The Vertex Splitting Model

Thordur Jonsson, University of Iceland

November 2010

Nordita Programme: Random geometry and applications

Outline

- Growing trees
- The vertex splitting model
- Vertex degree distribution
- Correlations of vertex degrees
- Subtree structure functions
- The Hausdorff dimension
- Mass distribution
- The infinte volume limit
- An exactly soluble special case - Markovian self-similarity

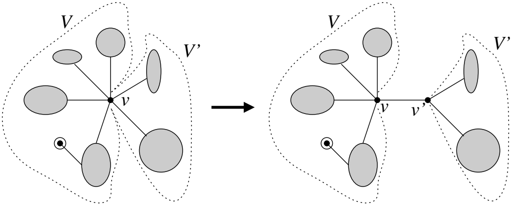
Growing trees

- Galton-Watson trees
- Preferential attachment trees
- Phylogenetic trees, α and $\alpha - \gamma$ models
- splitting vertex trees

The splitting vertex model

F. David, M. Dukes, S. Stefansson and T.J.: J. Stat. Mech. (2009) P04009

- A model of randomly growing rooted, planar trees



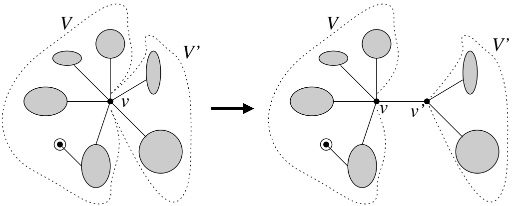
- Degree of vertices is bounded by an integer d (we also discuss the case $d = \infty$)
- Nonnegative splitting weights $w_1, w_2, \ldots, w_d$
- $n_j(T)$ = the number of vertices of degree j in a tree T
  p_j = Probability of choosing a vertex $v \in T$ of degree j

$$p_j = \frac{w_j}{\sum_i w_i n_i(T)}$$

The splitting vertex model

F. David, M. Dukes, S. Stefansson and T.J.: J. Stat. Mech. (2009) P04009

- A model of randomly growing rooted, planar trees



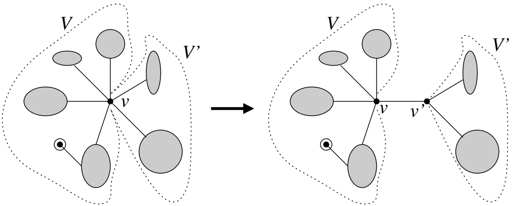
- Degree of vertices is bounded by an integer d (we also discuss the case $d = \infty$)
- Nonnegative splitting weights $w_1, w_2, \ldots, w_d$
- $n_j(T)$ = the number of vertices of degree j in a tree T
  p_j = Probability of choosing a vertex $v \in T$ of degree j

$$p_j = \frac{w_j}{\sum_i w_i n_i(T)}$$

The splitting vertex model

F. David, M. Dukes, S. Stefansson and T.J.: J. Stat. Mech. (2009) P04009

- A model of randomly growing rooted, planar trees



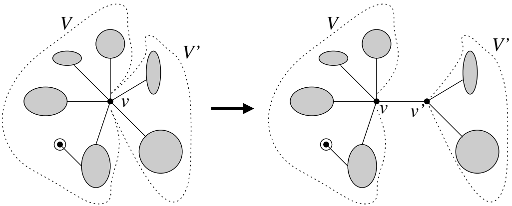
- Degree of vertices is bounded by an integer d (we also discuss the case $d = \infty$)
- Nonnegative splitting weights $w_1, w_2, \ldots, w_d$
- $n_j(T)$ = the number of vertices of degree j in a tree T
  p_j = Probability of choosing a vertex $v \in T$ of degree j

$$p_j = \frac{w_j}{\sum_i w_i n_i(T)}$$

The splitting vertex model

F. David, M. Dukes, S. Stefansson and T.J.: J. Stat. Mech. (2009) P04009

- A model of randomly growing rooted, planar trees



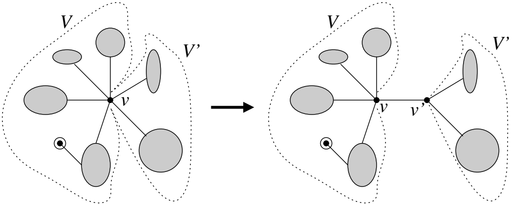
- Degree of vertices is bounded by an integer d (we also discuss the case $d = \infty$)
- Nonnegative splitting weights $w_1, w_2, \ldots, w_d$
- $n_j(T) =$ the number of vertices of degree j in a tree T
  $p_j =$ Probability of choosing a vertex $v \in T$ of degree j

$$p_j = \frac{w_j}{\sum_i w_i n_i(T)}$$

The splitting vertex model

F. David, M. Dukes, S. Stefansson and T.J.: J. Stat. Mech. (2009) P04009

- A model of randomly growing rooted, planar trees



- Degree of vertices is bounded by an integer d (we also discuss the case $d = \infty$)
- Nonnegative splitting weights $w_1, w_2, \ldots, w_d$
- $n_j(T)$ = the number of vertices of degree j in a tree T
  p_j = Probability of choosing a vertex $v \in T$ of degree j

$$p_j = \frac{w_j}{\sum_i w_i n_i(T)}$$

Splitting rules

- The parameters of the model are

$$\begin{bmatrix} 0 & w_{1,2} & w_{1,3} & \cdots & w_{1,d-1} & w_{1,d} \\ w_{2,1} & w_{2,2} & w_{2,3} & \cdots & w_{2,d-1} & w_{2,d} \\ w_{3,1} & w_{3,2} & w_{3,3} & \cdots & w_{3,d-1} & 0 \\ w_{4,1} & w_{4,2} & w_{4,3} & & 0 & 0 \\ \vdots & \vdots & \vdots & \iddots & \vdots & \vdots \\ w_{d,1} & w_{d,2} & 0 & \cdots & 0 & 0 \end{bmatrix}$$

  a symmetric matrix of non–negative partitioning weights
- Split a vertex of degree i into vertices of degree k and $i+2-k$ with probability $w_{k,i+2-k}/w_i$ – all such splittings equally probable
- The splitting weights $w_1, w_2, \ldots, w_d$ are related to the partitioning weights by

