RNA folding and tree growth models

François David [1], C. Hagendorf [2] & K. J. Wiese [2]

arXiv:0711.3421 [q-bio.BM]

[1] SPhT, CEA/Saclay, France - [2] LPTENS, Ecole Normale, Paris, France

Plan of the talk

1. RNA folding & the random RNA model
2. Greedy approximation
3. Arch deposition model
4. Tree growth model
5. Conclusion

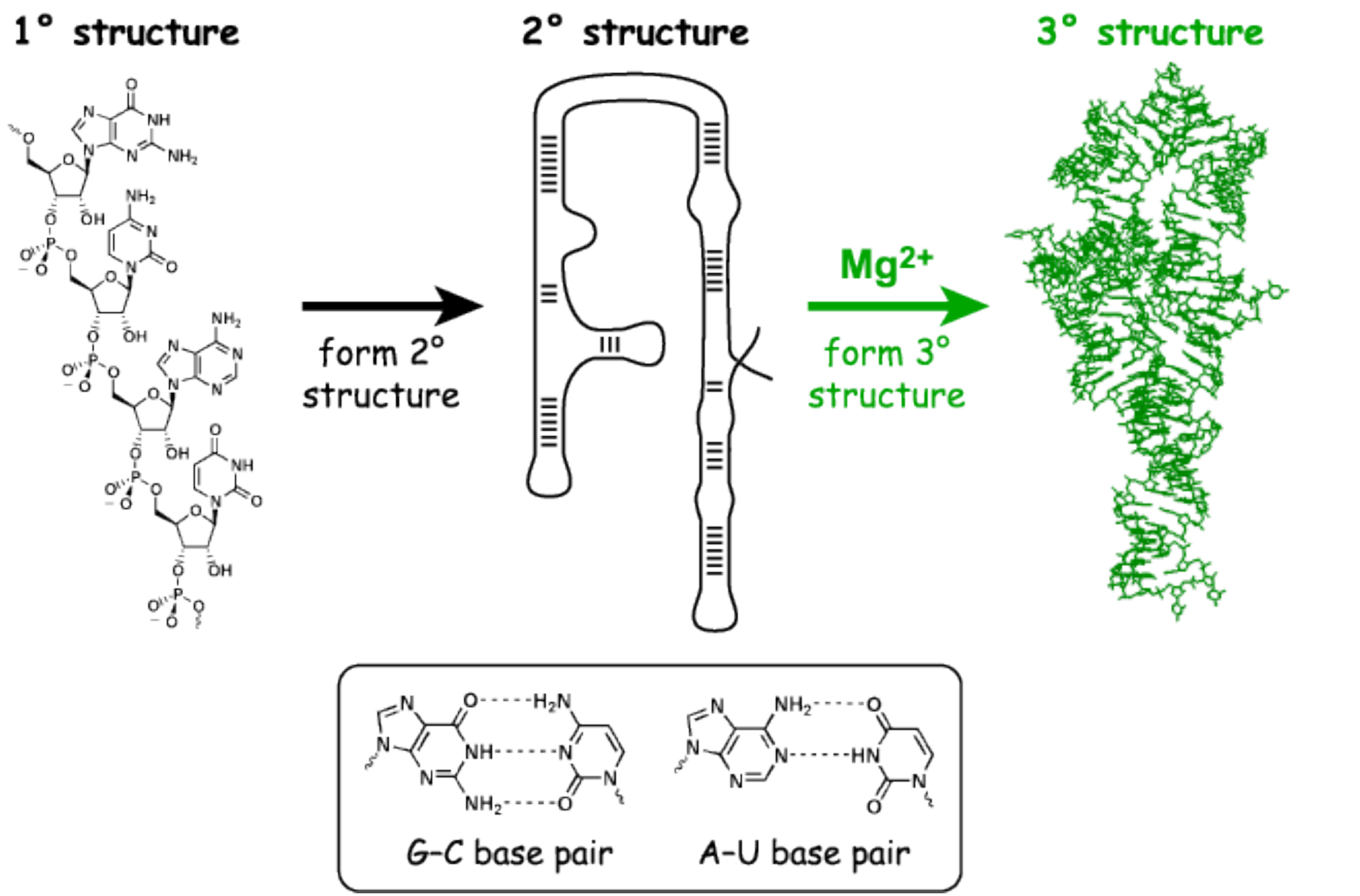
RNA folding

biological(ly inspired) physics: secondary structure of RNA strand

one-dimensional disordered system with long range interactions (pairing) & frustration (topology: planar or almost-planar structure)

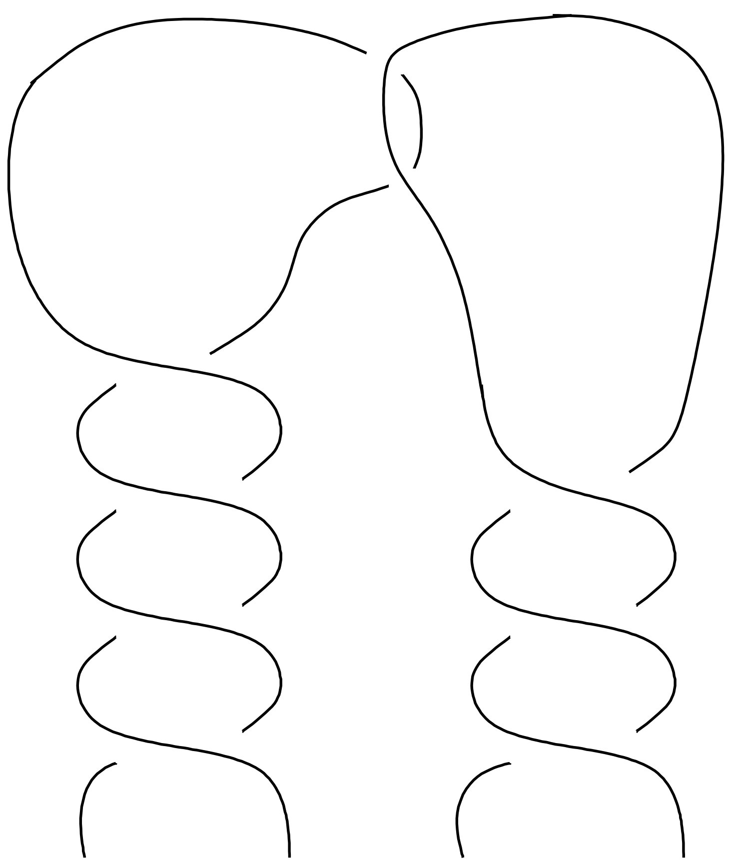
freezing transition, glass-like phase with non-trivial ground-state(s)

RNA primary, secondary and tertiary structure

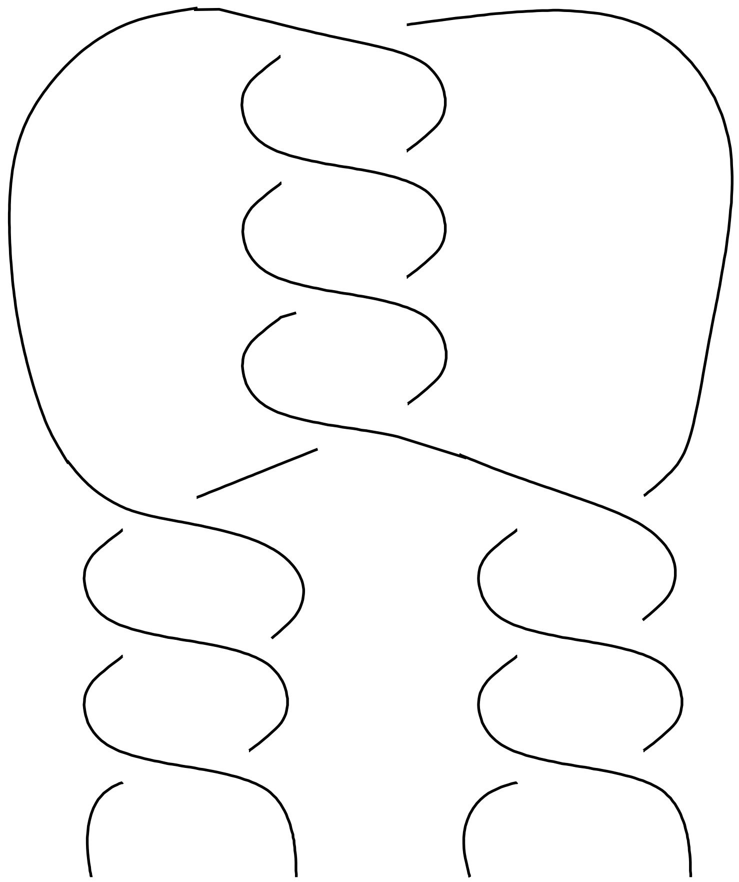


(borrowed from Scott K. Silverman scott@scs.uiuc.edu)

