# ARN et arbres: problèmes de croissance et de repliement


François David
IPhT CEA-Saclay & CNRS

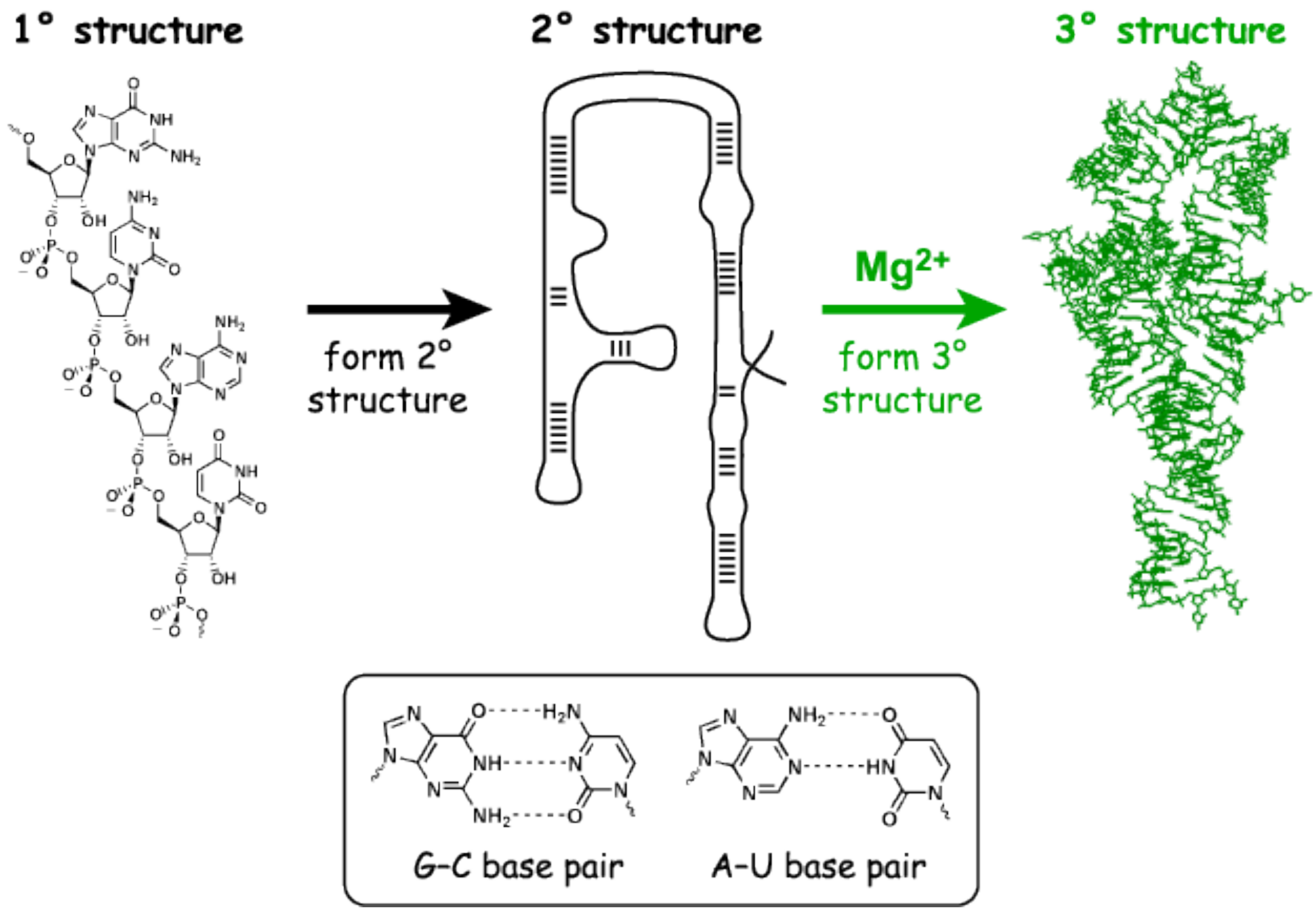
*Emmanuel Guitter, Philippe Di Francesco*
*IPhT*
Mark Dukes, Þorður Jónsson, Sigurður Örn Stefánsson
University of Iceland
Christian Hagendorf, Kay Jörg Wiese
LPTENS

# Why trees?

- Population dynamics & genealogy
- Biology, Evolution
- Growth processes
- Branched polymers
- Statistical mechanics
- Spanning trees, random matrices
- Quantum gravity (2d & 3d)
- Combinatorics
- Computer science, data bases, XML, etc.
- Optimisation problems
- Social sciences, economics
- etc.

# First problem: RNA folding & random splitting trees

## RNA primary, secondary and tertiary structure




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

# Problem: find the secondary (and the tertiary) structure from the base sequence

GCCUUAAUGCACAUGGGCAAGCCCACGUAGCUAGUCGCGACACCAGUCCCAAAUAAUGUUCACCCAACUCGCCUGACCGUCCCGCAGUA
GCUAUACUACCGACUCCUACGCGGUUGAAACUAGACUUUUCUAGCGAGCUGUCAUAGGUAUGGUGCACUGUCUUUAAUUUUGUAUUGGGCC
AGGCACGAAAGGCUUGGAAGUAAGGCCCCGCUUGACCCGAGAGGUGACAAUAGCGGCCAGGUGUAACGAUACGCGGGUGGCACGUACCCCA
AACAAUUAAUCACACUGCCCGGGCUCACAUUAAUCAUGCCAUUCGUUGCCGAUCCGACCCAUAUAGGAUGUGUAUGCCUCAUUCCCGGUCG
GGGCGGCGACUGUUAACGCAUGAGAACUGAUUAGAUCUCGUGGUAGUGCUUGUCAAAUAGAAUGAGGCCAUUCCACAGACAUAGCGUUUCC
CAUGAGCUAGGGGUCCCAUGUCCAGGUCCCCUAAAUAAAAGAGUC

1. “kissing hairpins” & interlaced strands are rare (unfavored by kinematics & topology)

2. RNA can (to some extent) be considered as a **planar tree**

# Random bond RNA model (R. Bundshuh and T. Hwa)

Question: Which features are sequence dependent (biological functions) and which are generic?

Long chain with random sequence of bases

Allows numerical and analytical studies (in particular at IPhT...)

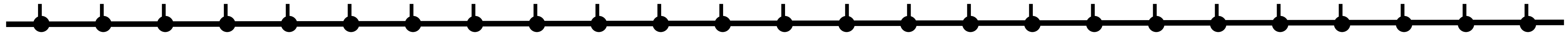
Exhibit a freezing transition at a finite temperature $T_g$

High temperature molten phase (many equivalent microstates - generic tree with fractal dimension 2)

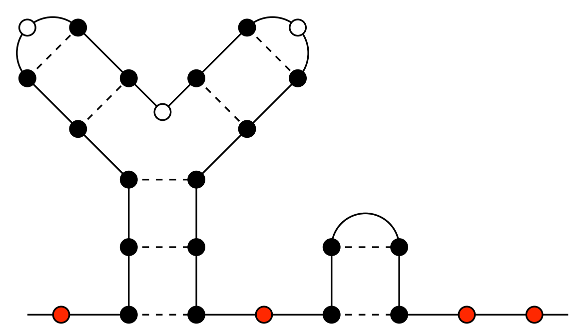
Low temperature frozen state (the system is frozen in one of the minimal energy configurations, like in a glass or a spin-glass)

# The (planar) random bond RNA folding model

1) Start from a strand made of L random bases $b_1 \; b_2 \; b_3 \; \cdots \; b_L$



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

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

... instead of a function of the random bases $e(b_i, b_j)$

5) And find the lowest energy configuration (ground state at zero temperature)

......This is a difficult problem: infinite range forces & topological frustration

# Greedy algorithm

# Greedy algorithm

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

# Greedy algorithm

A simple 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 simple 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 simple 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

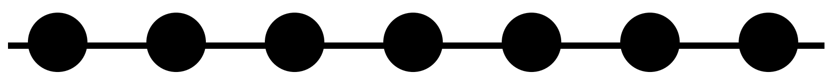
A simple step-by-step strategy to find a (hopefully not too bad) approximation of the ground state

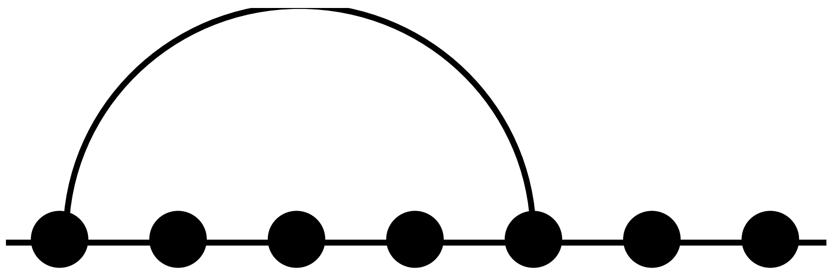
Chose as first pairing the pairing with lowest energy