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

Splitting rules

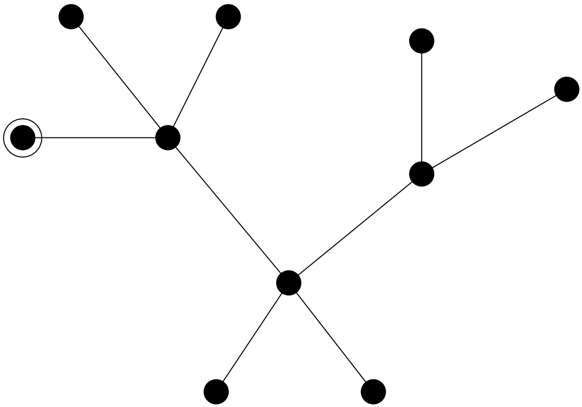
- The parameters of the model are

$$\begin{bmatrix} 0 & w_{1,2} & w_{1,3} & \cdots & w_{1,d-1} & w_{1,d} \\ w_{2,1} & w_{2,2} & w_{2,3} & \cdots & w_{2,d-1} & w_{2,d} \\ w_{3,1} & w_{3,2} & w_{3,3} & \cdots & w_{3,d-1} & 0 \\ w_{4,1} & w_{4,2} & w_{4,3} & & 0 & 0 \\ \vdots & \vdots & \vdots & \iddots & \vdots & \vdots \\ w_{d,1} & w_{d,2} & 0 & \cdots & 0 & 0 \end{bmatrix}$$

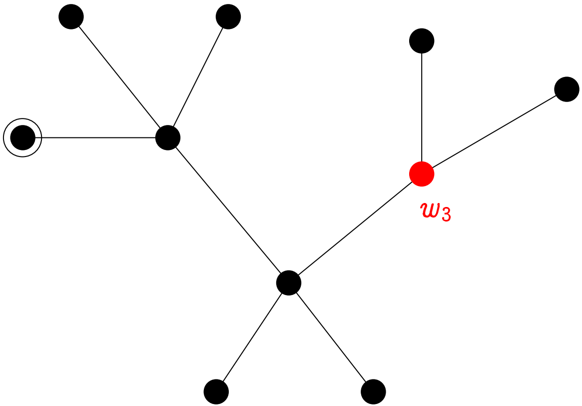
  a symmetric matrix of non–negative partitioning weights
- Split a vertex of degree i into vertices of degree k and $i+2-k$ with probability $w_{k,i+2-k}/w_i$ – all such splittings equally probable
- The splitting weights $w_1, w_2, \ldots, w_d$ are related to the partitioning weights by

