vertex degree distributions and master equations for various correlations functions, like subtree structure probabilities and vertex degree correlations. The mean field results were proved in some special cases and checked by numerical simulations in many general cases. Under a natural scaling hypothesis we were able to compute



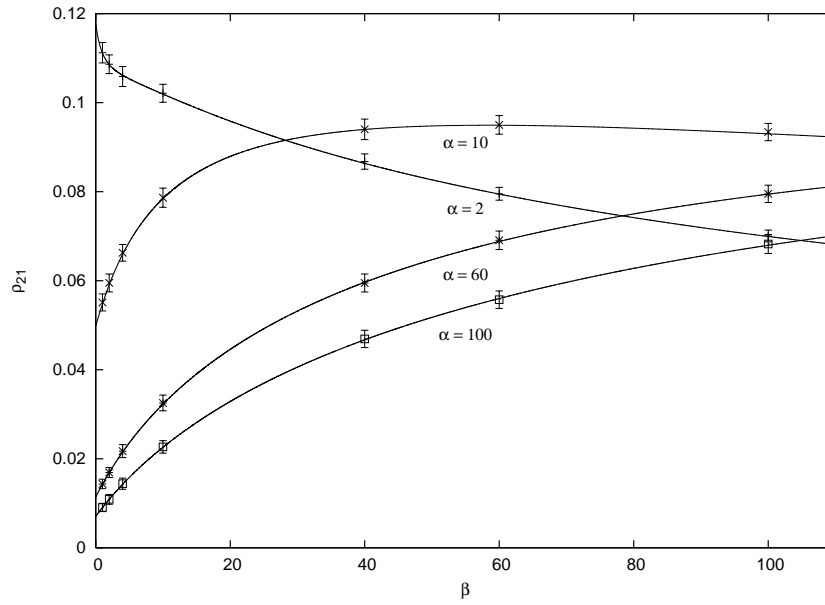


FIGURE 14. The above solution for the density  $\rho_{21}$  plotted as a function of  $\beta$  for a few values of  $\alpha$ . Each datapoint is calculated from simulations of 100 trees on 10000 vertices.

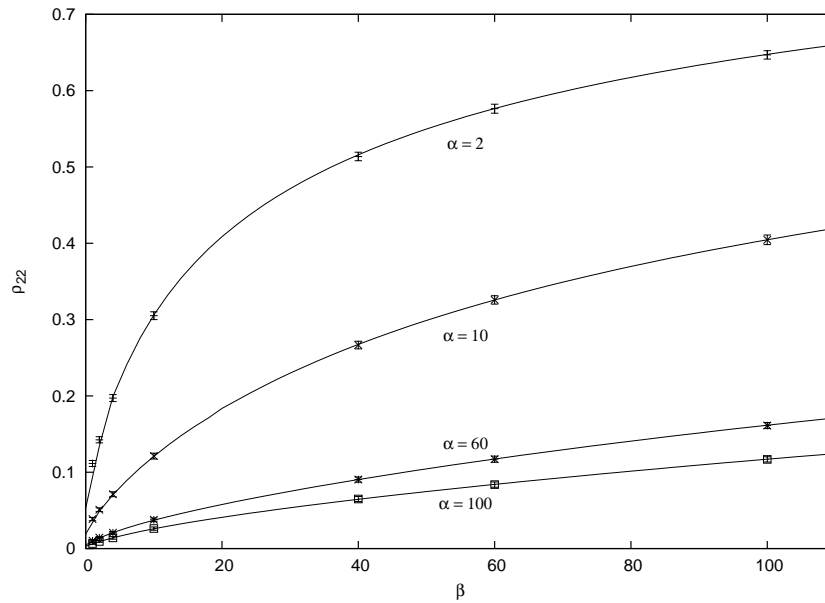


FIGURE 15. The above solution for the density  $\rho_{22}$  plotted as a function of  $\beta$  for a few values of  $\alpha$ . Each datapoint is calculated from simulations of 100 trees on 10000 vertices.

the intrinsic Hausdorff dimension  $d_H$  of the tree. This Hausdorff dimension may take values between  $d_H = 1$  and  $d_H = \infty$ , depending on the parameters  $w_{i,j}$ . These predictions for  $d_H$  were checked by numerical simulations.