Not so poor approximation

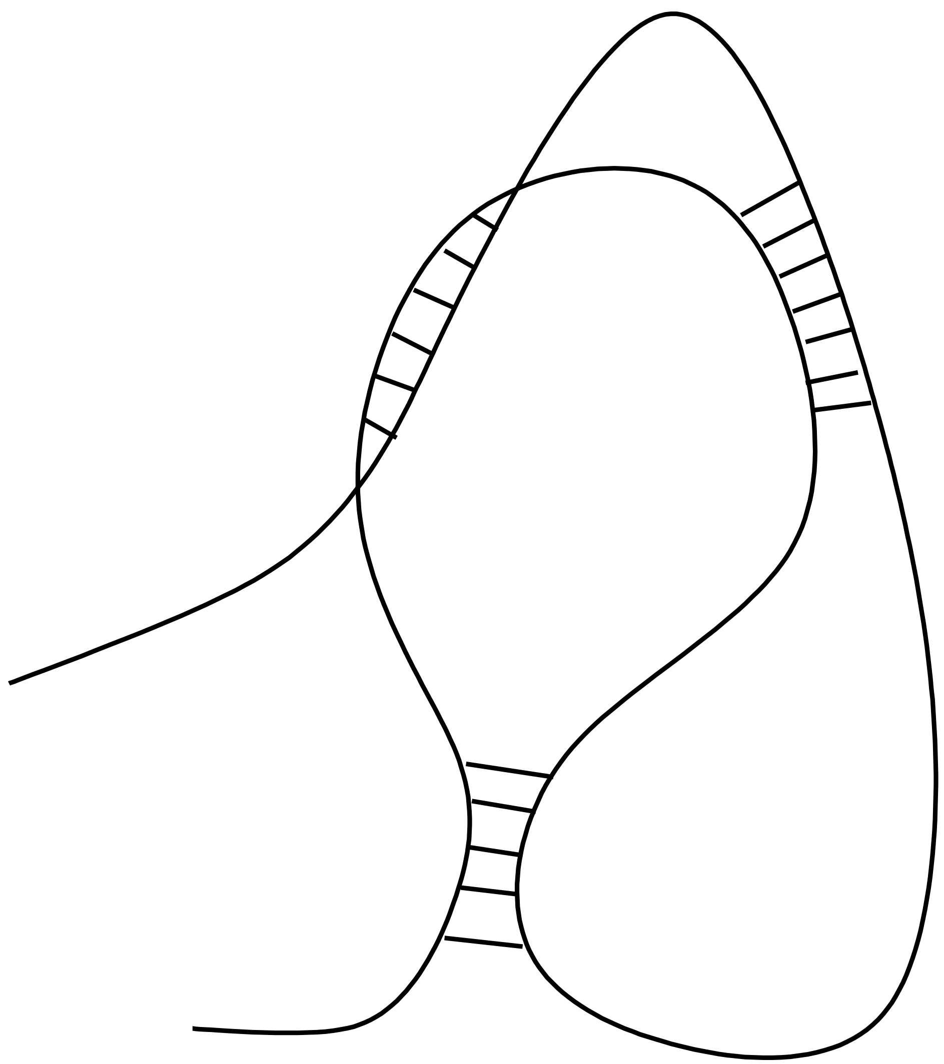


no pseudo-knot

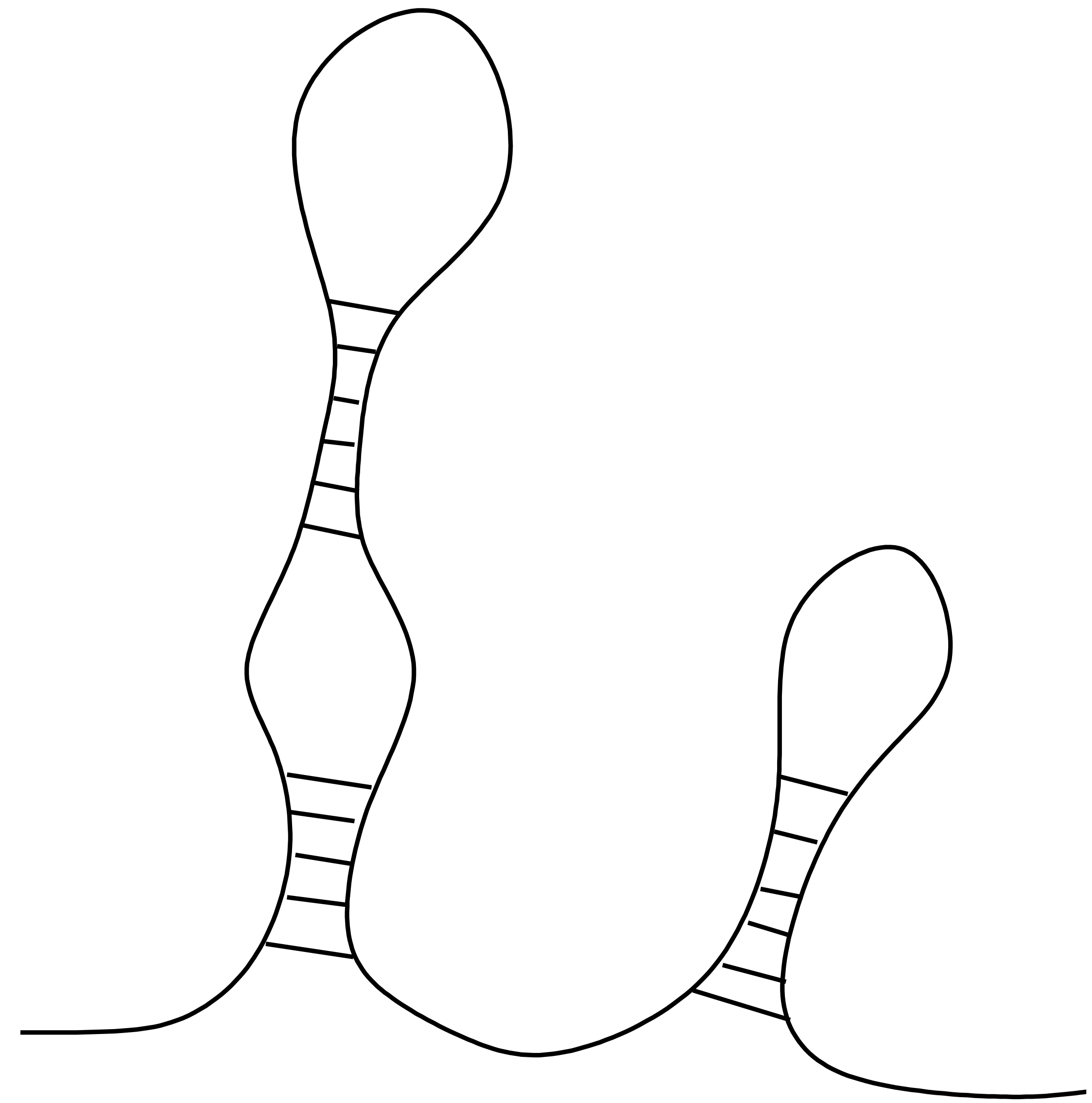


no knot

only planar configurations



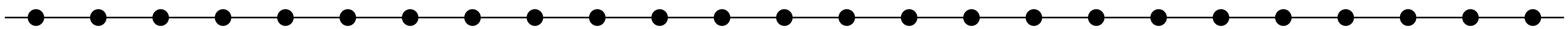
~~non-planar~~



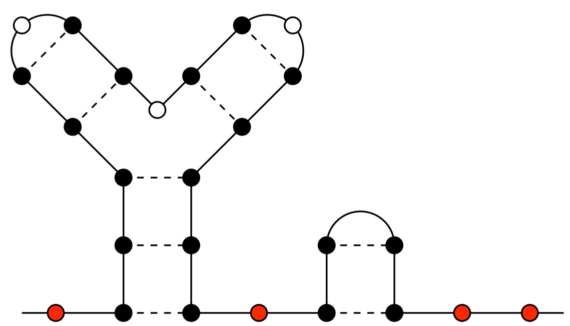
planar = tree

The random bond RNA folding model

1) Start from a strand made of L monomers



2) Fold it in a planar way (neither knots nor pseudo-knots)



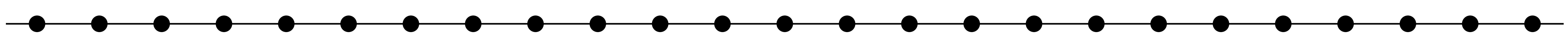
3) Assign to each configuration an energy sum of an energy for each bond

$$\mathcal{E} = \sum_{\text{pairs}\,(ij)} e(i,j)$$

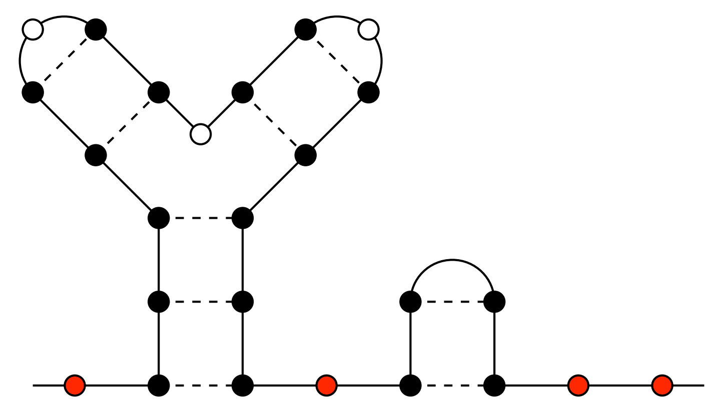
4) Take the bond energies to be independent Gaussian quenched random variables

$$\overline{e(i,j)e_(k,l)} = \sigma\, \delta_{ik}\delta_{jl}$$

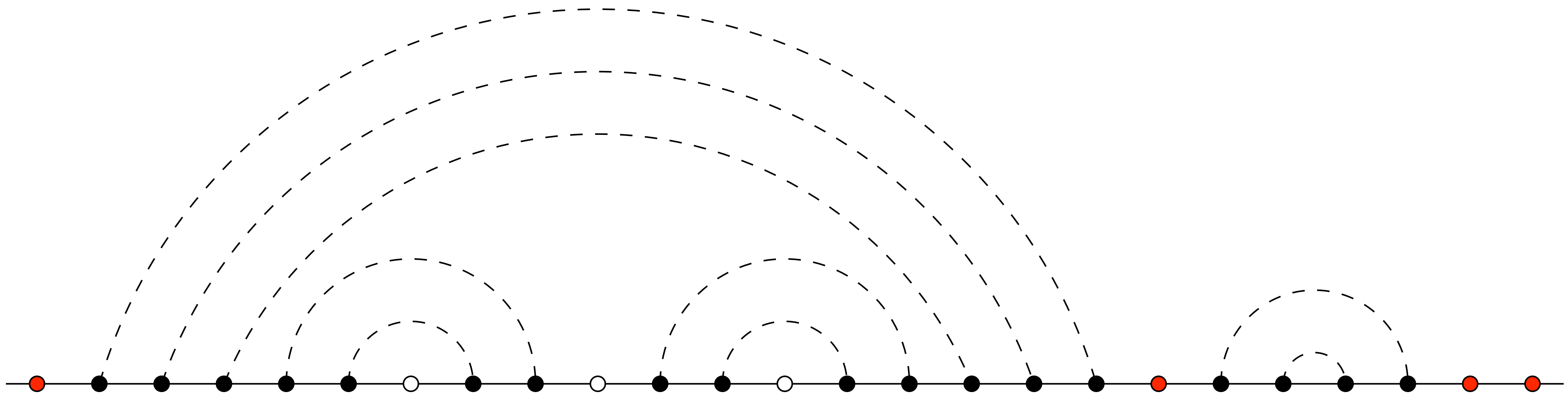
Pairings = Trees = Arch systems = Dyck-like paths



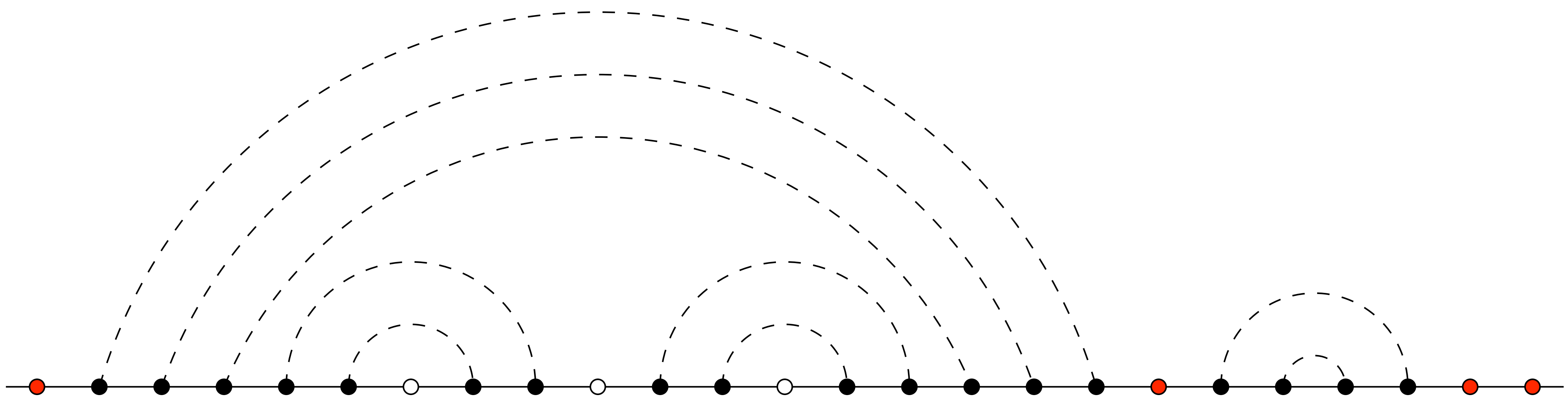
Pairings = Trees = Arch systems = Dyck-like paths



Pairings = Trees = Arch systems = Dyck-like paths

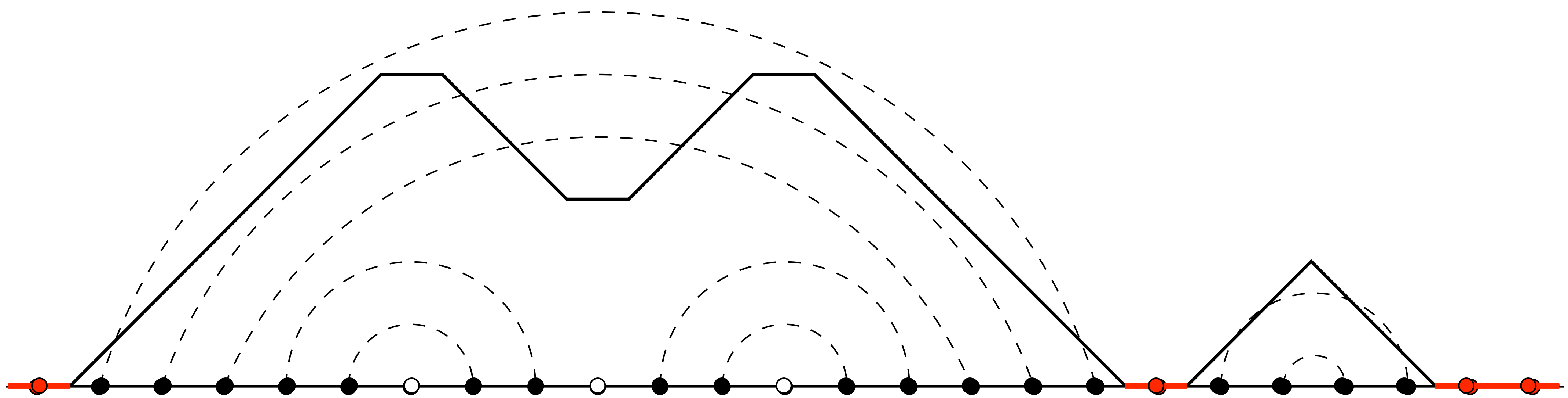


Pairings = Trees = Arch systems = Dyck-like paths



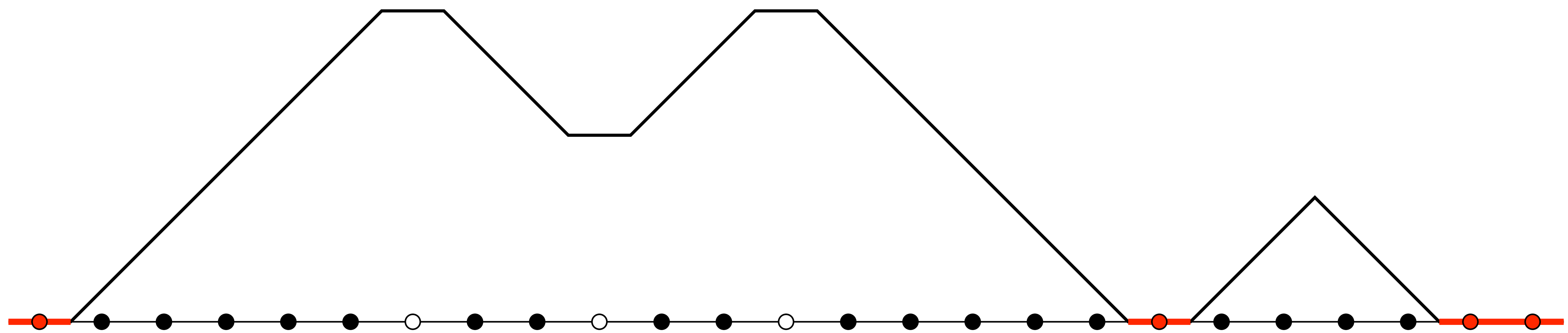
pairing probabilities $\Phi(i, j) \propto |i - j|^{-\rho}$

Pairings = Trees = Arch systems = Dyck-like paths



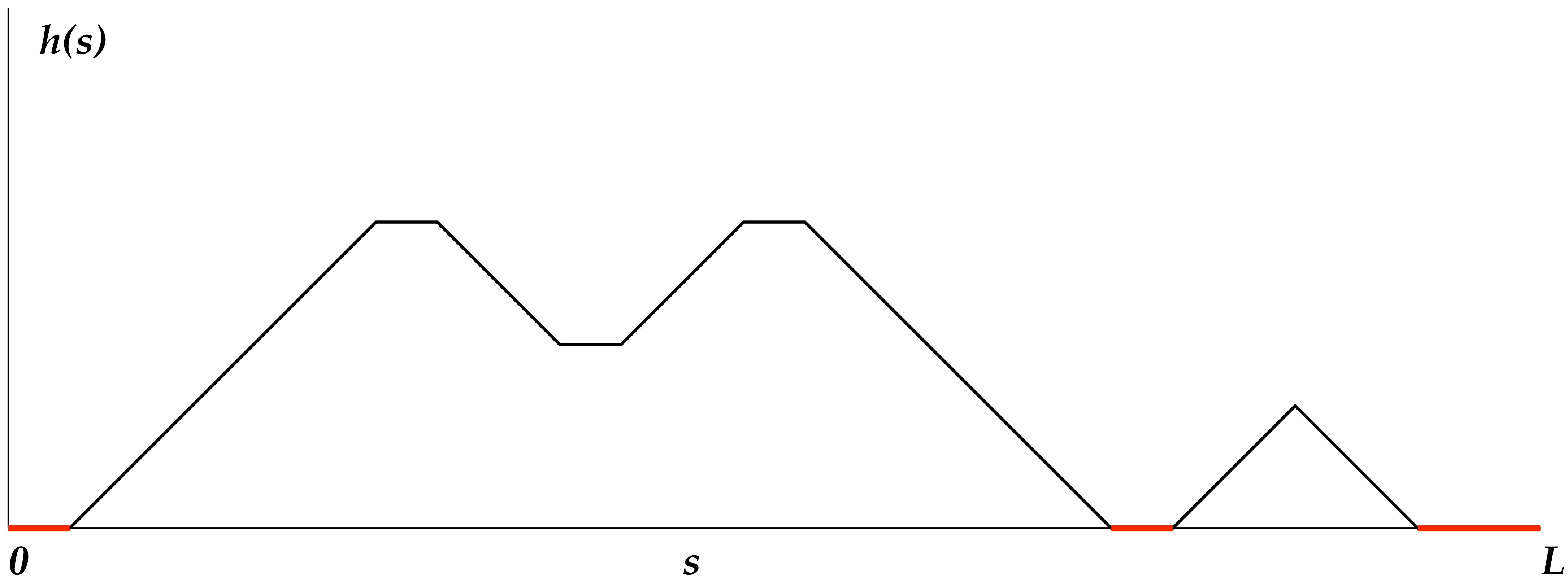
pairing probabilities $\Phi(i, j) \propto |i - j|^{-\rho}$