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

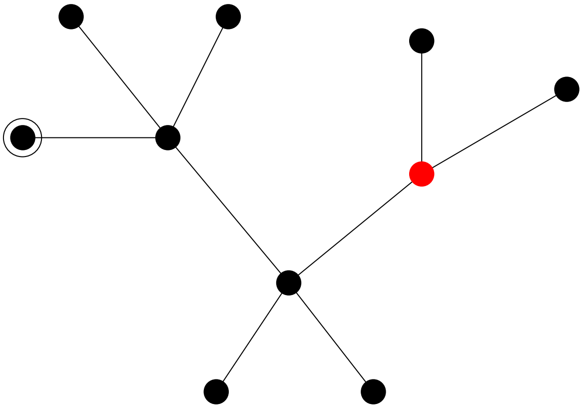
A tree



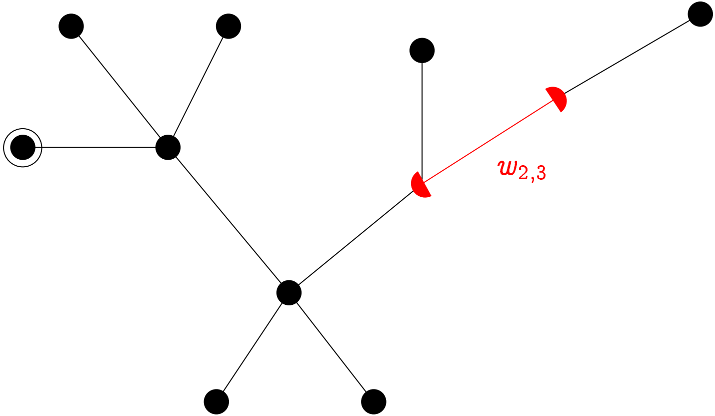
Vertex splitting rules



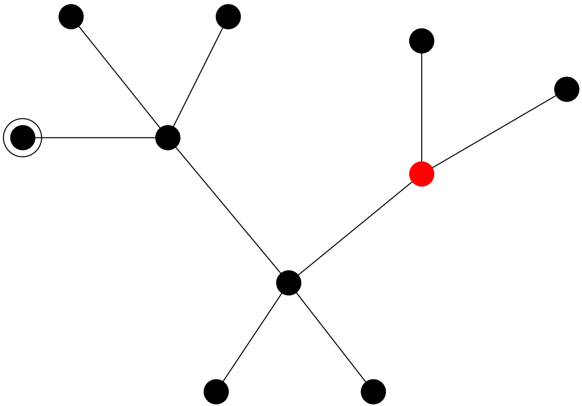
Vertex splitting rules



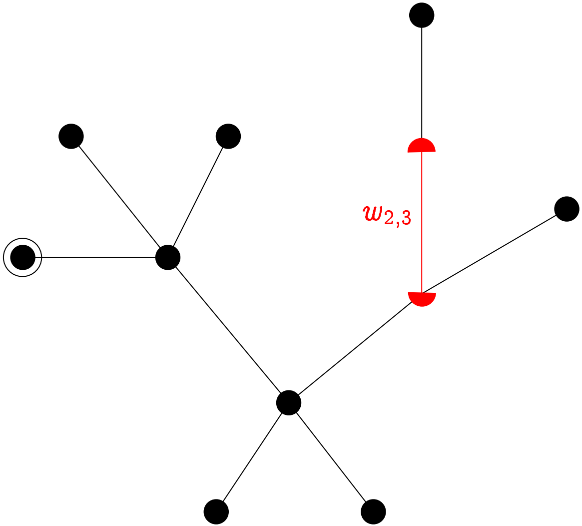
Vertex splitting rules



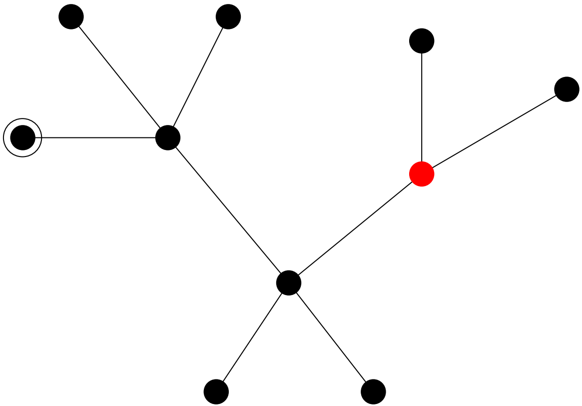
Vertex splitting rules



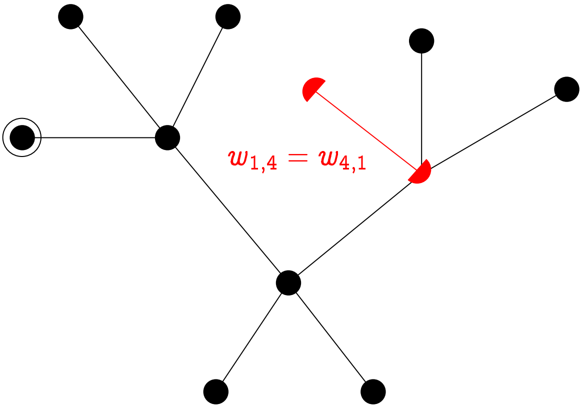
Vertex splitting rules



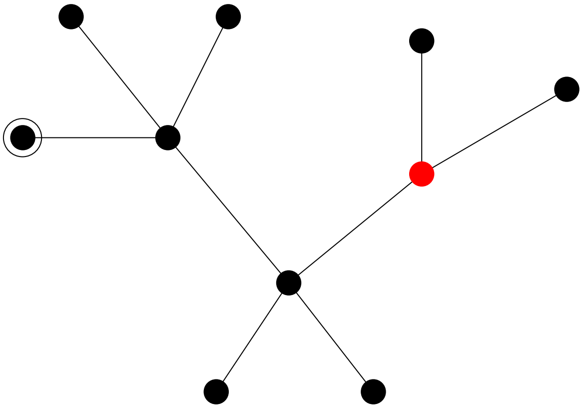
Vertex splitting rules



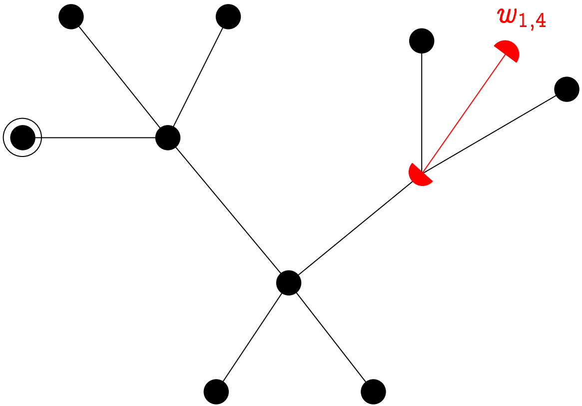
Vertex splitting rules



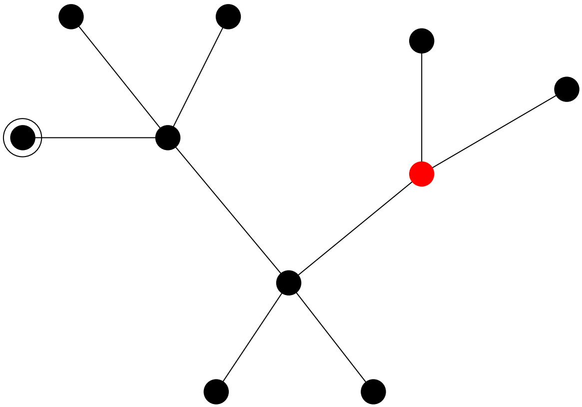
Vertex splitting rules



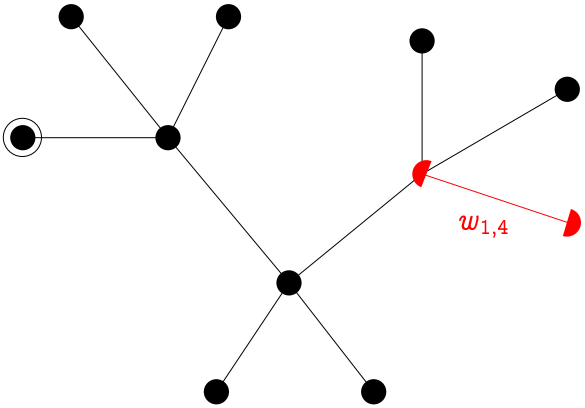
Vertex splitting rules



Vertex splitting rules

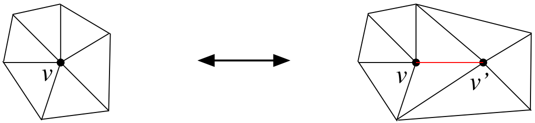


Vertex splitting rules



Relation to other models

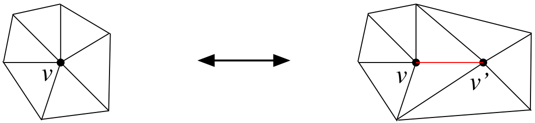
- Ergodic moves in Monte Carlo simulations of triangulations in 2d-quantum gravity J. Ambjorn, J. Jurkiewicz et al.



- A tree growth process which is equivalent to a simplified model for RNA secondary structures F. David, C. Hagendorf, K. J. Wiese
- When $w_{i,j} = 0$ unless $j = 1$ or $i = 1$ it reduces to the preferential attachment model R. Albert and A. L. Barabasi et al.
- For $d = 3$, in a certain limit, it reduces to Ford's alpha model for phylogenetic trees

Relation to other models

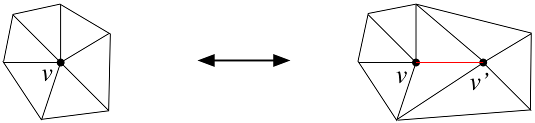
- Ergodic moves in Monte Carlo simulations of triangulations in 2d-quantum gravity J. Ambjorn, J. Jurkiewicz et al.



- A tree growth process which is equivalent to a simplified model for RNA secondary structures F. David, C. Hagendorf, K. J. Wiese
- When $w_{i,j} = 0$ unless $j = 1$ or $i = 1$ it reduces to the preferential attachment model R. Albert and A. L. Barabasi et al.
- For $d = 3$, in a certain limit, it reduces to Ford's alpha model for phylogenetic trees

Relation to other models

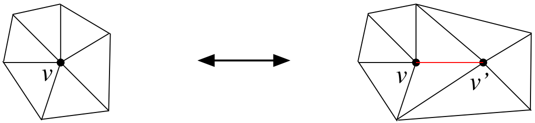
- Ergodic moves in Monte Carlo simulations of triangulations in 2d-quantum gravity J. Ambjorn, J. Jurkiewicz et al.



