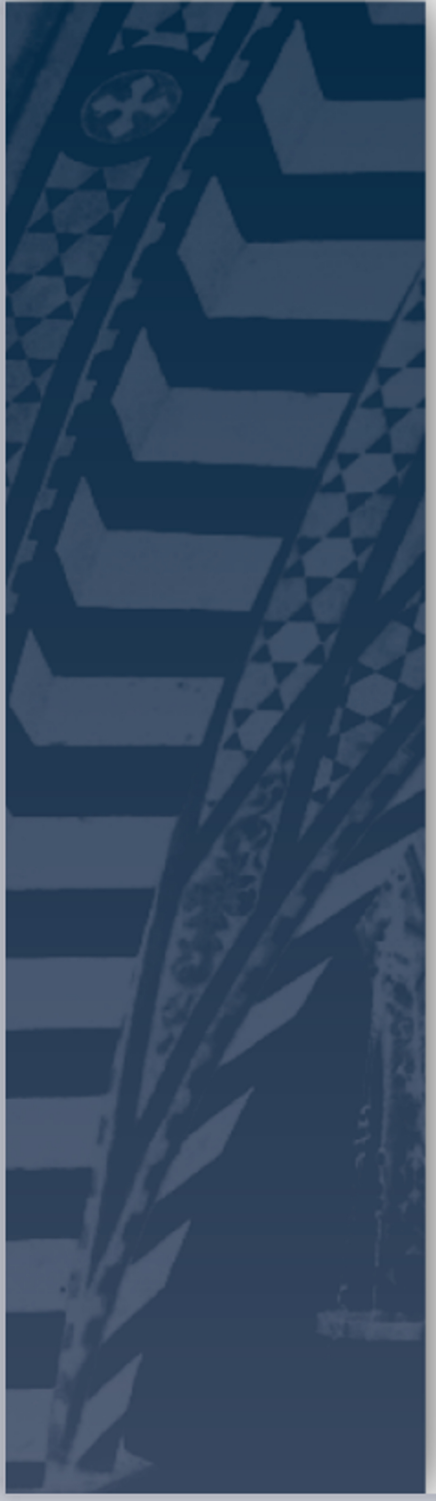


IMT SCHOOL FOR ADVANCED STUDIES LUCCA

Data Science Summer School

Part II: Network Science
Lecture 2/2
G. Caldarelli,

networks.imtlucca.it

September 2, 2019

IMT SCHOOL FOR ADVANCED STUDIES LUCCA

Motivation

Modelling

The art of Modelling is based on

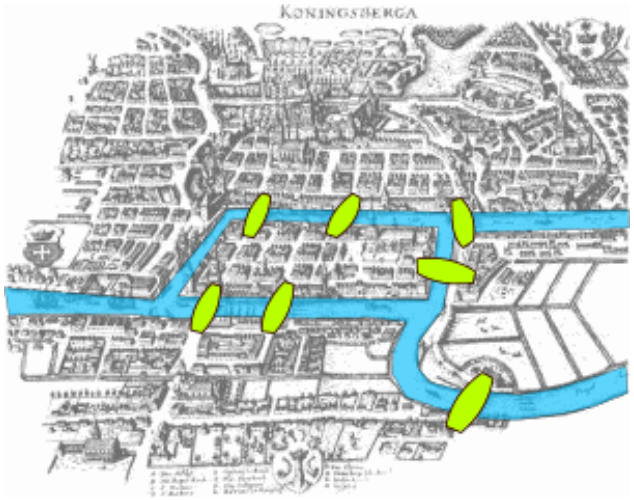
- Find the most important features
- Realize a synthetic system based on these features
- Check if the model can reproduce the real system
- Predict future behaviour of the system through the model

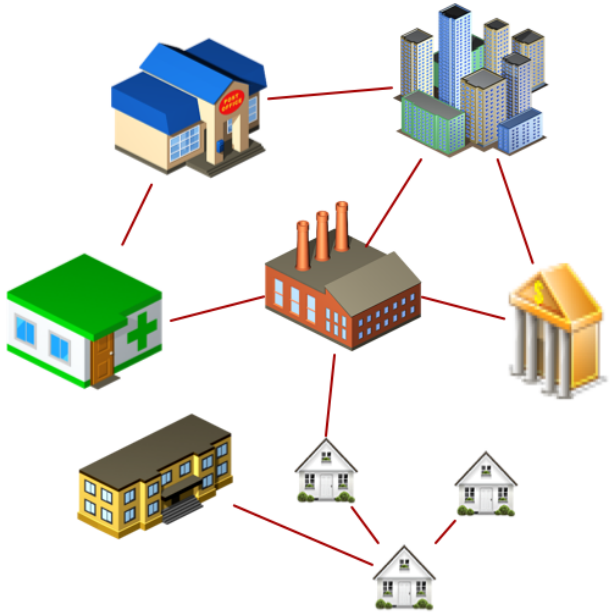
Hidden and Evident Hypotheses

Graphs connect

- part of cities across rivers
- buidings
- offices in the same building

Vertices are stable and edge creation has a finite and not negligible cost





History

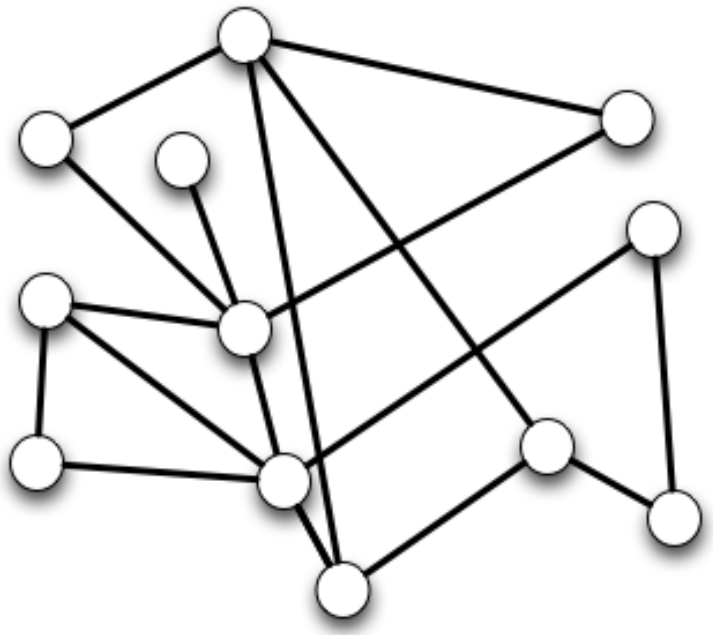
The main motivation in the creation of Random Graph theory was to provide

- a benchmark for the connection of various vertices
- in the case of connecting different buildings with costly phone lines

Definition

- Take a fixed number of vertices N
- no edge is present
- we draw a set of m edges out of the $N(N-1)/2$ available
- every edge is extracted with a fixed probability p

Such model is known as Random Graph model
[Erdős et al. 1959, Gilbert 1959].
No "particular" vertex can be found.



Common Definition

- Take N vertices
- For any couple of vertices draw a link with probability p

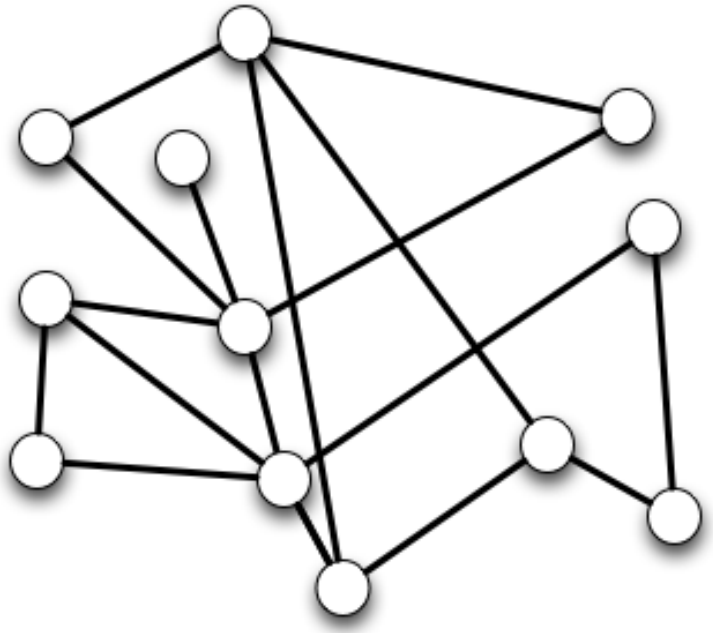
Expected value of Graph

The total number of edges m is a random variable with the expectation value E(m)=p[N(N-1)/2] .

