Random trees and vertex splitting

Thordur Jonsson, University of Iceland

September 2009

Statistical physics, combinatorics and probability: from discrete to continuous models

Institut Henri Poincare

Outline

- Introduction and background
- Definition of the vertex splitting model
- Vertex degree distribution
- Correlations
- Structure functions and the Hausdorff dimension
- Open problems

Random graphs: two main approaches

1. Equilibrium statistical mechanics

   $\mathcal{T}$= Set of graphs, μ a probability measure on $\mathcal{T}$

   $$\mu(T) = Z^{-1} e^{-\beta E(T)}$$

2. Growing trees $T_n \mapsto T_{n+1}$, time discrete

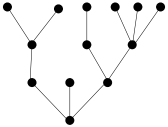
   Stochastic growth rules induce a probability measure on $\mathcal{T}_t$, the trees that can arise in t steps

- Sometimes (1) is more natural
- Sometimes (2) is more natural
- Sometimes (1) and (2) are known to be equivalent

Galton-Watson trees

- p_n = probability of having n descendents,

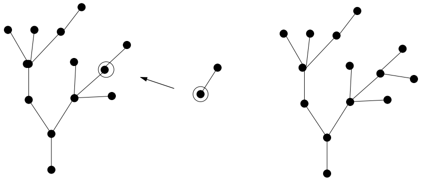
$$\sum_n p_n = 1, \quad m = \sum_n n p_n$$



- $m < 1$ subcritical, $m > 1$ supercritical, $m = 1$ critical
- n generations at time $t = n - 1$ if no extinction
- Well understood

Preferential attachment trees

- In each timestep one new edge is attached to a preexisting tree



- Probability of attaching to a vertex v of degree k

$$P_v = \frac{w_k}{\sum_k n_k w_w}, \quad w_k \geq 0,$$

$n_k =$ the number of vertices of degree k.

- Growth rule induces a probability measure on $\mathcal{T}_t$.

Local trees

- Weight factor of a tree T

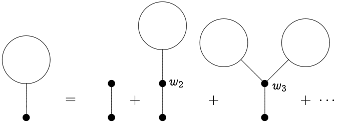
$$W(T) = \prod_{i \in T} w_{\sigma(i)}$$

$\sigma(i) =$ degree of the vertex i

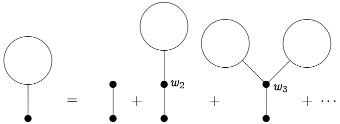
- Partition functions

$$Z_N = \sum_{T:|T|=N} W(T), \quad Z = \sum_N \zeta^N Z_N, \quad |\zeta| < \zeta_0$$

- Generating function $g(z) = \sum_n w_n z^{n-1}$, radius of convergence ρ
- Main equation



Generic trees



- Algebraically

$$Z(\zeta) = \zeta g(Z(\zeta)) = \zeta \sum_{i=0}^{\infty} w_{i+1} Z^i(\zeta)$$

- Define $Z_0 = \lim_{\zeta \to \zeta_0} Z(\zeta)$
- If $Z_0 < \rho$ then we say that the trees are *generic*.
- $Z(\zeta) - Z_0 \sim \sqrt{\zeta_0 - \zeta}$

- Define

$$\mu(T) = Z_0^{-1}\zeta_0^{|T|}\prod_{i\in T} w_{\sigma(i)}$$

Probability measure

- This measure is the same as the one obtained from a Galton-Watson process with

$$p_n = \zeta_0 w_{n+1} Z_0^{n-1}$$

- Define

$$\mu(T) = Z_0^{-1} \zeta_0^{|T|} \prod_{i \in T} w_{\sigma(i)}$$

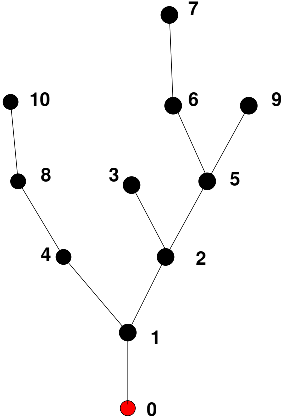
  Probability measure

- This measure is the same as the one obtained from a Galton-Watson process with

$$p_n = \zeta_0 w_{n+1} Z_0^{n-1}$$

Another equivalence

- Preferential attachment trees $\approx$ Causal trees
- Weight proportional to the number of *causal labelings*



- More branchings, more ways to grow

Properties of generic trees

- Let $V_T(R)$=volume of a ball of radius R centered on the root

$$\langle V_T(R)\rangle \sim R^{d_H}, \quad R\to\infty \quad \text{defines } d_H$$