- A tree growth process which is equivalent to a simplified model for RNA secondary structures F. David, C. Hagendorf, K. J. Wiese
- When $w_{i,j} = 0$ unless $j = 1$ or $i = 1$ it reduces to the preferential attachment model R. Albert and A. L. Barabasi et al.
- For $d = 3$, in a certain limit, it reduces to Ford's alpha model for phylogenetic trees

Relation to other models

- Ergodic moves in Monte Carlo simulations of triangulations in 2d-quantum gravity J. Ambjorn, J. Jurkiewicz et al.



- A tree growth process which is equivalent to a simplified model for RNA secondary structures F. David, C. Hagendorf, K. J. Wiese
- When $w_{i,j} = 0$ unless $j = 1$ or $i = 1$ it reduces to the preferential attachment model R. Albert and A. L. Barabasi et al.
- For $d = 3$, in a certain limit, it reduces to Ford's alpha model for phylogenetic trees

Degree distribution

If we consider linear splitting weights

$$w_i = ai + b.$$

the analysis simplifies due to the Euler relation for trees

$$\sum_{i=1}^{d} n_i(T) = |T|, \quad \sum_{i=1}^{d} i n_i(T) = 2(|T| - 1).$$

The normalization factor $\sum_i w_i n_i(T)$ depends only on the size of the tree T.

Generating function

Let $p_t(n_1,\ldots,n_d)$ be the probability that the tree T at time t has $(n_1(T),\ldots,n_d(T)) = (n_1,\ldots,n_d)$.
The probability genereting function

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

satisfies the equation

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

where $\mathbf{c}(\mathbf{x}) = (c_1(\mathbf{x}), c_2(\mathbf{x}),\ldots,c_d(\mathbf{x}))$ with

$$c_i(\mathbf{x}) = \frac{i}{2}\sum_{j=1}^{i+1} w_{j,i+2-j} x_j x_{i+2-j} \quad \text{and} \quad \nabla = \Big(\partial/\partial x_1,\ldots,\partial/\partial x_d\Big).$$

Recurrence

- Begin with a tree T_0 at time 0
- At time $t > 0$ we have a tree T_t with $n_i(T_t)$ vertices of degree i
- Let $\bar{n}_{t,i}$ denote the average of $n_i(T)$ over all trees that can arise at time t, i.e.

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

- Define

$$\rho_{t,i} = \frac{\bar{n}_{t,i}}{t}$$

and we will use the notation

$$\rho(t) = (\rho_{t,1}, \ldots, \rho_{t,d})$$

Recurrence

- Begin with a tree T_0 at time 0
- At time $t > 0$ we have a tree T_t with $n_i(T_t)$ vertices of degree i
- Let $\bar{n}_{t,i}$ denote the average of $n_i(T)$ over all trees that can arise at time t, i.e.
$$\overline{n}_{t,k} \; = \sum_{n_1+\ldots+n_d=t} p_t(n_1,\ldots,n_d)n_k \; = \; \partial_k \mathcal{H}_t(\mathbf{x})|_{\mathbf{x}=\mathbf{1}},$$
- Define
$$\rho_{t,i} = \frac{\bar{n}_{t,i}}{t}$$
and we will use the notation
$$\rho(t) = (\rho_{t,1}, \ldots, \rho_{t,d})$$

- The recurrence for $\mathcal{H}_t$ gives rise to a recurrence for $\rho(t)$.

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

$$\Longrightarrow$$

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

- Under mild conditions on the $w_{i,j}$ the limits

$$\lim_{t\to\infty} \rho_{t,i} = \rho_i$$

exist and are the unique positive solution to the linear equations

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

- These values are independent of the initial tree.
- The proof uses the Perron–Frobenius theorem for "positive" matrices.

- Under mild conditions on the $w_{i,j}$ the limits

$$\lim_{t\to\infty} \rho_{t,i} = \rho_i$$

exist and are the unique positive solution to the linear equations

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

- These values are independent of the initial tree.
- The proof uses the Perron–Frobenius theorem for "positive" matrices.

- Under mild conditions on the $w_{i,j}$ the limits

$$\lim_{t\to\infty} \rho_{t,i} = \rho_i$$

exist and are the unique positive solution to the linear equations

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

- These values are independent of the initial tree.
- The proof uses the Perron–Frobenius theorem for "positive" matrices.

Perron-Frobenius

Theorem. Let A be a matrix with nonnegative matrix elements such that all the matrix elements of A^p are positive (A primitive) for some integer p. Then the eigenvalue of A with the largest absolute value is positive and simple. The corresponding eigenvector can be taken to have positive entries.

Iterating the recurrence equation for $\rho(t)$ we find

$$\rho(t) = \frac{1}{t} \prod_{i=1}^{t-1} \left(1 + \frac{1}{(2a+b)i - 2a} B \right) \rho_0$$

where B is a matrix with nonnegative entries except on the diagonal. If B is primitive and diagonalizable, then $\rho(t)$ converges to the normalized Perron-Frobenius eigenvector of B.

The B matrix

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

Examples

- $d = 3$ The matrix B is diagonalizable and

$$\rho_1 = \rho_3 = 2/7, \quad \rho_2 = 3/7$$

if the partitioning weights are chosen to be uniform, i.e.

$$w_{i,j} = w_{i+j-2} \frac{2}{(i+j-2)(i+j-1)}.$$

- $d = 4$ Can again solve explicitly with uniform partitioning weights and get ρ_i's which vary with a and b.
- $d = \infty$ Do not have a proof of convergence but can solve the equation for the ρ_i's

$$\rho_k \sim \frac{1}{k!} 2^{k-1} k^{-1-x}, \quad x = b/a$$

$$\rho_k = \frac{1}{e(k-1)!}, \quad a = 0.$$

Examples

- $d = 3$ The matrix B is diagonalizable and

$$\rho_1 = \rho_3 = 2/7, \quad \rho_2 = 3/7$$

if the partitioning weights are chosen to be uniform, i.e.