If G_0 is a graph with N nodes and m edges, the probability of obtaining it by this graph construction process is
$P(G_0) = p^m(1-p)^{N(N-1)/2-m}$

First use

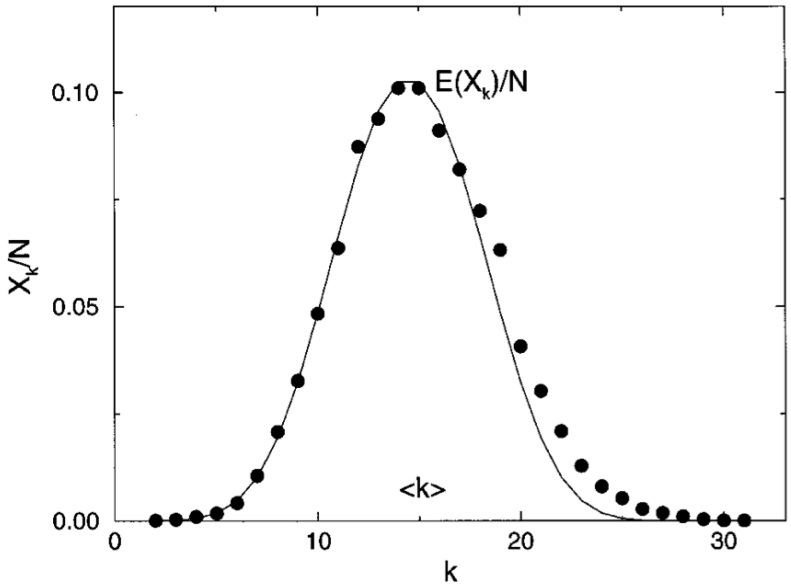
- a benchmark for the connection of various vertices
- in the case of connecting different buildings with costly phone lines



Degree Distribution

Similarly it is possible to determine the degree distribution[Bollobas 1985]. To have degree k

- an edge must be drawn k times $p^k(1-p)^{(N-1)-k}$
- this can happen in $\binom{N-1}{k} = \frac{(N-1)!}{(N-1-k)!k!}$ combinations



This distribution is automatically normalized since

$$\sum_{k=1,n-1} P_k = (p + (1-p))^{N-1} = 1.$$

Degree Distribution II

This distribution is usually approximated by means of the Poisson distribution in the two limits $N \to \infty$ and $p \to 0$ (when Np is kept constant and $N - 1 \simeq N$) we have:

$$P_k = \frac{N!}{(N-k)!k!} p^k (1-p)^{N-k} \simeq \frac{(Np)^k e^{-pN}}{k!}.$$

Since the mean value $\langle k \rangle$ of the above distribution is given by np we can write

$$P_k = \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}.$$

Degree Distribution III

- The above results are telling us that a characteristic degree exists
- This corresponds to the mean value $\langle k \rangle = Np$.
- Both larger and smaller values are less probable.
- On this respect the random graph model does not reproduce complex networks

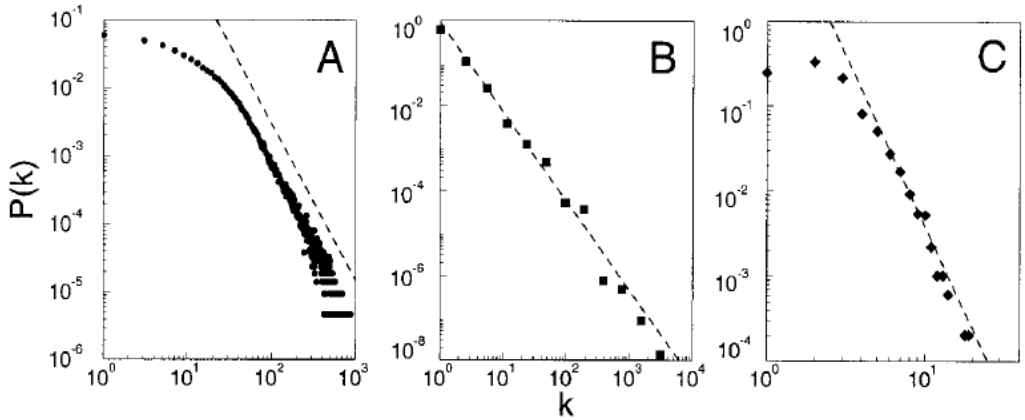
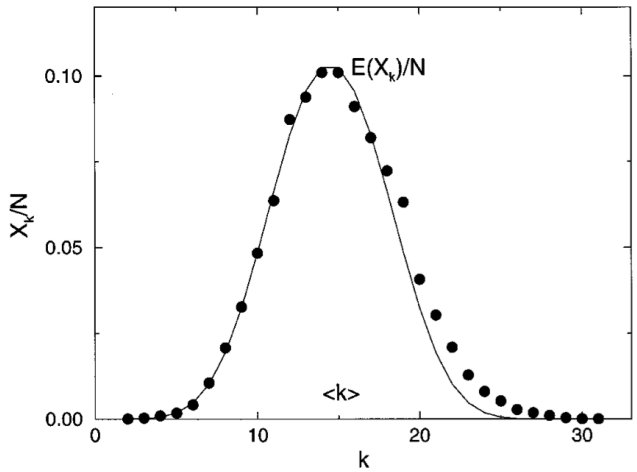


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (*6*). (**C**) Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.



Clustering

We can give an estimate of the Clustering Coefficient: for a complete graph it must be 1. If the graph is enough sparse then two points link each other with probability p

Expected value

$$E(C) \simeq p = \frac{\langle k \rangle}{N}$$

Diameter

Same estimate can be given for the average distance l between two vertices.If a graph has $\langle k \rangle$ average degree then

- the first neighbours will be $\langle k \rangle$
- the second neighbours will be at most $\langle k \rangle^2$
- the n-th neighbours will be at most $\langle k \rangle^n$
- For the Diameter D, we assume $\langle k \rangle^D$ of order N

Expected values

$$\langle l \rangle \leq D \simeq \frac{\log N}{\log k}$$

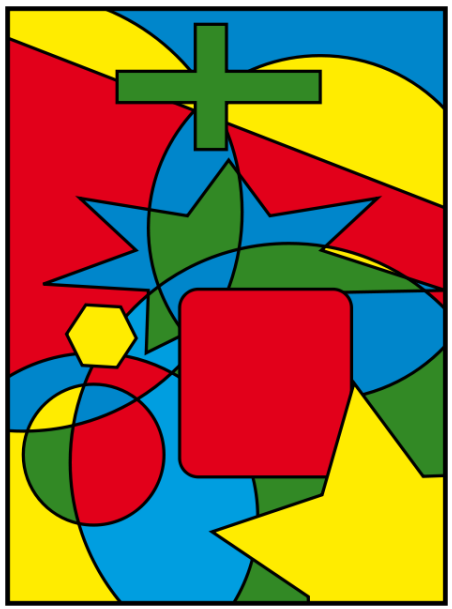
Connectedness

- If $\langle k \rangle = pN < 1$, a typical graph is composed of isolated trees and its diameter equals the diameter of a tree.
- If $\langle k \rangle > 1$, a giant cluster appears. The diameter of the graph equals the diameter of the giant cluster if $\langle k \rangle > 3.5$, and is proportional to $ln(N)/ln(\langle k \rangle)$.
- If $\langle k \rangle > ln(N)$, almost every graph is totally connected. The diameters of the graphs having the same N and $\langle k \rangle$ are concentrated on a few values around $ln(N)/ln(\langle k \rangle)$

Coloring of a map

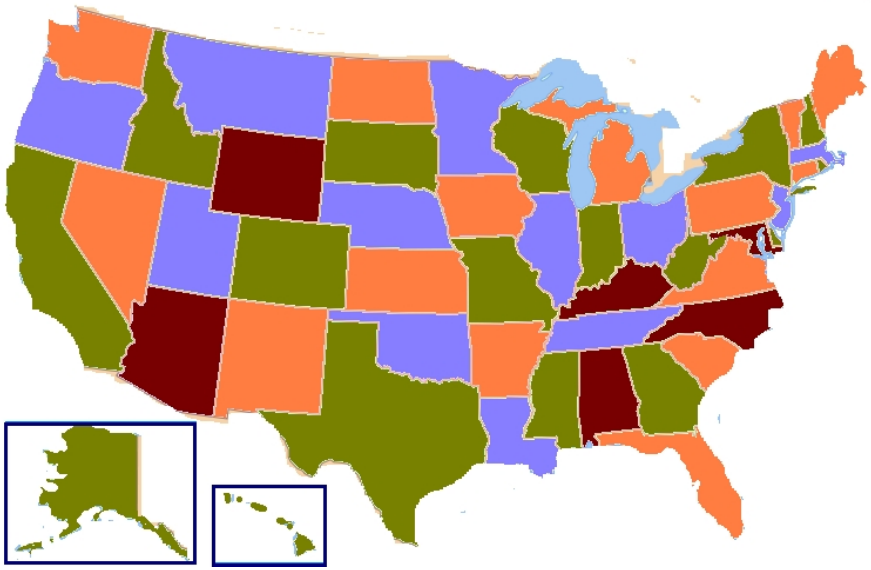
The theorem

Given any separation of a plane into contiguous regions, producing a figure called a map, no more than four colors are required to color the regions of the map so that no two adjacent regions have the same color.



