Degree distribution in random Apollonian networks structures

Alexis Darrasse
joint work with Michèle Soria

26 February 2007

Systèmes Polynomiaux LIP6 SPIRAL Implantations et Résolutions Algébriques

Plan

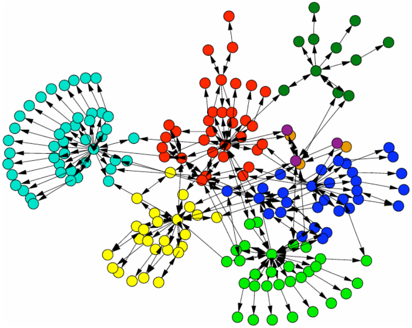
Real-life Networks

Application domains

- Computer Science
- Biology
- Sociology
- ...

Models

- Needed to simulate real-life networks
- Simple classes of random graphs not a good model



Web site

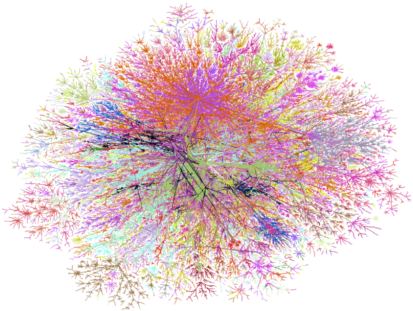
Real-life Networks

Application domains

- Computer Science
- Biology
- Sociology
- ...

Models

- Needed to simulate real-life networks
- Simple classes of random graphs not a good model



Internet

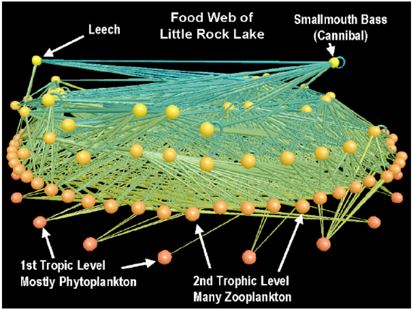
Real-life Networks

Application domains

- Computer Science
- Biology
- Sociology
- . . .

Models

- Needed to simulate real-life networks
- Simple classes of random graphs not a good model



Food web

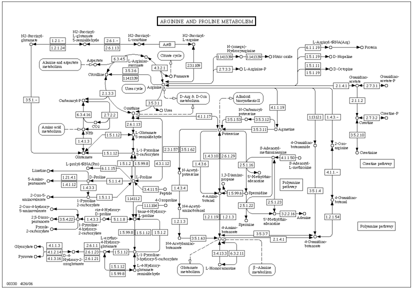
Real-life Networks

Application domains

- Computer Science
- Biology
- Sociology
- ...

Models

- Needed to simulate real-life networks
- Simple classes of random graphs not a good model



Metabolism

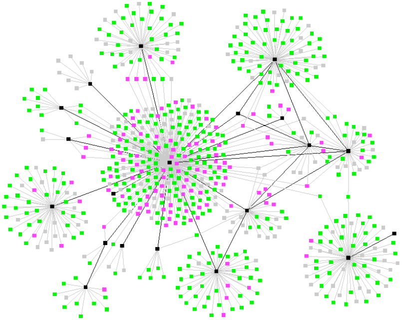
Real-life Networks

Application domains

- Computer Science
- Biology
- Sociology
- ...

Models

- Needed to simulate real-life networks
- Simple classes of random graphs not a good model



Contagion of diseases

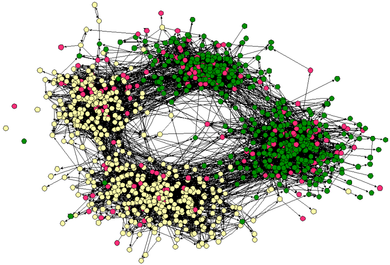
Real-life Networks

Application domains

- Computer Science
- Biology
- Sociology
- ...

Models

- Needed to simulate real-life networks
- Simple classes of random graphs not a good model



Friendship

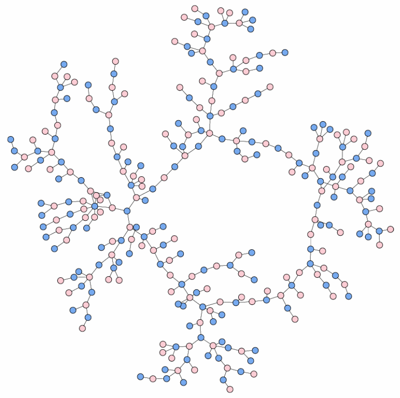
Real-life Networks

Application domains

- Computer Science
- Biology
- Sociology
- ...

Models

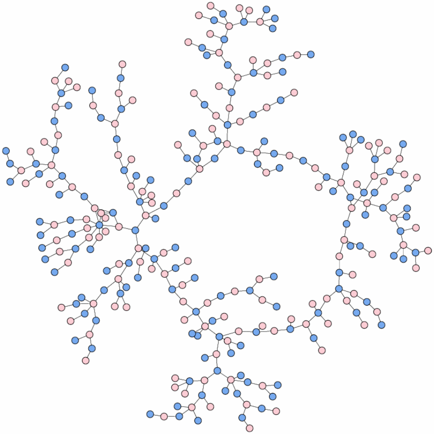
- Needed to simulate real-life networks
- Simple classes of random graphs not a good model



Dating

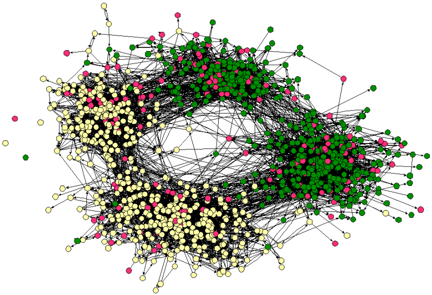
Distinctive properties of real-life graphs

- Number of edges
  - Of the same order as the number of vertices
- Connectivity
- Degree distribution
- Mean distance
- Clustering