$$w_{i,j} = w_{i+j-2} \frac{2}{(i+j-2)(i+j-1)}.$$

- $d = 4$ Can again solve explicitly with uniform partitioning weights and get ρ_i's which vary with a and b.
- $d = \infty$ Do not have a proof of convergence but can solve the equation for the ρ_i's

$$\rho_k \sim \frac{1}{k!} 2^{k-1} k^{-1-x}, \quad x = b/a$$

$$\rho_k = \frac{1}{e(k-1)!}, \quad a = 0.$$

Examples

- $d = 3$ The matrix B is diagonalizable and
$$\rho_1 = \rho_3 = 2/7, \quad \rho_2 = 3/7$$
if the partitioning weights are chosen to be uniform, i.e.
$$w_{i,j} = w_{i+j-2} \frac{2}{(i+j-2)(i+j-1)}.$$
- $d = 4$ Can again solve explicitly with uniform partitioning weights and get ρ_i's which vary with a and b.
- $d = \infty$ Do not have a proof of convergence but can solve the equation for the ρ_i's
$$\rho_k \sim \frac{1}{k!} 2^{k-1} k^{-1-x}, \quad x = b/a$$
$$\rho_k = \frac{1}{e(k-1)!}, \quad a = 0.$$

General splitting weights

- Use mean field theory for the normalization factor $\sum_i n_i(T) w_i \to t \sum_i \rho_i w_i$.
  Equation for a steady state vertex distribution

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

  subject to the constraints

$$\rho_1 + \ldots + \rho_d = 1, \quad w_1\rho_1 + \ldots + w_d\rho_d = w.$$

- There is a unique positive solution by the Perron-Frobenius theorem.
- For $d = 3$ and uniform partitioning weights we find

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

  where $\alpha = w_2/w_1$ and $\beta = w_3/w_1$.

Comparison with simulations

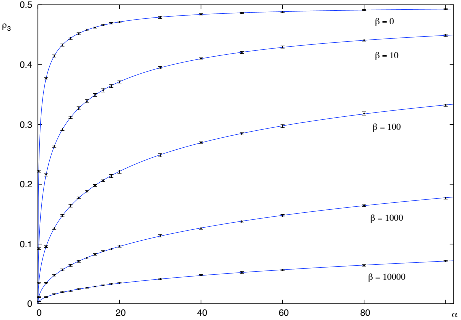


FIGURE 4. The value of ρ_3 as given in (2.45) compared to results from simulations. Each point is calculated from 20 trees on 10000 vertices.

A comparison of the theoretical prediction with simulations in the case d=3 and uniform partitioning weights.

$$\alpha = \frac{w_2}{w_1}, \quad \beta = \frac{w_3}{w_1}$$

Correlations

In a typical infinite tree, what is the proportion of edges whose endpoints have degrees j and k ?

Let $n_{j,k}$ = number of such edges in a finite tree of size t, where the vertex of degree j is closer to the root

Let $\rho_{j,k} = \lim_{t\to\infty} \frac{n_{j,k}}{t}$. Then (for linear splitting weights)

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

assuming the existence of the limit.

Correlations

In a typical infinite tree, what is the proportion of edges whose endpoints have degrees j and k ?

Let $n_{j,k}$ = number of such edges in a finite tree of size t, where the vertex of degree j is closer to the root

Let $\rho_{j,k} = \lim_{t\to\infty} \frac{n_{j,k}}{t}$. Then (for linear splitting weights)

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

assuming the existence of the limit.

Correlations

In a typical infinite tree, what is the proportion of edges whose endpoints have degrees j and k ?

Let $n_{j,k} =$ number of such edges in a finite tree of size t, where the vertex of degree j is closer to the root

Let $\rho_{j,k} = \lim_{t\to\infty} \frac{n_{j,k}}{t}$. Then (for linear splitting weights)

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

assuming the existence of the limit.

Correlations

In a typical infinite tree, what is the proportion of edges whose endpoints have degrees j and k ?

Let $n_{j,k}$ = number of such edges in a finite tree of size t, where the vertex of degree j is closer to the root

Let $\rho_{j,k} = \lim_{t\to\infty} \dfrac{n_{j,k}}{t}$. Then (for linear splitting weights)

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

assuming the existence of the limit.

Explicit solutions

Can solve in simple cases and find nontrivial correlations: In general "disassortative" i.e. vertices of high degree are unlikely to be neighbours.

Take $d = 3$, linear splitting weights and uniform partitioning weights. Then $\rho_1 = \rho_3 = 2/7$ and $\rho_2 = 3/7$. Let $y = w_3/w_2$.

Then the solutions to the correlation equation are

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

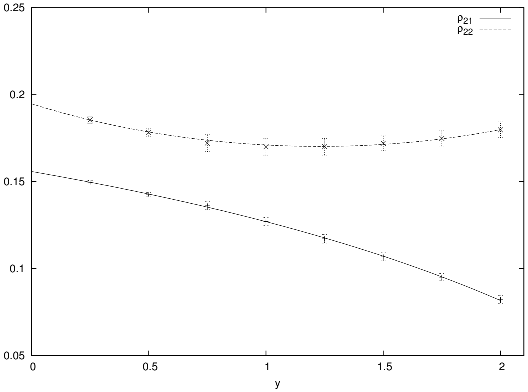
Sum rules

The following sum rules hold:

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

These relations show that there are only two independent link densities, e.g. ρ_{21} and ρ_{22}.

Comparison with simulations

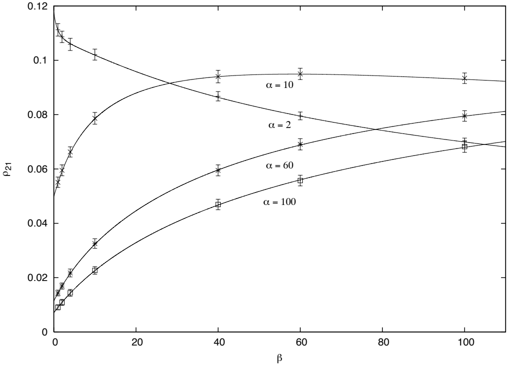


Nonlinear splitting weights

Taking $d = 3$ and general nonlinear splitting weights

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

where $\alpha = w_2/w_1$, $\beta = w_3/w_1$ and $\gamma = \sqrt{\alpha\,(\alpha+24\,\beta+24)}$.