Pairings = Trees = Arch systems = Dyck-like paths



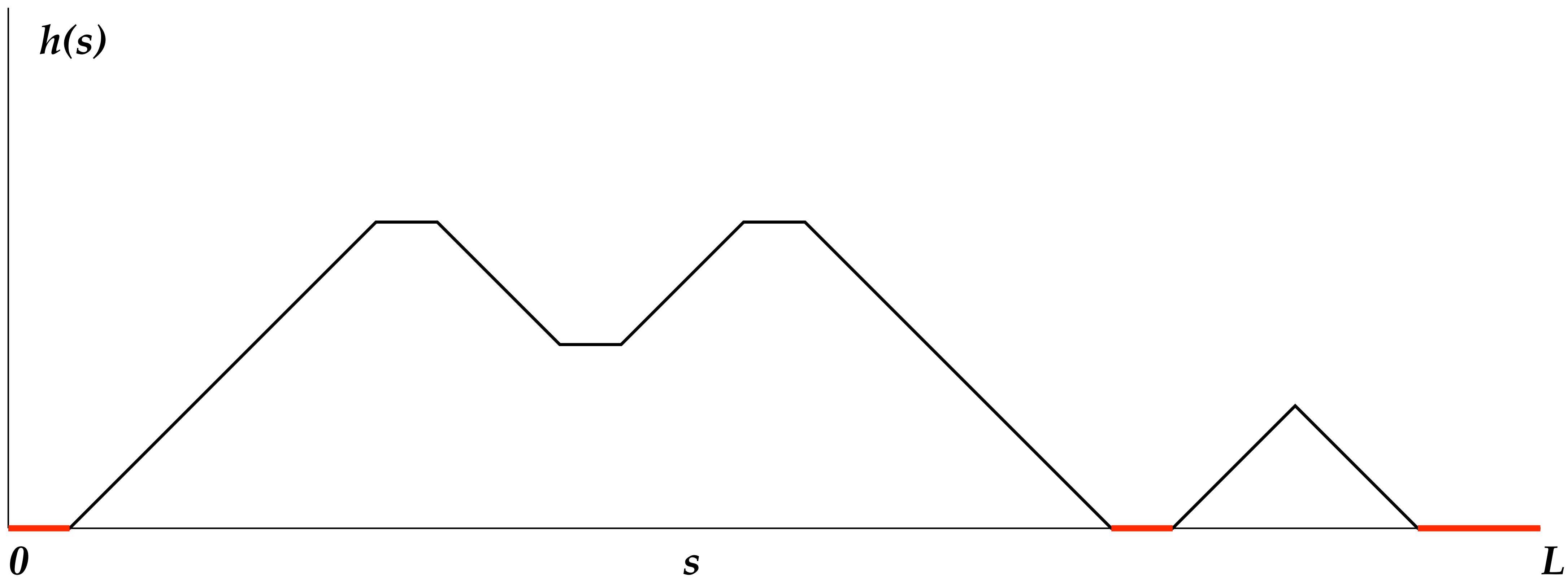
pairing probabilities $\Phi(i, j) \propto |i - j|^{-\rho}$

Pairings = Trees = Arch systems = Dyck-like paths



pairing probabilities $\Phi(i, j) \propto |i - j|^{-\rho}$

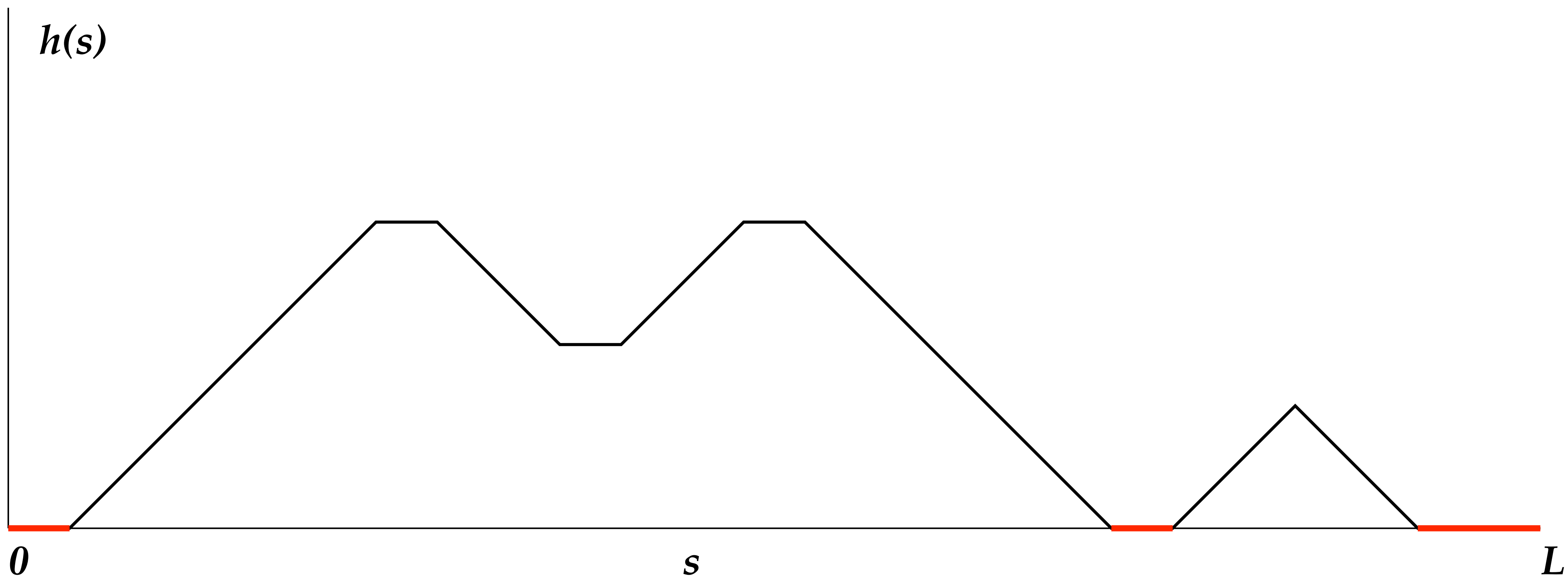
Pairings = Trees = Arch systems = Dyck-like paths



pairing probabilities $\Phi(i,j) \propto |i-j|^{-\rho}$

height function $h(i) \propto i^{\zeta}$ $\rho + \zeta = 2$

Pairings = Trees = Arch systems = Dyck-like paths



pairing probabilities $\Phi(i,j) \propto |i-j|^{-\rho}$

height function $h(i) \propto i^{\zeta}$ $\rho + \zeta = 2$

fractal dimension of the tree $d_f = \dfrac{1}{\zeta}$

Greedy algorithm*

Greedy algorithm*

A single-minded step-by-step strategy to find a (hopefully not too bad) approximation of the ground state

Greedy algorithm*

A single-minded step-by-step strategy to find a (hopefully not too bad) approximation of the ground state

Chose as first pairing the pairing with lowest energy

Greedy algorithm*

A single-minded step-by-step strategy to find a (hopefully not too bad) approximation of the ground state

Chose as first pairing the pairing with lowest energy

Chose as second pairing the planar pairing left with the lowest energy

Greedy algorithm*

A single-minded step-by-step strategy to find a (hopefully not too bad) approximation of the ground state

Chose as first pairing the pairing with lowest energy

Chose as second pairing the planar pairing left with the lowest energy

Iterate until we cannot add any pairing

Greedy algorithm*

A single-minded step-by-step strategy to find a (hopefully not too bad) approximation of the ground state

Chose as first pairing the pairing with lowest energy

Chose as second pairing the planar pairing left with the lowest energy

Iterate until we cannot add any pairing

Since the pair energies are independent, this amounts to build an arch system by successive random deposition of arches

Greedy algorithm*

A single-minded step-by-step strategy to find a (hopefully not too bad) approximation of the ground state

Chose as first pairing the pairing with lowest energy

Chose as second pairing the planar pairing left with the lowest energy

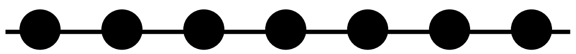
Iterate until we cannot add any pairing

Since the pair energies are independent, this amounts to build an arch system by successive random deposition of arches

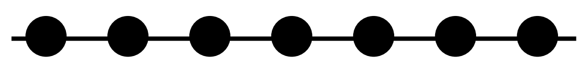
* already considered in Markus Müller's thesis (2002)