One motivation for this study came from a very specific model that arises from RNA folding problems [10] and more precisely when applying a simple greedy algorithm to find the ground state of the random bond RNA model. In this case, it was shown that the problem was equivalent to a random tree growth model, slightly more complicated than the model considered here, with splitting weights fixed by the geometry and dynamics of the process. The scaling properties were found to be universal. This is not the case for the general model.

The model presented here is also related to other tree growth models. Firstly, it contains as a special subclass the so called preferential attachment models, which have been extensively studied in the literature (see e.g. [1, 9]), when all the weights are zero except the  $w_{k,1} = w_{1,k}$ .

Secondly, it also contains as a special limit case the so called alpha-model introduced by D. Ford [12], in connection with phylogenetic trees and cladograms. This model is related to the so called beta-models introduced by Aldous [2, 3]. The alpha-model has been studied in [15, 17] and generalised into the alpha-gamma model in [8] (a mixture of the alpha-model and of the preferential attachment model). It is beyond the scope of this paper to discuss in detail the relationship between these different models. However, we discuss briefly the basic idea.

The alpha model is a growth model for rooted binary trees, with only 1-vertices (the leaves) and 3-vertices (internal vertices), so that  $n_1 = n_3 + 1$  (excluding the root). The growth process grafts a new leaf to a given edge, with probability weight  $\alpha$  if the edge is internal (connects two internal vertices) and weight  $1 - \alpha$  if the edge is external (connects a leaf to an internal vertex). Thus at each step a leaf and an internal vertex are created.

This leaf-edge attachment process can be easily reproduced in our vertex-splitting model by considering the general model with  $d = 3$  (we allow only 1-, 2- and 3-vertices) and by taking the weight  $w_{3,1} \rightarrow +\infty$ . Indeed, in this case, if we start from a binary tree  $T$  with no 2-vertices, any splitting will produce a tree  $T'$  with a 2-vertex, either by splitting a leaf into a 1-vertex and a 2-vertex (with probability weight  $w_{2,1}$ ), or by splitting an internal vertex into a 3-vertex and a 2-vertex (with probability weight  $w_{3,2}$ ). But since  $w_{3,1} = +\infty$ , at the next step, with probability 1, the new 2-vertex will split into a 1-vertex and a 3-vertex, thus producing a new binary tree where a new 1-vertex has been grafted on an edge. These two two-step processes are depicted in Fig. 16. For large binary trees, the density of 1-vertices and 3-vertices are equal ( $\rho_1 = \rho_3 = 1/2$ ,  $\rho_2 = 0$ ), as well as the density of internal and external edges ( $\rho_{3,1} = \rho_{3,3} = 1/2$ ). It is easy to see that our

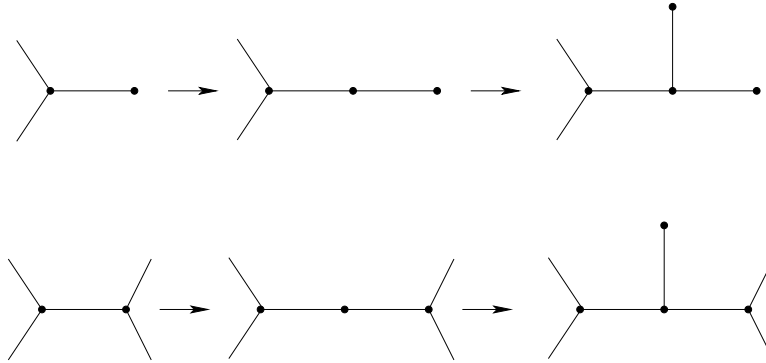


FIGURE 16. The two step processes in the  $d = 3$  model which reproduce the grafting processes of the  $\alpha$ -model

model for  $d = 3$  and  $w_{3,1} = +\infty$  should be equivalent to the  $\alpha$ -model with