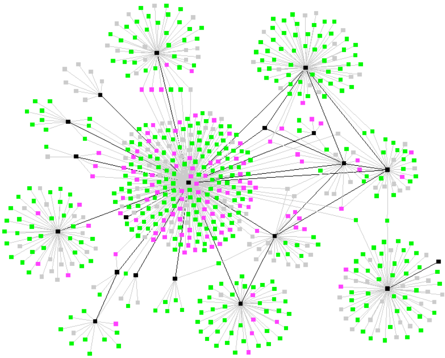
Distinctive properties of real-life graphs

- Number of edges
- Connectivity
  - Strong
    (Giant component)
- Degree distribution
- Mean distance
- Clustering



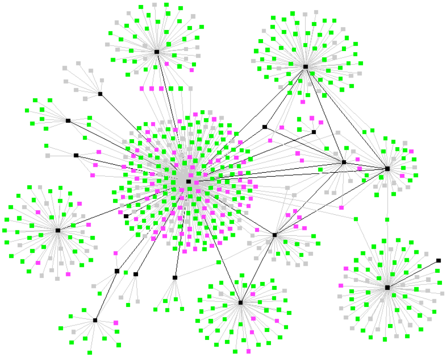
Distinctive properties of real-life graphs

- Number of edges
- Connectivity
- Degree distribution
  - Heavy tailed (Power law, Scale-free)
- Mean distance
- Clustering



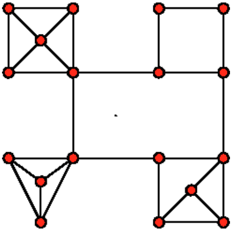
Distinctive properties of real-life graphs

- Number of edges
- Connectivity
- Degree distribution
- Mean distance
  - Small
- Clustering

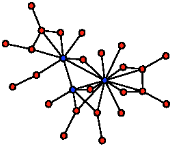


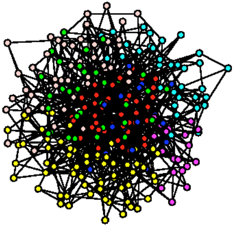
Distinctive properties of real-life graphs

- Number of edges
- Connectivity
- Degree distribution
- Mean distance
- Clustering
    - Strong



Scale-free networks



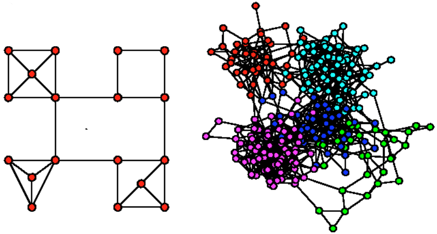


A.-L. Barabási & R. Albert
Emergence of scaling in random networks
Science **286**, 509 (1999)

- Number of edges
- Connectivity
- Degree distribution
- Mean distance
- Clustering

Existing models

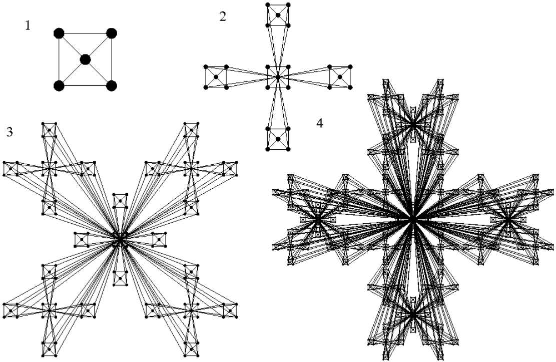
"Small world" networks



Watts D. J. & Strogatz S. H.
Collective dynamics of "small-world" networks
Nature **393**, 440 (1998)

- Number of edges
- Connectivity
- Degree distribution
- Mean distance
- Clustering

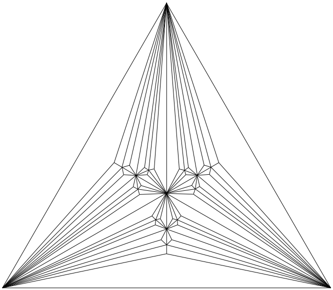
Hierarchical networks



- All properties satisfied
- A new property : hierarchical modularity
- Model is deterministic

E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A.-L. Barabási
Hierarchical Organization of Modularity in Metabolic Networks
Science **297**, 1551 (2002)

A similar model - Apollonian networks

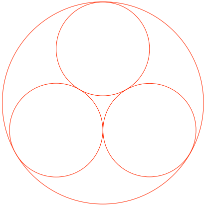


- Same properties as the hierarchical networks
- Inspired from the apollonian packings

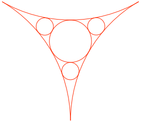
J. S. Andrade, Jr., H. J. Herrmann,
R. F. S. Andrade & L. R. da Silva
Apollonian Networks : Simultaneously Scale-Free, Small World, Euclidean, Space Filling, and with Matching Graphs
Phys. Rev. Lett. **94**, 018702 (2005)