- Let $p_T(t)$ = probability that a random walker is back at the root after t steps on T

$$\langle p_T(t)\rangle \sim t^{-d_s/2}, \quad t\to\infty \quad \text{defines } d_s$$

Properties of generic trees

- Let $V_T(R)$=volume of a ball of radius R centered on the root

$$\langle V_T(R)\rangle \sim R^{d_H}, \quad R \to \infty \ \text{ defines } d_H$$

- Let $p_T(t)$ = probability that a random walker is back at the root after t steps on T

$$\langle p_T(t)\rangle \sim t^{-d_s/2}, \quad t \to \infty \ \text{ defines } d_s$$

Properties of generic trees

- Let $V_T(R)$=volume of a ball of radius R centered on the root

$$\langle V_T(R)\rangle \sim R^{d_H}, \quad R \to \infty \quad \text{defines } d_H$$

- Let $p_T(t)$ = probability that a random walker is back at the root after t steps on T

$$\langle p_T(t)\rangle \sim t^{-d_s/2}, \quad t \to \infty \quad \text{defines } d_s$$

Properties of generic trees

- Let $V_T(R)$=volume of a ball of radius R centered on the root

$$\langle V_T(R)\rangle \sim R^{d_H}, \quad R \to \infty \;\; \text{defines } d_H$$

- Let $p_T(t)$ = probability that a random walker is back at the root after t steps on T

$$\langle p_T(t)\rangle \sim t^{-d_s/2}, \quad t \to \infty \;\; \text{defines } d_s$$

- Averages taken w.r.t. a measure on infinite trees

$$\nu_N = Z_N^{-1} \prod_{i \in T} w_{\sigma(i)}$$

$$\nu_N \to \nu_\infty \quad \text{as} \quad N \to \infty$$

- Properties:

  (i) $d_H = 2$

  (ii) $d_s = 4/3$

  (iii) There is a unique infinite simple path whose outgrowths are critical GW-trees

- Averages taken w.r.t. a measure on infinite trees

$$\nu_N = Z_N^{-1} \prod_{i \in T} w_{\sigma(i)}$$

$$\nu_N \to \nu_\infty \quad \text{as} \quad N \to \infty$$

- Properties:
  (i) $d_H = 2$
  (ii) $d_s = 4/3$
  (iii) There is a unique infinite simple path whose outgrowths are critical GW-trees

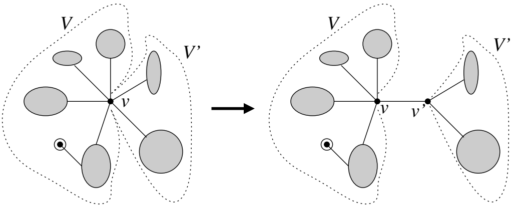
Preferential attachment trees

- In general many infinite simple paths
- $d_H = \infty$ in many cases (all cases?)
- Can calculate vertex degree distribution - and fluctuations. Independent of the initial tree.
- Broad distribution of sizes of subtrees, depends on the initial tree

The vertex splitting model

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

- A model of randomly growing rooted, planar trees



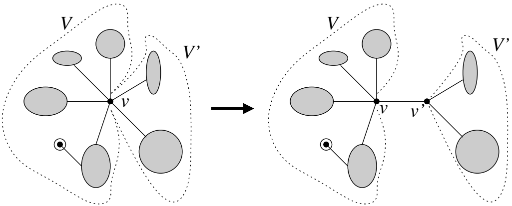
- Degree of vertices is bounded by an integer d (we also discuss the case $d = \infty$)
- Nonnegative splitting weights $w_1, w_2, \ldots, w_d$
- $n_j(T)$ = the number of vertices of degree j in a tree T
  p_j = Probability of choosing a vertex $v \in T$ of degree j

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

The vertex splitting model

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

- A model of randomly growing rooted, planar trees



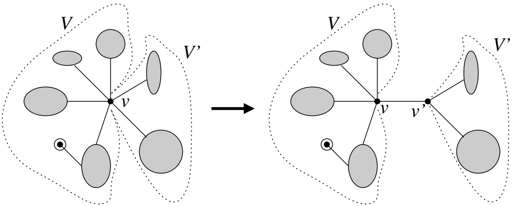
- Degree of vertices is bounded by an integer d (we also discuss the case $d = \infty$)
- Nonnegative splitting weights $w_1, w_2, \ldots, w_d$
- $n_j(T)$ = the number of vertices of degree j in a tree T
  p_j = Probability of choosing a vertex $v \in T$ of degree j