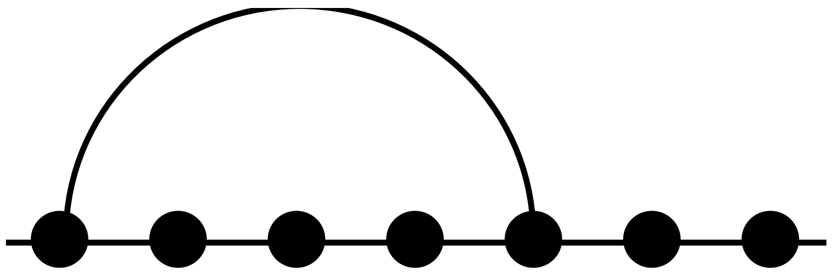
Chose as second pairing the planar pairing left with the lowest energy

Iterate until we cannot add any pairing

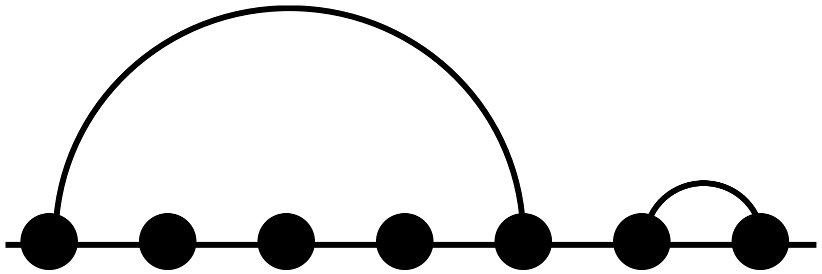
Since the pair energies are independent, this amounts to build an arch system by successive random deposition of arches



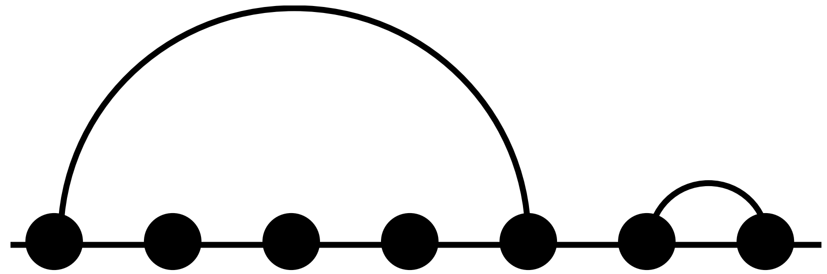




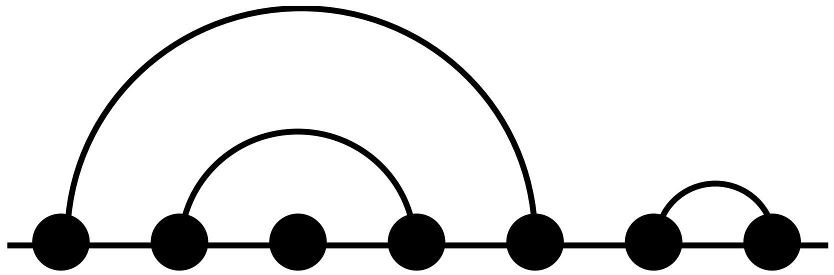
$$\frac{1}{21}$$



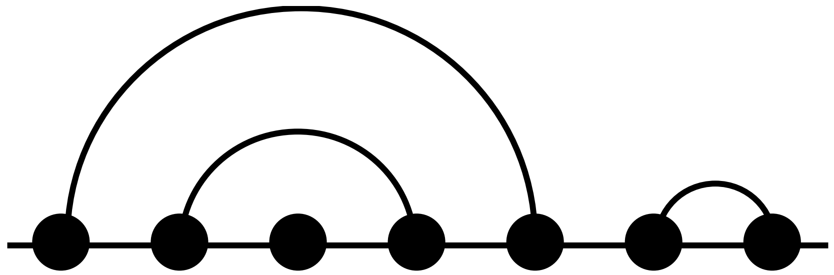
$$\frac{1}{21}$$



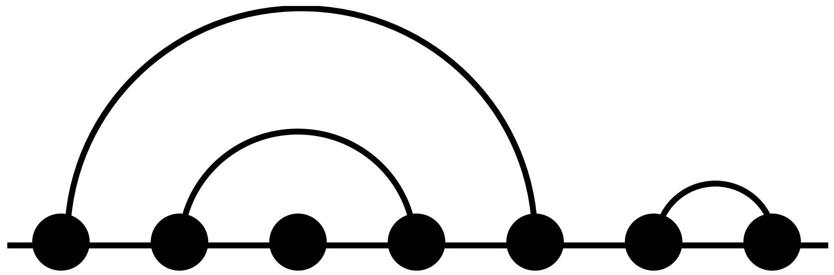
$$\frac{1}{21} \times \frac{1}{4}$$



$$\frac{1}{21} \times \frac{1}{4}$$



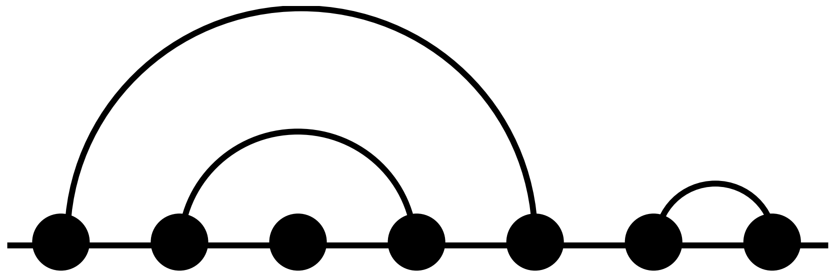
$$\frac{1}{21} \times \frac{1}{4} \times \frac{1}{3}$$



$$\frac{1}{21} \times \frac{1}{4} \times \frac{1}{3}$$

At each step, choose a new arch with uniform probability among all possible arches.
A final configuration is constructed via several deposition processes

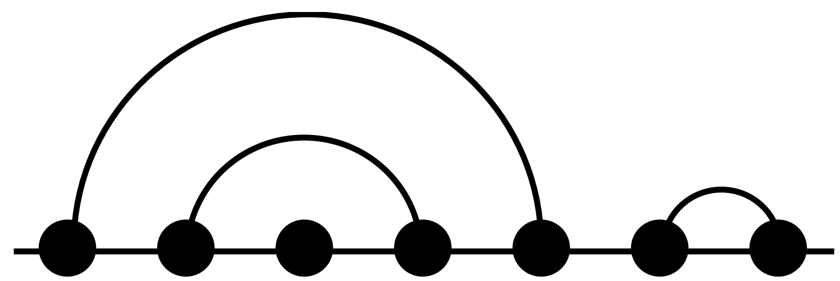
Total probability



$$\frac{1}{21} \times \frac{1}{4} \times \frac{1}{3}$$

At each step, choose a new arch with uniform probability among all possible arches.
A final configuration is constructed via several deposition processes

Total probability $\frac{1}{84}+\frac{1}{630}+\frac{1}{210}+\frac{1}{252}+2\,\frac{1}{126}=\frac{4}{105}$



$$\frac{1}{21} \times \frac{1}{4} \times \frac{1}{3}$$

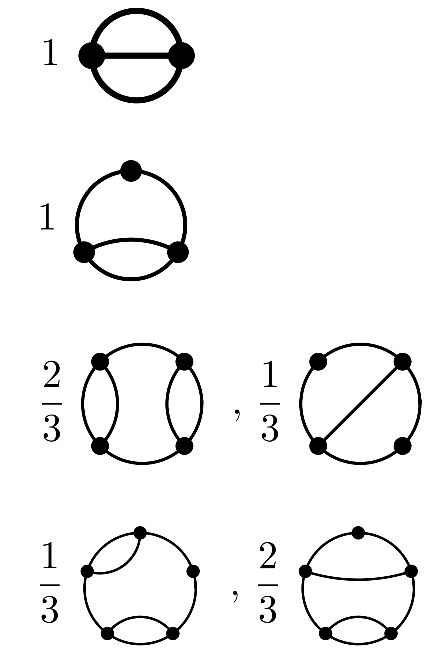
At each step, choose a new arch with uniform probability among all possible arches.
A final configuration is constructed via several deposition processes