$$\alpha = \frac{2w_{3,2}}{w_{2,1} + 3w_{3,2}} . \quad (6.1)$$

As expected,  $w_{2,2}$  is irrelevant in the limit  $w_{3,1} \rightarrow +\infty$ . This equivalence can be checked explicitly. For instance it is proven in [12] that the so-called Sackin's index  $S(t)$  scales with the size  $t$  of the binary tree as  $S(t) \sim t^{1+\alpha}$ . Here  $S(t)$  is defined as the sum over the leaves  $v$  of the number of edges between  $v$  and the root (minus 1) and it is expected to scale as  $t^{1+1/d_H}$ , where  $d_H$  is the Hausdorff dimension of the tree. Therefore we expect that

$$\alpha = 1/d_H \quad (6.2)$$

in the  $\alpha$ -model. Taking the limit  $w_{3,1} \rightarrow +\infty$  in the general formula (4.33) for  $d_H$  in the  $d = 3$  model, we obtain

$$\lim_{w_{3,1} \rightarrow \infty} d_H = \frac{w_{2,1} + 3w_{3,2}}{2w_{3,2}} . \quad (6.3)$$

It is clear that the general splitting model also encompasses the alpha-gamma model of [8].

There are still many open questions for the general model presented here. The most urgent is to prove rigourously that the mean field theory is valid in the general case where the splitting weights  $w_i$  are not linear with the vertex degrees  $i$ . Another goal is to get a better understanding of the degree-degree correlations for vertices which are at a finite distance greater than one, and of the general substructure probabilities. It would also be very interesting to get general distribution functions for correlations involving the geodesic distances on the trees.

One would also like to understand the continuum limit of the model, that is the relation between these trees and continuous random trees. A further clarification of the relation between this model and models of phylogenetic trees is also desirable.

**Acknowledgements.** This research is supported in part by the French Agence Nationale de la Recherche program GIMP ANR-05-BLAN-0029-01 (F.D.), the ENRAGE European network, MRTN-CT-2004-5616, the Programme d'Action Intégrée J. Verne “Physical applications of random graph theory”, the University of Iceland Research Fund and the Eimskip Research Fund at the University of Iceland.

#### APPENDIX A. ILLUSTRATION OF RECURSION EQUATIONS FOR CORRELATIONS

To make the notation more compact we will write for  $i \leq j$

$$\ell_{i,j} = \ell_i, \dots, \ell_j, \quad \text{and} \quad |\ell_{i,j}| = \ell_i + \dots + \ell_j.$$

We can write the following recursions for going from time  $\ell - 1$  to time  $\ell$ . Note that  $s < \ell$  in (A.1), (A.4) and (A.5).

$$\begin{aligned} p_{1k}(\ell''_{1,k-1}; s) &= \\ & \frac{1}{W(\ell-1) + w_1} \left( \sum_{i=1}^{k-1} W(\ell''_i - 1) p_{1k}(\ell''_{1,i-1}, \ell''_i - 1, \ell''_{i+1,k-1}; s) \right. \\ & \quad \left. + \delta_{k2} w_1 p_R(\ell-1; s) + \delta_{\ell 1} \delta_{k1} w_1 \right). \end{aligned} \tag{A.1}$$

$$\begin{aligned} p_{1k}(\ell''_{1,k-1}; \ell) &= \\ & \frac{1}{W(\ell-1) + w_1} \sum_{s=0}^{\ell-1} \sum_{i=k-1}^d (k-1) w_{k,i+2-k} \sum_{|\tilde{\ell}_{1,i+1-k}| = \ell'_1 - 1} p_{1i}(\tilde{\ell}_{1,i+1-k}, \ell''_{1,k-1}; s). \end{aligned} \tag{A.2}$$