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

The vertex splitting model

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

- A model of randomly growing rooted, planar trees



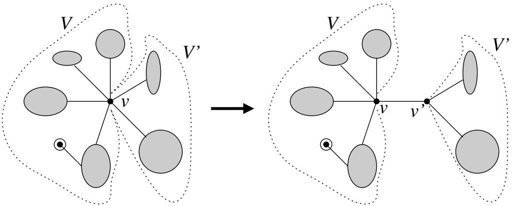
- Degree of vertices is bounded by an integer d (we also discuss the case $d = \infty$)
- Nonnegative splitting weights $w_1, w_2, \ldots, w_d$
- $n_j(T)$ = the number of vertices of degree j in a tree T
  p_j = Probability of choosing a vertex $v \in T$ of degree j

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

The vertex splitting model

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

- A model of randomly growing rooted, planar trees



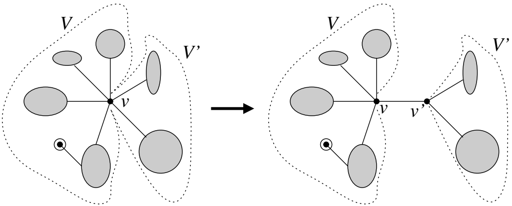
- Degree of vertices is bounded by an integer d (we also discuss the case $d = \infty$)
- Nonnegative splitting weights $w_1, w_2, \ldots, w_d$
- $n_j(T)$ = the number of vertices of degree j in a tree T
  p_j = Probability of choosing a vertex $v \in T$ of degree j

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

The vertex splitting 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

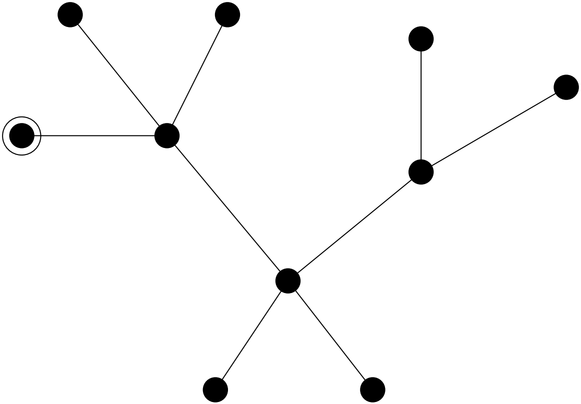
- The parameters of the model are

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

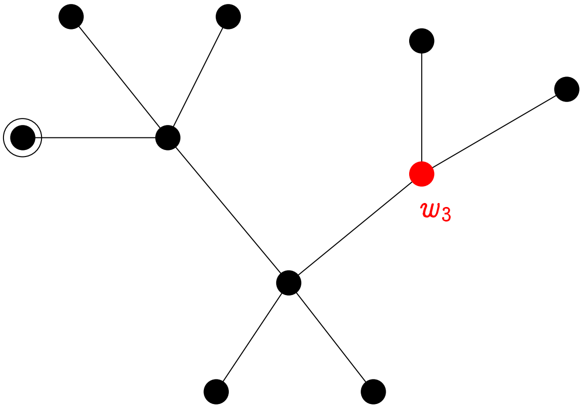
  a symmetric matrix of non–negative partitioning weights
- Split a vertex of degree i into vertices of degree k and $i+2-k$ with probability $w_{k,i+2-k}/w_i$ – all such splittings equally probable
- The splitting weights $w_1, w_2, \ldots, w_d$ are related to the partitioning weights by