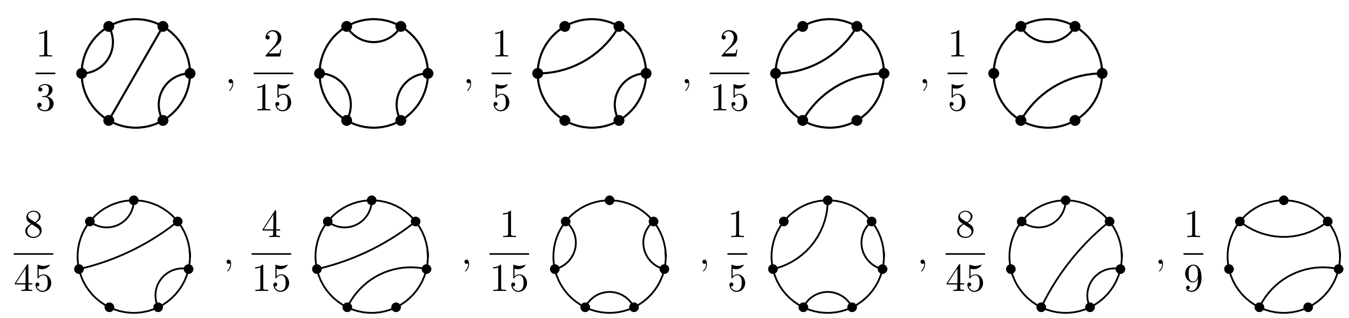
Total probability $$\frac{1}{84}+\frac{1}{630}+\frac{1}{210}+\frac{1}{252}+2\,\frac{1}{126}=\frac{4}{105}$$

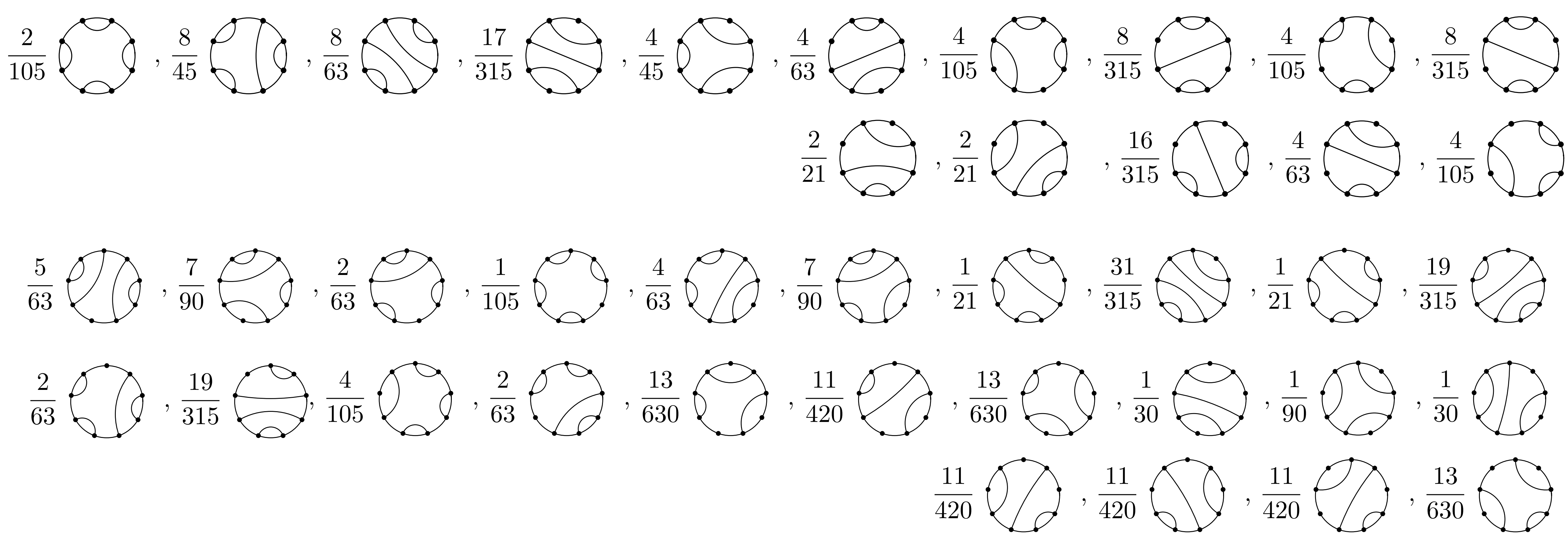
Relation to fragmentation models?

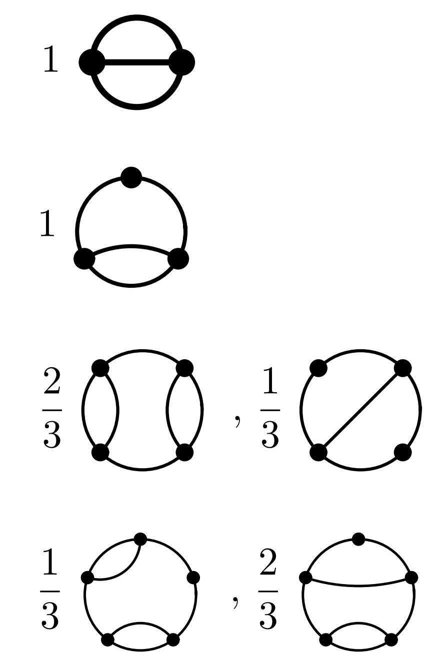
Not a model for real-time RNA folding

We generate an ensemble of configurations with non-trivial weights (different from those obtained by local processes, no combinatorics...)