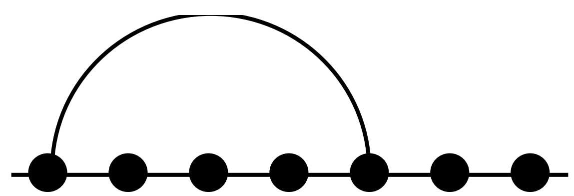
Random planar arch deposition

Random planar arch deposition

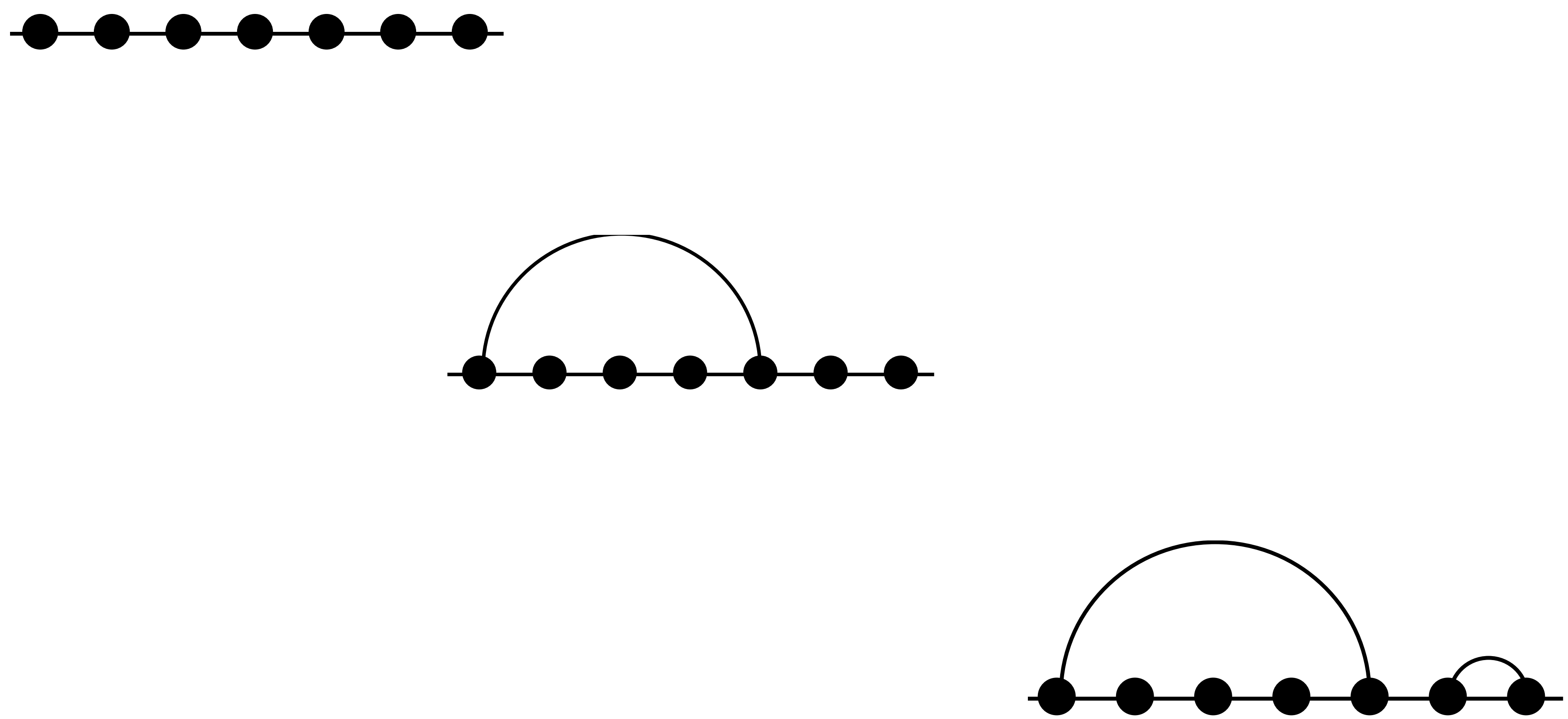


Random planar arch deposition

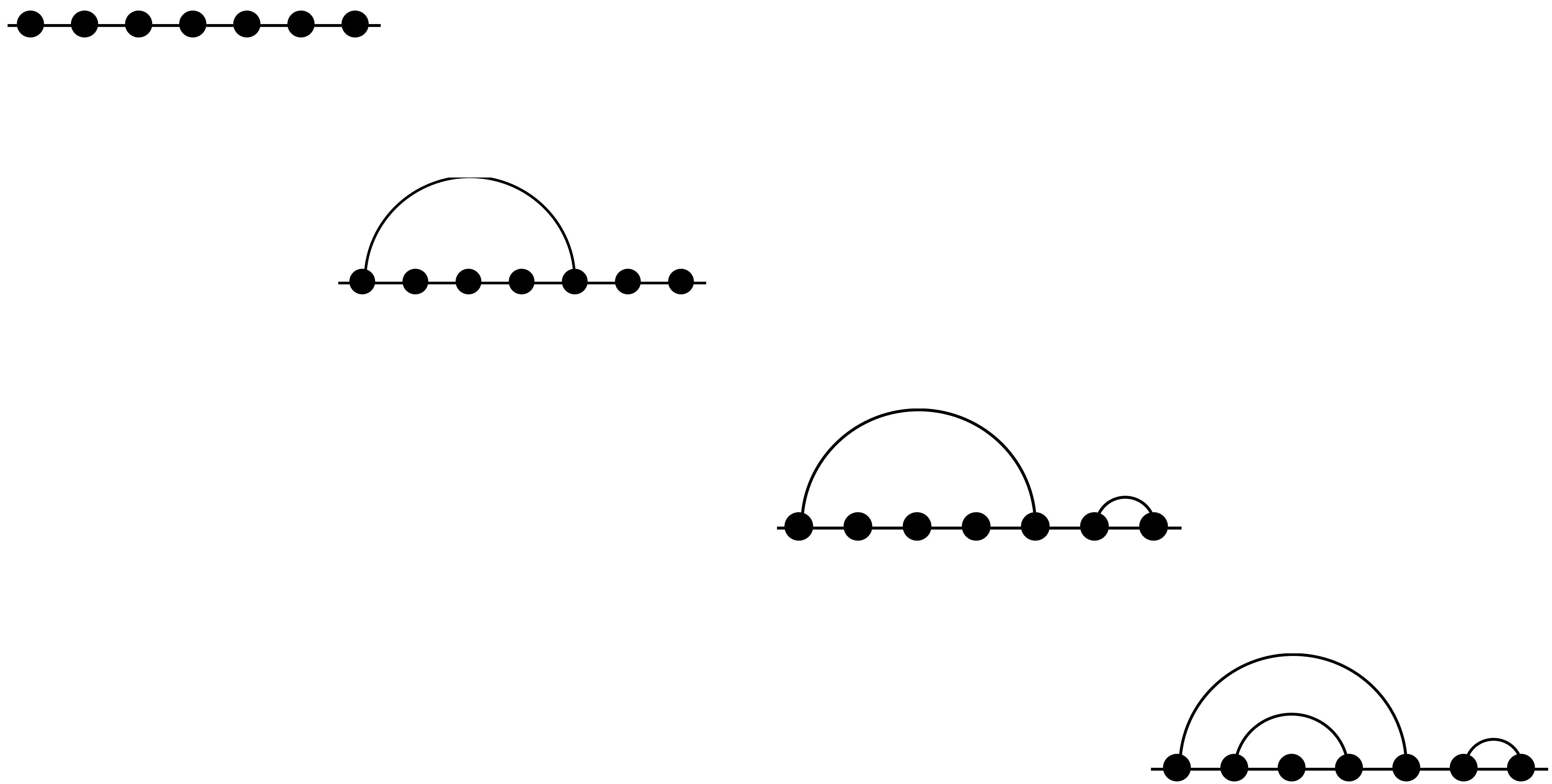




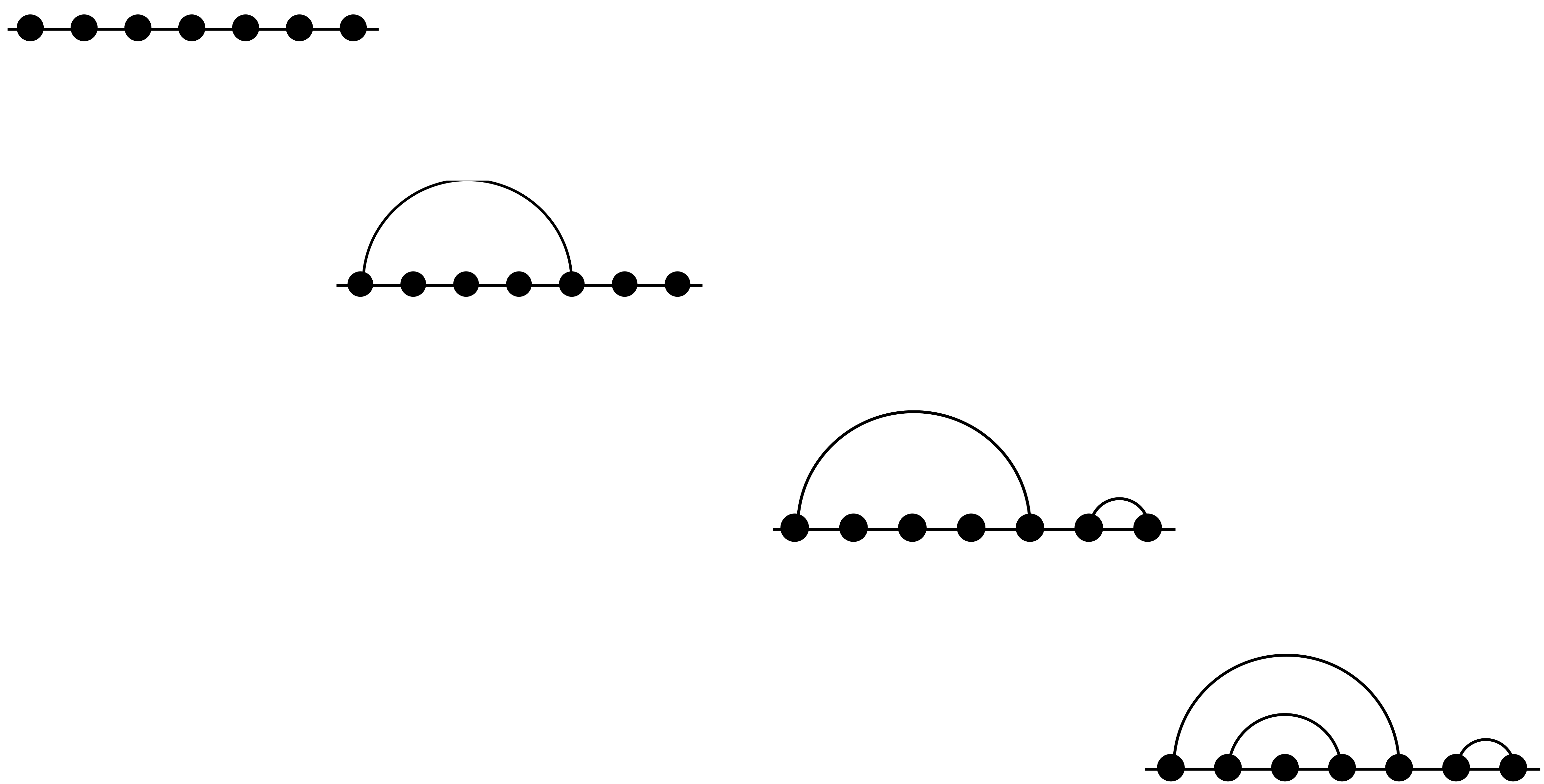
Random planar arch deposition



Random planar arch deposition



Random planar arch deposition



Relation to fragmentation models?

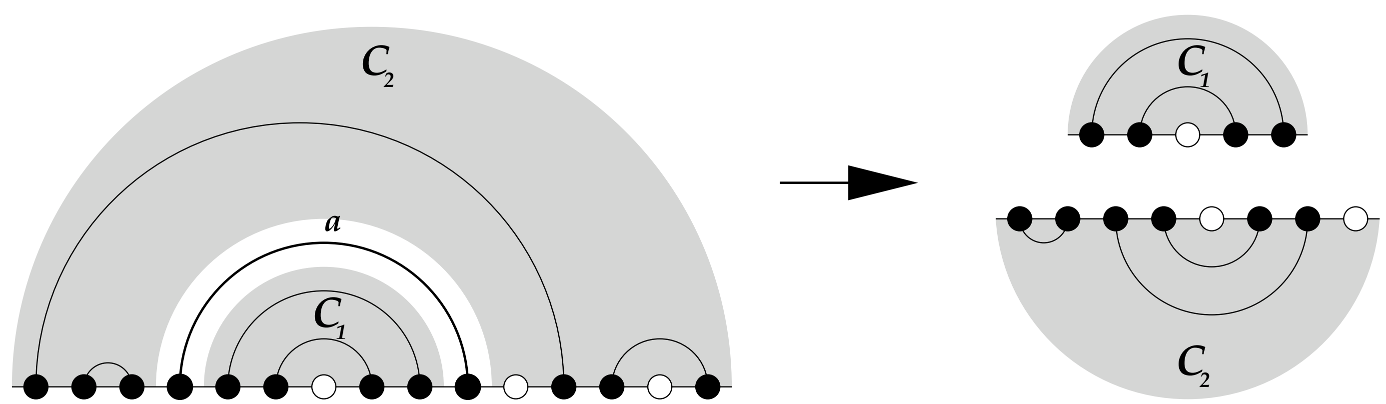
Recursion relations

Recursion relations

for configuration probabilities of arch structures

Recursion relations

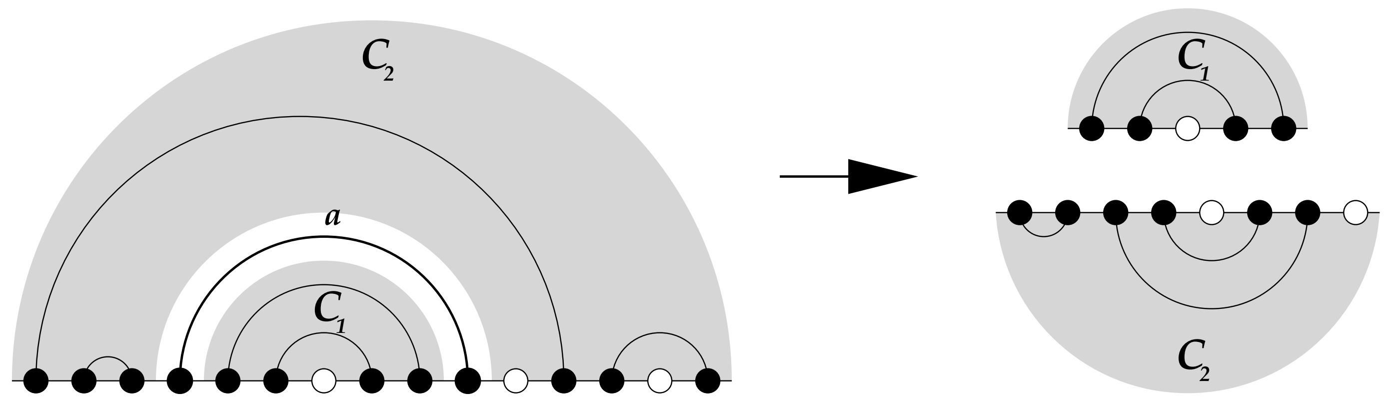
for configuration probabilities of arch structures



$$P(\mathcal{C}) = \sum_{\substack{\text{first arch}\\ a\in\mathcal{C}}} \sum_{\substack{\text{substructures}\\ \mathcal{C}_1\,\&\,\mathcal{C}_2}} \frac{2}{n(n-1)}\, P(\mathcal{C}_1)\, P(\mathcal{C}_2)$$

Recursion relations

for configuration probabilities of arch structures

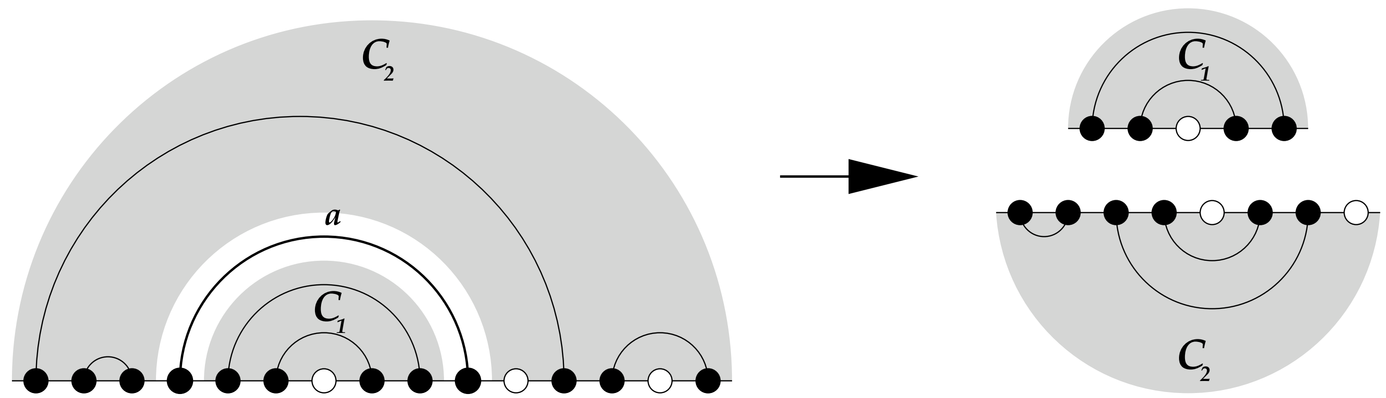


$$P(\mathcal{C}) = \sum_{\substack{\text{first arch} \\ a \in \mathcal{C}}} \sum_{\substack{\text{substructures} \\ \mathcal{C}_1 \,\&\, \mathcal{C}_2}} \frac{2}{n(n-1)} P(\mathcal{C}_1)\, P(\mathcal{C}_2)$$

for heights = number of arches above a point

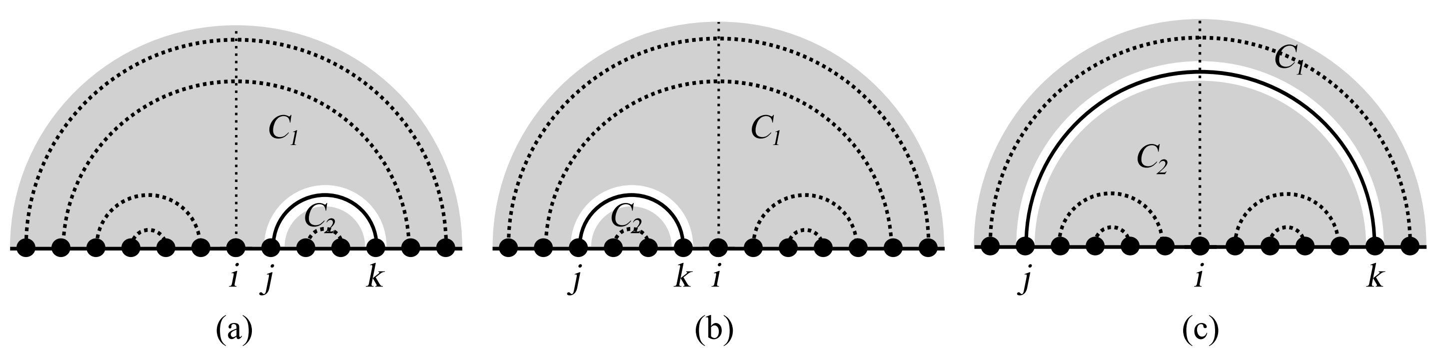
Recursion relations

for configuration probabilities of arch structures



$$P(\mathcal{C}) = \sum_{\substack{\text{first arch} \\ a\in\mathcal{C}}} \sum_{\substack{\text{substructures} \\ \mathcal{C}_1\,\&\,\mathcal{C}_2}} \frac{2}{n(n-1)} P(\mathcal{C}_1)\, P(\mathcal{C}_2)$$

for heights = number of arches above a point



$$\begin{aligned}\frac{n(n-1)}{2} h(i,n) &= \sum_{i<j<k<n} h(i, n-k+j-1) + \sum_{0<j<k\leq i} h(i-k+j-1, n-k+j-1) \\ &+ \sum_{0<j\leq i<k<n} [h(j-1, n-k+j-1) + h(i-j, k-j-1) + 1]\end{aligned}$$