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

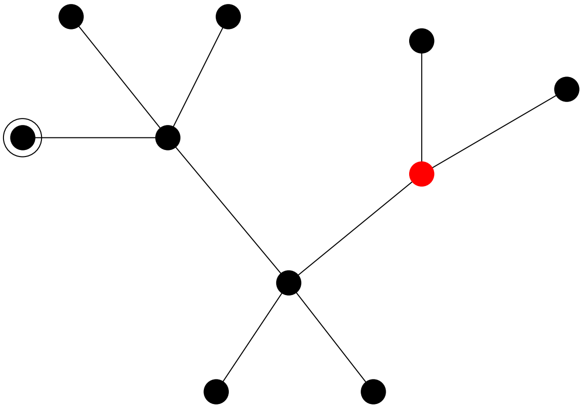
A tree



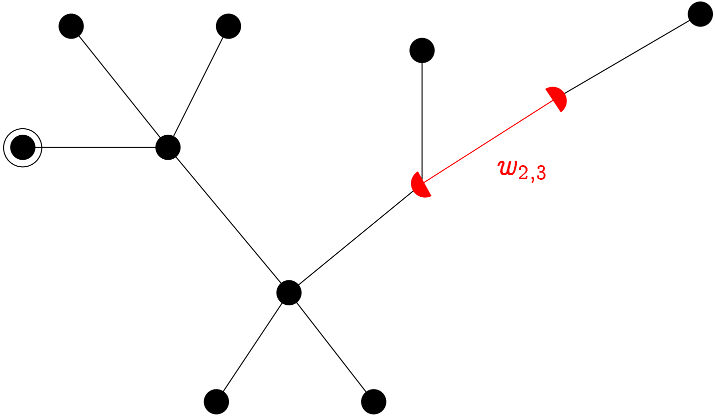
Vertex splitting rules



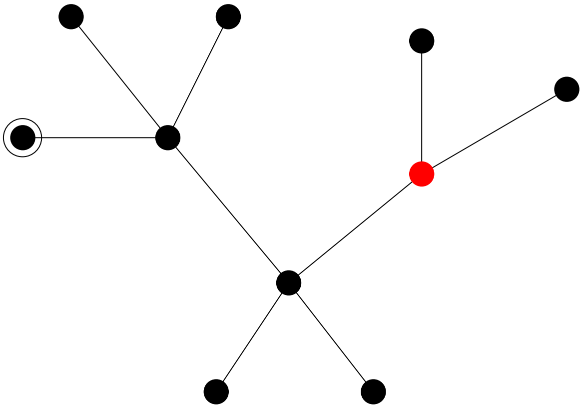
Vertex splitting rules



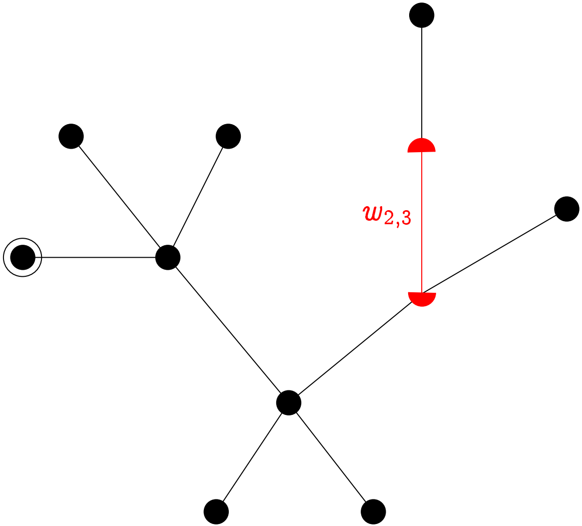
Vertex splitting rules



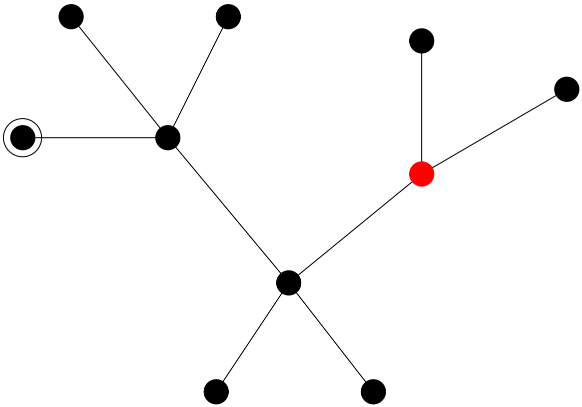
Vertex splitting rules



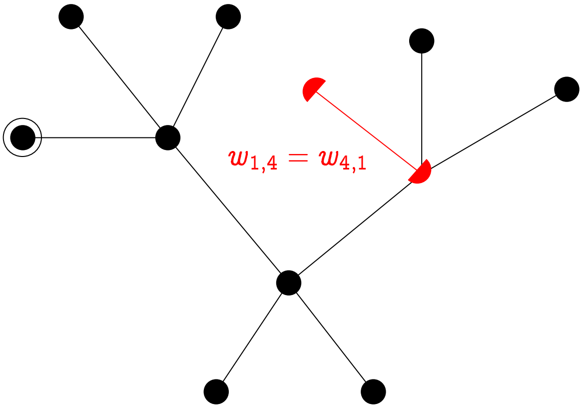
Vertex splitting rules



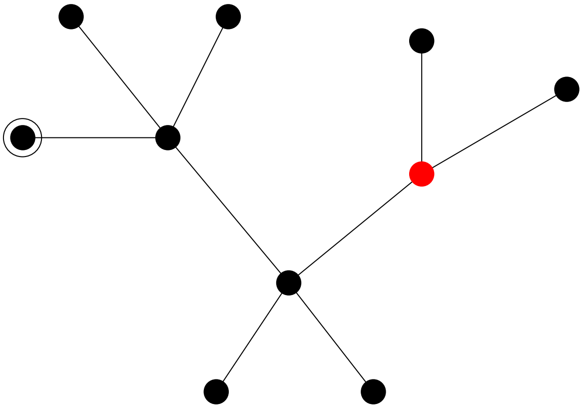
Vertex splitting rules



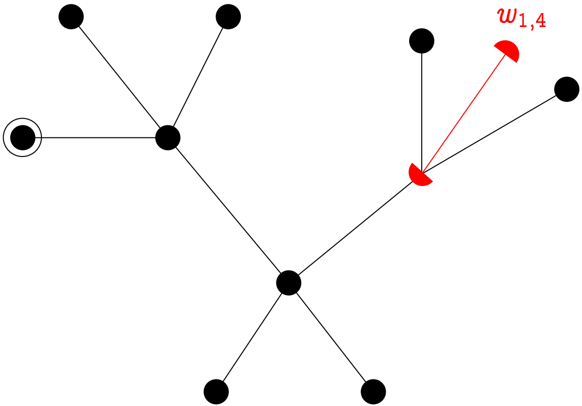
Vertex splitting rules



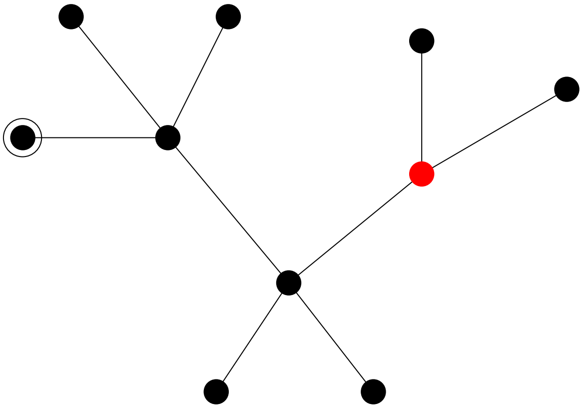
Vertex splitting rules



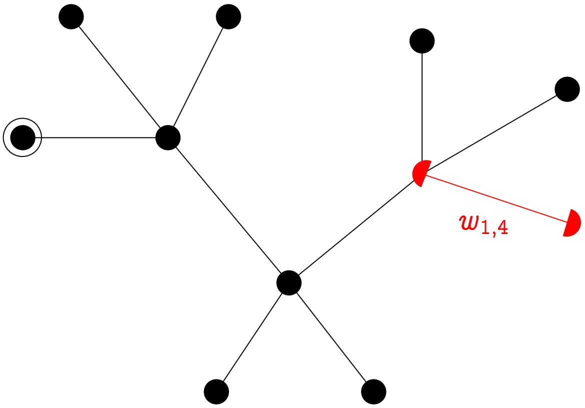
Vertex splitting rules



Vertex splitting rules

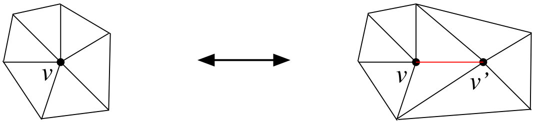


Vertex splitting rules



Relation to other models

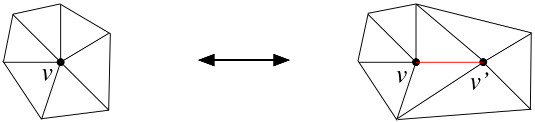
- Ergodic moves in Monte Carlo simulations of triangulations in 2d-quantum gravity



- A tree growth process which arises in the study of 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
- For $d = 3$, in a certain limit, it reduces to Ford's alpha model for phylogenetic trees

Relation to other models

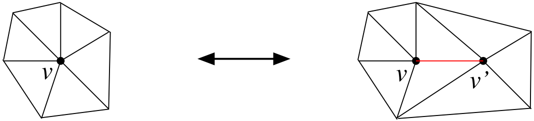
- Ergodic moves in Monte Carlo simulations of triangulations in 2d-quantum gravity



- A tree growth process which arises in the study of 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
- For $d = 3$, in a certain limit, it reduces to Ford's alpha model for phylogenetic trees

Relation to other models

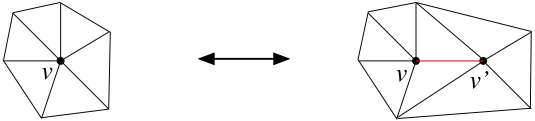
- Ergodic moves in Monte Carlo simulations of triangulations in 2d-quantum gravity



- A tree growth process which arises in the study of 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
- 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



- A tree growth process which arises in the study of 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
- For $d = 3$, in a certain limit, it reduces to Ford's alpha model for phylogenetic trees

Main results

- Distribution of vertex degrees in a large tree
- Correlations between the degrees of vertices
- "Shape" of trees – Hausdorff dimension

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.