Comparison with simulations

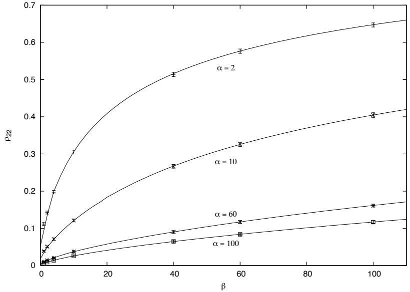


FIGURE 17. The solution (5.5) for the density ρ_{22} plotted as a function of β for a few values of α. Each datapoint is calculated from simulations of 100 trees on 10000 vertices.

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

Subtree probabilities

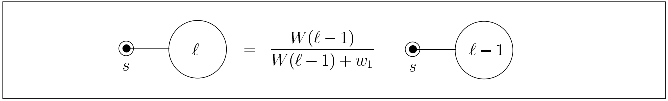
- Label vertices in the tree by their time of creation
- Use linear weights
- Derive expressions for the probabilistic structure of the tree as seen from the vertex created at a given time
- Average over the creation time
- Introduce a scaling assumption
- Extract the Hausdorff dimension
- Get results which agree with simulations

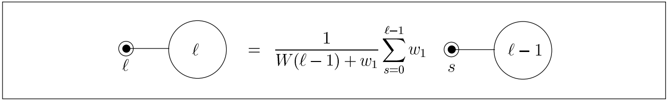
- Begin with a tree consisting of a single vertex at time $t = 0$
- In a tree of size ℓ let $p_R(\ell; s)$ be the probability that the vertex created at time $s \leq \ell$ is the root
- We find

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

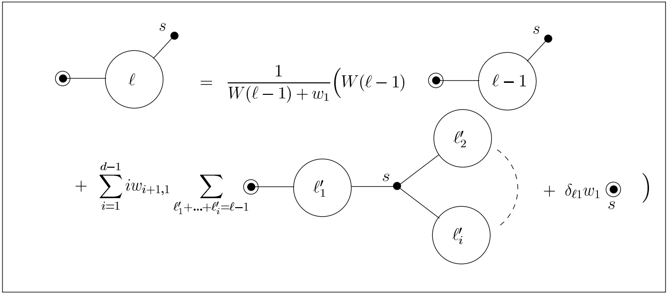
$$p_R(\ell; \ell) = \frac{1}{W(\ell-1)+w_1} \sum_{s=0}^{\ell-1} w_1 p_R(\ell-1; s), \quad s = \ell$$

$W(\ell) = (2a+b)\ell - a$ is a normalization factor.

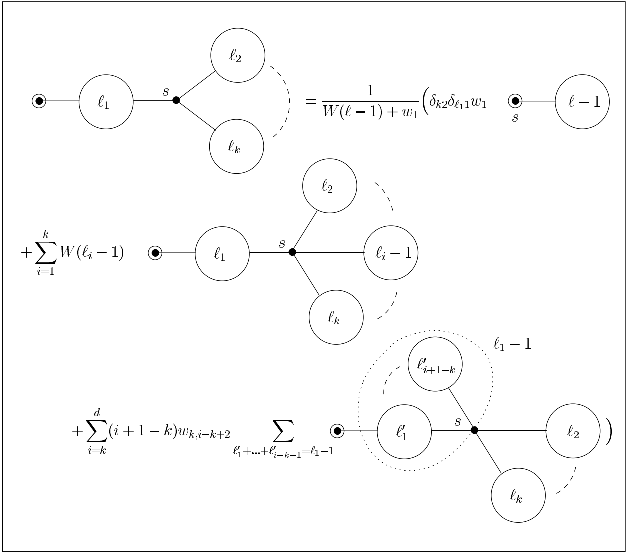




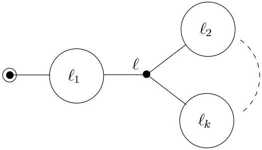
Let $p_k(\ell_1, \ell_2, \ldots, \ell_k; s)$ be the probability that the vertex v created at time s has degree k, the root subtree has ℓ_1 links and the other subtrees incident on v have size $\ell_2, \ldots, \ell_k$. Denote the sum of the ℓ_i's by ℓ. Then for $k = 1$ and $s < \ell$



and for $k > 1$ and $s < \ell$

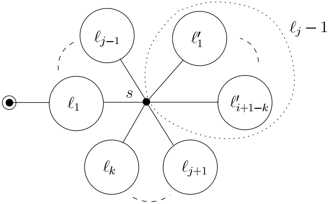


Finally $k > 1$ and $s = \ell$



$$= \frac{1}{W(\ell-1)+w_1}\sum_{s=0}^{\ell-1}\sum_{j=2}^{k}\sum_{i=k-1}^{d-1} w_{k,i-k+2}\sum_{\ell'_1+\ldots+\ell'_{i+1-k}=\ell_j-1}$$

$\times$



We average over s to get simpler recursions:

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

$$\begin{aligned} & p_1(\ell+1) \\ = \; & \frac{\ell+1}{\ell+2}\frac{1}{W(\ell)+w_1}\Big[W(\ell)p_1(\ell) + \sum_{i=1}^{d-1} i w_{i+1,1} \sum_{\substack{\ell_1'+\ldots+\ell_i' \\ =\ell}} p_i(\ell_1', \ldots, \ell_i') + 2\delta_{\ell 0} w_1\Big]. \end{aligned} \tag{3.11}$$

$$\begin{aligned} & p_k(\ell_1, \ldots, \ell_k) \\ = \; & \frac{\ell+1}{\ell+2}\frac{1}{W(\ell)+w_1}\Big[\delta_{k2}\delta_{\ell_1 1} w_1 p_R(\ell) + \sum_{i=1}^{k} W(\ell_i - 1) p_k(\ell_1, \ldots, \ell_i - 1, \ldots, \ell_k) \\ & + \sum_{i=k}^{d} (i-k+1)\, w_{k,i-k+2} \sum_{\substack{\ell_1'+\ldots+\ell_{i+1-k}' \\ =\ell_1-1}} p_i(\ell_1', \ldots, \ell_{i+1-k}', \ell_2, \ldots, \ell_k) \qquad (3.12) \\ & + \sum_{j=2}^{k} \sum_{i=k-1}^{d} w_{k,i-k+2} \sum_{\substack{\ell_1'+\ldots+\ell_{i+1-k}' \\ =\ell_j-1}} p_i(\ell_1, \ldots, \ell_{j-1}, \ell_1', \ldots, \ell_{i+1-k}', \ell_{j+1}, \ldots, \ell_k)\Big] \end{aligned}$$

Finally we define the "two point functions" that are needed to calculate the Hausdorff dimension:

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

which is the probability that i trees of total volume ℓ_1, none of which contains the root, are attached to a vertex of order k in a tree of total volume $\ell = \ell_1 + \ell_2$. There are $d(d-1)/2$ such functions, $1 \leq i \leq k-1$.