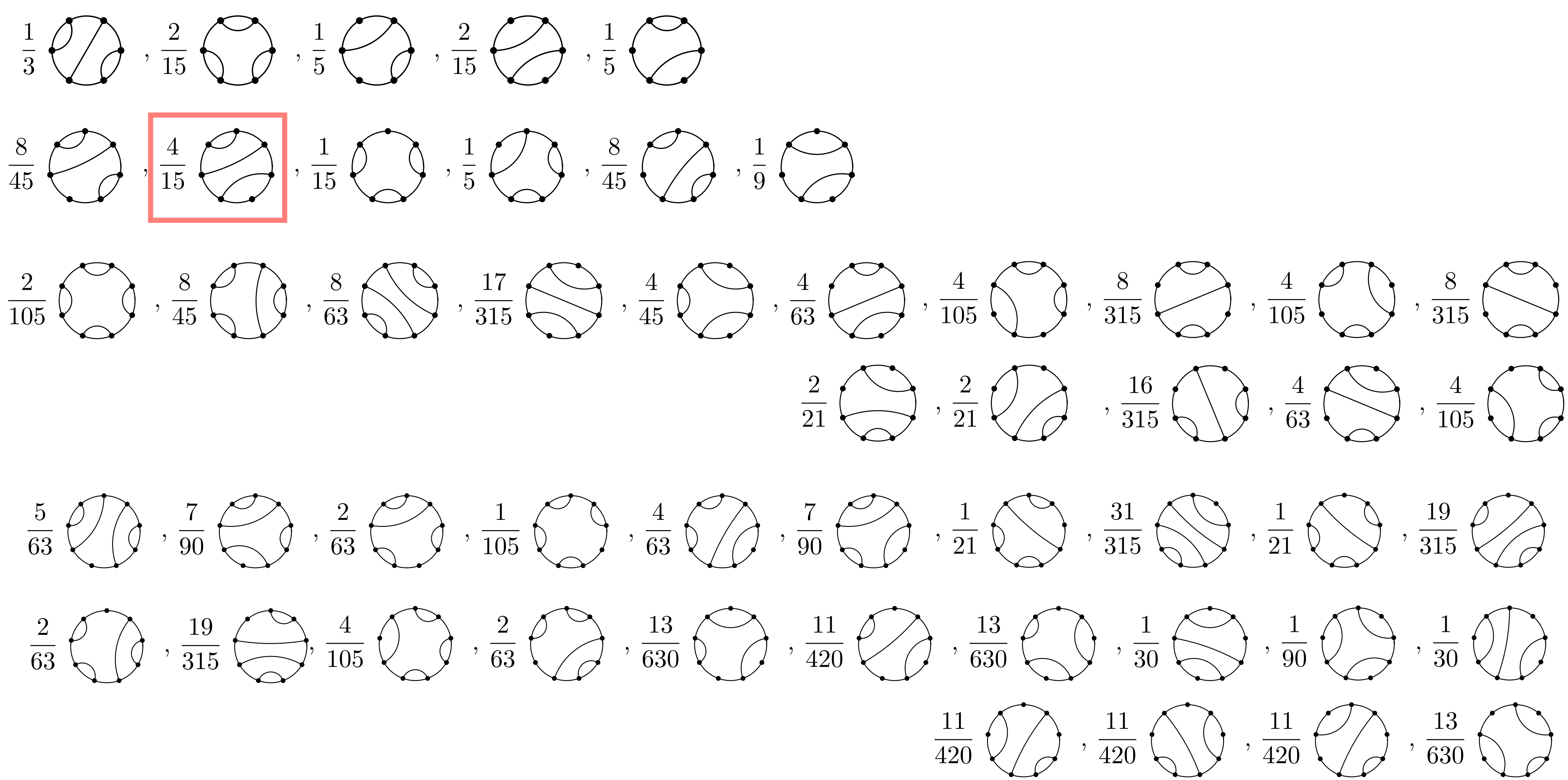


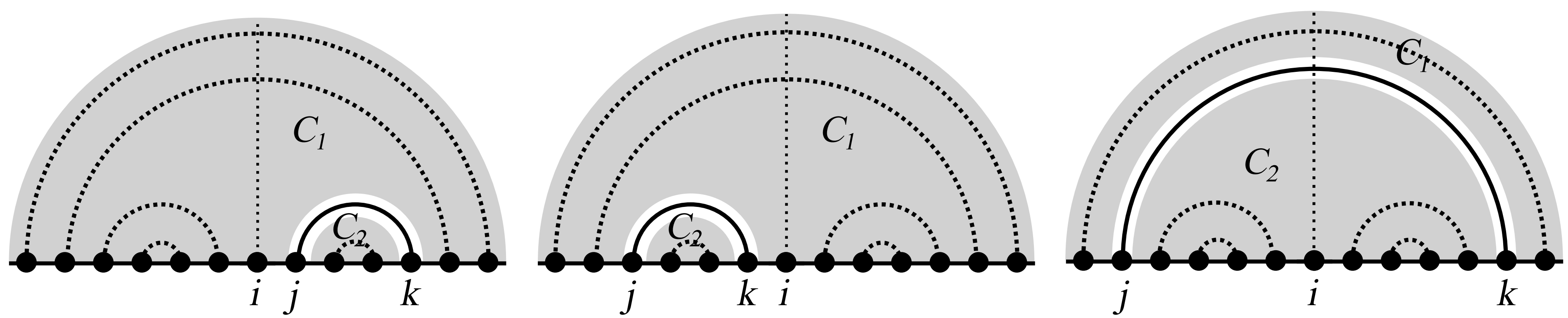



# We generate an ensemble of configurations with non-trivial weights (different from those obtained by local processes, no combinatorics...)



# Tools: recursion relations for probabilities and height

heights = number of arches above a point (Dyck & Motzkin paths)




Many explicit results

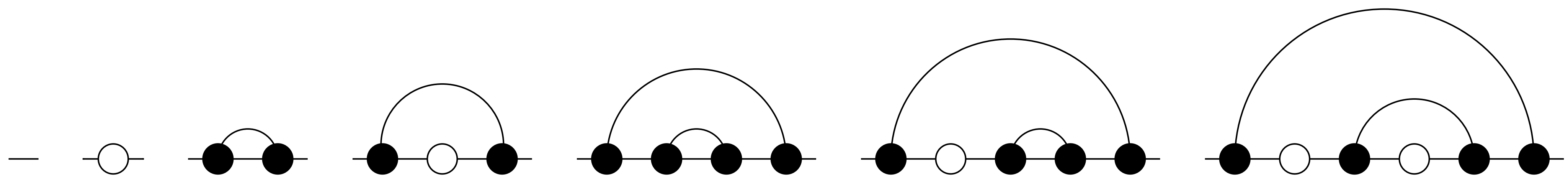
Hausdorff dimension

$$\zeta^2 + 3\zeta = 2\,, \quad \zeta = \frac{\sqrt{17}-3}{2} = 0.561\ldots\,, \quad \mathrm{d_F} = \frac{1}{\zeta} = 1.78\ldots < 2$$

Scaling functions in term of hypergeometric functions

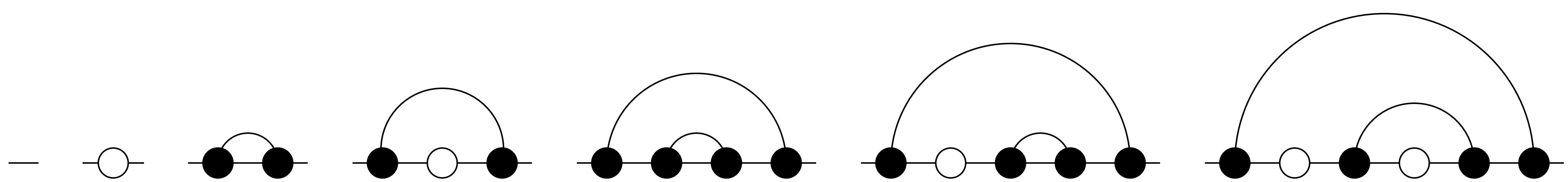
# This model is equivalent to a tree growth model !