$$\begin{aligned} p_{j1}(\ell'_{1,j-1}; s) &= \\ & \frac{1}{W(\ell-1) + w_1} \left( \sum_{i=1}^{j-1} W(\ell'_i - 1) p_{j1}(\ell'_{1,i-1}, \ell'_i - 1, \ell'_{i+1,j-1}; s) \right. \\ & \quad \left. + (j-1) w_{j,1} p_{j-1}(\ell'_{1,j-1}; s) \right. \\ & \quad \left. + \sum_{i=j-1}^d \frac{2w_{j,i+2-j}}{i-1} \sum_{p=1}^{j-2} \sum_{n=p+1}^{j-1} \sum_{|\tilde{\ell}_{1,i+1-j}| = \ell'_n - 1} p_{i1}(\ell'_{1,n-1}, \tilde{\ell}_{1,i+1-j}, \ell'_{n+1,j-1}; s), \right. \\ & \quad \left. + (j-1) \sum_{i=j}^d \frac{(i-j+1)w_{j,i+2-j}}{i-1} \sum_{|\tilde{\ell}_{1,i+1-j}| = \ell'_1 - 1} p_{i1}(\tilde{\ell}_{1,i+1-j}, \ell'_{2,j-1}; s) \right), \end{aligned} \tag{A.3}$$

$$p_{j1}(\ell'_{1,j-1}; \ell) = 0.$$

In deriving equations (A.4) and (A.5) and the corresponding figures, note that the index  $p$  is introduced in the second last diagram in each figure. The

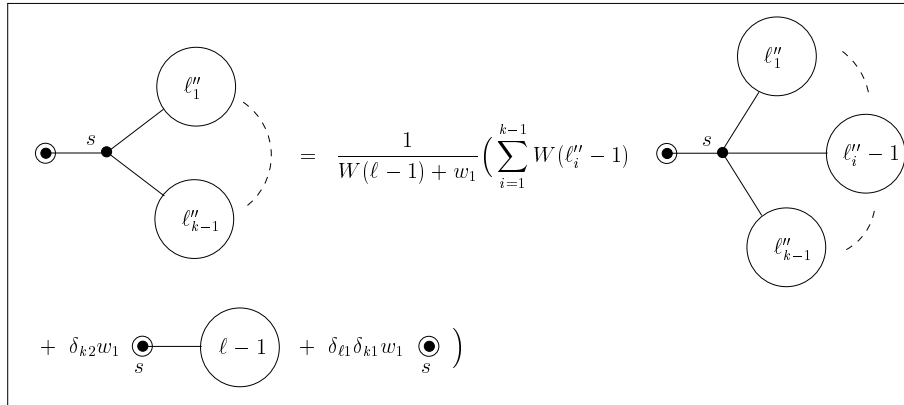


FIGURE 17. Illustration of equation (A.1).

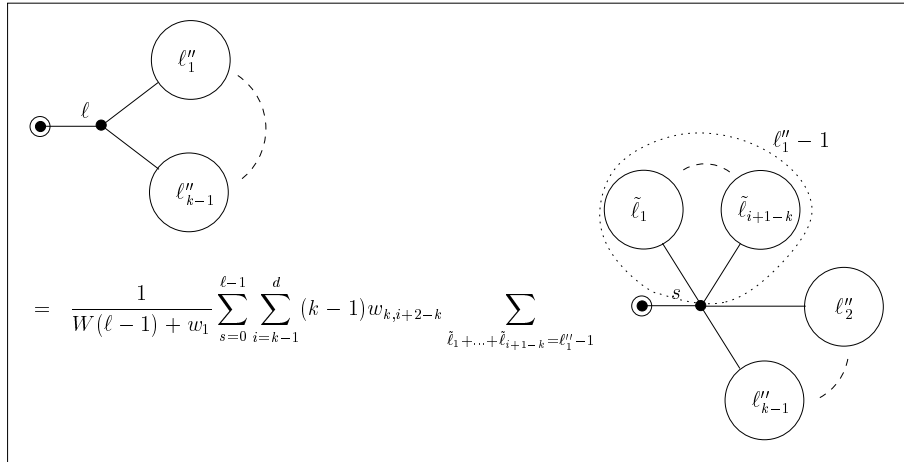


FIGURE 18. Illustration of equation (A.2).

reason for it is the following: even though the functions  $p_{j1}(\ell'_1, \dots, \ell'_{j-1}; s)$  and  $p_{jk}(\ell'_1, \dots, \ell'_{j-1}; \ell''_1, \dots, \ell''_{k-1}; s)$  are symmetric under permutations of  $(\ell'_2, \dots, \ell'_{j-1})$  it does matter where the edge going from  $s$  towards the root, is located.