The two point functions satisfy the recursion relation

$$
\begin{aligned}
q_{ki}(\ell_1,\ell_2) \quad = \quad & \frac{\ell+1}{\ell+2}\,\frac{1}{W(\ell)+w_1}\Big[\\
& \sum_{j=k-1}^{d} w_{k,j+2-k}\Big((j-i)q_{ji}(\ell_1-1,\ell_2)+iq_{j,j-(k-i)}(\ell_1,\ell_2-1)\Big) \\
& +\Big(W(\ell_1-1)+(k-i-1)(w_2-w_3)\Big)q_{ki}(\ell_1-1,\ell_2) \\
& +\Big(W(\ell_2-1)+(i-1)(w_2-w_3)\Big)q_{ki}(\ell_1,\ell_2-1) \\
& +\delta_{k2}\delta_{\ell_1 1}w_1 p_R(\ell_2)+\delta_{i1}\delta_{\ell_2 1}w_{k,1}\sum_{\ell_1'+\ldots+\ell_{k-1}'=\ell_1} p_{k-1}(\ell_1',\ldots,\ell_{k-1}')\Big]
\end{aligned}
$$

An almost closed system of linear equations.

Hausdorff dimension

- Let T be a tree with ℓ edges and v, w vertices of T.
- Denote the graph distance between v and w by $d_T(v, w)$.
- We define the radius of T as

$$R_T = \frac{1}{(2\ell)} \sum_{v \in T} d_T(r, v)\, \sigma(v),$$

- We define the Hausdorff dimension of the tree, d_H, by the scaling law for large trees

$$\langle R_T \rangle \;\sim\; \ell^{1/d_H} \qquad \ell \to \infty$$

This definition is different from the one usually given for infinite trees but is expected to be equivalent.

Hausdorff dimension

- Let T be a tree with ℓ edges and v, w vertices of T.
- Denote the graph distance between v and w by $d_T(v, w)$.
- We define the radius of T as

$$R_T = \frac{1}{(2\ell)} \sum_{v \in T} d_T(r, v)\, \sigma(v),$$

- We define the Hausdorff dimension of the tree, d_H, by the scaling law for large trees

$$\langle R_T \rangle \;\sim\; \ell^{1/d_H} \qquad \ell \to \infty$$

This definition is different from the one usually given for infinite trees but is expected to be equivalent.

Hausdorff dimension

- Let T be a tree with ℓ edges and v, w vertices of T.
- Denote the graph distance between v and w by $d_T(v, w)$.
- We define the radius of T as

$$R_T = \frac{1}{(2\ell)} \sum_{v \in T} d_T(r, v)\, \sigma(v),$$

- We define the Hausdorff dimension of the tree, d_H, by the scaling law for large trees

$$\langle R_T \rangle \sim \ell^{1/d_H} \qquad \ell \to \infty$$

This definition is different from the one usually given for infinite trees but is expected to be equivalent.

Hausdorff dimension

- Let T be a tree with ℓ edges and v, w vertices of T.
- Denote the graph distance between v and w by $d_T(v, w)$.
- We define the radius of T as

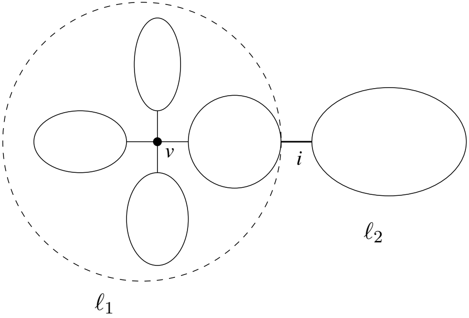
$$R_T = \frac{1}{(2\ell)} \sum_{v \in T} d_T(r, v)\, \sigma(v),$$

- We define the Hausdorff dimension of the tree, d_H, by the scaling law for large trees

$$\langle R_T \rangle \;\sim\; \ell^{1/d_H} \qquad \ell \to \infty$$

This definition is different from the one usually given for infinite trees but is expected to be equivalent.

Combinatorics



- Cutting the tree at an edge i we get two subtrees of size ℓ_1 and ℓ_2
- One can prove the following identity:

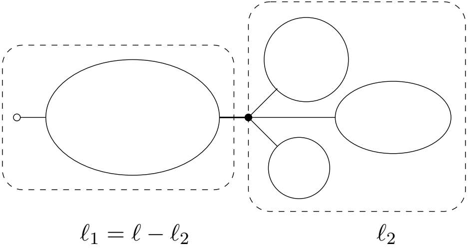
$$\sum_w d_T(v,w)\sigma(w) \;=\; \sum_i (2\ell_2(v;i)+1)$$

valid for any vertex v. We use it for $v = r$.

- The identity implies:

$$\langle R_T \rangle = \frac{1}{2\ell} \sum_T P(T) \sum_i (2\ell_2(r;i)+1)$$

$$= \frac{\ell+1}{2\ell} \sum_{\ell_2=0}^{\infty} (2\ell_2+1) \sum_{k=1}^{d} q_{k,k-1}(\ell-\ell_2;\ell_2)$$



- We use a scaling assumptions about the q functions

$$q_{ki}(\ell_1, \ell - \ell_1) = \ell^{-\rho}\omega_{ki}(\ell_1/\ell) + O(\ell^{\rho+1})$$

  This scaling hypothesis has been tested by direct calculation for trees up to size 50.000.

- Inserting into the recurrence equation for q_{ki} keeping leading order terms in ℓ^{-1} gives

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

- This is a Perron-Frobenius type equation. Gives ρ in principle.
- Can solve in simple cases and prove some bounds in more general cases.

Hausdorff dimension

Linear weights and $d = 3$

$$d_H = \frac{3(1+\sqrt{1+16y})}{8y}, \quad y = w_3/w_2$$

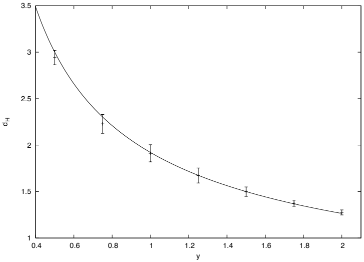


FIGURE 13. Equation (4.25) compared to simulations. The Hausdorff dimension, d_H, is plotted against $y = w_3/w_2$. The leftmost datapoint is calculated from 50 trees on 50000 vertices and the others are calculated from 50 trees on 10000 vertices.