Apollonian packings, Apollonian networks

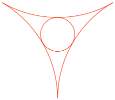
1



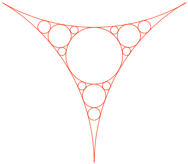
3



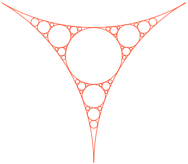
2



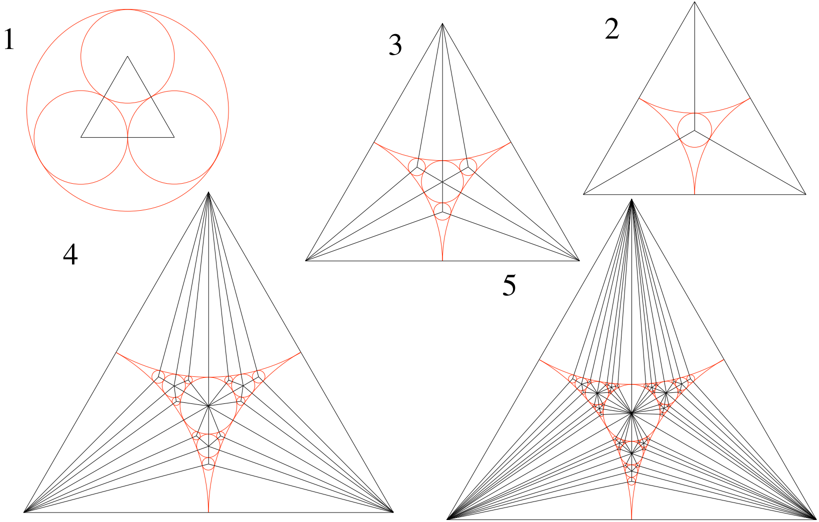
4



5



Apollonian packings, Apollonian networks

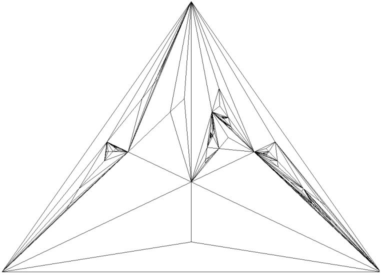


random Apollonian networks

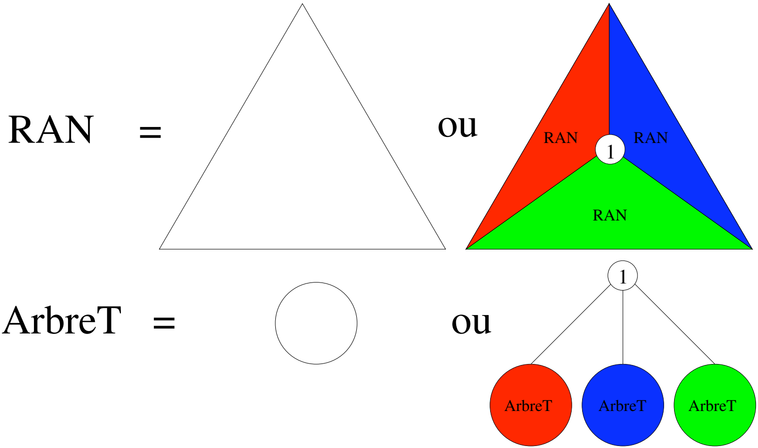
Tao Zhou, Gang Yan, & Bing-Hong Wang
Maximal planar networks with large clustering coefficient and power-law degree distribution
Physical Review E **71**, 046141 (2005)

Algorithm

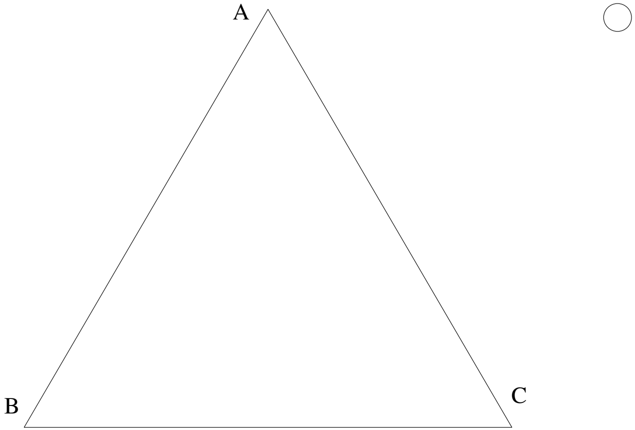
- Initial state : a triangle
- Iterative state : Choose a triangle and add to it a point and link it to the three vertices of the triangle



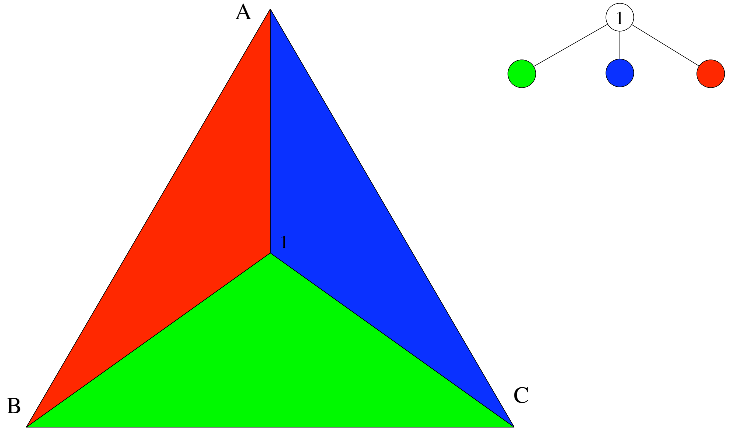
A bijection with ternary increasing trees



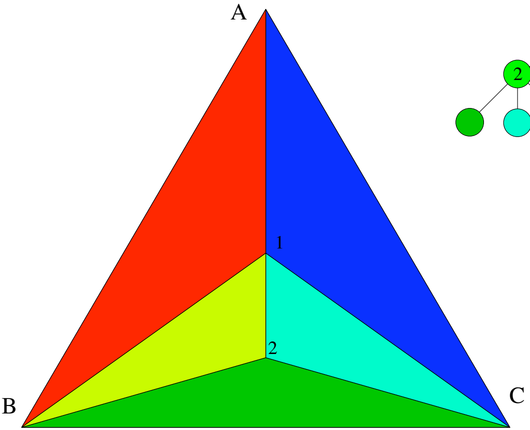
The bijection
Random Apollonian Networks $\leftrightarrow$ Ternary Increasing Trees

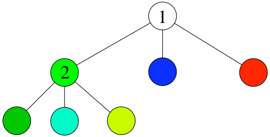


The bijection
Random Apollonian Networks $\leftrightarrow$ Ternary Increasing Trees