PDE for generating functions

$$G(u,v;z) = \sum_{n=0}^{\infty} \sum_{i=0}^{n} \left\langle e^{z\, h_{\mathcal{C}}(i,n)} \right\rangle u^i v^{n-i}$$

$$\left[\frac{1}{2} \left(u^2 \frac{\partial^2}{\partial u^2} + v^2 \frac{\partial^2}{\partial v^2} \right) + uv \frac{\partial^2}{\partial u \partial v} \right] G(u,v;z)$$

$$= \left[\frac{u^2}{(1-u)} \left(u \frac{\partial}{\partial u} + 1 \right) + \frac{v^2}{(1-v)} \left(v \frac{\partial}{\partial v} + 1 \right) \right] G(u,v;z) + uv\, e^z\, G(u,v;z)^2$$

PDE for generating functions

$$G(u,v;z) = \sum_{n=0}^{\infty} \sum_{i=0}^{n} \left\langle e^{z\, h_{\mathcal{C}}(i,n)} \right\rangle u^i v^{n-i}$$

$$\left[\frac{1}{2} \left(u^2 \frac{\partial^2}{\partial u^2} + v^2 \frac{\partial^2}{\partial v^2} \right) + uv \frac{\partial^2}{\partial u \partial v} \right] G(u,v;z)$$

$$= \left[\frac{u^2}{(1-u)} \left(u \frac{\partial}{\partial u} + 1 \right) + \frac{v^2}{(1-v)} \left(v \frac{\partial}{\partial v} + 1 \right) \right] G(u,v;z) + uv\, e^z\, G(u,v;z)^2$$

for the average mean height

$$K(t) = \left. \frac{\partial}{\partial z} G(t,t,z) \right|_{z=0} = \sum_n t^n \sum_{i=0}^{n} \langle h(i,n) \rangle$$

$$(1-t)^2 K''(t) - 2t(1-t)K'(t) - 4(2-t)K(t) = \frac{2}{(1-t)^2}$$

PDE for generating functions

$$G(u,v;z)=\sum_{n=0}^{\infty}\sum_{i=0}^{n}\left\langle e^{z\,h_{\mathcal{C}}(i,n)}\right\rangle u^i v^{n-i}$$

$$\left[\frac{1}{2}\left(u^2\frac{\partial^2}{\partial u^2}+v^2\frac{\partial^2}{\partial v^2}\right)+uv\frac{\partial^2}{\partial u\partial v}\right]G(u,v;z)$$
$$=\left[\frac{u^2}{(1-u)}\left(u\frac{\partial}{\partial u}+1\right)+\frac{v^2}{(1-v)}\left(v\frac{\partial}{\partial v}+1\right)\right]G(u,v;z)+uv\,e^z\,G(u,v;z)^2$$

for the average mean height

$$K(t)=\left.\frac{\partial}{\partial z}G(t,t,z)\right|_{z=0}=\sum_n t^n\sum_{i=0}^{n}\langle h(i,n)\rangle$$

$$(1-t)^2K''(t)-2t(1-t)K'(t)-4(2-t)K(t)=\frac{2}{(1-t)^2}$$

Scaling

$$K(t)\underset{t\to 1}{\propto}(1-t)^{-2-\zeta}\quad\Longrightarrow\quad\zeta^2+3\zeta=2\quad\Longrightarrow\quad\zeta=\frac{\sqrt{17}-3}{2}=0.561\ldots$$

NB: value for ζ obtained in M. Müller's PhD thesis (2002), assuming scaling

Scaling function for average height

$$\langle h(i,n)\rangle \underset{n\to\infty}{=} n^{\zeta}\,\mathcal{H}(x)\ ,\qquad x = i/n$$

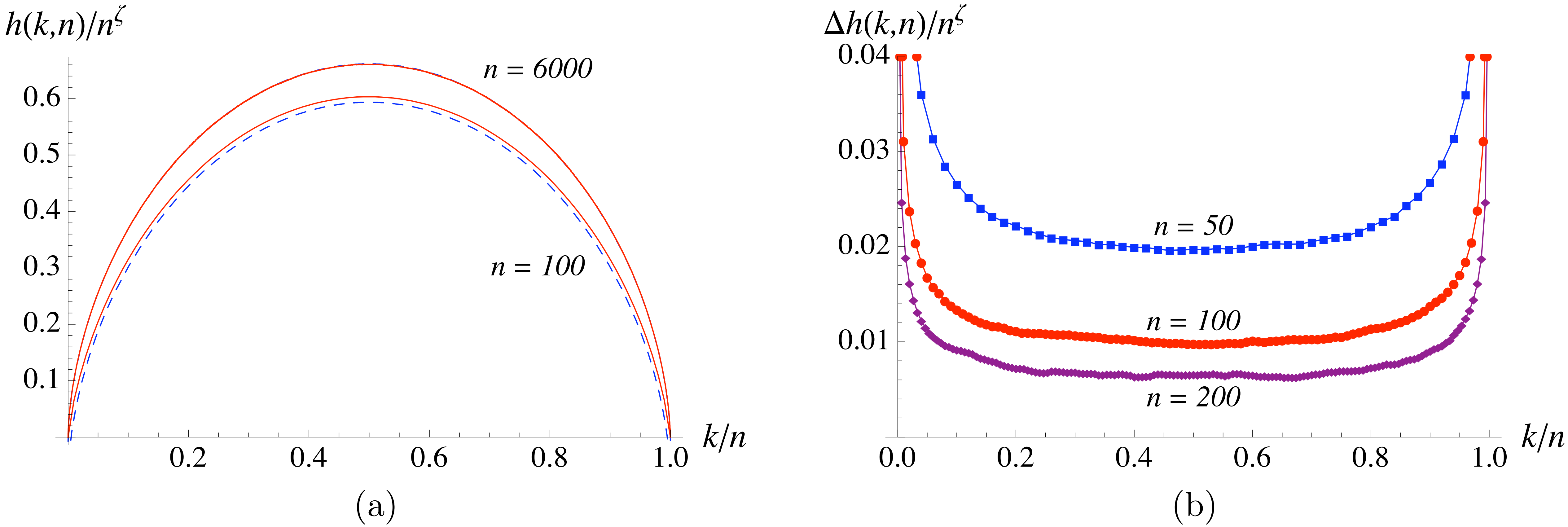
$$\mathcal{H}(x) = \mathbf{E}\,x^{\zeta}(1-x)^{\zeta}\ ,\qquad \zeta = \frac{\sqrt{17}-3}{2}\ ,\qquad \mathbf{E} = 0.888309\ldots$$

Scaling function for average height

$$\langle h(i,n)\rangle \underset{n\to\infty}{=} n^{\zeta}\,\mathcal{H}(x)\ , \quad x = i/n$$

$$\mathcal{H}(x) = \mathbf{E}\,x^{\zeta}(1-x)^{\zeta}\ , \quad \zeta = \frac{\sqrt{17}-3}{2}\ , \quad \mathbf{E} = 0.888309\ldots$$

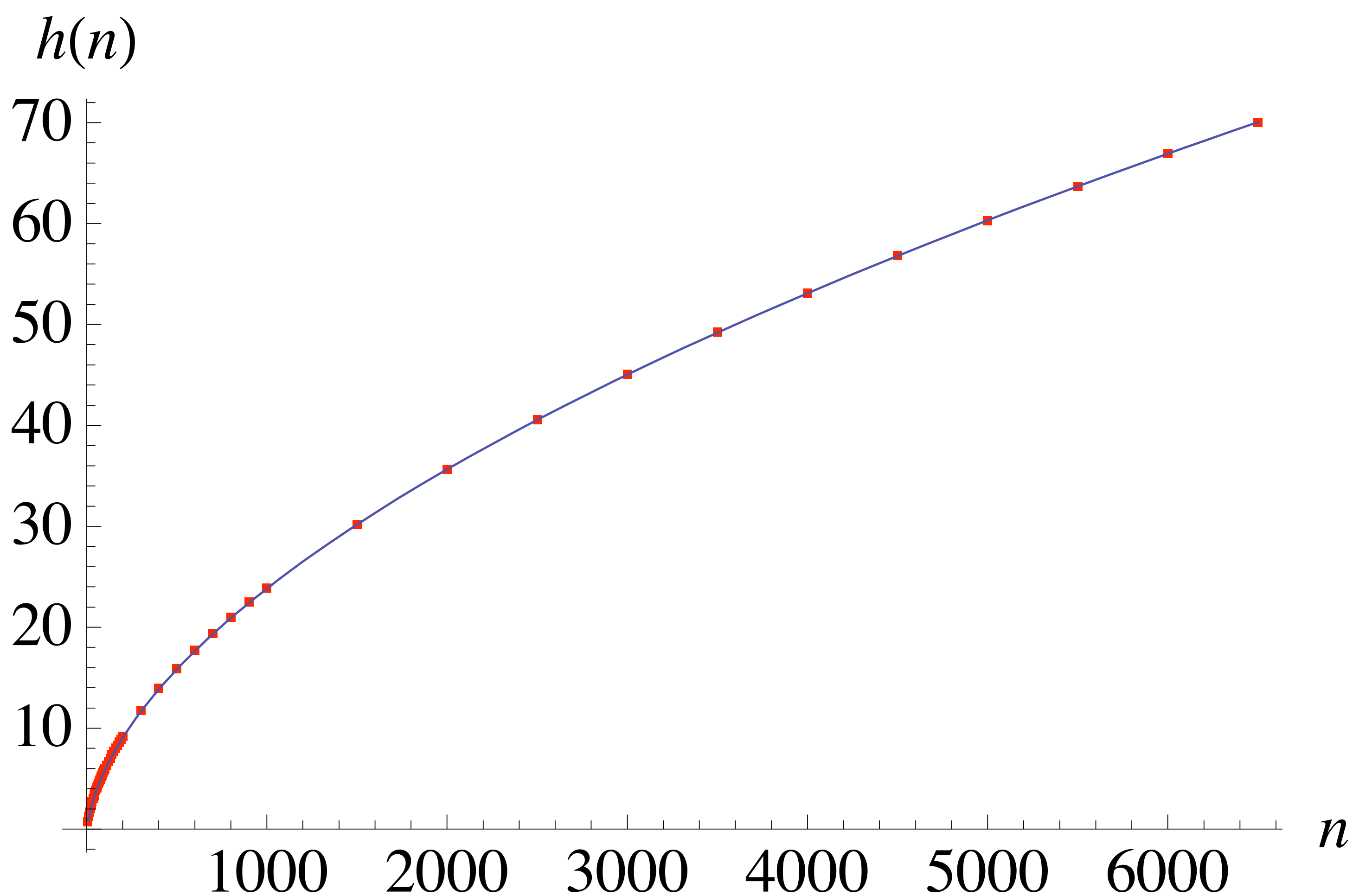
Numerical checks



Hypergeometric equations for many observables
example: mean height

$$K(v) = \frac{\mathbf{C}_+}{(1-v)^{\zeta+2}}\, M(-\zeta, -2-2\zeta; 2-2v) - \frac{1}{(1-v)^2} + \frac{\mathbf{C}_-}{(1-v)^{-1-\zeta}} M(\zeta+3, 2\zeta+4; 2-2v)$$

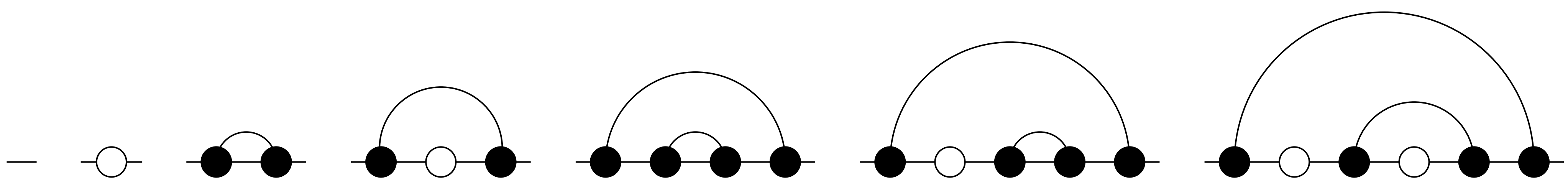
$\mathbf{C}_+ = 0.713263\ldots$ $\quad \mathbf{C}_- = 0.519299\ldots$ $\quad M$ confluent hypergeometric function (Kummer)



Allows to estimate finite size effects

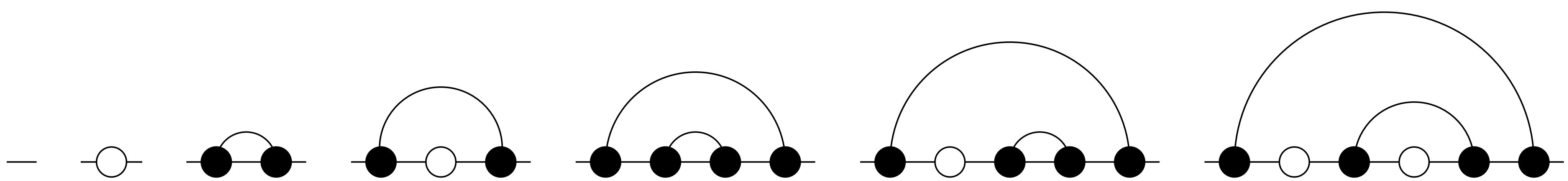
This model is equivalent to a tree growth model !