Recurrence for generating functions

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 generating 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 recurrence

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

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

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

Vertex degree distribution

- 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

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

Vertex degree distribution

- 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 $\rho(t)$.

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

$$\Longrightarrow$$

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

- 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.

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

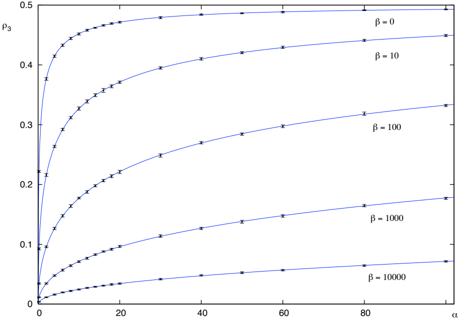


FIGURE 4. The value of ρ_3 as given in (2.45) compared to results from simulations. Each point is calculated from 20 trees on 10000 vertices.

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)

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

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)

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

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)

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

assuming the existence of the limit.

Explicit solutions

Can solve in simple cases and find nontrivial correlations
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

$$\rho_{21} = \frac{4(3-y)}{7(11-2y)}, \qquad \rho_{31} = \frac{10}{7(11-2y)},$$

$$\rho_{22} = \frac{4y^2 - 12y + 105}{7(2y+7)(11-2y)}, \qquad \rho_{32} = \frac{2(-8y^2 + 18y + 63)}{7(2y+7)(11-2y)},$$

$$\rho_{23} = \frac{2(-4y^2 + 20y + 21)}{7(2y+7)(11-2y)}, \qquad \rho_{33} = \frac{8(3y-14)}{7(2y+7)(2y-11)}.$$

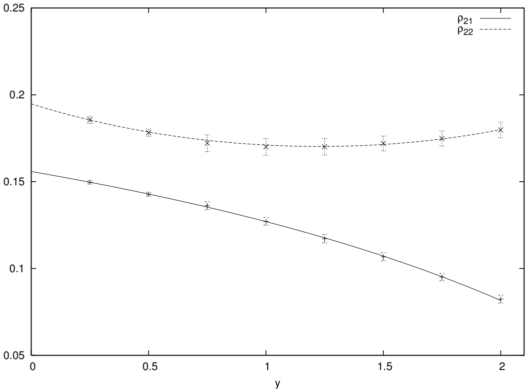
Sum rules

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

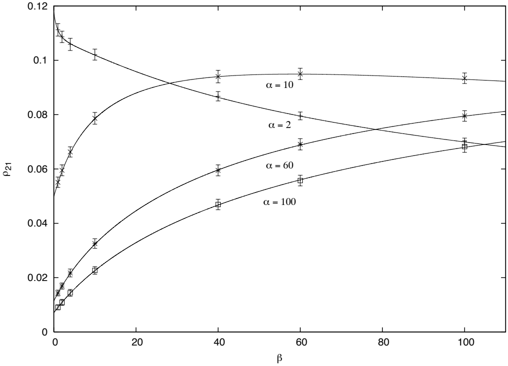


Nonlinear splitting weights

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

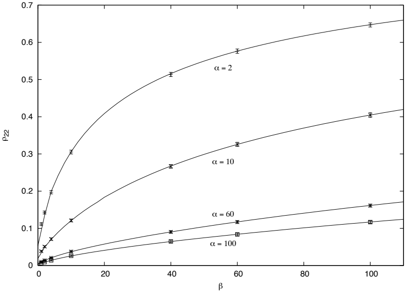


FIGURE 17. The solution (5.5) for the density ρ_{22} plotted as a function of β for a few values of α. Each datapoint is calculated from simulations of 100 trees on 10000 vertices.