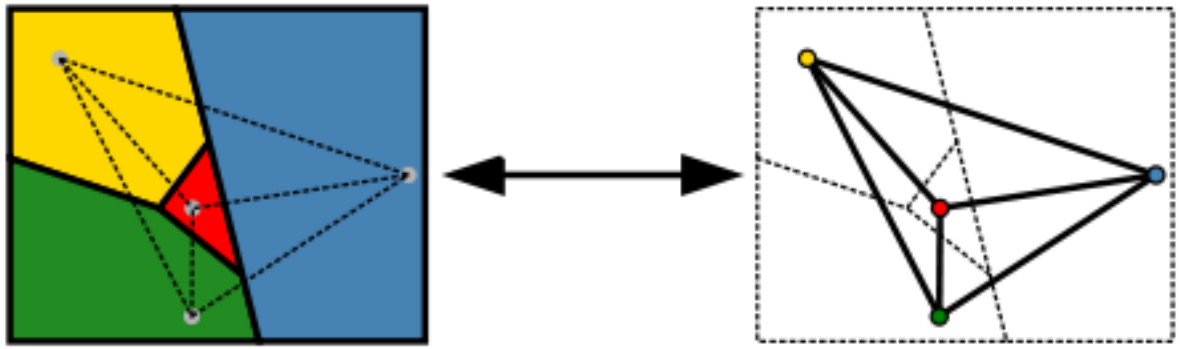
Counterexamples

Two regions are called adjacent if they share a common boundary that is not a corner, where corners are the points shared by three or more regions. For example, in the map of the United States of America, Utah and Arizona are adjacent, but Utah and New Mexico, which only share a point that also belongs to Arizona and Colorado, are not



Graph theory

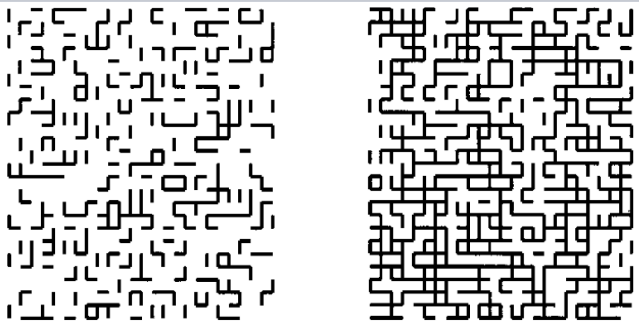
This problem can be easily visualized with planar graphs. The set of regions of a map can be represented more abstractly as an undirected graph that has a vertex for each region and an edge for every pair of regions that share a boundary segment



The Percolation model

Percolation

Sites (or bonds) of a lattice are chosen with probability p.
By varying p we have different clusters [Stauffer 2009].



- Bond percolation on a 2D latttice (25×25).
- Two nodes are connected by an edge with probability p.
- Two realizations: left p=0.315, right p=0.525

At $p = p_c = 0.5$, the bonds form a single cluster.
This value is indicated as *percolation threshold*.

The Percolation model

Percolation arise in a quantity of systems

- coffee (with percolator),
- water into rocks to extract oil (invasion percolation)
- certain types of fractures (mud cracking)
- networks (robustness to random and targeted attacks)
- wildfire propagation
- Epidemic spreading

how it is possible?

Universality

there are properties for a large class of systems that are independent of the dynamical details of the system. Systems display universality in a scaling limit, when a large number of interacting parts come together.

Percolation and Random Graphs

For $p < p_c = 1/N$

- The probability of a giant cluster in a graph, and of an infinite cluster in percolation, is equal to 0.
- The clusters of a random graph are trees, while the clusters in percolation have a fractal structure and a perimeter proportional with their volume.
- The largest cluster in a random graph is a tree with $\ln(N)$ nodes, while in general for percolation $P_p(|C| = s) \simeq e^{-s/\xi}$, suggesting that the size of the largest cluster scales as $\ln(N)$.

Percolation and Random Graphs

For $p = p_c = 1/N$

- A unique giant cluster or an infinite cluster appears.
- The size of the giant cluster is $N^{2/3}$ while for infinite dimensional percolation $P_p(|C| = s)\ s^{-3/2}$, thus the size of the largest cluster scales as $N^{2/3}$.

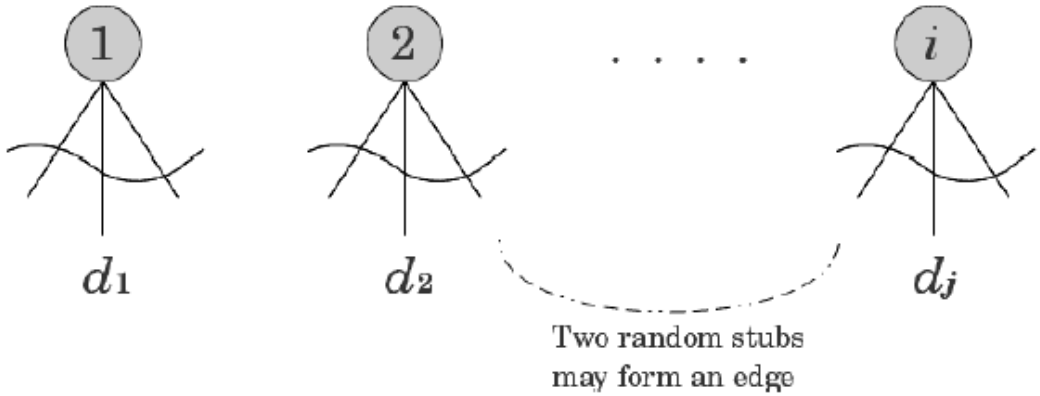
Percolation and Random Graphs

For $p > p_c = 1/N$

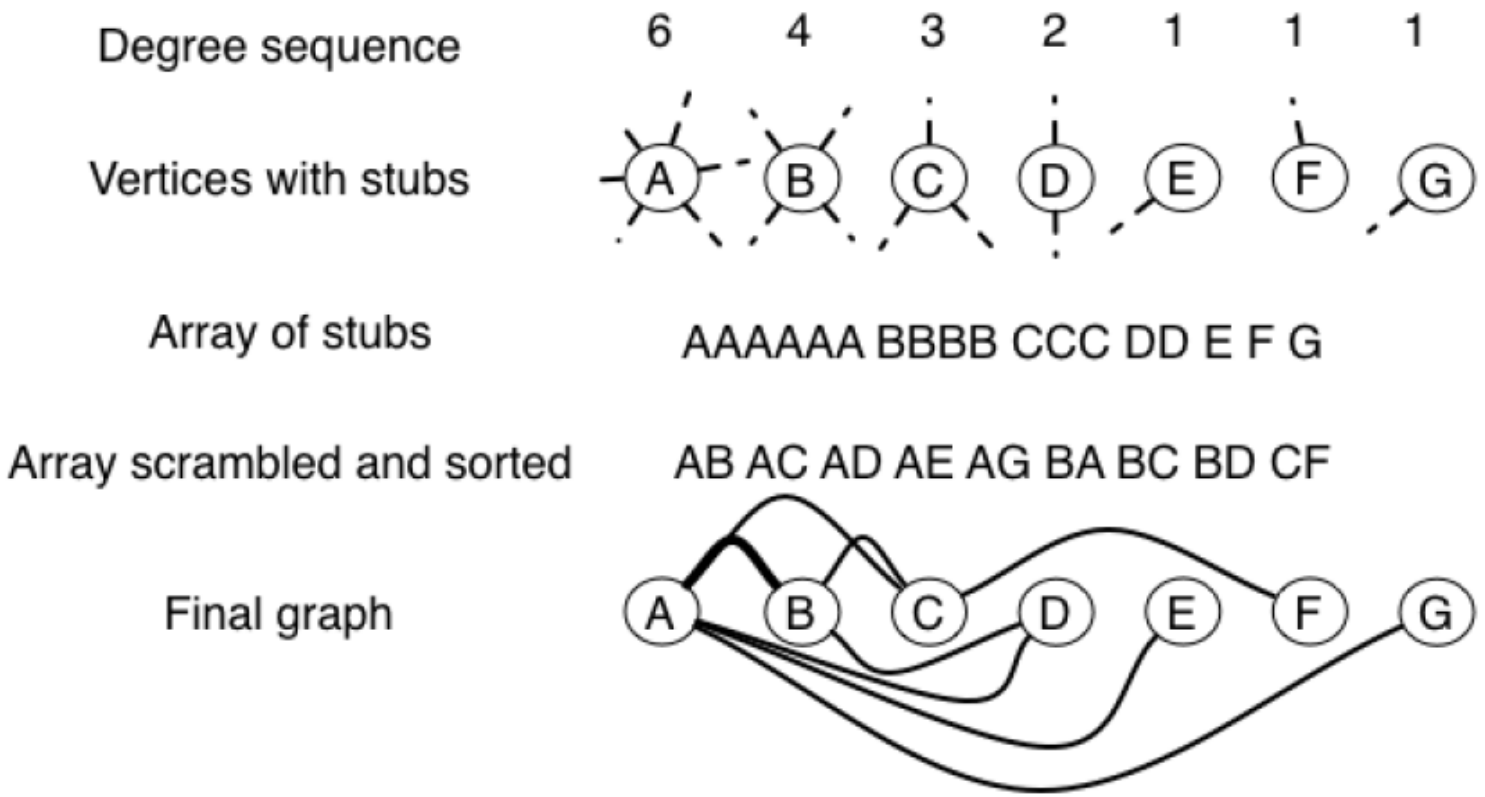
- The size of the giant cluster is $(f(p_c N) - f(pN))N$, where f is an exponentially decreasing function with $f(1) = 1$. The size of the infinite cluster is $\propto (p - p_c)N$.
- The giant cluster has a complex structure containing cycles, while the infinite cluster is no longer fractal, but compact.

Configuration model

- Let's start with the degree sequence.
- imagine that each node has edge "stubs" attached to it [Bender et al. 1978, Molloy et al. 1995].
- Edges are then assigned by randomly choosing two stubs and drawing an edge between them.