Therefore, we group together the balloons counter-clockwise from  $s$  towards the rooted balloon and the balloons clockwise from  $s$  towards the rooted balloon into another group, one of the groups is possibly empty. If the total number of balloons in the groups is  $i - 2$  then there are  $i - 1$  such possible configurations. In the equations we therefore divide by  $i - 1$  and sum over all the configurations which contribute to the configuration on the left of the equality sign.

The index  $p$  in the sum is the location of  $s$  counter-clockwise from the rooted balloon. Note that  $p$  can be no larger than  $j - 2$  since if it were larger,

there would be no space for the balloons inside the dotted circle. Note that the balloons inside the dotted circle are always drawn clockwise from the vertex  $s$ . To count the possibility that they are counter-clockwise from  $s$  we multiply by 2.

$$\begin{aligned}
 & \text{Graph} = \frac{1}{W(\ell-1) + w_1} \left( \sum_{i=1}^{j-1} W(\ell'_i - 1) \right) \\
 & + (j-1)w_{j,1} \text{Graph} \\
 & + \sum_{i=j-1}^d \frac{2w_{j,i+2-j}}{i-1} \sum_{p=1}^{j-2} \sum_{n=p+1}^{j-1} \sum_{l_1 + \dots + l_{i+1-j} = \ell_n - 1} \text{Graph} \\
 & + (j-1) \sum_{i=j}^d \frac{i-j+1}{i-1} w_{j,i+2-j} \sum_{l_1 + \dots + l_{i+1-j} = \ell'_1 - 1} \text{Graph}
 \end{aligned}$$

FIGURE 19. Illustration of equation (A.4).

$$\begin{aligned}
p_{jk}(\ell'_{1,j-1}; \ell''_{1,k-1}; \mathbf{s}) &= \\
&\frac{1}{W(\ell-1) + w_1} \left( \sum_{i=1}^{j-1} W(\ell'_i - 1) p_{jk}(\ell'_{1,i-1}, \ell'_i - 1, \ell'_{i+1,j-1}; \ell''_{1,k-1}; \mathbf{s}) \right. \\
&+ \sum_{i=1}^{k-1} W(\ell''_i - 1) p_{jk}(\ell'_{1,j-1}; \ell''_{1,i-1}, \ell''_i - 1, \ell''_{i+1,k-1}; \mathbf{s}) \\
&+ (j-1) w_{j,k} p_{j+k-2}(\ell'_{1,j-1}, \ell''_{1,k-1}; \mathbf{s}) \\
&+ \sum_{i=j-1}^d \frac{2w_{j,i+2-j}}{i-1} \sum_{p=1}^{j-2} \sum_{n=p+1}^{j-1} \sum_{|\tilde{\ell}_{1,i+1-j}|=\ell''_n-1} p_{ik}(\ell'_{1,n-1}, \tilde{\ell}_{1,i+1-j}, \ell'_{n+1,j-1}; \ell''_{1,k-1}; \mathbf{s}) \\
&\left. + (j-1) \sum_{i=j}^d \frac{i-j+1}{i-1} w_{j,i+2-j} \sum_{|\tilde{\ell}_{1,i+1-j}|=\ell'_1-1} p_{ik}(\tilde{\ell}_{1,i+1-j}, \ell'_{2,j-1}; \ell''_{1,k-1}; \mathbf{s}) \right), \tag{A.5}
\end{aligned}$$

$$\begin{aligned}
p_{jk}(\ell'_{1,j-1}; \ell''_{1,k-1}; \ell) &= \frac{1}{W(\ell-1) + w_1} \times \\
&\sum_{s=0}^{\ell-1} \sum_{n=1}^{k-1} \sum_{i=k-1}^d w_{k,i+2-k} \sum_{|\tilde{\ell}_{1,i+1-k}|=\ell''_n-1} p_{ji}(\ell'_{1,j-1}; \ell''_{1,n-1}, \tilde{\ell}_{1,i+1-k}, \ell''_{n+1,k-1}; \mathbf{s}). \tag{A.6}
\end{aligned}$$