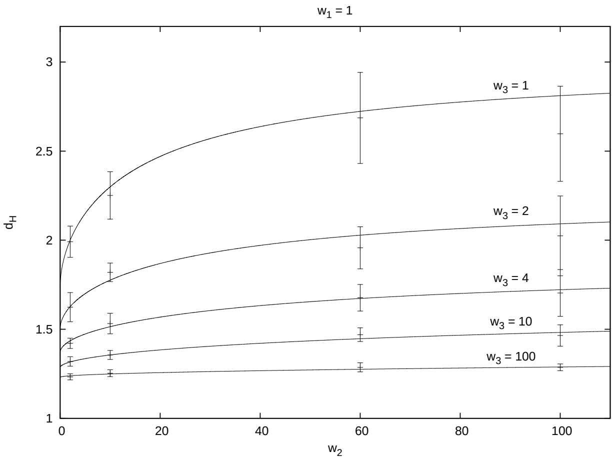
Hausdorff dimension

General solution for $d = 3$

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

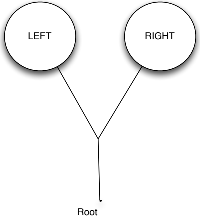
$$w_{3,2} = w_3/3$$

$$w_{3,1} = w_{2,2} = w_2/3$$



The mass distribution

Consider trees with vertices of order $1, 2$ and 3, i.e. $d = 3$.



What is the distribution of the size of the left (or right) tree as the total size of the tree gets large?

Has been studied for preferential attachment trees and is well understood for generic trees.

We study this for the SV model with linear splitting weights.

Definitions

- Let $\nu_0(N) =$ Probability that the left (or right) subtree is empty in a tree of size N
- Let $\nu_1(N_1, N_2) =$ Probability that the left and right subtrees have sizes N_1 and N_2
- Put $F_0(z) = \sum_N z^N \nu_0(N)$
- Put $F(x, y) = \sum_{N_1, N_2} x^{N_1} y^{N_2} \nu_1(N_1, N_2)$

Equations

- The splitting rules give linear relations between $\nu_0(N)$, $\nu_1(N_1, N_2)$ with $N_1 + N_2 = N - 1$ and the same functions with N replaced by $N + 1$.
- Find linear PDEs for the generating functions F_0 and F.

$$\partial_z F_0(z) + \frac{(\beta - \alpha + 1)z + \alpha - 1}{(1-z)z} F_0(z) = \frac{(Az + B)z}{(1-z)^2}$$

$$M(x)\partial_x F(x,y) + M(y)\partial_y F(x,y) + C(x,y)F(x,y) = D(x,y)$$

$$\begin{aligned} M(x) &= w_2(1-x)x, \quad C(x,y) = 2w_2 + b - (w_2 + b - w_1 + w_{2,3})(x+y), \\ D(x,y) &= w_{3,1}\left(x^{-2}F_0(x) + y^{-2}F_0(y)\right) \end{aligned}$$

- These equations can be solved in closed form.

Results

Use the scaling assumptions

$$\nu_1(xN, (1-x)N) \sim N^{-\lambda}$$

for $0 < x < 1$ as $N \to \infty$ and

$$\nu_1(N_1, N_2) \sim N_1^{-\lambda_1}$$

for $N_2 \gg N_1 \gg 1$

- Scaling assumptions agree with explicit calculations of the ν-functions
- $\lambda = \lambda_1 = b/3 + 4/3 \ (w_i = ai + b,\ 2a + b = 1)$
- In the infinite volume limit one of the subtrees is infinite with probability 1 and the size of the other one is distributed as $N^{-\lambda}$. Proven in a special case.

The infinite volume limit

Let Γ_N be rooted planar trees with vertices of degree at most D. Let μ_N be the probability measure on Γ_N induced by the vertex splitting procedure.

Theorem *Assume that the partitioning weights $w_{1,k}$ are nonzero for $k = 2, \ldots, D$. Then the measures μ_N converge weakly as $N \to \infty$ to a probability measure μ on the set of infinite trees which is independent of the initial tree.*

The infinite volume limit

Let Γ_N be rooted planar trees with vertices of degree at most D. Let μ_N be the probability measure on Γ_N induced by the vertex splitting procedure.

Theorem *Assume that the partitioning weights $w_{1,k}$ are nonzero for $k = 2, \ldots, D$. Then the measures μ_N converge weakly as $N \to \infty$ to a probability measure μ on the set of infinite trees which is independent of the initial tree.*

The proof uses a result on the convergence of inhomogeneous Markov Chains and standard properties of the set of trees with the metric

$$d(T_1, T_2) = \inf\{R^{-1} : B_R(T_1) = B_R(T_2)\}$$

where $B_R(T)$ is the subtree of T spanned by vertices within distance R from the root.

The main problem is to prove the convergence of

$$\mu_N(\{T : B_R(T) = T_0\}$$

as $N \to \infty$.

The proof uses a result on the convergence of inhomogeneous Markov Chains and standard properties of the set of trees with the metric

$$d(T_1, T_2) = \inf\{R^{-1} : B_R(T_1) = B_R(T_2)\}$$

where $B_R(T)$ is the subtree of T spanned by vertices within distance R from the root.

The main problem is to prove the convergence of

$$\mu_N(\{T : B_R(T) = T_0\}$$

as $N \to \infty$.

Exactly soluble special case (S. Stefansson)

For linear splitting weights, $D = 3$ and if the splitting weight w_r of the root is defined in a special way, then the model becomes Markovian self-similar if $w_{2,2} = 2w_{3,2} = w_r$.
This means that

$$P_N(T) = Q(N_1, N_2) P_{N_1}(T_1) P_{N_2}(T_2)$$

where P_N is a probability distribution on the set of trees of size N, Q is a positive function and $N_{1,2}$ are the sizes of the left and right subtrees $T_{1,2}$ of T.

One can show explicitly that

- there is a unique spine
- the probability distribution for outgrowths from the spine can be described explicitly
- the Hausdorff dimension can be calculated and varies continuously with the parameters of the model from 1 to ∞
- There is a closed form for the ν functions.
- Can be generalized to trees with vertices of higher order.

Conclusions and problems

- The vertex splitting model encompasses a very large class of random growing trees as well as generic trees – in limiting cases at least.
- Can one generalize the exact solution in the MSS case to the general case?
- What are the spectral properties?
- How can one characterize the infinite volume limits?