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

Subtree probabilities

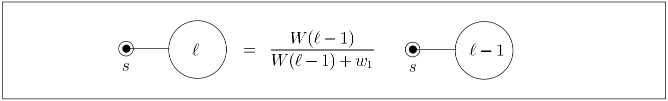
- Label vertices in the tree by their time of creation
- Use linear weights
- Derive expressions for the probabilistic structure of the tree as seen from the vertex created at a given time
- Average over the creation time
- Introduce a scaling assumption
- Extract the Hausdorff dimension
- Get results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

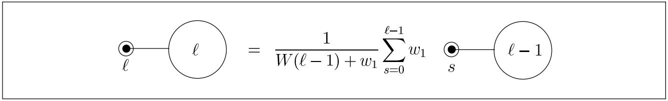
- Begin with a tree consisting of a single vertex at time $t = 0$
- In a tree of size ℓ let $p_R(\ell; s)$ be the probability that the vertex created at time $s \leq \ell$ is the root
- We find

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

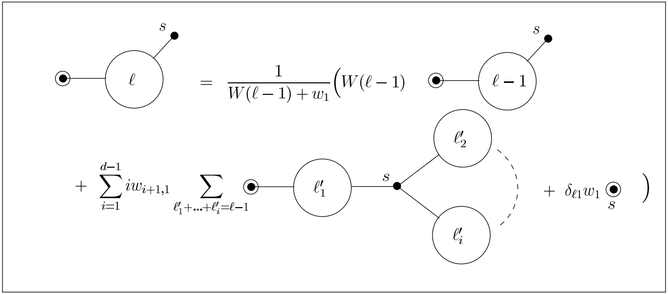
$$p_R(\ell; \ell) = \frac{1}{W(\ell-1)+w_1} \sum_{s=0}^{\ell-1} w_1 p_R(\ell-1; s), \quad s = \ell$$

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

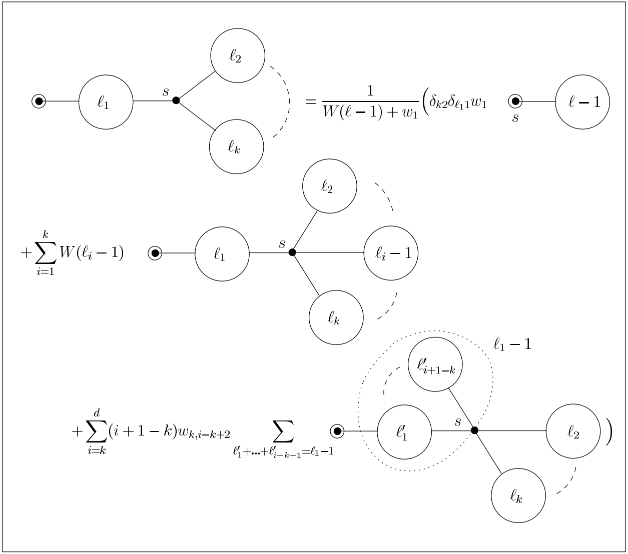




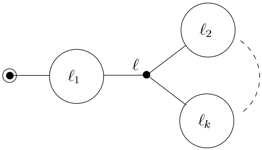
Let $p_k(\ell_1, \ell_2, \ldots, \ell_k; s)$ be the probability that the vertex v created at time s has degree k, the root subtree has ℓ_1 links and the other subtrees incident on v have size $\ell_2, \ldots, \ell_k$. Denote the sum of the ℓ_i's by ℓ. Then for $k = 1$ and $s < \ell$



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

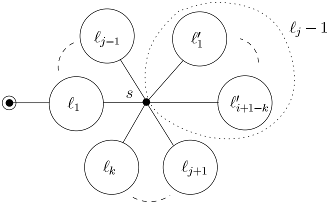


Finally $k > 1$ and $s = \ell$



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

$\times$



We average over s to get simpler recursions:

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

$$p_1(\ell+1) \tag{3.11}$$
$$= \frac{\ell+1}{\ell+2}\frac{1}{W(\ell)+w_1}\Big[W(\ell)p_1(\ell) + \sum_{i=1}^{d-1} i w_{i+1,1} \sum_{\substack{\ell'_1+\ldots+\ell'_i \\ =\ell}} p_i(\ell'_1,\ldots,\ell'_i) + 2\delta_{\ell 0} w_1\Big].$$

$$p_k(\ell_1,\ldots,\ell_k)$$
$$= \frac{\ell+1}{\ell+2}\frac{1}{W(\ell)+w_1}\Big[\delta_{k2}\delta_{\ell_1 1} w_1 p_R(\ell) + \sum_{i=1}^{k} W(\ell_i-1) p_k(\ell_1,\ldots,\ell_i-1,\ldots,\ell_k)$$
$$+ \sum_{i=k}^{d} (i-k+1)\, w_{k,i-k+2} \sum_{\substack{\ell'_1+\ldots+\ell'_{i+1-k} \\ =\ell_1-1}} p_i(\ell'_1,\ldots,\ell'_{i+1-k},\ell_2,\ldots,\ell_k) \tag{3.12}$$
$$+ \sum_{j=2}^{k}\sum_{i=k-1}^{d} w_{k,i-k+2} \sum_{\substack{\ell'_1+\ldots+\ell'_{i+1-k} \\ =\ell_j-1}} p_i(\ell_1,\ldots,\ell_{j-1},\ell'_1,\ldots,\ell'_{i+1-k},\ell_{j+1},\ldots,\ell_k)\Big]$$