Now average over the label  $s$  as before and get the following recursion, going from time  $\ell$  to  $\ell+1$

$$\begin{aligned}
p_{1k}(\ell''_{1,k-1}) &= \\
&\frac{\ell+1}{\ell+2} \frac{1}{W(\ell) + w_1} \left( \sum_{i=1}^{k-1} W(\ell''_i - 1) p_{1k}(\ell''_{1,i-1}, \ell''_i - 1, \ell''_{i+1,k-1}) + \delta_{k2} w_1 p_R(\ell) \right. \\
&+ \delta_{\ell,0} \delta_{k1} w_1 + (k-1) \sum_{i=k-1}^d w_{k,i+2-k} \sum_{|\tilde{\ell}_{1,i+1-k}|=\ell''_1-1} p_{1i}(\tilde{\ell}_{1,i+1-k}, \ell''_{1,k-1}) \left. \right).
\end{aligned}$$

$$\begin{aligned}
& \bullet \text{---} \ell'_1 \text{---} s \text{---} \ell''_{k-1} \\
& \quad \ell'_{j-1} \text{---} \ell''_1 \\
& = \frac{1}{W(\ell-1) + w_1} \left( \sum_{i=1}^{j-1} W(\ell'_i - 1) \bullet \text{---} \ell'_1 \text{---} s \text{---} \ell''_{k-1} \right. \\
& \quad \ell'_{i-1} \text{---} \ell'_{j-1} \text{---} \ell''_1 \\
& + \sum_{i=1}^{k-1} W(\ell''_i - 1) \bullet \text{---} \ell'_1 \text{---} s \text{---} \ell''_{k-1} \\
& \quad \ell'_{j-1} \text{---} \ell''_1 \text{---} \ell''_{i-1} \\
& + (j-1)w_{j,k} \bullet \text{---} \ell'_1 \text{---} s \text{---} \ell''_{k-1} \\
& \quad \ell'_{j-1} \text{---} \ell''_1 \\
& + \sum_{i=j-1}^d \frac{2}{i-1} w_{j,i+2-j} \sum_{p=1}^{j-2} \sum_{n=p+1}^{j-1} \sum_{\tilde{\ell}_1 + \dots + \tilde{\ell}_{i+1-j} = \ell'_n - 1} \bullet \text{---} \ell'_1 \text{---} p \text{---} s \text{---} \ell''_{k-1} \\
& \quad \ell'_{n-1} \text{---} \ell''_1 \text{---} \ell''_{k-1} \\
& \quad \ell_{j-1} \text{---} \ell_{n+1} \text{---} \tilde{\ell}_1 \text{---} \tilde{\ell}_{i+1-j} \\
& \quad \ell'_n - 1 \\
& + (j-1) \sum_{i=j}^d \frac{i-j+1}{i-1} w_{j,i+2-j} \sum_{\tilde{\ell}_1 + \dots + \tilde{\ell}_{i+1-j} = \ell'_1 - 1} \bullet \text{---} \ell'_1 \text{---} s \text{---} \ell''_{k-1} \\
& \quad \ell'_1 - 1 \text{---} \tilde{\ell}_1 \text{---} \tilde{\ell}_{i+1-j} \text{---} \ell''_1 \text{---} \ell''_{k-1} \\
& \quad \ell'_2 \text{---} \ell'_{j-1}
\end{aligned}$$

FIGURE 20. Illustration of equation (A.5).