How to build the graph



As we see here, it happens that we end up with multiple edges

Probability of connections

Let k_i, k_j denote the non-zero degrees of two particular vertices i, j in a network of m edges.
For a particular stub attached to vertex i, there are k_j possible stubs, out of $2m - 1$ possible ones

probability that i and j are connected

is given by

$$\frac{k_i k_j}{2m-1} \simeq \frac{k_i k_j}{2m}$$

Number of multiple edges

The probability that a second edge appears between i, j is

$$\frac{(k_i - 1)(k_j - 1)}{2m}$$

Thus, the probability of both a first and a second edge is

$$\frac{k_i k_j (k_i - 1)(k_j - 1)}{(2m)^2}$$

.

We can now need obtain the number of multiple edges summing up on all the possible couples

Total multiple edges

$$
\begin{aligned}
\sum_{ij} \frac{k_i k_j}{2m} \frac{(k_i-1)(k_j-1)}{(2m)} &= \frac{1}{2}\frac{1}{(2m)^2}\sum_{i=1}^{n} k_i(k_i-1)\sum_{j=1}^{n} k_j(k_j-1) \\
&= \frac{1}{2}\frac{1}{\langle k\rangle^2 n^2}\sum_{i=1}^{n}(k_i^2-k_i)\sum_{j=1}^{n}(k_j^2-k_j) \\
&= \frac{1}{2}\frac{1}{\langle k\rangle^2}\left(\frac{1}{n}\sum_{i=1}^{n} k_i^2 - \frac{1}{n}\sum_{i=1}^{n} k_i\right)^2 \\
&= \frac{1}{2}\left[\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\right]^2 \qquad (1)
\end{aligned}
$$

Self-Loops

The number of self-loops can be computed similarly. We have that

- The number of pair of possible connections is $\binom{k_i}{2}$ and not $k_i k_j$.
- Thus, the probability of a self-loop is $p_{ii} = \frac{k_i(k_i-1)}{4m}$,
- The expected number of self-loops is (the constant)

$$\frac{\langle k^2 \rangle - \langle k \rangle}{2\langle k \rangle}$$

Just as with multi-edges, self-loops are a vanishingly small O(1/n) fraction of all edges in the large-n limit.

CM with expected degree

A generalization consists in considering the expected degree sequence and not the actual one.

Chung-Lu model

Every node i has an expected degree w_i, each possible edge exists independently with probability $p_{ij} = \frac{w_i w_j}{\sum_k w_k}$

the expected degree of a node is given by

$$\langle k \rangle = \sum_j p_{ij} = \frac{w_i \sum_j w_j}{\sum_k w_k} = w_i$$

Simple case ERGM

Let us consider a simple case. The only observable is the number of edges E

- $H(G) = \theta E(G)$
- The partition function is

$$\begin{aligned} Z &= \sum_{G \in \mathcal{G}} e^{\theta E(G)} \\ &= \sum_{G \in \mathcal{G}} \prod_{i=1,n} \prod_{j=i+1,n} e^{\theta A_{ij}(G)} \quad (A_{ij}(G) = 0, 1) \\ &= \prod_{i=1,n} \prod_{j=i+1,n} (1 + e^{\theta}) = (1 + e^{\theta})^{\binom{n}{2}} \end{aligned}$$

ERGM and RG

We can now compute the probability to observe a graph with E edges

$$P(G) = \frac{e^{H(G)}}{Z} = e^{\theta E}(1+e^{\theta})^{\binom{n}{2}} = \left(\frac{e^{\theta}}{1+e^{\theta}}\right)^{E}\left(1-\frac{e^{\theta}}{1+e^{\theta}}\right)^{\binom{n}{2}-E}$$

but also

$$P(G) = p^{E}(1-p)^{\binom{n}{2}-E}$$

from which we recognise that the two coincide if $p = \frac{e^{\theta}}{1+e^{\theta}}$

Conclusion

Random Graph

- Do **not** reproduce the degree distribution
- do reproduce the distance distribution
- are **less** clustered
- are **more** robust to target attack

than real networks

Small-World Definition

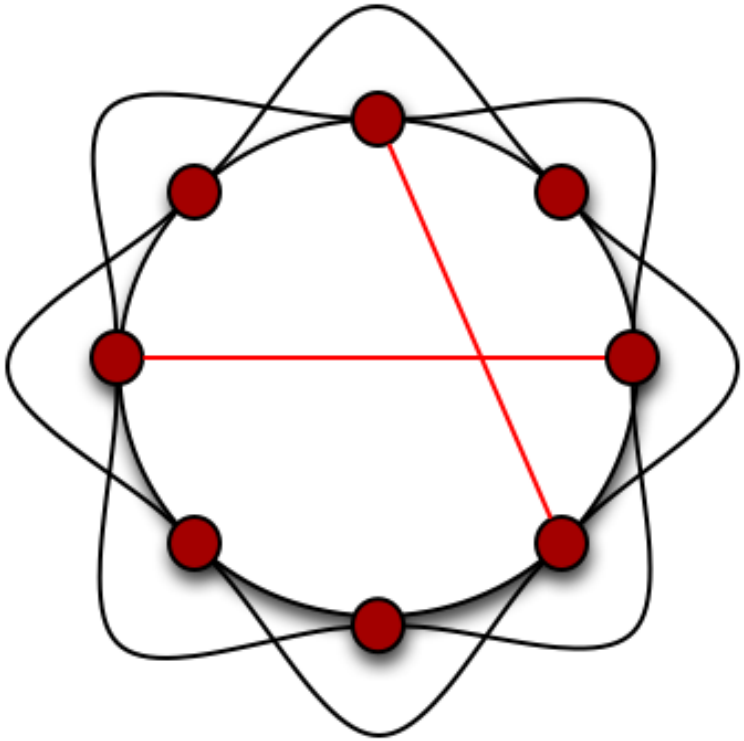
small-world effect

The small-world model explains why the diameter of real graphs can remain very small when the number of vertices increases (small-world effect).

We have seen in the previous section that in the random graph model the diameter increases logarithmically with respect to the number of vertices. This is a common feature in most if not all graph models.

Model Definition

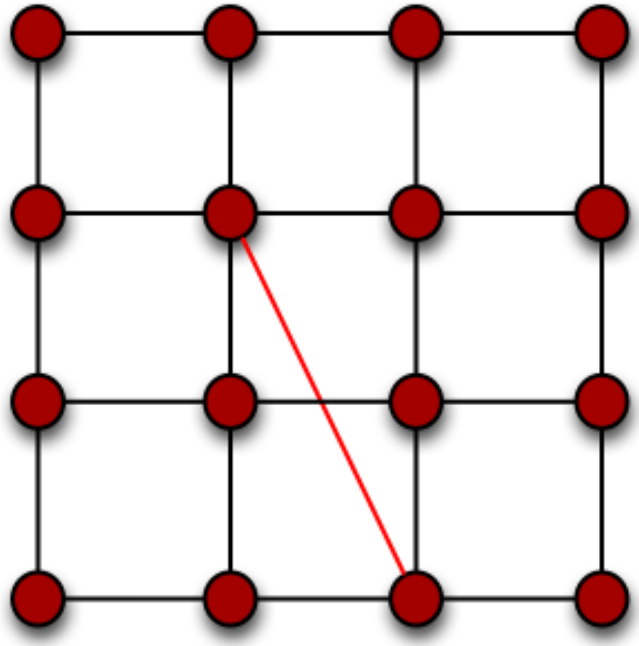
- start with a portion of an ordered grid;
- vertices at one (and two) grid units are connected;
- they form the set of the first neighbours;
- Add shortcuts with probability p.



Basic Ingredients

Basic idea

On top of every-day links, random connections are also established with probability p between vertices.



Model Parameters

Tunable quantities

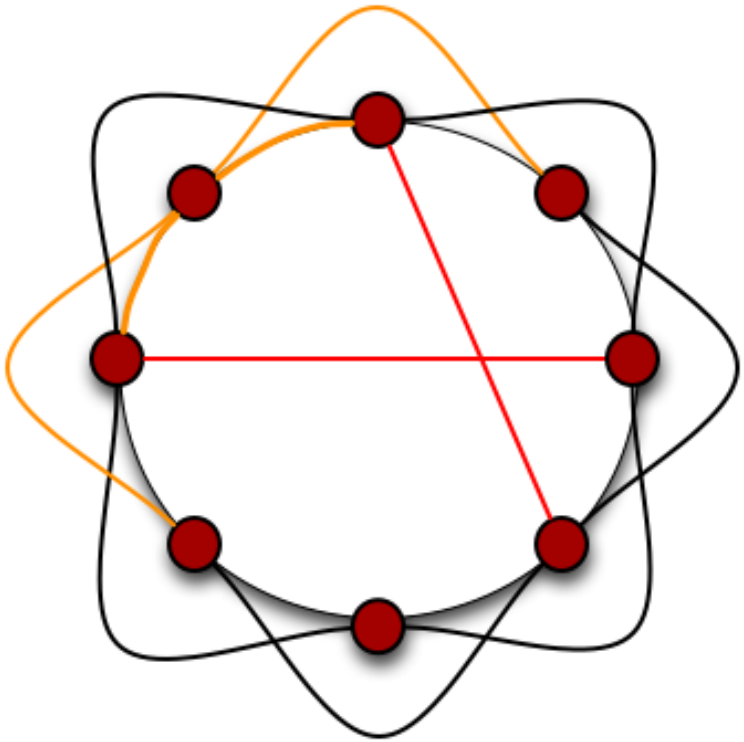
There are two main quantities that can be changed in the model

- The **coordination number z** that gives the number of vertices directly connected in the regular structure.
- The **probability of rewiring p** that gives the probability per existing edge to draw a new edge (shortcut) between two random vertices.