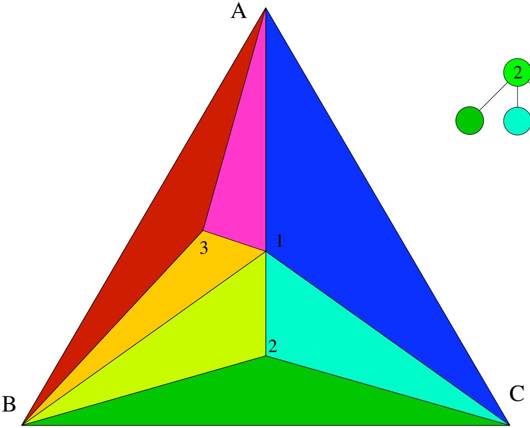


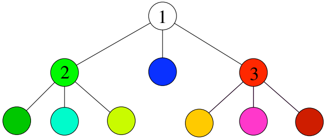
The bijection Random Apollonian Networks ↔ Ternary Increasing Trees



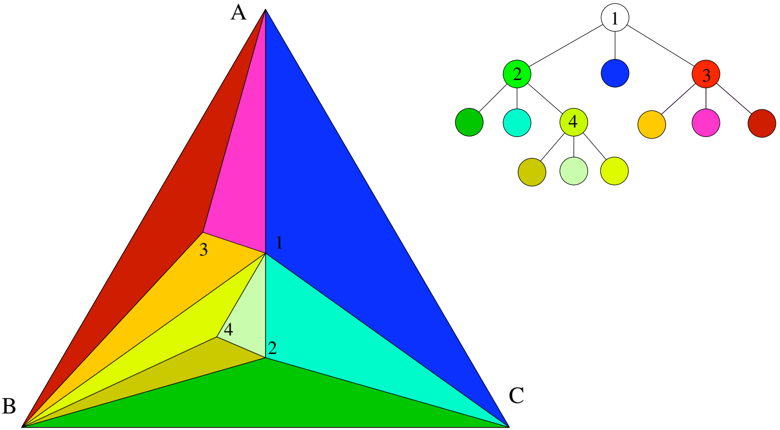


The bijection Random Apollonian Networks ↔ Ternary Increasing Trees

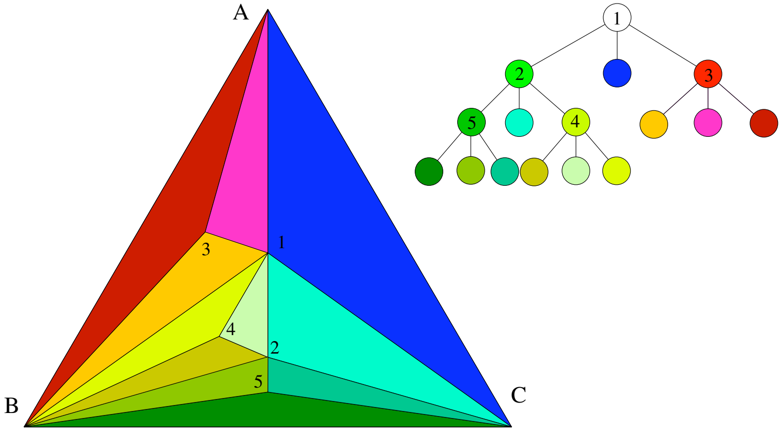




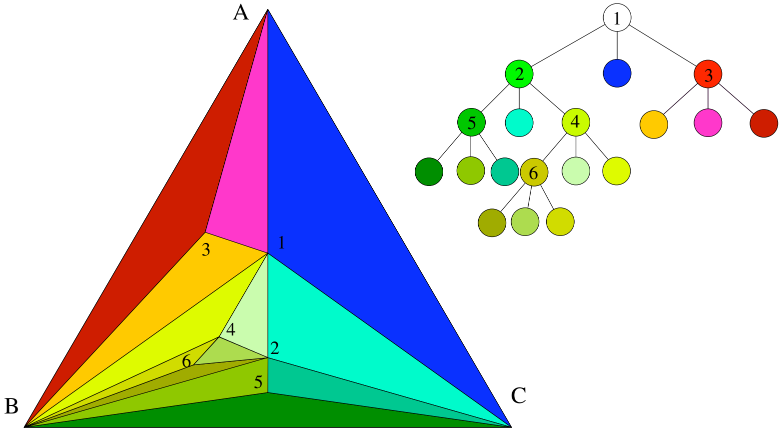
The bijection Random Apollonian Networks ↔ Ternary Increasing Trees



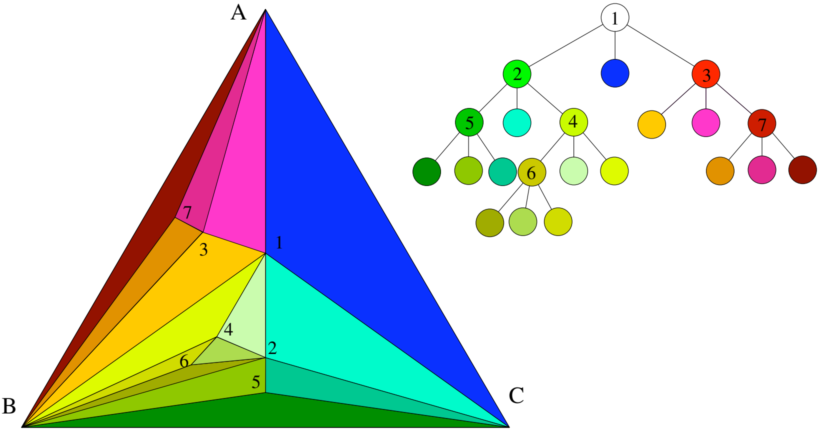
The bijection
Random Apollonian Networks $\leftrightarrow$ Ternary Increasing Trees