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

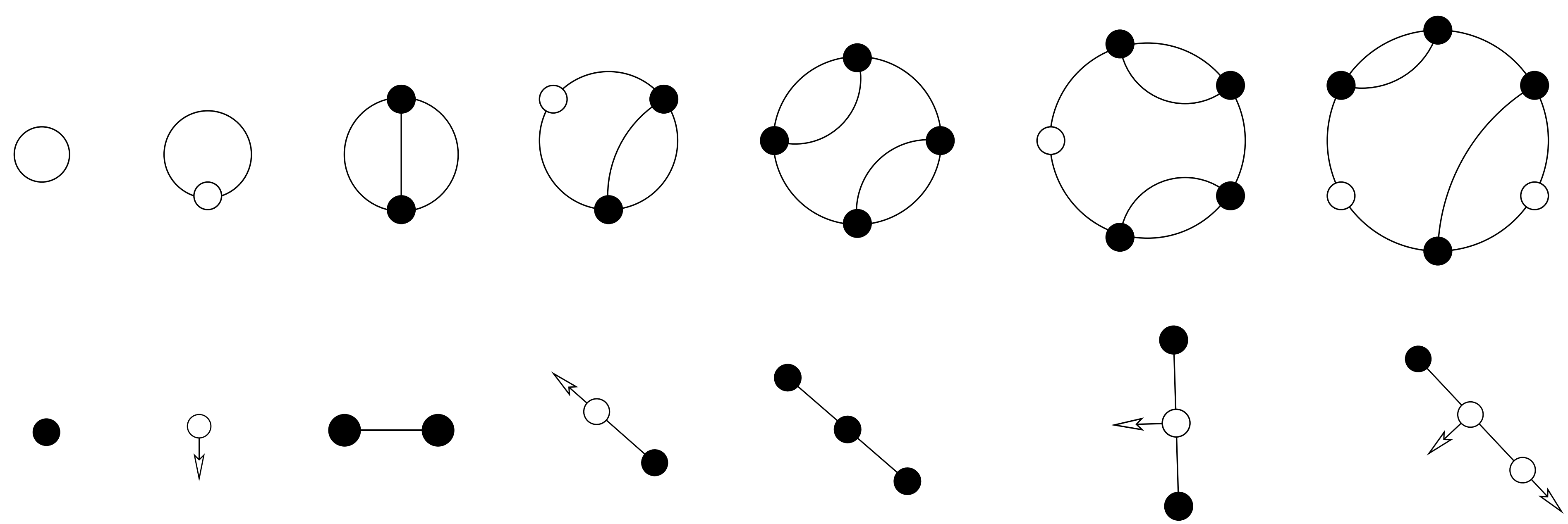


# This model is equivalent to a tree growth model !

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

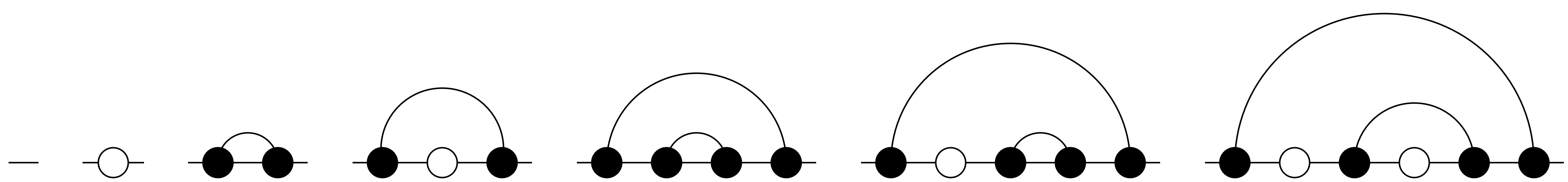


Closing + duality gives a decorated tree growth model

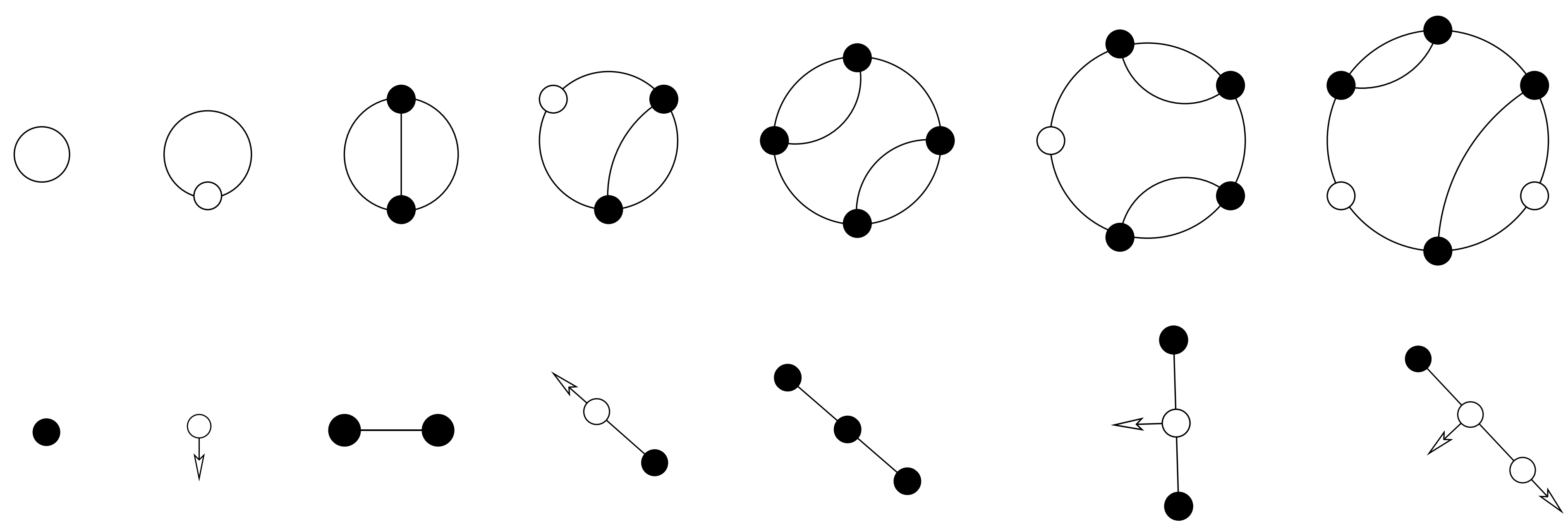


# This model is equivalent to a tree growth model !

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



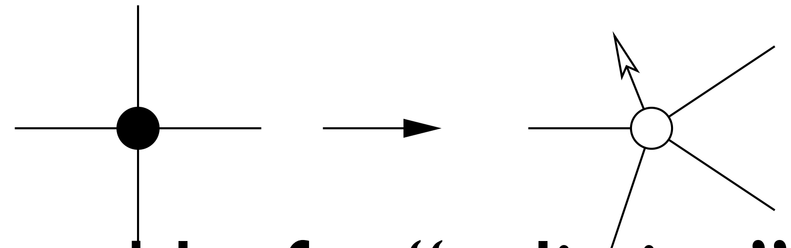
Closing + duality gives a decorated tree growth model



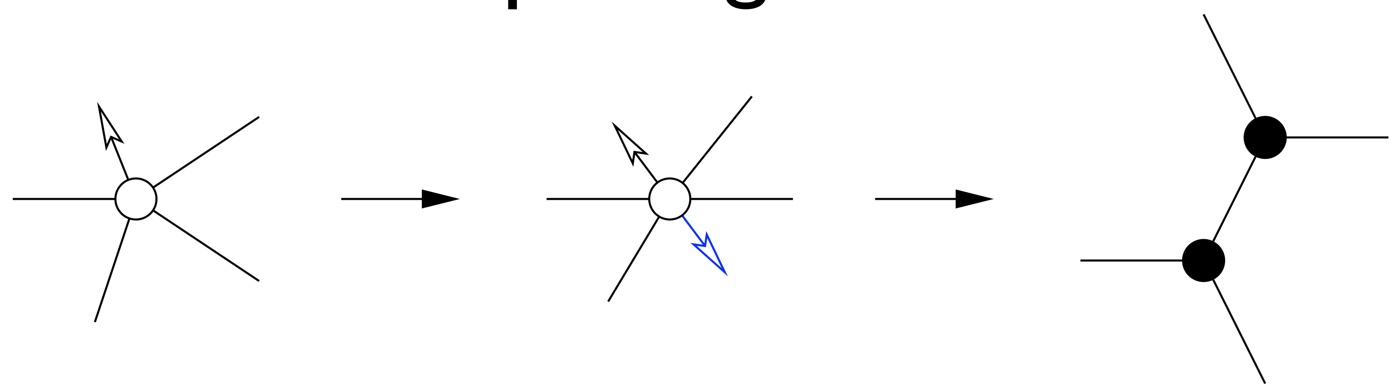
arch/tree growth at time $t$ = arch deposition for size $n=t$ !

2 steps growth process:

- add a leaf to a vertex: 1 black → 1 white



- add a second leaf = "splitting" 1 white → 2 black

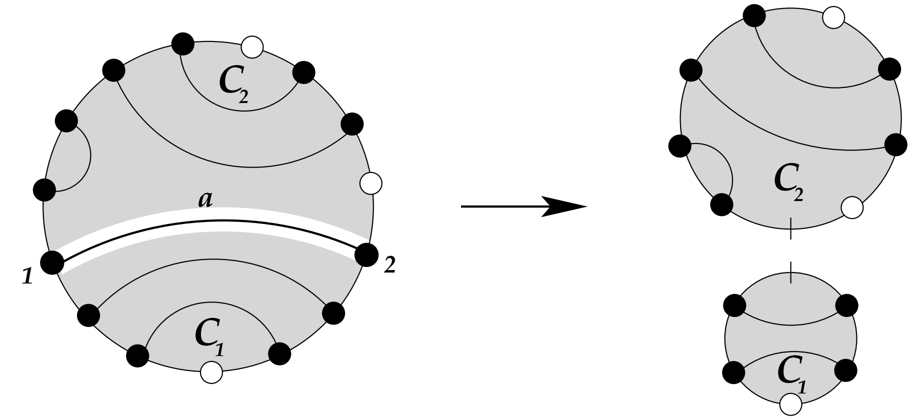


$$k \, \textit{white} \; \to (k_1, k_2) \; \text{black} \qquad k_1 + k_2 = k + 2$$

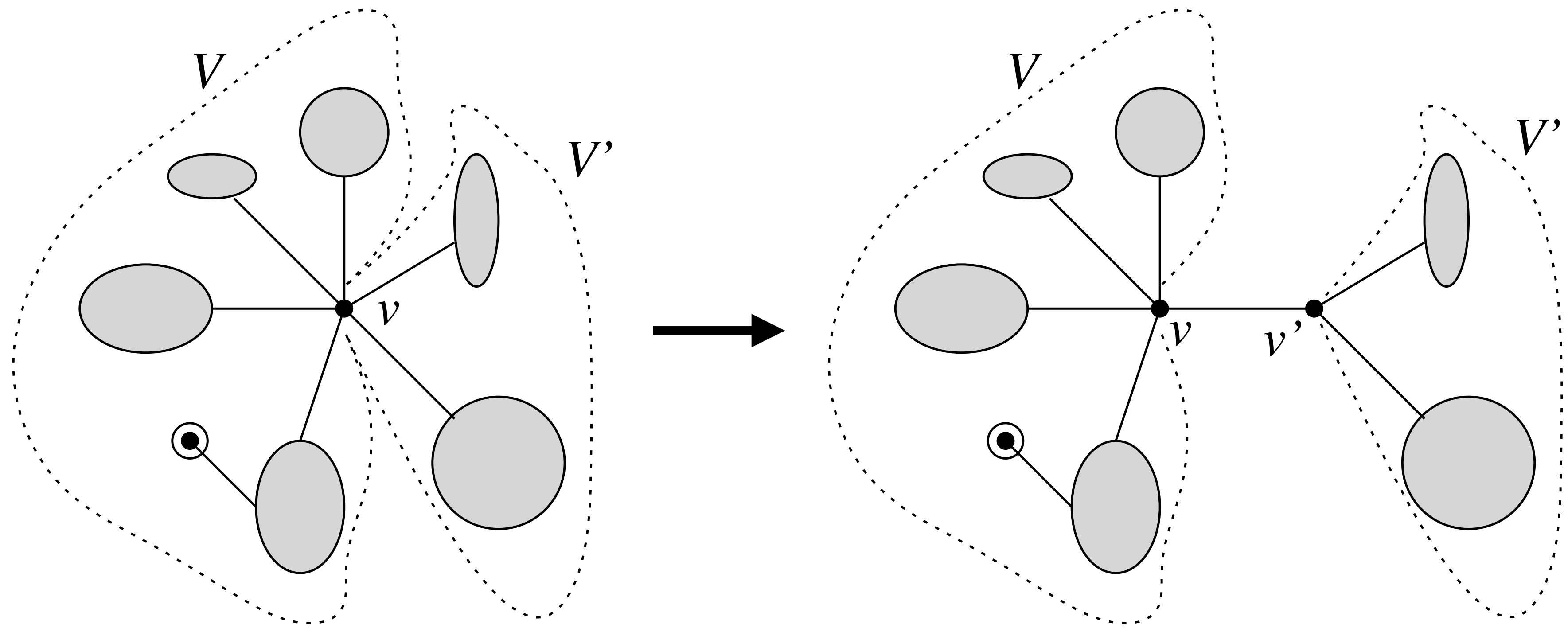
NB: attachment is just a special case

$$k_1 = 1 \,, \;\; k_2 = k + 1$$

Proof of the equivalence
deposition = growth by splitting
Use the recursion relation for
probabilities for both models