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

$$\begin{aligned} q_{ki}(\ell_1,\ell_2) \;&=\; \frac{\ell+1}{\ell+2}\,\frac{1}{W(\ell)+w_1}\Big[\\ &\sum_{j=k-1}^{d} w_{k,j+2-k}\Big((j-i)q_{ji}(\ell_1-1,\ell_2)+iq_{j,j-(k-i)}(\ell_1,\ell_2-1)\Big) \\ &+\Big(W(\ell_1-1)+(k-i-1)(w_2-w_3)\Big)q_{ki}(\ell_1-1,\ell_2) \\ &+\Big(W(\ell_2-1)+(i-1)(w_2-w_3)\Big)q_{ki}(\ell_1,\ell_2-1) \\ &+\delta_{k2}\delta_{\ell_1 1}w_1 p_R(\ell_2)+\delta_{i1}\delta_{\ell_2 1}w_{k,1}\sum_{\ell'_1+\ldots+\ell'_{k-1}=\ell_1} p_{k-1}(\ell'_1,\ldots,\ell'_{k-1})\Big] \end{aligned}$$

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 we wrote down earlier 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 we wrote down earlier 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

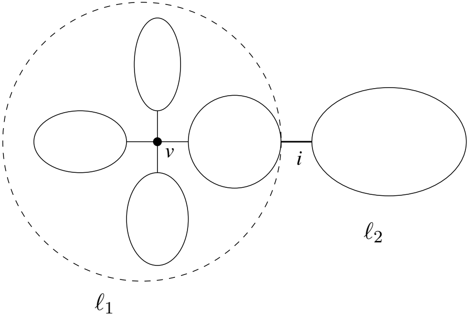
$$R_T = \frac{1}{(2\ell)} \sum_{v \in T} d_T(r, v)\, \sigma(v),$$

- 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 we wrote down earlier 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:

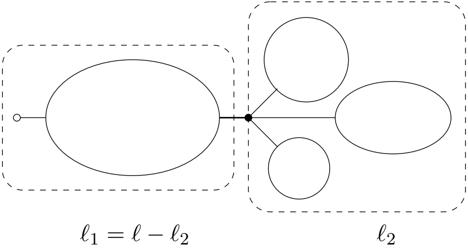
$$\sum_w d_T(v,w)\sigma(w) \;=\; \sum_i (2\ell_2(v;i)+1)$$

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

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

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

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

- 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 $d = 3$

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

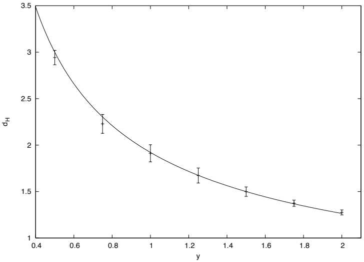


FIGURE 13. Equation (4.25) compared to simulations. The Hausdorff dimension, d_H, is plotted against $y = w_3/w_2$. The leftmost datapoint is calculated from 50 trees on 50000 vertices and the others are calculated from 50 trees on 10000 vertices.

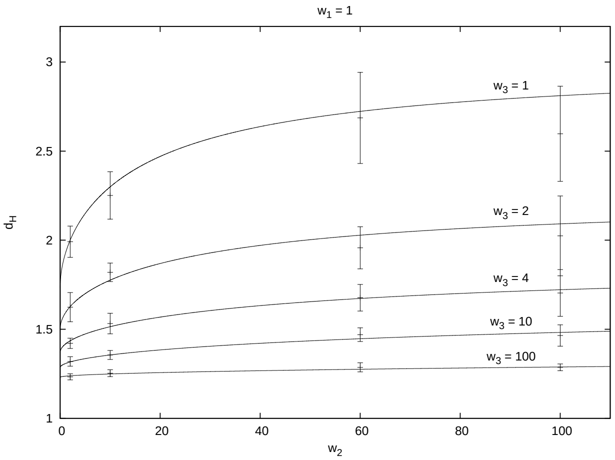
Hausdorff dimension

General solution for $d = 3$

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

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

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



Open problems

- Description in terms of equilibrium statistical mechanics
- The infinite volume limit
- A continuum limit
- Spectral properties
- Properties of the finite volume measures