Coordination number

In a one-dimensional ($d = 1$) system with $j = 2$ connectivity every vertex has $z = 4$ connections with other vertices (two from one side and two from the other). This number of connections also grows with the dimensionality. In general we can write

$$z = 2jd.$$

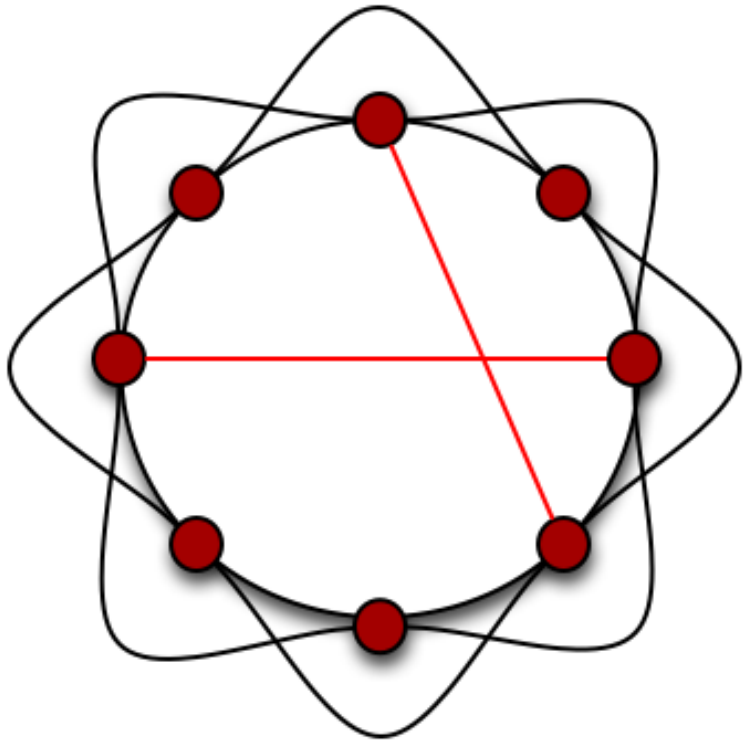


Shortcuts Probability

If p is the probability to draw a shortcut, the expected value of total number of shortcuts is

$$mp = nzp/2.$$

To remove the 2 in this formula, we can define the coordination number as $z' = z/2$. In this way the total number of shortcuts becomes $nz'p$.



The Length Distribution

- On a regular grid the average distance grows with the number of vertices N
- In small world model the shortcuts keep distances small
- Using numerical simulation we can compute the variation on the **diameter**. Take $N = 1,000$ vertices ($d = 1$), a coordination number $z = 10$,
  - with a rewiring probability $p = 1/4 = 0.25$ we have a diameter as small as $d = 3.6$.
  - with p as small as $p = 1/64 = 0.015625$ we still find a small diameter $d = 7.6$.
- With no rewiring at all, the diameter of the same system is $d = 50$.

Phase transition

It has been proposed an analytical expression for the mean distance l

$$l = \frac{n}{z'} f(npz')$$

where $z' = z/2$ and the function $f(x)$ is

$$f(x) = \frac{1}{2\sqrt{x^2 + 2x}} \tanh^{-1} \frac{x}{\sqrt{x^2 + 2x}}.$$

Clustering Coefficient

The **clustering coefficient** of the whole network is usually very high and it is reminiscent of the regular connection of the underlying grid. As long as z stays reasonably small and in particular $z < \frac{2}{3}n$ (as is the case when $n \to \infty$), we have:
For the original formulation (with rewiring)

$$C = \frac{3(z-1)}{2(2z-1)}(1-p)^3$$

while for the formulation without rewiring

$$C = \frac{3(z-1)}{2(2z-1)+4zp(p+2)}.$$

The Degree Distribution

- We have that the **degree distribution** is a function peaked around the fixed value z characteristic of the regular grid.
- With no shortcuts, the distribution is not even a regular function, but it is zero elsewhere apart from z (it is a delta function different from zero only in z and zero otherwise).
- When shortcuts are many and there is no more underlying grid we must expect a behaviour similar to that of random graph.

The Degree Distribution

- We have that the **degree distribution** is a function peaked around the fixed value z characteristic of the regular grid.
- With no shortcuts, the distribution is not even a regular function, but it is zero elsewhere apart from z (it is a delta function different from zero only in z and zero otherwise).
- When shortcuts are many and there is no more underlying grid we must expect a behaviour similar to that of random graph.

Motivation

Growth

The Barabási-Albert model wants to reproduce the time growth of many real networks (e.g. Internet and WWW)

To reproduce this feature the graph is built through successive time-steps when new vertices are added to the system. Also the number of edges increases time, since the new vertices connect to the old ones.

Preferential Attachment

The vertices destination (those already present) are chosen with a probability that is proportional to their degree at the moment.

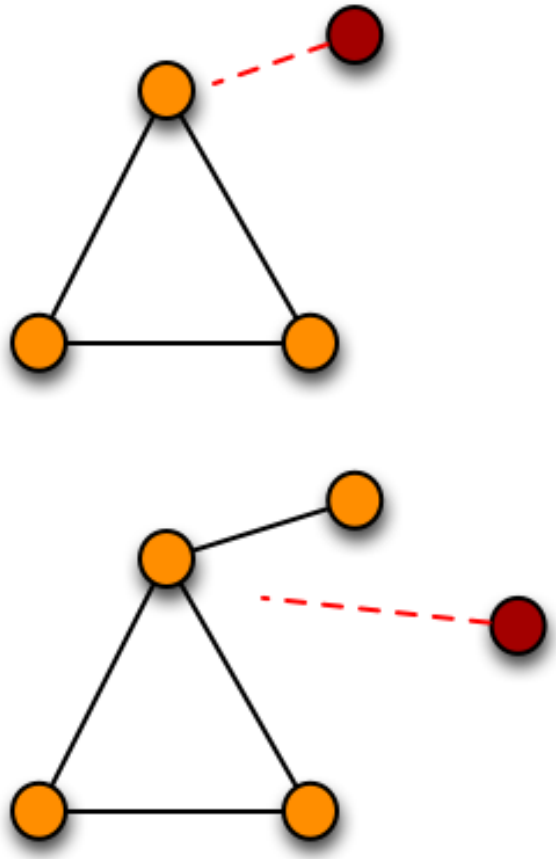
The two ingredients

Growth implies that new vertices enter the network at some rate. Preferential attachment means that these newcomers establish their connections preferentially with vertices that already have a large degree (*rich-get-richer*). This latter rule is in the spirit of the Matthew effect More quantitatively, this model can be reconnected to a Yule process described in the same section. Growth and preferential attachment are specifically suited to model the Internet and the World Wide Web (though the latter is directed while the former is not), two networks that in a relatively short timespan (fifteen to twenty years) have seen a huge growth of their elements.

The rules for the construction

1. We start with a disconnected set of n_0 vertices (no edges are present).
2. New vertices enter the system at any time step. For any new vertex m_0 new edges are drawn.
3. The m_0 new edges connect the newcomers' vertices with the old ones. The latter are extracted with a probability $\Pi(k_i)$ proportional to their degree, that is

$$\Pi(k_i) = \frac{k_i}{\sum_{j=1,n} k_j}.$$



The rules for the construction

Note that, since at every time step only one vertex enters, we have for the order and the size of the network respectively

$$\begin{aligned} n &= n_0 + t \\ m &= 1/2 \sum_{j=1,n} k_i = m_0 t. \end{aligned} \tag{2}$$

Growth

Analytical indications

Here we consider the degree as a continuous variable. New vertices enter the network at a constant rate.

At time t the old ones are $n = n_0 + t - 1$. The first quantity we can derive is the variation of the degree with time.

$$\frac{\partial k_i}{\partial t} = A\Pi(k) = A\frac{k_i}{\sum_{j=1,n} k_j} = \frac{Ak_i}{2m_0 t}.$$

The constant A is the change of connectivity in one time step, therefore $A = m_0$. Since at initial time t_i the initial degree is $k(t_i) = m_0$ we have

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \rightarrow k_i(t) = m_0 \left(\frac{t}{t_i}\right)^{1/2}.$$

Degree Distribution I

Analytical indications

This simple computation shows that in a Barabási-Albert model the degree grows with the square root of time. This relation allows us to compute the exponent of the degree distribution:

The probability $P(k_i < k)$ that a vertex has a degree lower than k is $P(k_i < k) = P(t_i > \frac{m_0^2 t}{k^2})$. Since vertices enter at a constant rate, their distribution is uniform in time, that is $P(t) = A$, where A is a constant. The value of A can be determined by imposing normalization of the distribution. This means $\int_0^n A = 1$, which gives

$$A = P(t) = 1/n = 1/(n_0 + t).$$

Degree Distribution II

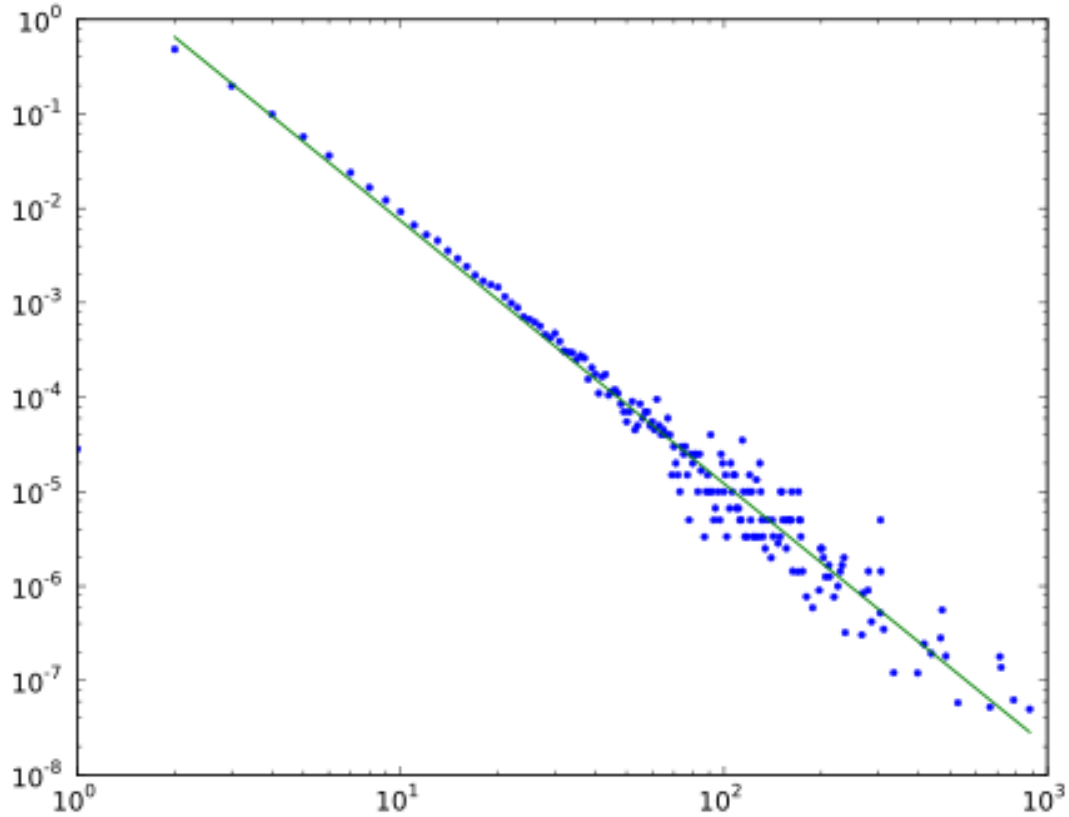
In this way, we can write

$$P(t_i > \frac{m_0^2 t}{k^2}) = 1 - P(t_i \leq \frac{m_0^2 t}{k^2}) = 1 - \frac{m_0^2 t}{k^2} \frac{1}{(n_0 + t)}$$

from which we have

$$P(k) = \frac{\partial P(k_i > k)}{\partial k} = \frac{2 m_0^2 t}{(n_0 + t)} \frac{1}{k^3}.$$

Therefore, we find that the degree distribution is a power law with a value of the exponent $\gamma = 3$.



Plot of the degree distribution for a Barabási-Albert model

Properties of the Barabási-Albert

For this model some results have been obtained:

- The **degree distribution** is scale invariant only if the preferential attachment rule is perfectly linear; otherwise the degree is distributed according to a stretched exponential function.
- As regards the **diameter** D of Barabási-Albert networks, an analytical computation shows that $D \propto \ln(n)/\ln(\ln(n))$.
- The **clustering coefficient** of a Barabási-Albert model is five times larger than those of a random graph with comparable size and order. It decreases with the network order (number of vertices). Some analytical results are available in the particular limit of large and dense graphs.

1st change: growth of edges

Edges Growth

Not only the vertices but also the edges can 'grow'. In particular, we can allow new edges to be added between existing vertices.

The motivation of this model was to provide a more realistic model for study of the World Wide Web. Indeed in this specific network (as turned out to be the case also for Wikipedia) most of the modifications are addition or rewiring of edges. In the model, the edges are directed, therefore every vertex i is determined by both the in-degree k_i^{in} and out-degree k_i^{out}.

Rules

1. With probability p *a new vertex is added to the system*. Edges are drawn according to the preferential attachment rule. The key quantity is the in-degree of the target vertex j. In this case the preferential attachment probability is given by $\Pi(k_j^{in}) = (k_j^{in} + \lambda)$
2. With probability $q = (1 - p)$ *a new directed edge is added to the system*. The choice of the end vertices depends upon the out-degree k_i^{out} of the originating vertex i and the in-degree k_j^{in} of the target vertex j. This creation function is assumed to be of the form

$$C(k_i^{out}, k_j^{in}) = (k_j^{in} + \lambda)(k_i^{out} + \mu).$$

Form of Distribution

It is possible to derive analytically the form of the two distributions $P(k^{in})$ and $P(k^{out})$, they are

$$\begin{aligned} P(k^{in}) \propto k^{-\gamma^{in}} &\rightarrow \gamma^{in} = 2 + p\lambda, \\ P(k^{out}) \propto k^{-\gamma^{out}} &\rightarrow \gamma^{out} = 1 + \frac{1}{q} + \frac{\mu p}{q}. \end{aligned}$$

Form of Distribution

Motivation

Actors can retire or die and do not attract any more edge. Similar consideration apply to the networks of scientific citations.

These effects can be put into the model by introducing an ageing effect. Vertices in the network can be either *active* of *inactive*. In the first state they can still receive edges and modify their state. Otherwise their dynamics is frozen and they no longer take part in the evolution of the system. At any time step, the number m of active vertices is kept constant.

Rules

1. growth mechanism remains and new vertices enter the system at any time step. Newcomers are always in the active state.
2. A number m_0 of new edges are drawn between the newcomer vertex and *every one* of the active vertices.
3. One vertex i is selected from the set of active ones. This vertex is deactivated and removed from the evolution of the system. This happens with a probability

$$P_i^{deact} = \frac{1}{N}\frac{1}{(k_i + a)} = \frac{1}{\sum_{j=1,N_a}(k_j + a)^{-1}}\frac{1}{(k_i + a)}.$$

Where k_i is the degree of vertex i, a is a constant, and $1/N$ is the normalization constant given by $1/N = 1/\sum_{j=1,N_a}(k_j - a)^{-1}$.

Results

The degree distribution can computed and it is still a power law $P(k) \propto (k + a)^{-\gamma}$.
The clustering coefficient of this model is larger than that of random graphs and fits nicely the data of some real networks. An analytical estimate gives the value $C = 5/6$ while from computer simulations we find $C = 0.83$.

Motivations

An example of how a specific case study could inspire the definition of a network model is again given by the World Wide Web. If you want to add your web page to the system (i.e. add a vertex and some edges to the graph) one common procedure is to take one template (a page that you like) and to modify it a little bit. In this way most of the old hyperlinks are kept
The same mechanism is in agreement with the current view of genome evolution. When organisms reproduce, the duplication of their DNA is accompanied by mutations. Those mutations can sometimes entail a complete duplication of a gene. A protein can now be produced by two different copies of the same gene; this means that point-like mutations on one of them can accumulate at a rate faster than normal, since a weaker selection pressure is applied.

Results

The rate of change of the in-degree of a node is then given by

$$\frac{\partial k_{in,i}(t)}{\partial t} = (1-\alpha)\frac{k_{in,i}(t)}{n} + m_0\frac{\alpha}{n} \tag{3}$$

where the first term on the right-hand side of eqn 3 is the probability that a vertex pointing to vertex i is duplicated and its edges toward i retained. The second term on the right-hand side represents the probability that the duplicated vertex points toward i by one of its rewired out-going edges. For linearly growing networks we have that $n \simeq t$. The solution of eqn 3 is

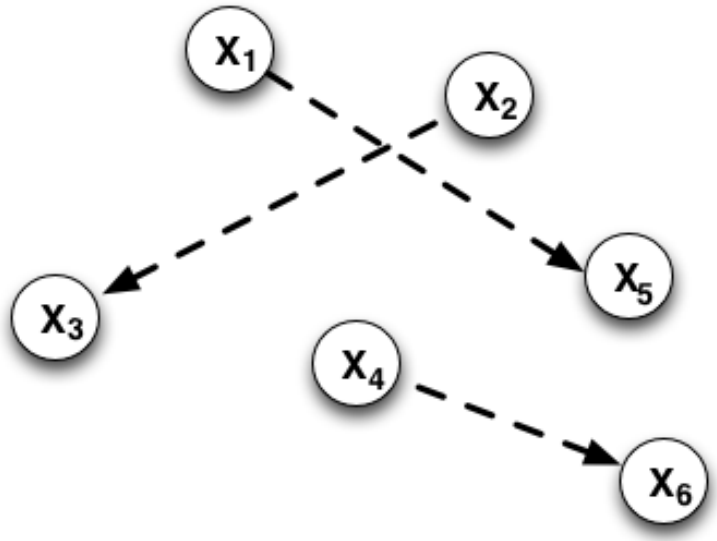
$$k_i^{in}(t) = \frac{m_0\alpha}{1-\alpha}\left[\left(\frac{t}{t_i}\right)^{1-\alpha} - 1\right] \tag{4}$$

Motivation

FItness Model

Not necessarily all the vertices are created equal. Likely this affects the network creation

We must assign a scalar quantity (indicated by η_i or x_i) for every vertex and modify the models accordingly