# Simpler but more general model: growth of planar trees by splitting




$$k \to (k_1, k_2) \quad , \qquad k_1 + k_2 = k + 2$$

Relative probability for specific splitting of a vertex

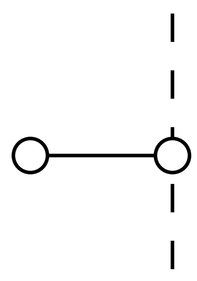
$$w_{k_1,k_2}$$

Relative probability for splitting of a vertex

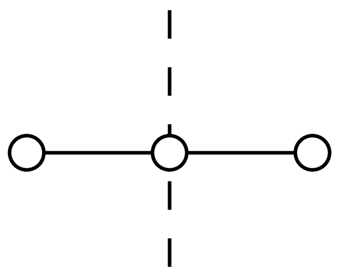
$$w_k = \frac{k}{2} \sum_{k_1+k_2=k+2} w_{k_1,k_2}$$

# This includes attachement processes

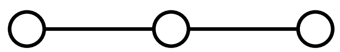
# This includes attachement processes



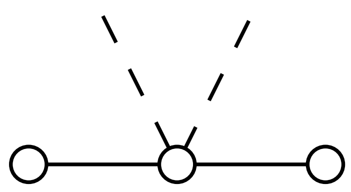
# This includes attachement processes



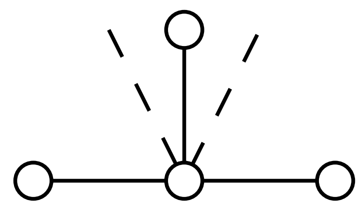
# This includes attachement processes



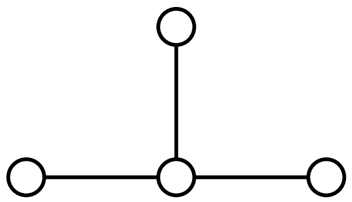
# This includes attachement processes



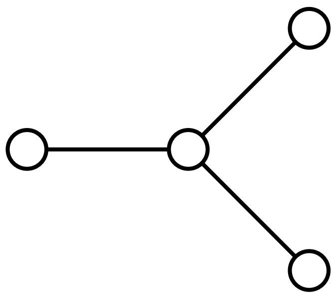
# This includes attachement processes



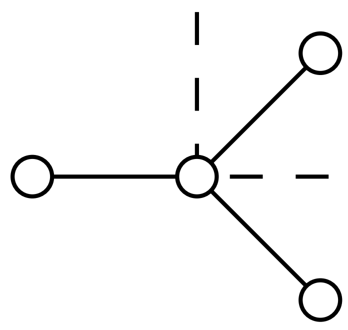
# This includes attachement processes



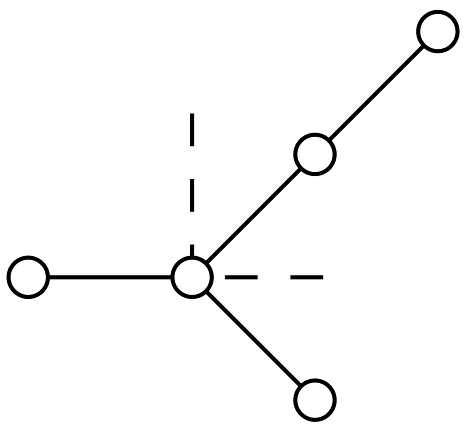
# This includes attachement processes



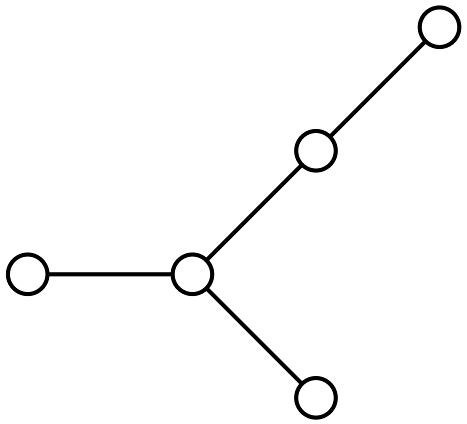