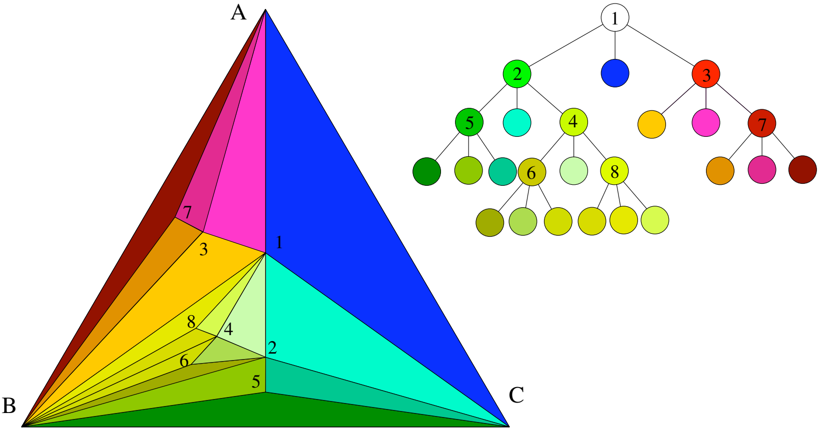
The bijection
Random Apollonian Networks ↔ Ternary Increasing Trees



The bijection Random Apollonian Networks $\leftrightarrow$ Ternary Increasing Trees



The bijection
Random Apollonian Networks ↔ Ternary Increasing Trees



Random Apollonian network structures

Replace Ternary Increasing Trees with Ternary Trees

Properties

- Same bijection
- Same class of graphs
- Different probability distribution
- Properties preserved
- Simple combinatorial description of the model

What for ?

- General methods for sampling
- Efficient generation (Boltzmann)
- Greater flexibility

Ternary tree generation using the Boltzmann model

The Boltzmann model

Specifiable combinatorial classes

- Basic operations :
  Union, **Product**, Sequence, Cycle, Set
- **Recursive definitions**

Properties

- Uniform generation
- Approximate size
- Efficiency

P. Duchon, P. Flajolet, G. Louchard, G. Schaeffer
Boltzmann samplers for the random generation of combinatorial structures

Algorithm for the generation of a ternary tree

$$T(z) = z + zT(z)^3$$

Algorithm : TernaryTree(p)

if rand(0..1) < p **then**
  Leaf
else
  Node(TernaryTree(p),TernaryTree(p),TernaryTree(p))
end if

$p = x/T(x)$, $x \leq \rho$

- Aim at mean value : $x < \rho$
- Singular sampling : $x = \rho$
- Pointing

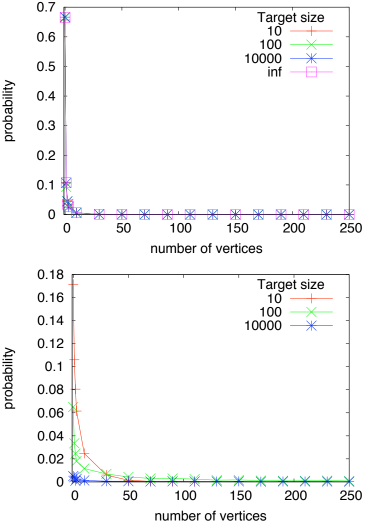
Distribution of the sizes of the generated trees

Using the straightforward algorithm

- Most generated trees are leaves
- A few very big trees
- Power law distribution

With pointing

- Many small trees still present, but
- Less disparity



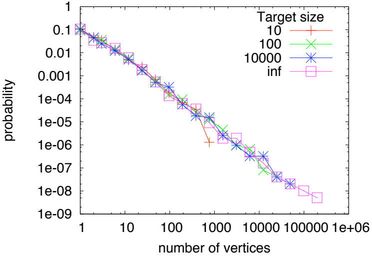
Distribution of the sizes of the generated trees

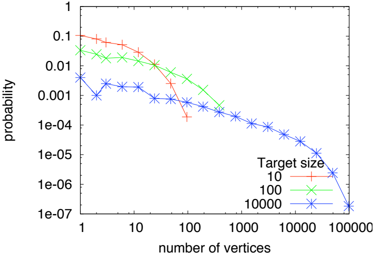
Using the straightforward algorithm

- Most generated trees are leaves
- A few very big trees
- Power law distribution

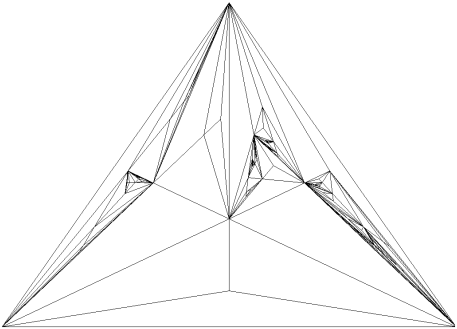
With pointing

- Many small trees still present, but
- Less disparity





Back to the network properties



Properties of the generated networks

By construction :

- Number of edges
  Equal to $3v - 6$, where v the number of vertices
- Connectivity
  A single component
- Mean degree
  $= 6$

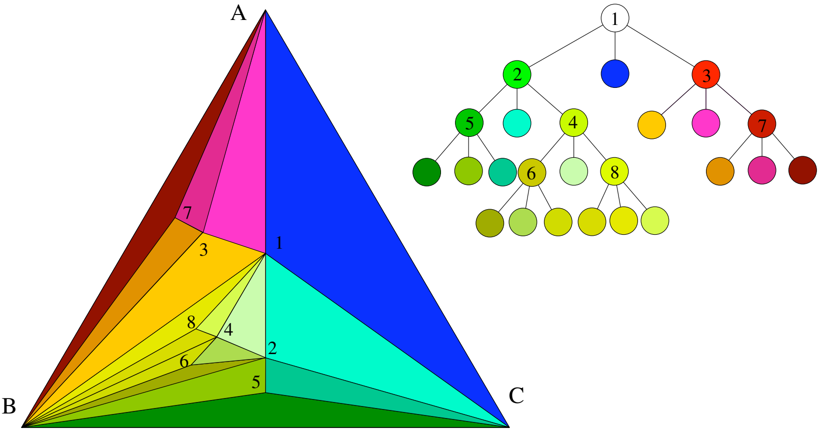
Needing further investigation :