Definition of Bianconi-Barabási model

- We start with n_0 different vertices characterised by a constant ability (*fitness*) η_i to attract new edges. The η_i are extracted from a probability distribution $\rho(\eta)$.
- The growth remains with new vertices entering the system with their new fitnesses η_i.
- The preferential attachment is slightly modified, taking into account the fitnesses. The edges are drawn towards the old vertices with a probability $\Pi(k_i, \eta_i)$

$$\Pi(k_i, \eta_i) = \frac{\eta_i k_i}{\sum_{j=1,n} \eta_j k_j}.$$

Model Definition

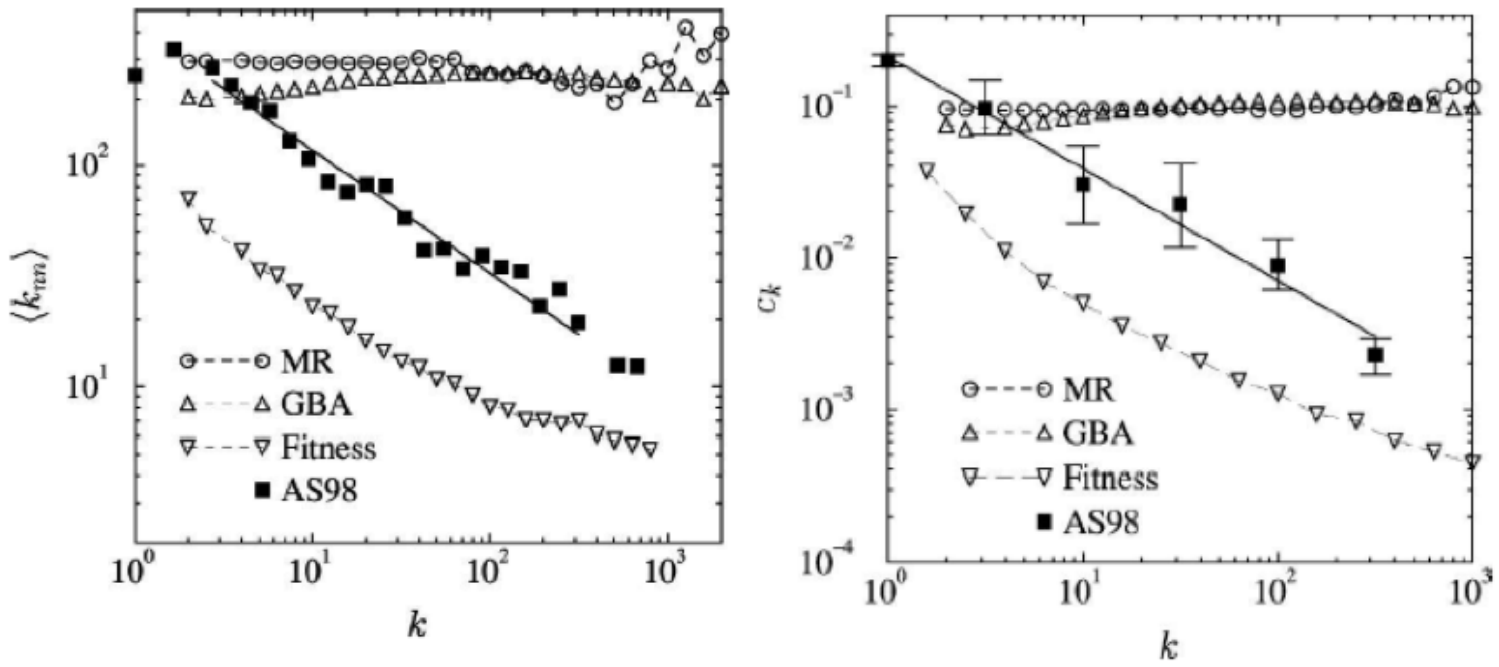
It is possible to derive analytically the form of the degree distribution that is now dependent upon the form of the fitness distribution $\rho(\eta)$. In the case of a uniform distribution (i.e. $\rho(\eta)$ constant), we have that

$$P(k) \propto \frac{k^{-(1+C^*)}}{\ln(k)} \tag{5}$$

where $C^* = 1.255$ is a constant whose value is determined numerically.

Fitness

While no particular result is known for the clustering, this model develops non-trivial disassortative properties that make it a very good model to reproduce Internet autonomous systems properties.



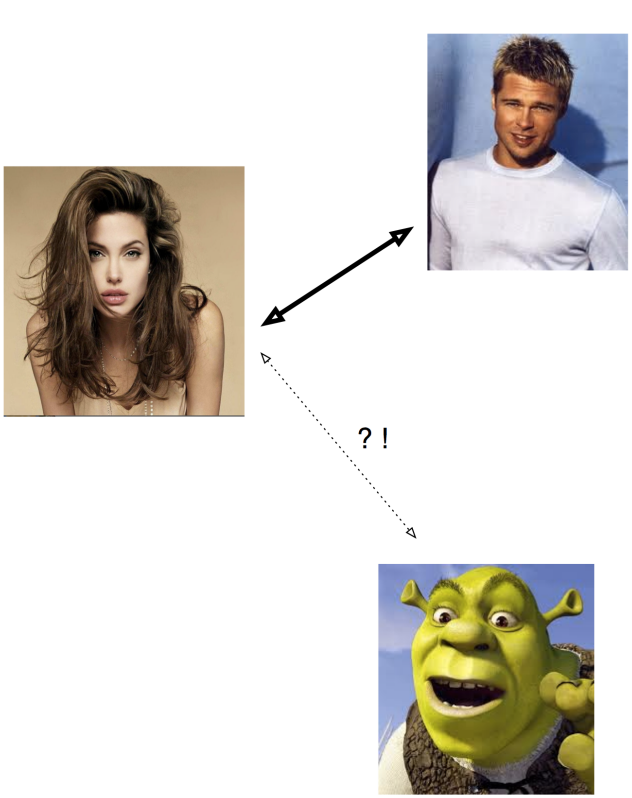
Beyond Preferential Attachment

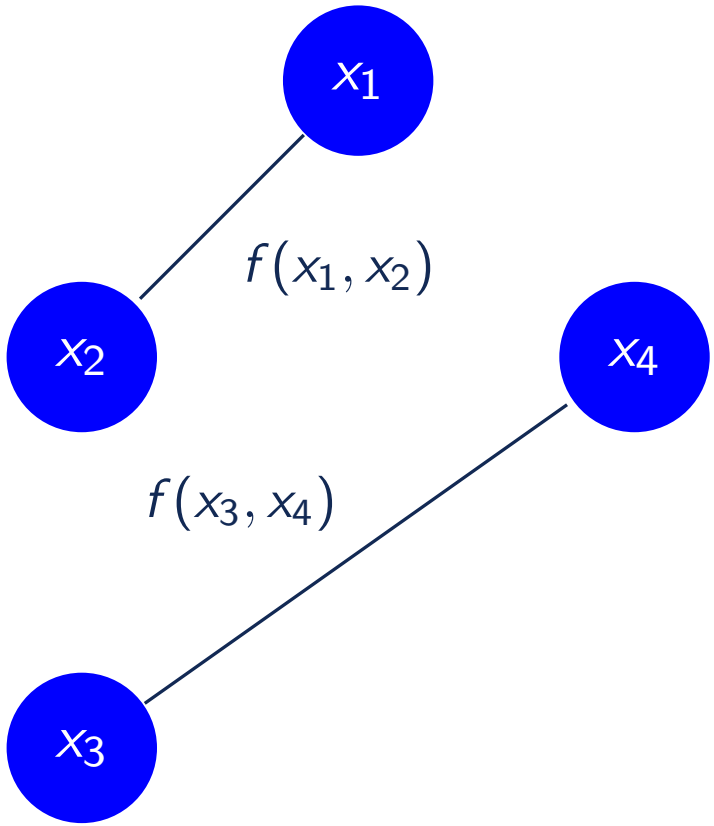
Although in some contexts preferential attachment can be a very reasonable assumption, in many others it is certainly not. Instead, it is reasonable to think that two vertices become connected when the edge creates a mutual benefit.
This benefit depends on some intrinsic properties (authoritativeness, friendship, social success, scientific relevance, interaction strength, etc) of the vertices.

Fitness Model

Basic principles

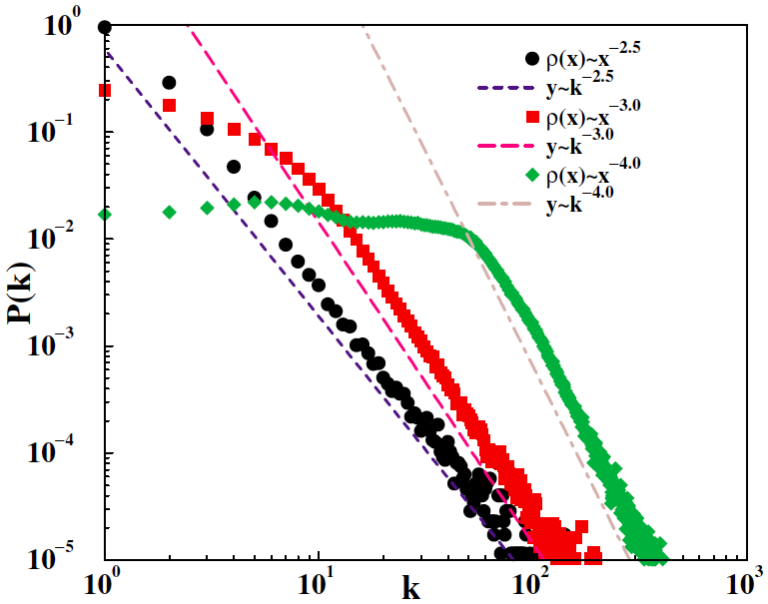
Vertices have state variable **(fitness)**
Edges drawn with **(fitness-dependent) probabilities**