# This includes attachement processes



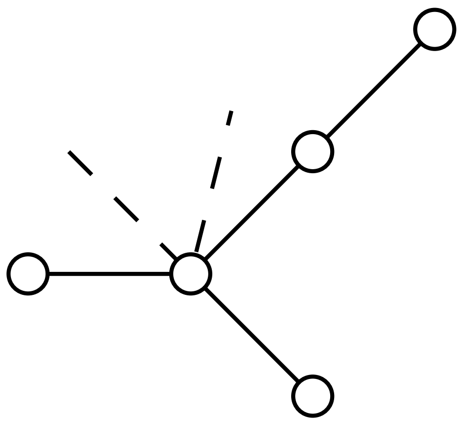
# This includes attachement processes



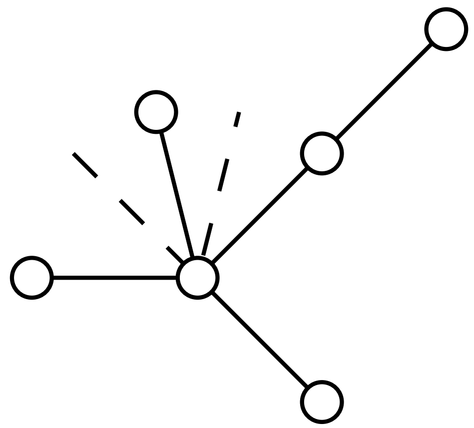
# This includes attachement processes



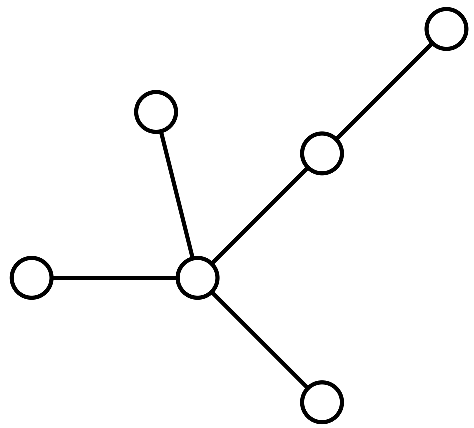
# This includes attachement processes



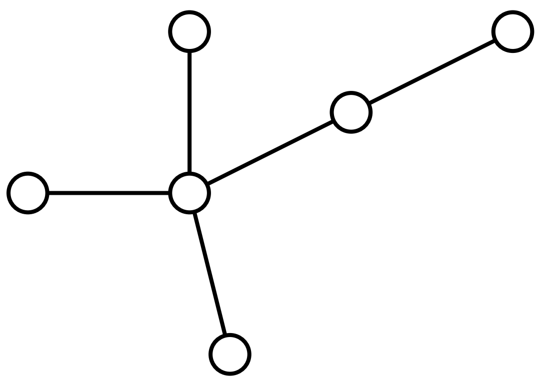
# This includes attachement processes



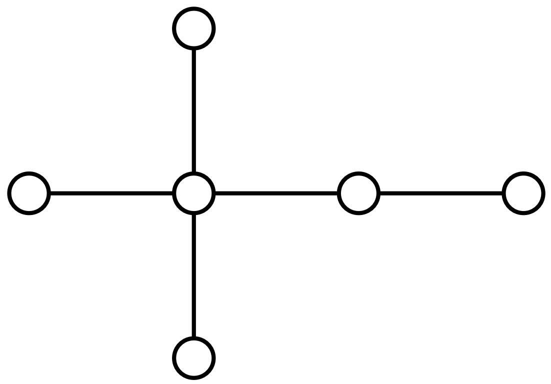
# This includes attachement processes



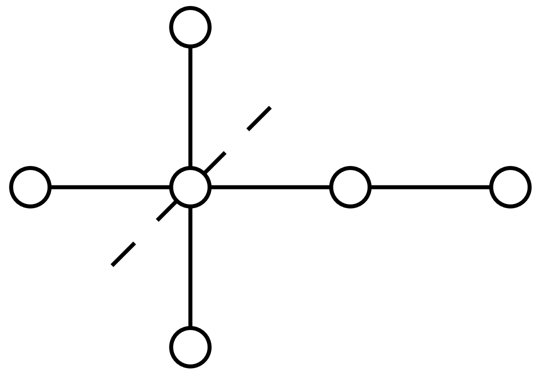
# This includes attachement processes



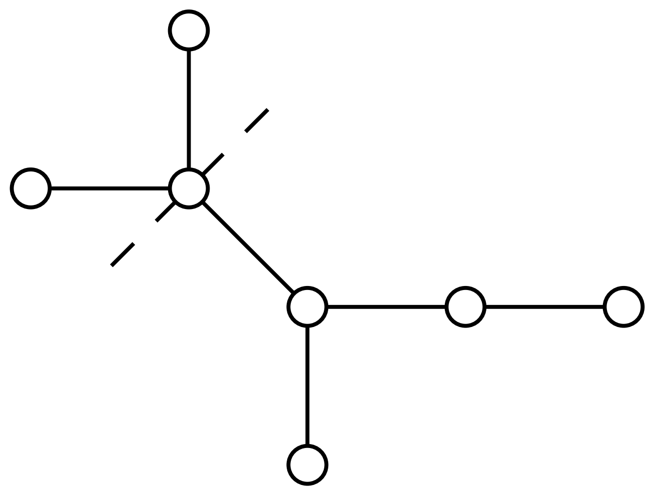
# This includes attachement processes



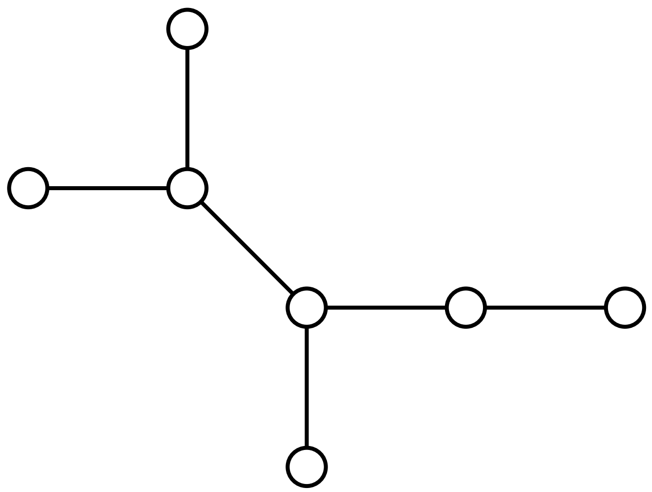
# This includes attachement processes



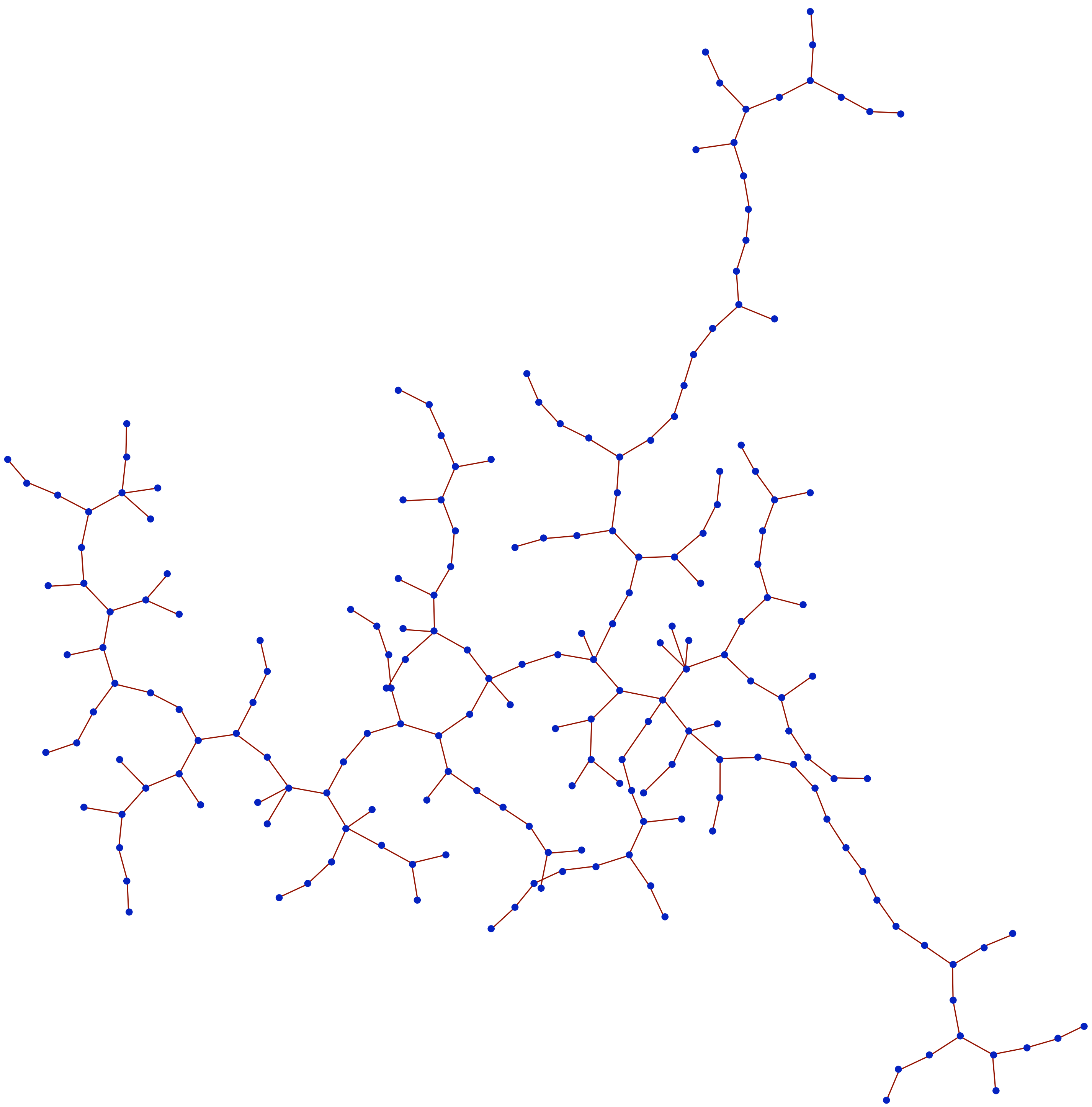
# This includes attachement processes



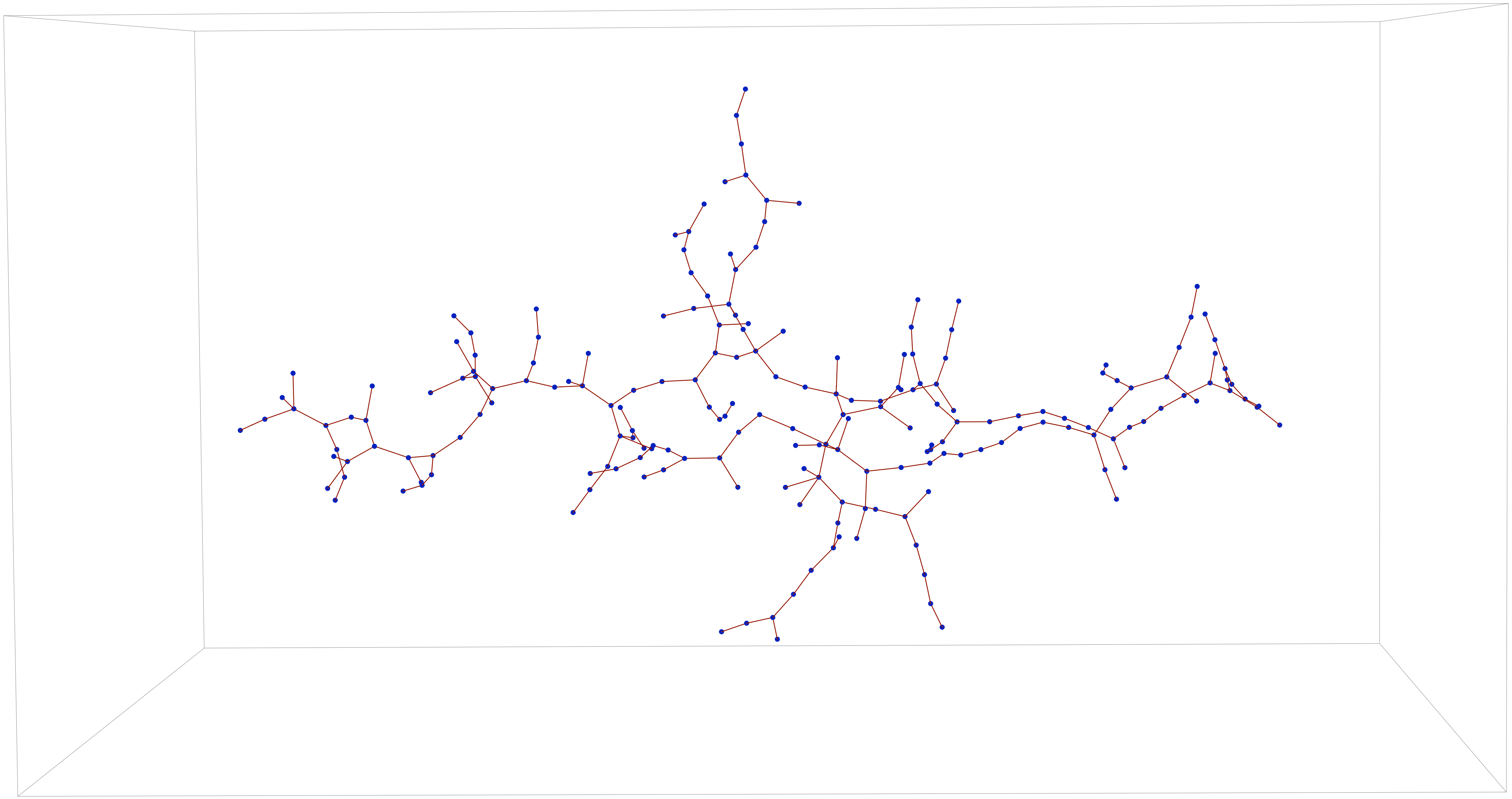
# This includes attachement processes



# This includes attachement processes



# This includes attachement processes



- What kind of trees do we get?