- Degree distribution
- Clustering
- Mean distance

Degree

Neighborhood of a vertex

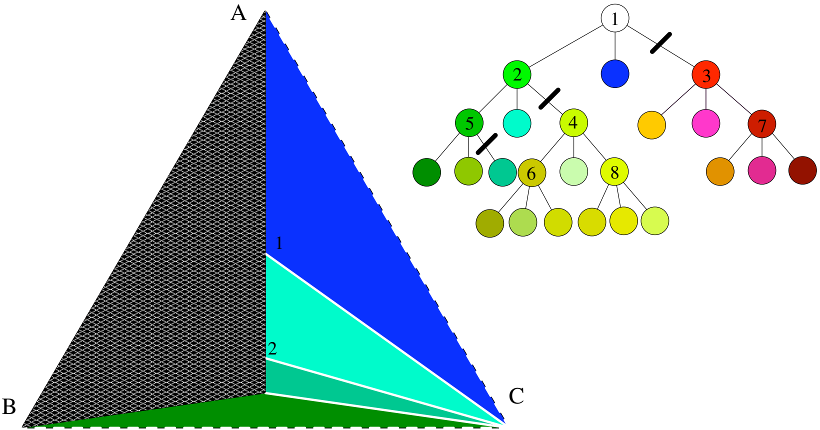
- 3 ancestors + root neighborhood



Degree

Neighborhood of a vertex

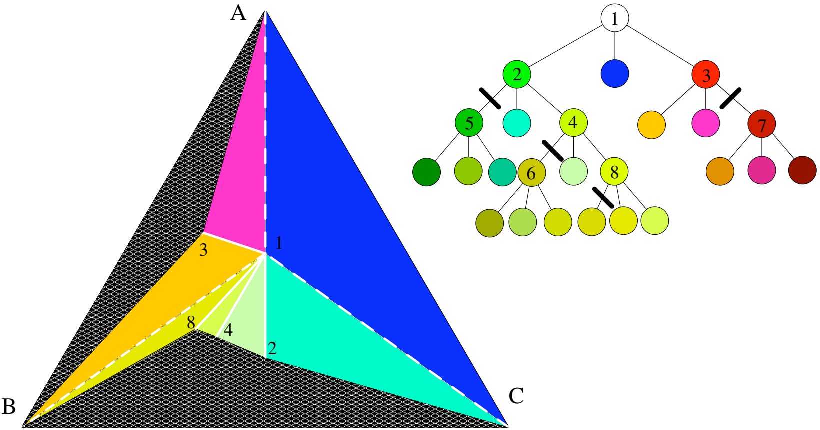
- 3 ancestors + root neighborhood



Degree

Neighborhood of a vertex

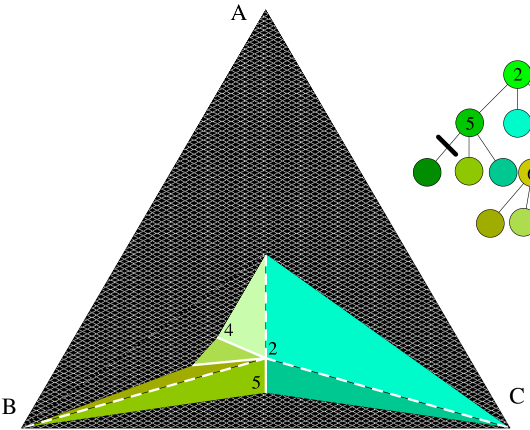
- 3 ancestors + root neighborhood

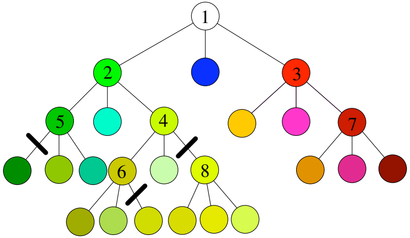


Degree

Neighborhood of a vertex

- 3 ancestors + root neighborhood





Bivariate generating functions

u marks the neighbors

- of the center (root) : $RD(z,u) = zu^3T^3(z,u)$

Bivariate generating functions

u marks the neighbors

- of the center (root) : $RD(z,u) = zu^3 T^3(z,u)$
- of an external node : $T(z,u) = 1 + zuT^2(z,u)T(z)$

Bivariate generating functions

u marks the neighbors

- of the center (root) : $RD(z,u) = zu^3 T^3(z,u)$
- of an external node : $T(z,u) = 1 + zuT^2(z,u)T(z)$

The distribution of the value of a parameter on Boltzmann generated objects

$$\Pr(\Omega = k) = \sum_n \Pr(\Omega = k / N = n) \times \Pr(N = n)$$

$$= \sum_n \frac{C_{n,k}}{C_n} \times \frac{C_n x^n}{C(x)} = \frac{\sum_n C_{n,k} x^n}{C(x)} = \frac{[u^k]C(x,u)}{C(x,1)}$$

Degree distribution

Proposition : Statistical properties

Same for :

- the set of all subtrees of a random tree
- a set of random trees independently generated with a Boltzmann sampler

Degree distribution

Proposition : Statistical properties

Same for :

- the set of all subtrees of a random tree
- a set of random trees independently generated with a Boltzmann sampler