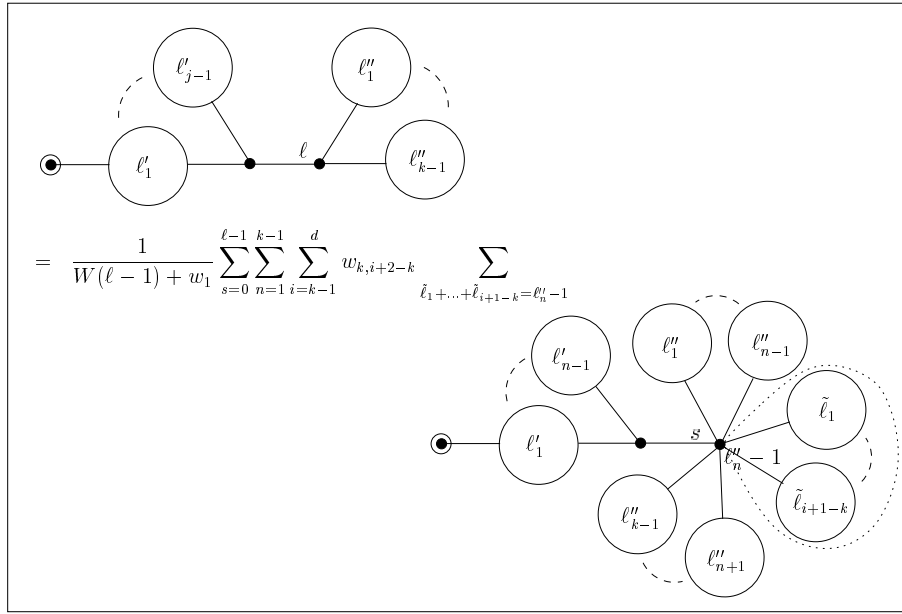


FIGURE 21. Illustration of equation (A.6).

$$\begin{aligned}
p_{j1}(\ell'_{1,j-1}) &= \\
&\frac{\ell+1}{\ell+2} \frac{1}{W(\ell) + w_1} \left( \sum_{i=1}^{j-1} W(\ell'_i - 1) p_{j1}(\ell'_{1,i-1}, \ell'_i - 1, \ell'_{i+1,j-1}) \right. \\
&\quad \left. + (j-1) w_{j,1} p_{j-1}(\ell'_{1,j-1}) \right) \\
&+ \sum_{i=j-1}^d \frac{2w_{j,i+2-j}}{i-1} \sum_{p=1}^{j-2} \sum_{n=p+1}^{j-1} \sum_{|\tilde{\ell}_{1,i+1-j}| = \ell''_n - 1} p_{i1}(\ell'_{1,n-1}, \tilde{\ell}_{1,i+1-j}, \ell'_{n+1,j-1}) \\
&+ (j-1) \sum_{i=j}^d \frac{i-j+1}{i-1} w_{j,i+2-j} \sum_{|\tilde{\ell}_{1,i+1-j}| = \ell'_1 - 1} p_{i1}(\tilde{\ell}_{1,i+1-j}, \ell'_{2,j-1}).
\end{aligned}$$

$$\begin{aligned}
p_{jk}(\ell'_{1,j-1}; \ell''_{1,k-1}) &= \\
&\frac{\ell+1}{\ell+2} \frac{1}{W(\ell) + w_1} \left( \sum_{i=1}^{j-1} W(\ell'_i - 1) p_{jk}(\ell'_{1,i-1}, \ell'_i - 1, \ell'_{i+1,j-1}; \ell''_{1,k-1}) \right. \\
&+ \sum_{i=1}^{k-1} W(\ell''_i - 1) p_{jk}(\ell'_{1,j-1}; \ell''_{1,i-1}, \ell''_i - 1, \ell''_{i+1,k-1}) \\
&+ (j-1) w_{j,k} p_{j+k-2}(\ell'_{1,j-1}, \ell''_{1,k-1}) \\
&+ \sum_{i=j-1}^d \frac{2w_{j,i+2-j}}{i-1} \sum_{p=1}^{j-2} \sum_{n=p+1}^{j-1} \sum_{|\tilde{\ell}_{1,i+1-j}|=\ell''_n-1} p_{ik}(\ell'_{1,n-1}, \tilde{\ell}_{1,i+1-j}, \ell'_{n+1,j-1}; \ell''_{1,k-1}) \\
&+ (j-1) \sum_{i=j}^d \frac{i-j+1}{i-1} w_{j,i+2-j} \sum_{|\tilde{\ell}_{1,i+1-j}|=\ell'_1-1} p_{ik}(\tilde{\ell}_{1,i+1-j}, \ell'_{2,j-1}; \ell''_{1,k-1}) \\
&+ \left. \sum_{n=1}^{k-1} \sum_{i=k-1}^d w_{k,i+2-k} \sum_{|\tilde{\ell}_{1,i+1-k}|=\ell''_n-1} p_{ji}(\ell'_{1,j-1}; \ell''_{1,n-1}, \tilde{\ell}_{1,i+1-k}, \ell''_{n+1,k-1}) \right).
\end{aligned}$$