- Like scale-free networks? (vertex with large coordination, infinite Hausdorff dimension)

- Or like generic trees (finite coordination, finite Hausdorff dimension)

- Universal properties?

- Do the Hausdorff dimension depend on the details of the transition rates matrix

Simple example:

Maximal coordination 3

Uniform transition rates

$$w_{1,2} = w_{2,1} = \frac{w_1}{2}$$
$$w_{1,3} = w_{2,2} = w_{3,1} = \frac{w_2}{3}$$
$$w_{3,2} = w_{2,3} = \frac{w_3}{2}$$

Linear k-dependence

$$w_k = ak + b$$

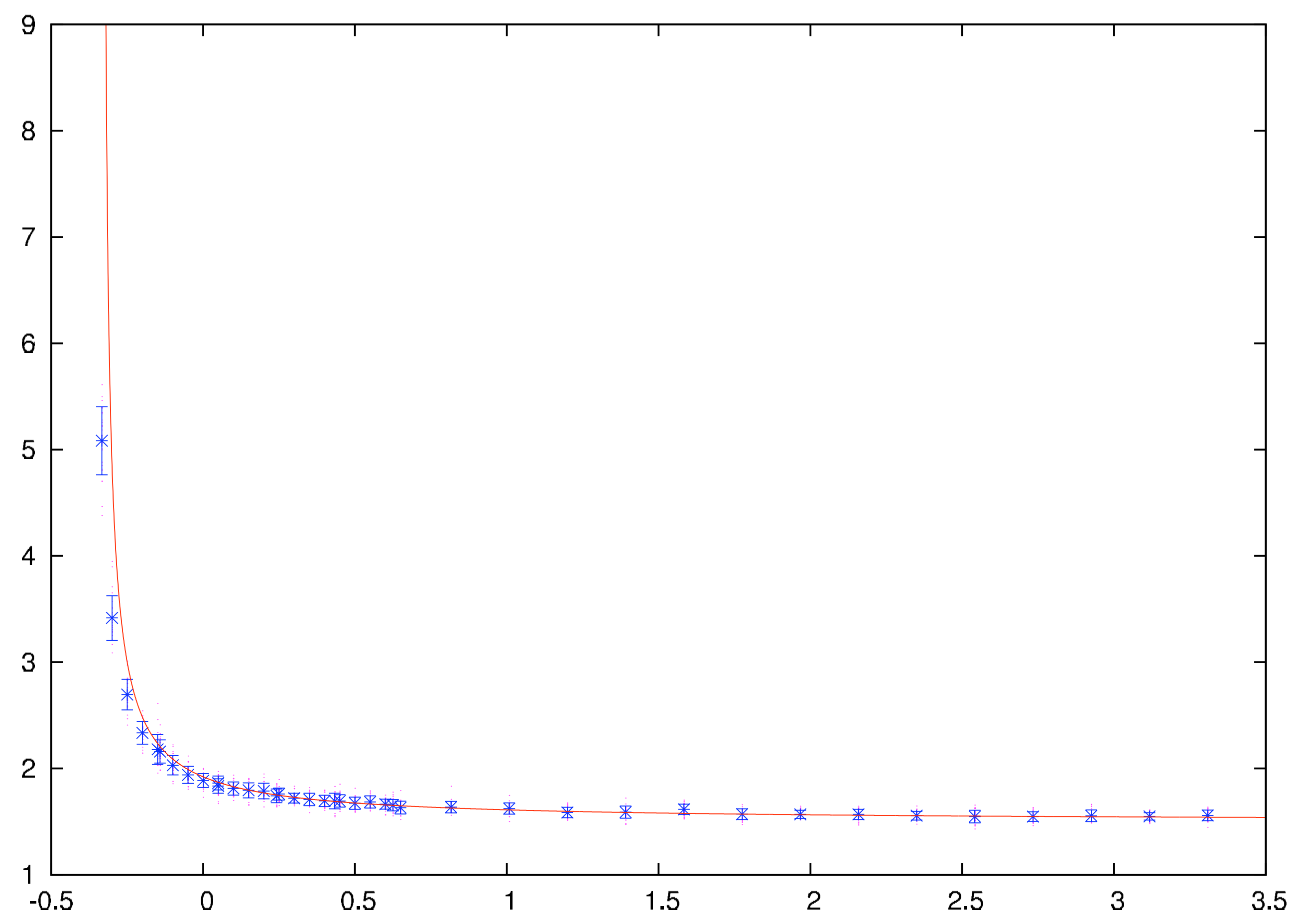
General approach: recursions for correlations

$$x = a/b$$

$$d_H = \frac{3(\sqrt{100x^2 + 84x + 17} + 2x + 1)}{8(3x + 1)}$$

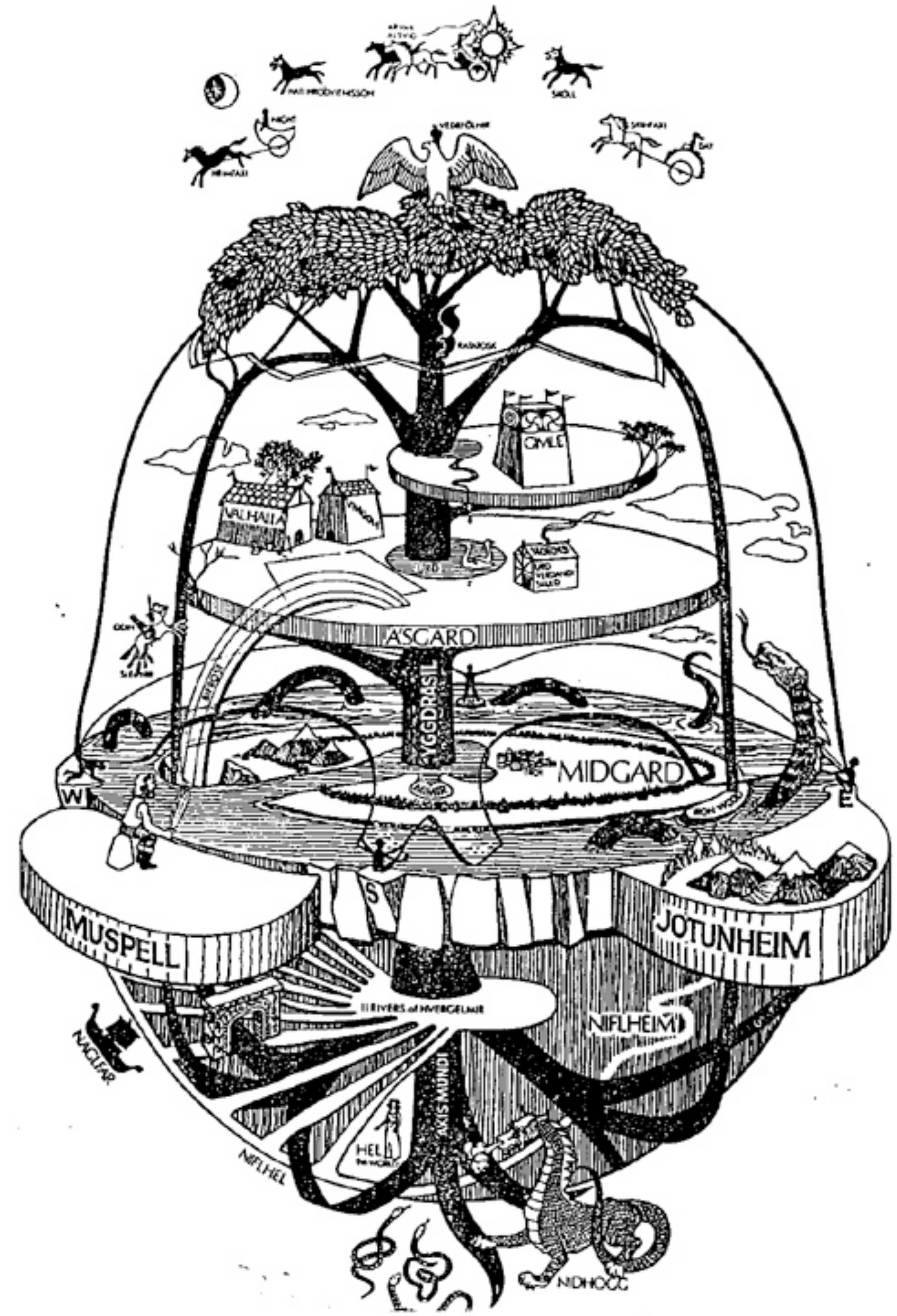
# Hausdorff dimension varies continuously

$$1 < d_H < \infty$$



# Conclusion: the world as a tree?

# Conclusion: the world as a tree?



# Thanks to the organizers!