Random deposition point model: successive deposition of points on intervals & create an arch whenever it is possible

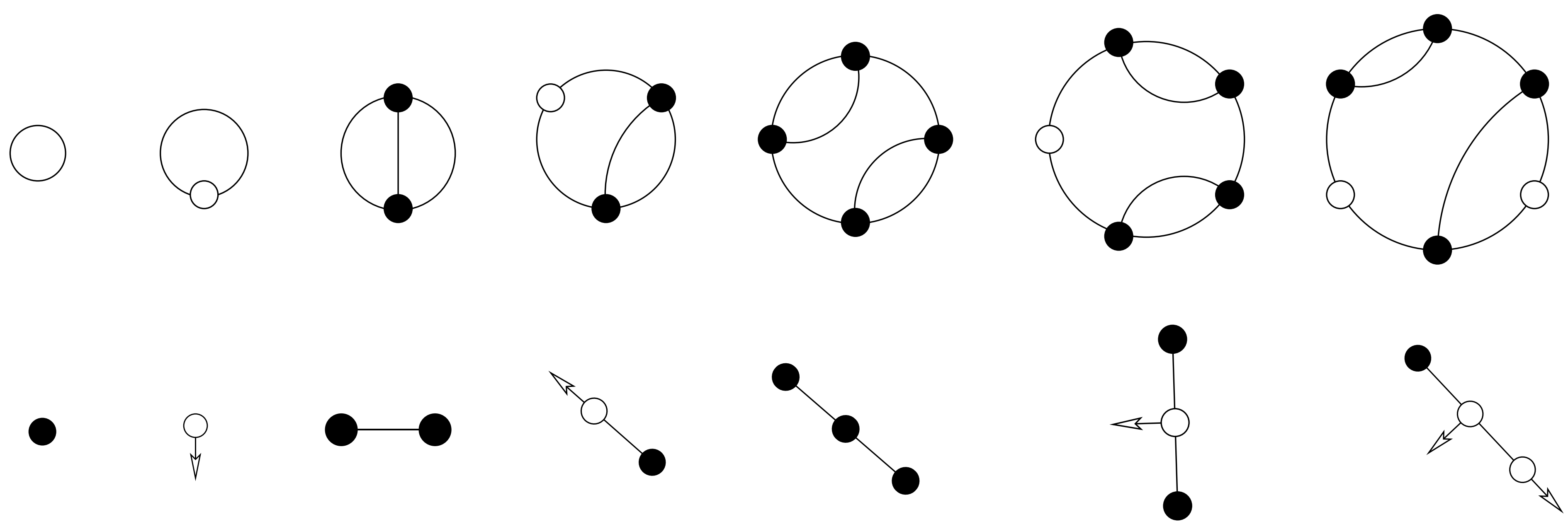


This model is equivalent to a tree growth model !

Random deposition point model: successive deposition of points on intervals & create an arch whenever it is possible

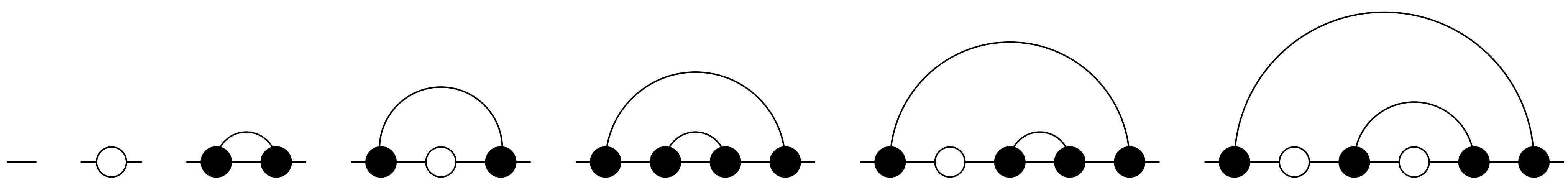


Closing + duality gives a decorated tree growth model

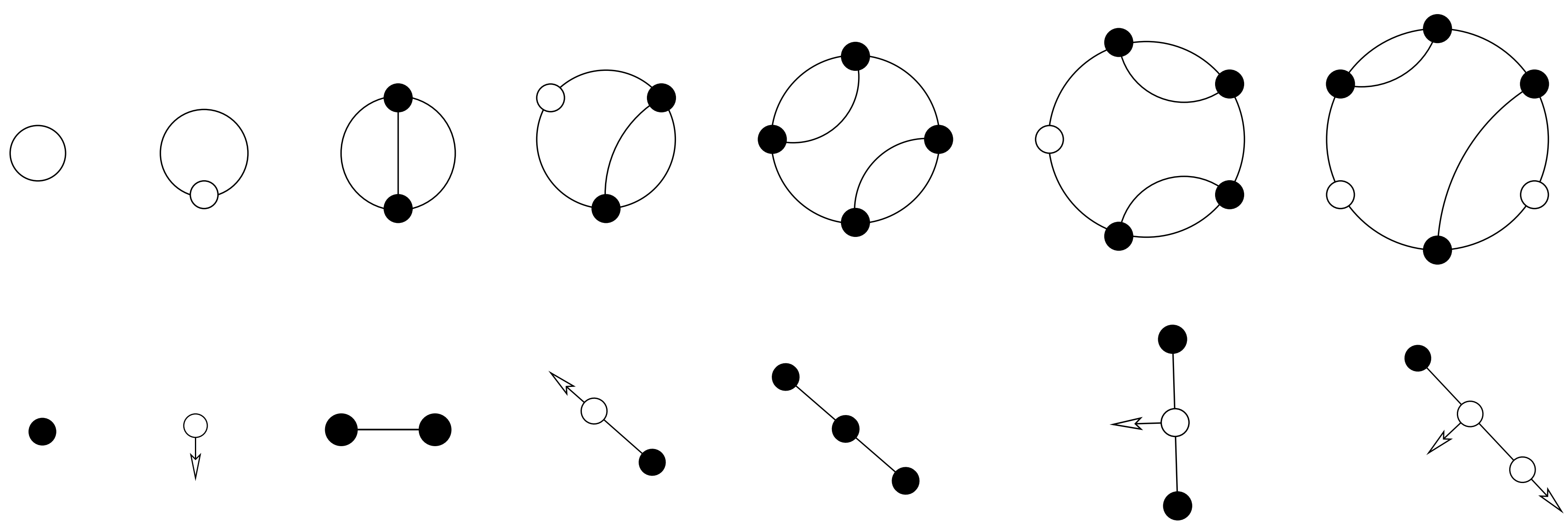


This model is equivalent to a tree growth model !

Random deposition point model: successive deposition of points on intervals & create an arch whenever it is possible



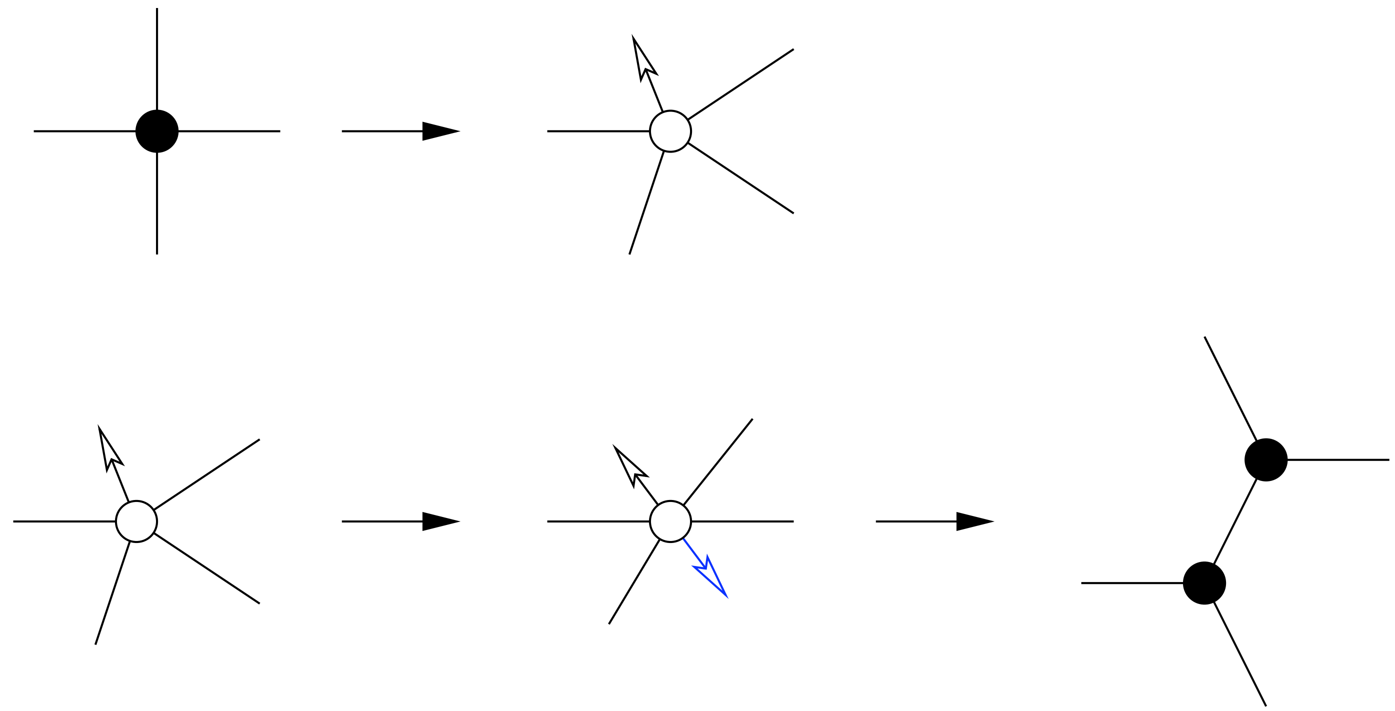
Closing + duality gives a decorated tree growth model



arch/tree growth at time t = arch deposition for size $n=t$!

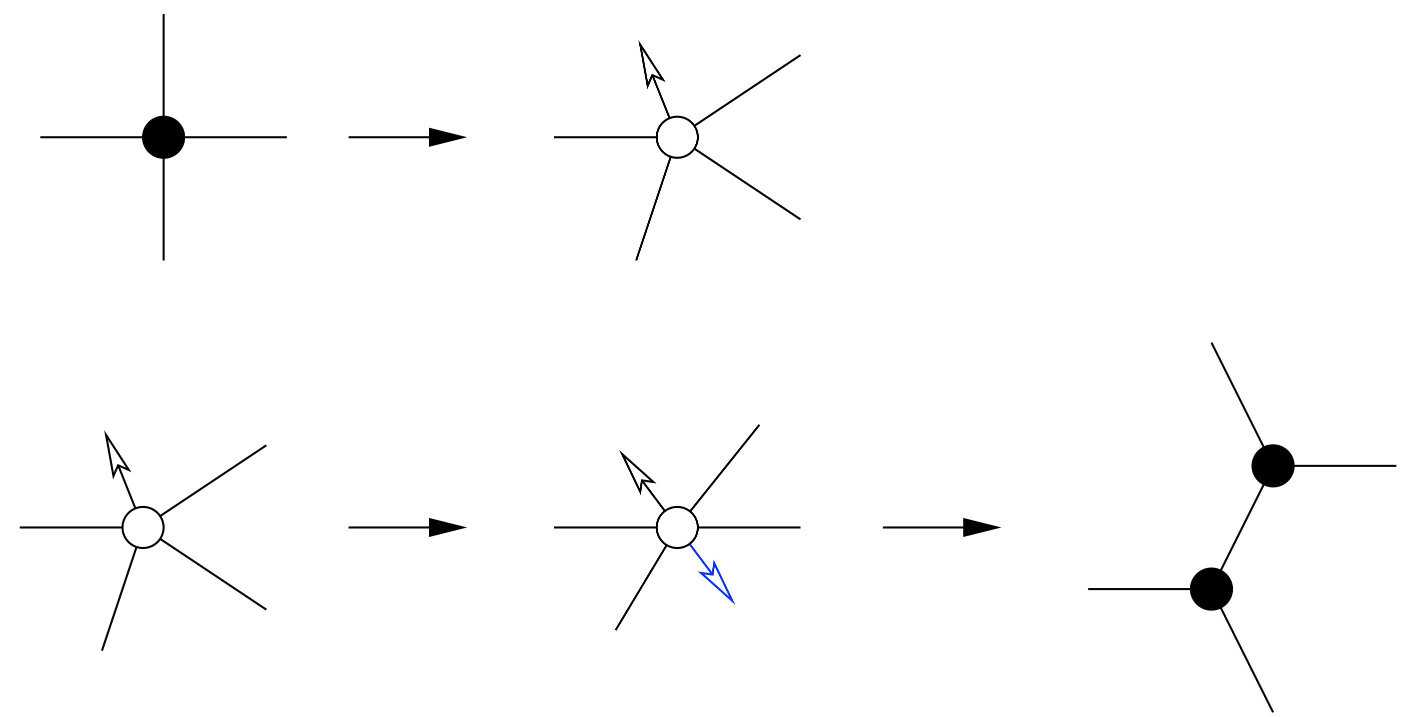
2 steps growth process:

- add a leaf to a vertex
- add a second leaf = “budding” i.e. 1 vertex $\longrightarrow$ 2 vertices



2 steps growth process:

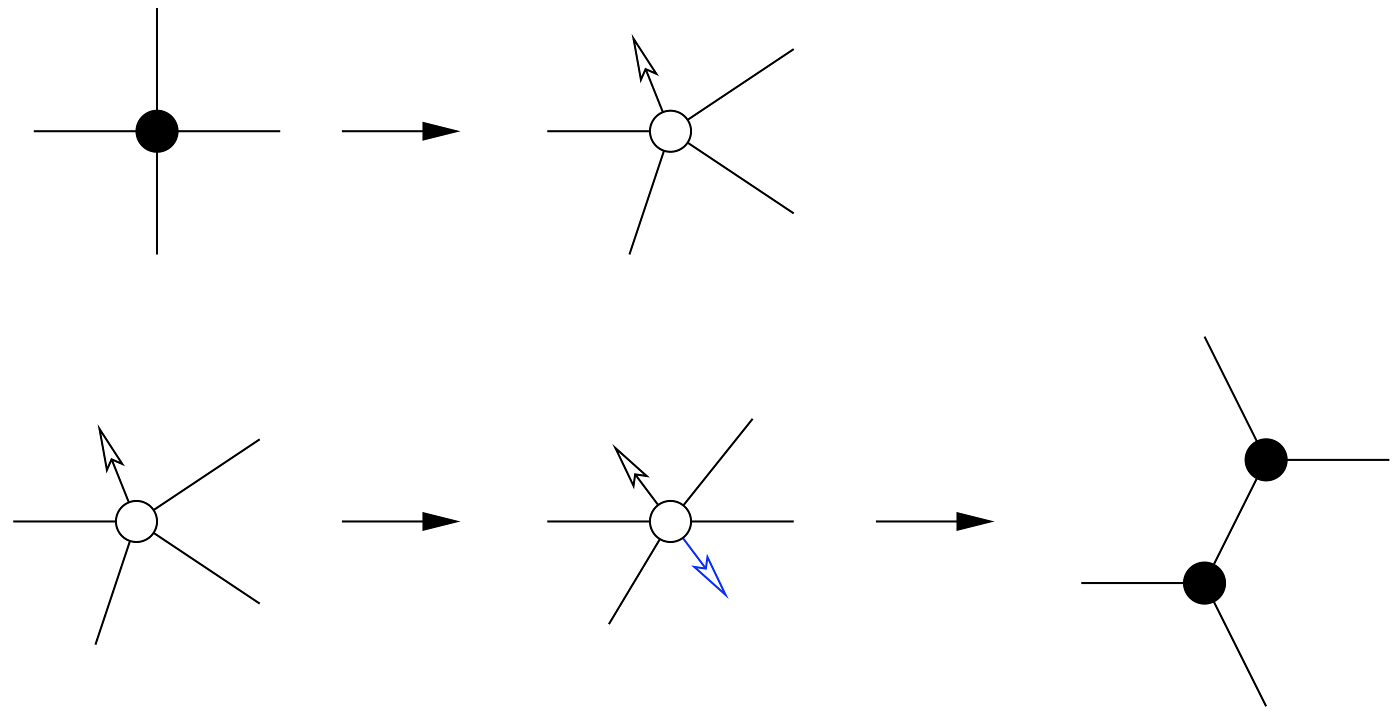
- add a leaf to a vertex
- add a second leaf = “budding” i.e. 1 vertex $\longrightarrow$ 2 vertices



This is different from usual tree and network growth processes

2 steps growth process:

- add a leaf to a vertex
- add a second leaf = "budding" i.e. 1 vertex $\longrightarrow$ 2 vertices

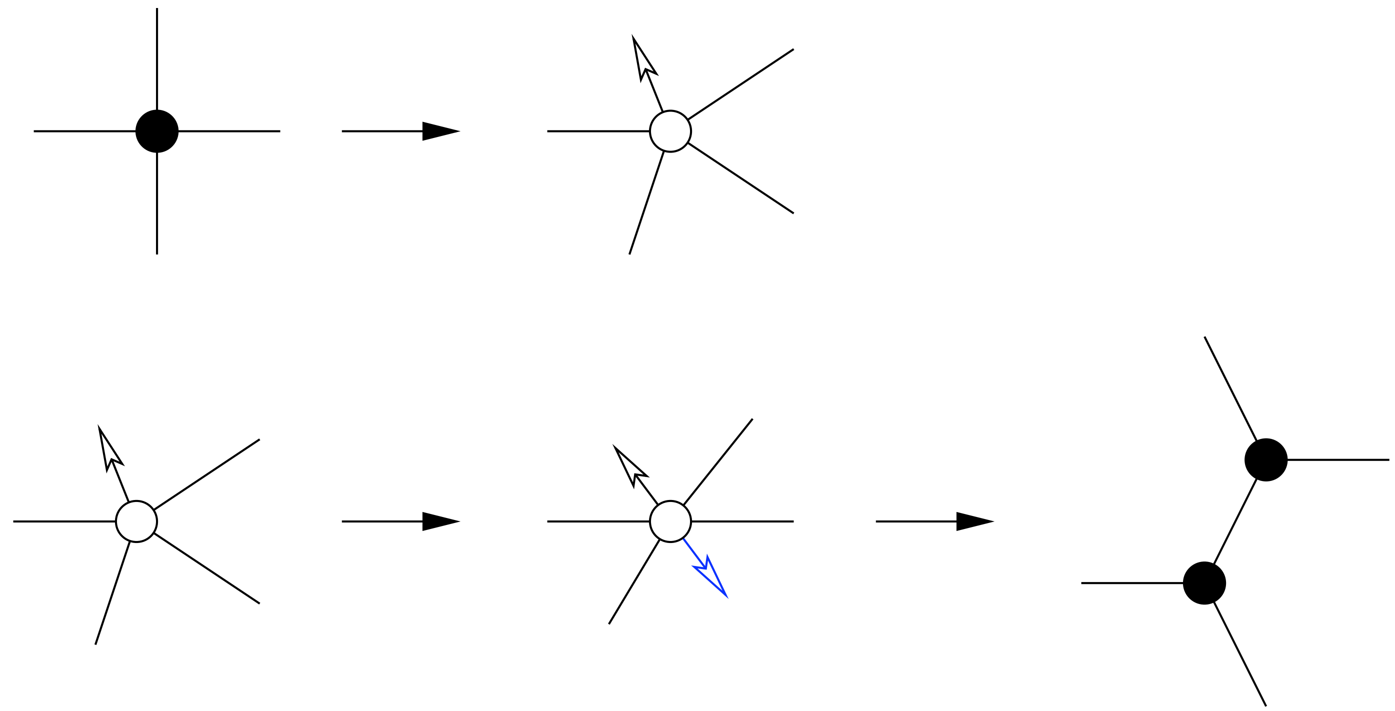


This is different from usual tree and network growth processes

This can be used to compute e.g. asymptotic proportion of vertices

2 steps growth process:

- add a leaf to a vertex
- add a second leaf = "budding" i.e. 1 vertex $\longrightarrow$ 2 vertices



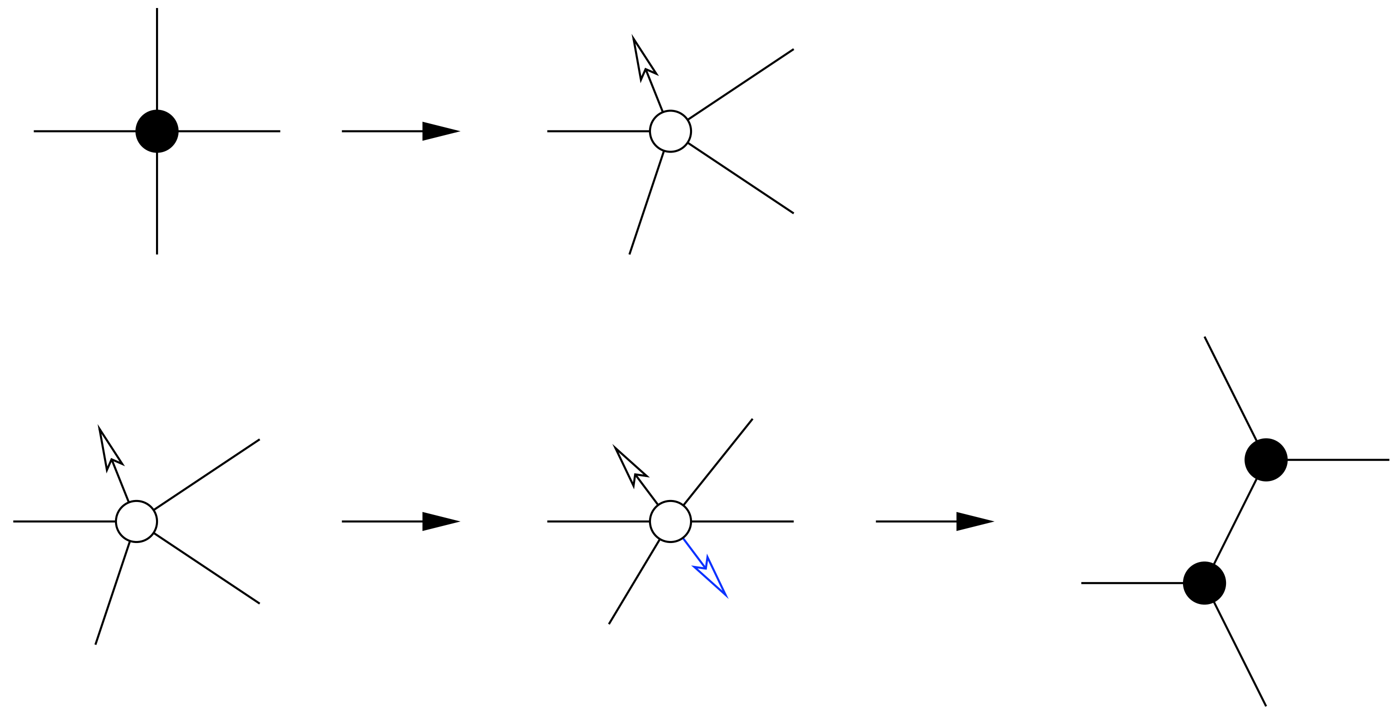
This is different from usual tree and network growth processes

This can be used to compute e.g. asymptotic proportion of vertices

$$\beta_k = \frac{1}{e^2}\frac{2^k}{(k+1)!} \quad , \qquad \omega_k = \frac{1}{e^2}\frac{k\ 2^k}{(k+2)!}$$

2 steps growth process:

- add a leaf to a vertex
- add a second leaf = "budding" i.e. 1 vertex $\longrightarrow$ 2 vertices



This is different from usual tree and network growth processes

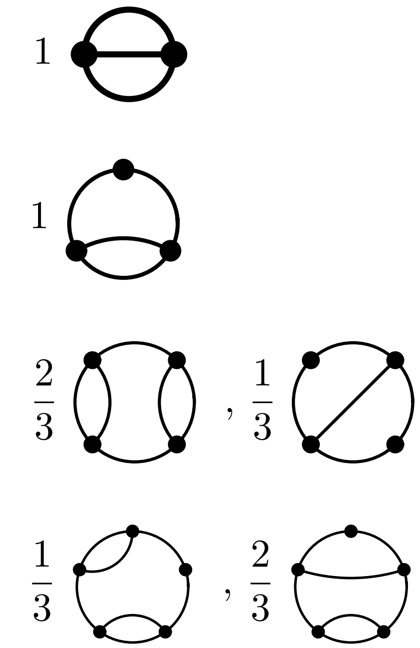
This can be used to compute e.g. asymptotic proportion of vertices

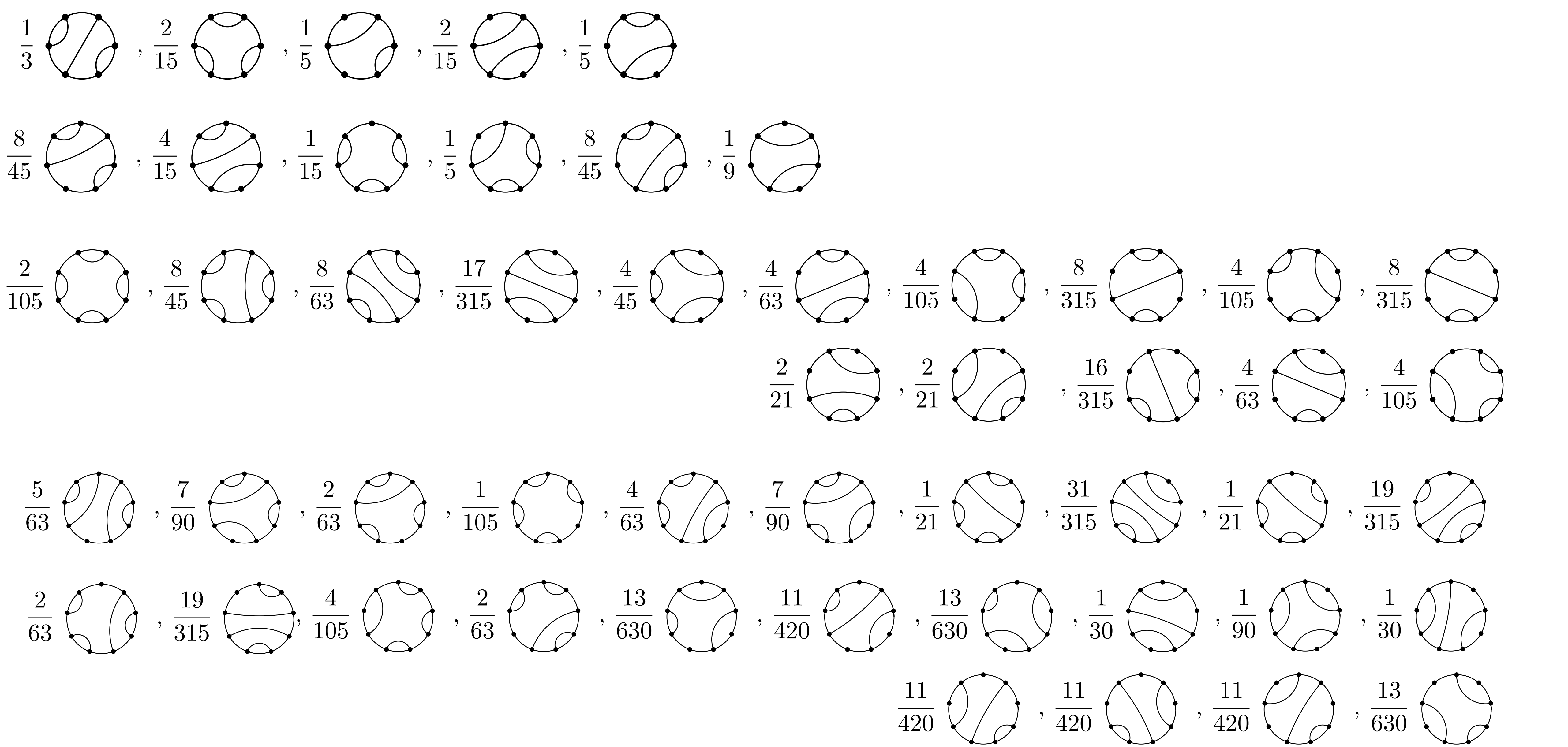
$$\beta_k = \frac{1}{e^2}\frac{2^k}{(k+1)!} \quad , \qquad \omega_k = \frac{1}{e^2}\frac{k\ 2^k}{(k+2)!}$$

and for what more ? ....