## REFERENCES

- [1] R. Albert and A. L. Barabási, *Statistical mechanics of complex networks*, Rev. Mod. Phys. **74** (2002) 47.
- [2] D. Aldous, Probability Distributions on Cladograms. In *Random Discrete Structures (Minneapolis, MN, 1993)*, volume 76 of *Vol. Math. Appl.*, pages 1-18. Springer, New York, 1996.
- [3] D. Aldous, *Stochastic models and descriptive statistics for phylogenetic trees, from Yule to today*, Statist. Sci., 16(1), 23-34, 2001.
- [4] J. Ambjørn, B. Durhuus and T. Jonsson, *Quantum geometry - A statistical field theory approach*, Cambridge University Press, Cambridge (1997).
- [5] P. Bialas, *Ensemble of causal trees*, arXiv:cond-mat/0403669v1 [cond-mat.stat-mech].
- [6] P. Bialas and A. K. Oles, *Correlations in connected random graphs*, arXiv:0710.3319 [cond-mat.stat-mech].
- [7] J. Bouttier, P. Di Francesco and E. Guitter, *Planar maps as labeled mobiles*, Electronic J. of Combinatorics **11** (2004) R69.
- [8] B. Chen, D. Ford and M. Winkel, *A new family of Markov branching trees, the alpha-gamma model*, Preprint, arXiv:0807:0554 [math.PR].
- [9] F. David, P. Di Francesco, E. Guitter and T. Jonsson, Mass distribution exponents for growing trees, *J. Stat. Mech.* (2007) P02011.
- [10] F. David, C. Hagendorf and K. J. Wiese, *A growth model for RNA secondary structures*, J. Stat. Mech. (2008) P04008.
- [11] B. Durhuus, T. Jonsson and J. F. Wheeler, *The spectral dimension of generic trees*, J. Stat. Phys. 128 (2007) 1237-1260.

- [12] D. J. Ford, *Probabilities on cladograms: introduction to the alpha model*, Preprint, arXiv:math.PR/0511246.
- [13] T. E. Harris, *The theory of branching processes*, Springer-Verlag, Berlin (1963).
- [14] D. J. Hartfiel, *Dense sets of diagonalizable matrices*, Proceedings of the American Mathematical Society. Vol. 123, No. 6 (Jun., 1995), pp. 1669-1672.
- [15] B. Hass, G. Miermont, J. Pitman and M. Winkel, *Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models*, Annals of Probability, 36(5), 1790-1837, 2008.
- [16] M. Kimmel and D. E. Axelrod, *Branching processes in biology*, Springer-Verlag, New York (2002).
- [17] J. Pitman and M. Winkel, *Regenerative tree growth: binary self-similar continuum random trees and Poisson-Dirichlet compositions*, Preprint, arXiv:0803.309[math.PR].
- [18] G. Schaeffer, *Conjugaison d'arbres et cartes combinatoires aléatoires*, Ph.D. Thesis, Université Bordeaux I (1998).
- [19] C. Semple and M. Steel, *Phylogenetics*, Oxford University Press, Oxford (2003).
- [20] E. Seneta, *Non-negative matrices and Markov chains*, Springer, New York (2006).

(Dukes , Jonsson , Stefánsson) SCIENCE INSTITUTE, UNIVERSITY OF ICELAND, DUNHAGA 3, 107 REYKJAVÍK, ICELAND.

*E-mail address:* dukes/thjons@raunvis.hi.is , siguste@hi.is

(David) INSTITUT DE PHYSIQUE THÉORIQUE, CNRS, URA 2306, F-91191 GIF-SUR-YVETTE, FRANCE CEA, IPHT, F-91191 GIF-SUR-YVETTE, FRANCE

*E-mail address:* francois.david@cea.fr