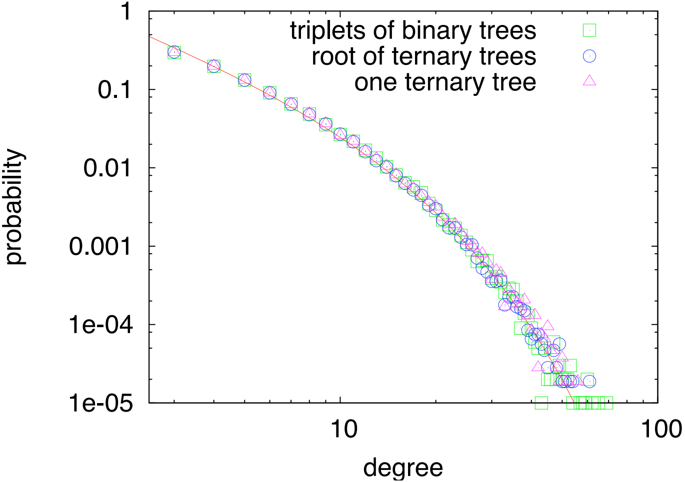
Theorem : degree distribution in RANS

Mean value 6 and a Catalan form for the pgf :

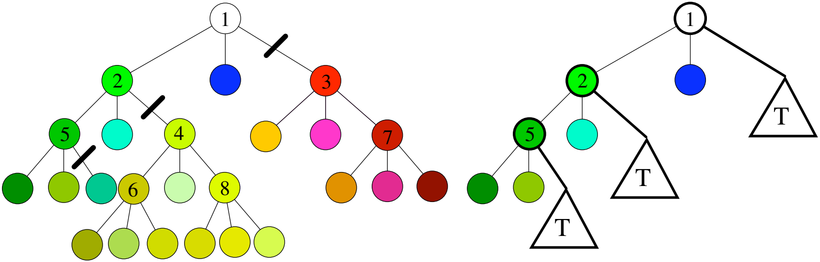
$$\Pr(D = 3 + k) = \frac{8}{9} \frac{1}{k+3} \binom{2k+2}{k} \sim C \left(\frac{8}{9}\right)^k (k+3)^{-3/2}$$

Degree distribution

$$P(D = 3 + k) \sim C \left(\frac{8}{9}\right)^k (k+3)^{-3/2}$$



Sketch of proof



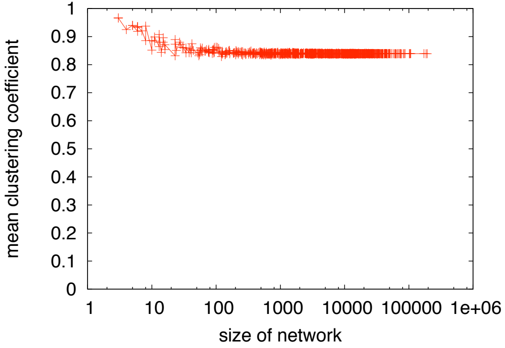
- Ternary trees marked for degree :
  $T(z,u) = 1 + zuT^2(z,u)T(z)$
- Simulated by binary trees :
  $T(z,u) = B(zuT(z))$, where $B(t) = \sum B_n t^n$
- Schema is subcritical : $\rho\tau < 1/4$
- $[u^k]B(zuT(z)) = \rho^k \tau^k \frac{1}{k+1}\binom{2k}{k}$

Clustering

Definition : Clustering coefficient of a vertex of degree k

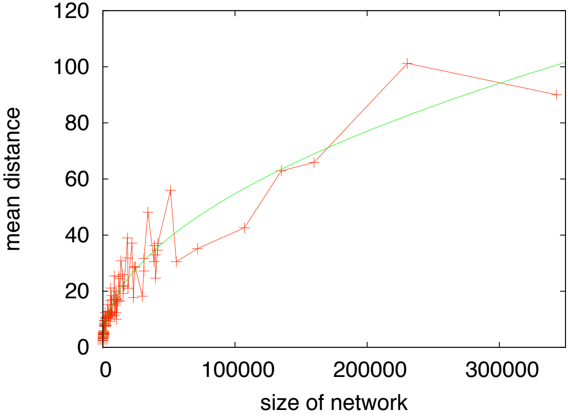
$$C(k) = \frac{\text{number of links between neighbors}}{k(k-1)}$$

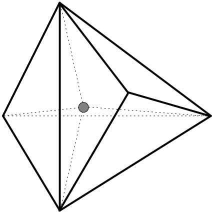
- $C(k) = 3\frac{2k-d-1}{k(k-1)}$
- Mean value over all vertices independent of size

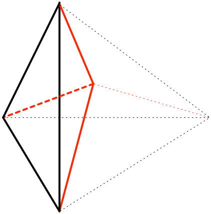


Mean distance

Simulation confirms a small mean distance (order $\sqrt{N}$)

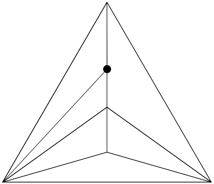


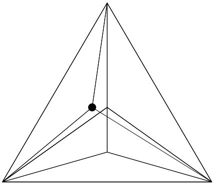




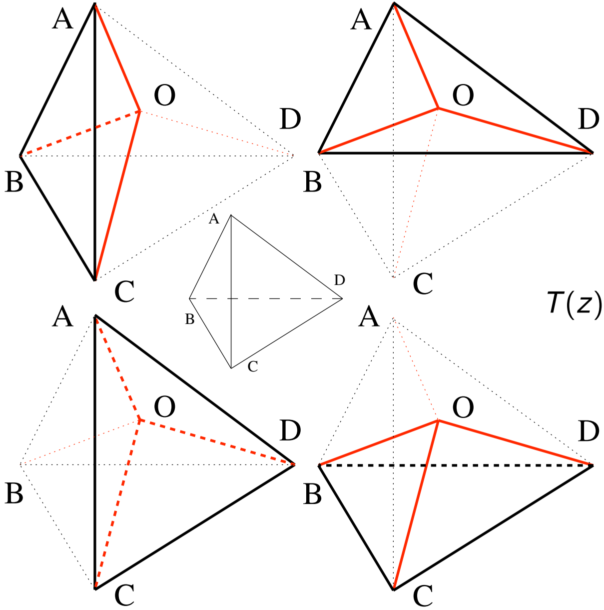
More flexibility :