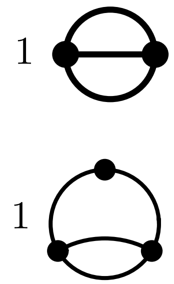

1
1
2/3 , 1/3
1/3 , 2/3
1/3 , 2/15 , 1/5 , 2/15 , 1/5
8/45 , 4/15 , 1/15 , 1/5 , 8/45 , 1/9
2/105 , 8/45 , 8/63 , 17/315 , 4/45 , 4/63 , 4/105 , 8/315 , 4/105 , 8/315
2/21 , 2/21 , 16/315 , 4/63 , 4/105
5/63 , 7/90 , 2/63 , 1/105 , 4/63 , 7/90 , 1/21 , 31/315 , 1/21 , 19/315
2/63 , 19/315 , 4/105 , 2/63 , 13/630 , 11/420 , 13/630 , 1/30 , 1/90 , 1/30
11/420 , 11/420 , 11/420 , 13/630

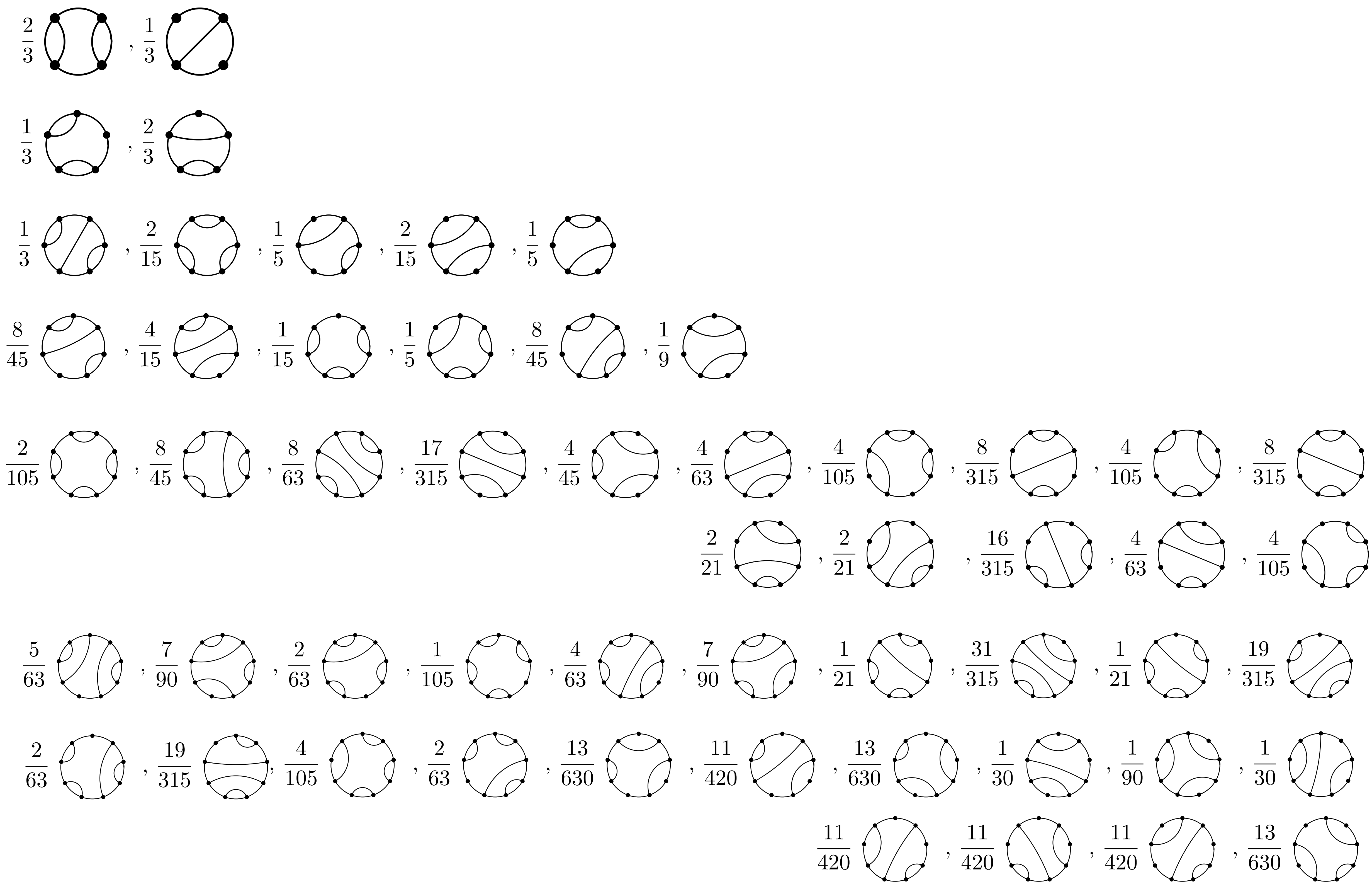
Interesting arch/tree statistics from growth process

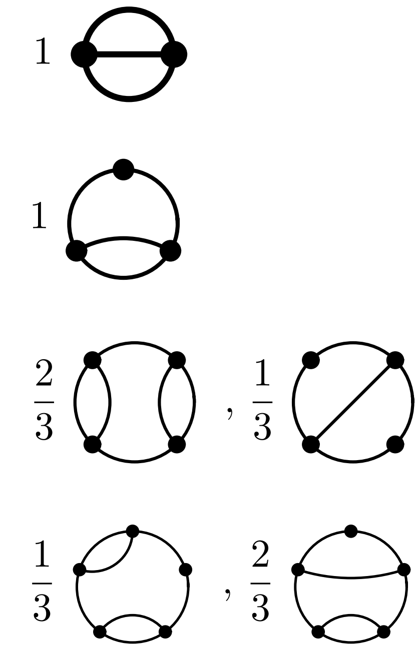




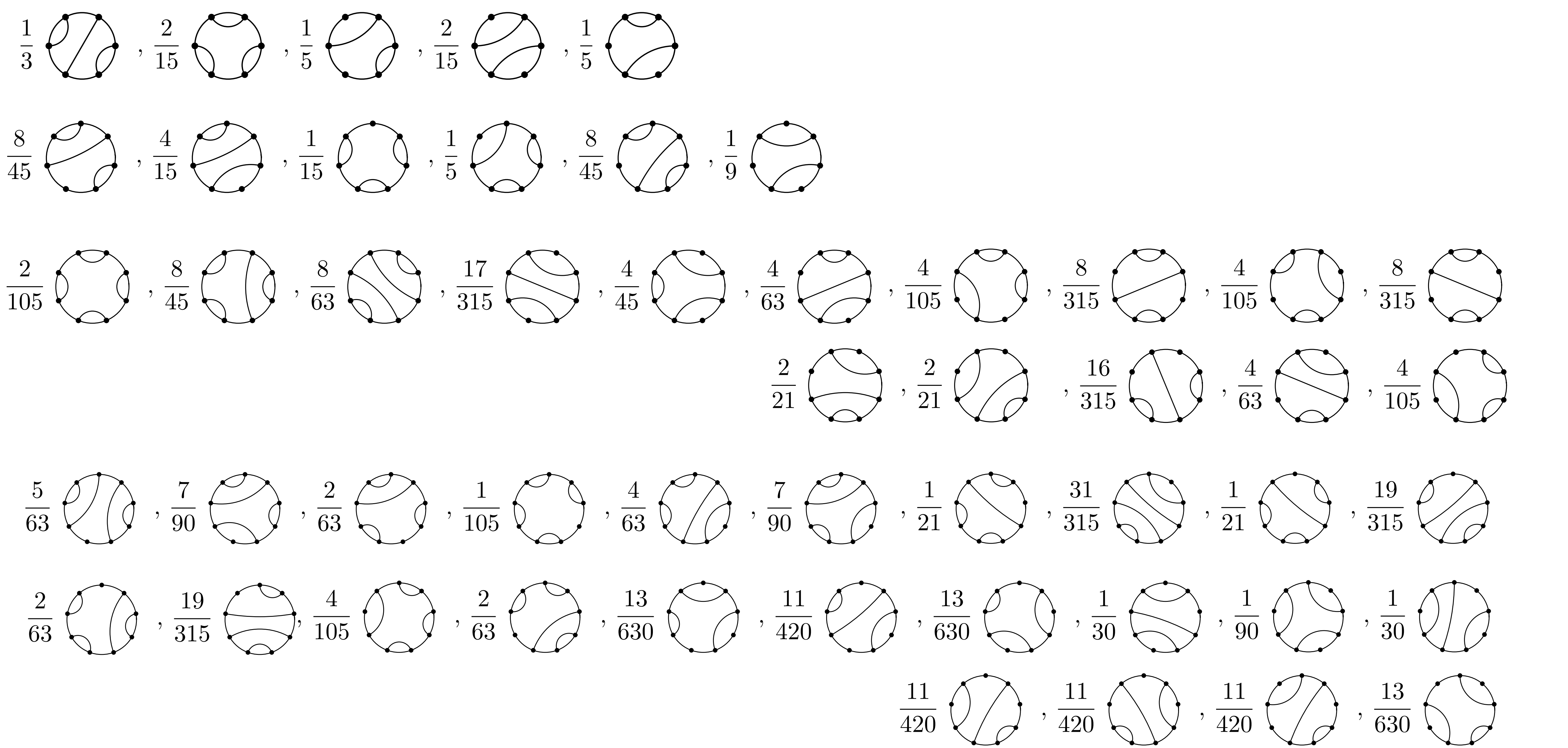
Interesting arch/tree statistics from growth process quite different from generic or causal trees

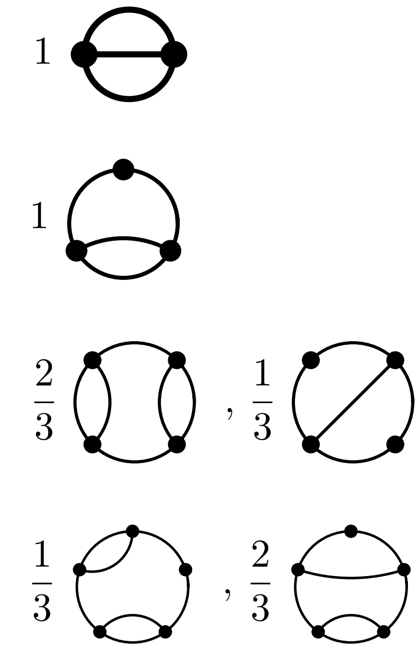




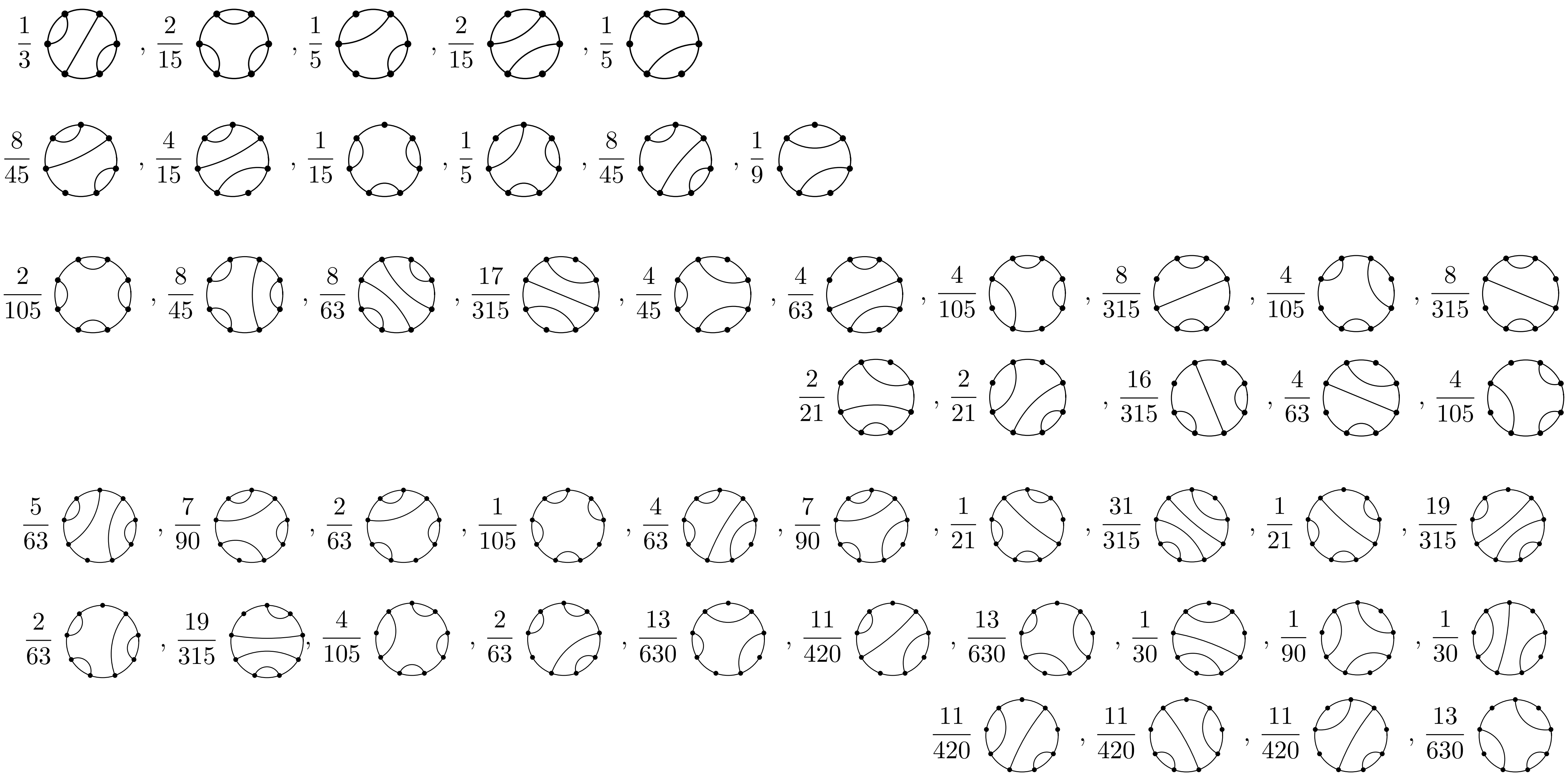


Interesting arch/tree statistics from growth process quite different from generic or causal trees and from growth by attachement trees





Interesting arch/tree statistics from growth process quite different from generic or causal trees and from growth by attachement trees
Is it new? is it universal?



Interesting arch/tree statistics from growth process
quite different from generic or causal trees
and from growth by attachement trees
Is it new? is it universal?
good for something?