Variants





Add a dimension

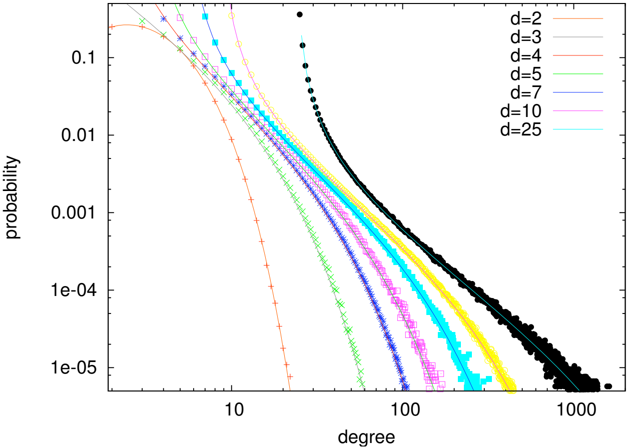


$$T(z) = z + zT(z)^4$$

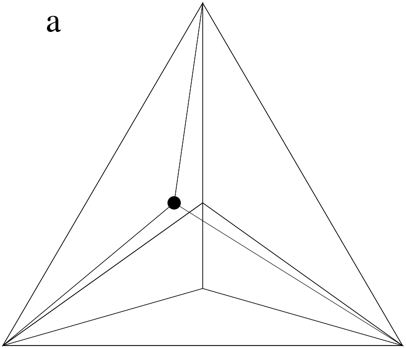
Higher dimension RANS

$$\Pr(D_d = d + k) \sim C\alpha^k \left(k + \frac{d}{d-2}\right)^{-\frac{3}{2}}$$

$$RD_d(z,u) = zu^d T_d^d(z,u) \qquad\qquad T_d(z,u) = T_{d-1}(uzT_d(z))$$

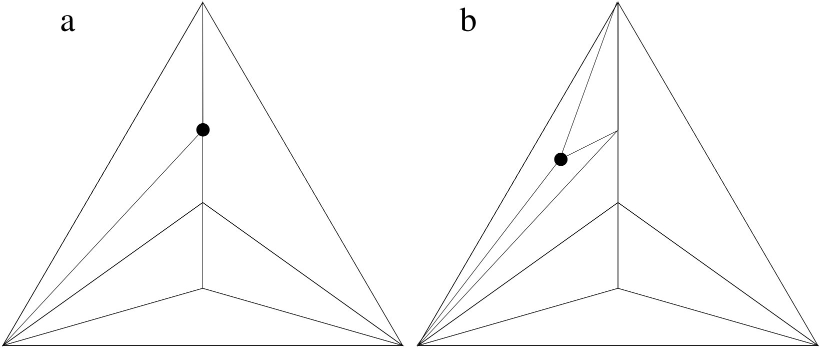


Reuse triangles



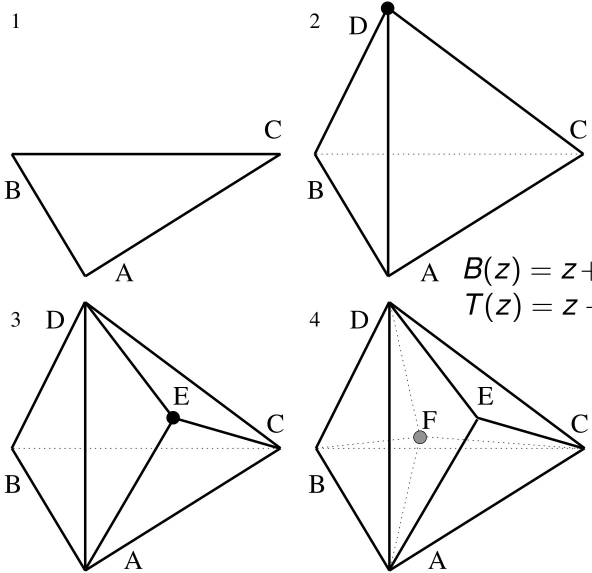
$$T(z) = z + zT(z)^4$$

Remove siblings



$$T(z) = z + 3zT(z) + 3zT(z)^2 + zT(z)^3$$

Combine models



$$B(z) = z + zB(z)^3 + zB(z)^3 T(z)$$
$$T(z) = z + zB(z)^4 T(z)^4$$

Implementation

$T(z) = z + zT(z)^d$

- 1000 lines of C
- Tree and Network generation
- Parameter computation

Sampling of

10^6 generated trees
10^6 maximum "usable" size
10^9 maximum size
in a few seconds time...

Simple families of trees $T_i(z) = z + \phi(< T(z) >)$

- Sampler compiler
- Written in Maple, using Combstruct
- Sampler in C, eventually using Maple as a co-routine

Conclusion and Perspectives

- A simple and extensible model
- Similar models :
  - *k*-trees
  - stacked triangulations
- More adequate to model real graphs ?
- Different strategies to generate trees
- Tree sampling has many more applications (eg. XML)

Image references

From http ://www-personal.umich.edu/ Emejn/networks/

Web site — M. E. J. Newman and M. Girvan, *Finding and evaluating community structure in networks*, Physical Review E 69, 026113 (2004).

Internet — Hal Burch and Bill Cheswick, Lumeta Corp.

Food web — Neo Martinez and Richard Williams.

Contagion of diseases — Valdis Krebs, www.orgnet.com.

Friendship — James Moody, *Race, school integration, and friendship segregation in America*, American Journal of Sociology 107, 679-716 (2001).

Dating — Data drawn from Peter S. Bearman, James Moody, and Katherine Stovel, *Chains of affection : The structure of adolescent romantic and sexual networks*, American Journal of Sociology 110, 44-91 (2004), image made by Mark Newman.

Scale-free & small world — E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, A.-L. Barabási *Hierarchical Organization of Modularity in Metabolic Networks* Science 297, 1551 (2002).