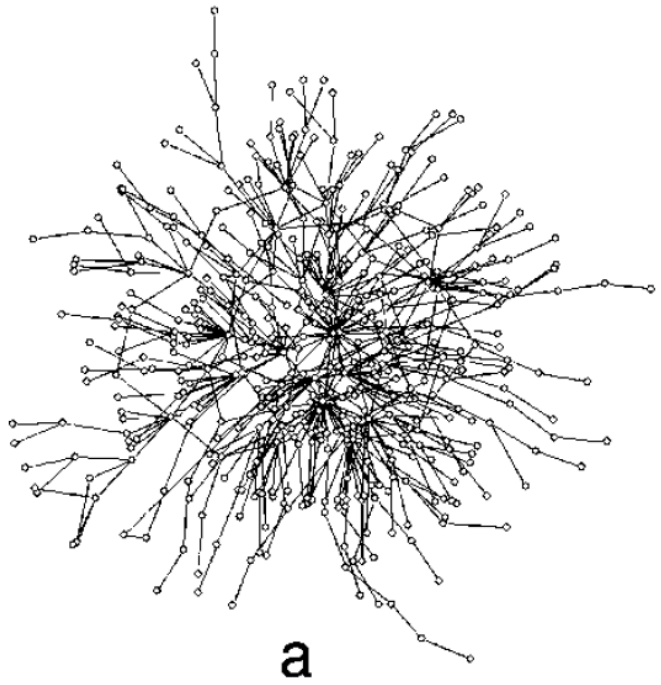




Fitness Models

This model is based on a modification of Random Graphs.
Vertices differ, edges are **not** equally likely[1] [2]





$P(k) = Ak^{-\gamma}$ for a variety of choices

Definition

- Start with n vertices. For every vertex i draw a real number x_i representing the fitness of the vertex. Fitnesses are supposed to measure the importance or rank of the vertex in the graph and they are extracted from a given probability distribution $\rho(x)$.
- For every couple of vertices, i, j, we can draw an edge with a probability given by the *linking function* $f(x_i, x_j)$ depending on the fitnesses of the vertices involved. If the network is not directed the function f is symmetric that is $f(x_i, x_j) = f(x_j, x_i)$.

Limit case

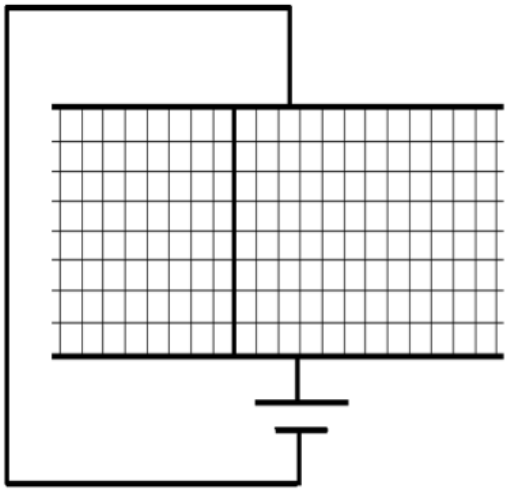
A trivial realization of the above rules is the model of Erdős and Rényi. In this case the $f(x_i, x_j)$ is constant and equal to p for all vertex couples. While this particular choice does not produce scale-free networks, as soon as random fitnesses are introduced, the situation changes completely.

static vs dynamic ?

This model can be considered static as well as dynamic. If the size of the graph is fixed, one checks all the possible couples of vertices as in the random graph model. Otherwise by adding new vertices at every time step, one can connect the new ones to the old ones. A general expression for the **degree distribution** $P(k)$ can be derived easily.

Physical Meaning

Without introducing growth or preferential attachment we can have power-laws We consider "disorder" in the Random Graph model (i.e. vertices differ one from the other).
This mechanism is responsible of self-similarity in Laplacian Fractals



Formulas

Parameters of the model

- $\rho(y)$ from which we extract fitnesses
- $f(x, y)$ to draw edges

For any choice

$$k(x) = N \int_0^\infty f(x, y)\rho(y)dy = NF(x)$$

Under suitable conditions we can write

$$P(k) = \rho\left[F^{-1}\left(\frac{k}{N}\right)\right]\frac{d}{dk}F^{-1}\left(\frac{k}{N}\right)$$

Example

As a particular example, consider $f(x, y) \propto xy$. Then

$$k(x) = ANx \int_0^\infty y\rho(y)dy = ANx\langle x \rangle$$

and we have the simple relation,

$$P(k) = \frac{1}{NA}\rho(\frac{k}{NA})$$

Whenever $\rho(x)$ is power law, we have $P(k)$ power law.

Note that...

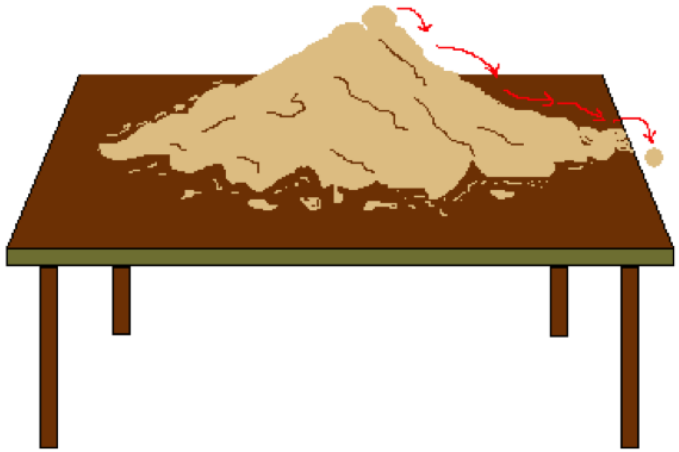
Power laws arise spontaneously in other cases.

SOC and Networks

Networks can arise from Self-Organised Processes

We start from a graph with arbitrary degree sequence We define a "local " (for the sites) rule of update We find a steady state characterised by power-law distribution of the degree This mechanism can maybe explain the onset of most of the observed Pareto's law in nature and consequently explain the ubiquity of scale-free networks. The behaviour of the model can be understood in terms of the Bak-Sneppen model of Self-Organised Criticality

Here we focus on the case when the two processes evolve over comparable timescales, by considering the interplay between topology and dynamics

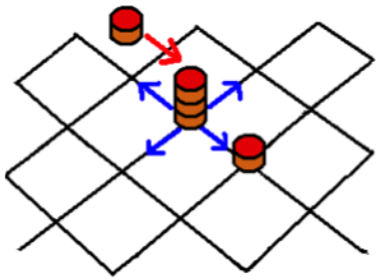


As a result, the process is self-organized and a non-equilibrium stationary state is reached, independently of (otherwise arbitrary) initial conditions

SOC Models

BTW Sandpile Model:

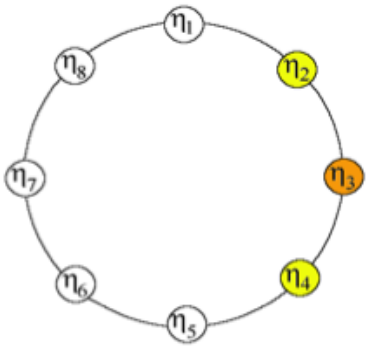
sand is added on the sites of a lattice. At a critical threshold, the site topples on the neighbours triggering other topplings



P. Bak, C. Tang, K. Weisenfeld PRL 52, 1033 (1984).

BS Bak and Sneppen Model:

A system of species i characterized by a fitness hi. Recursively the species with the minimum fitness and its neighbours are removed and changed with three new ones with random η_i



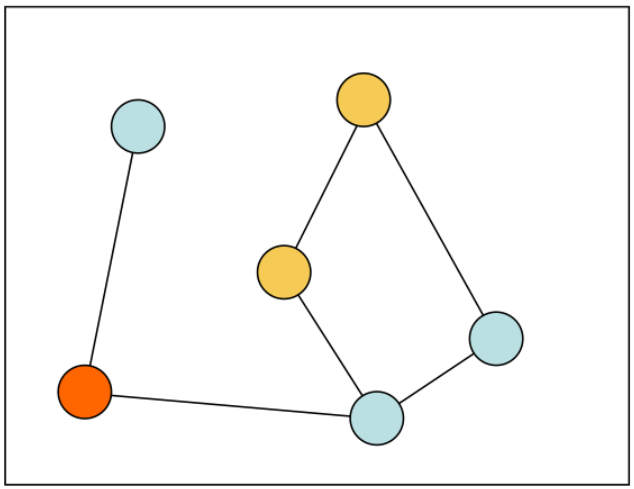
P. Bak, K. Sneppen PRL 71, 4083 (1993).

IP Invasion Percolation:

A fluid (water) is injected in a porous medium to extract oil. Amongst the different channels on the boundaries the one with the minimum diameter is selected to be invaded.

D. Wilkinson and J. F. Willemsen, J. Phys. A (London) 16, 3365 (1983).

Definitions



- We start from a graph with fitnesses on sites INITIAL DISTRIBUTION DOES NOT MATTER
- We select the minimum fitness and remove this site and all its neighbours
- We repeat the procedure many times ($> 10^5$)

The system approaches a steady state, where

- The fitness distribution is a power law
- The degree distribution is a power law

Definitions 2

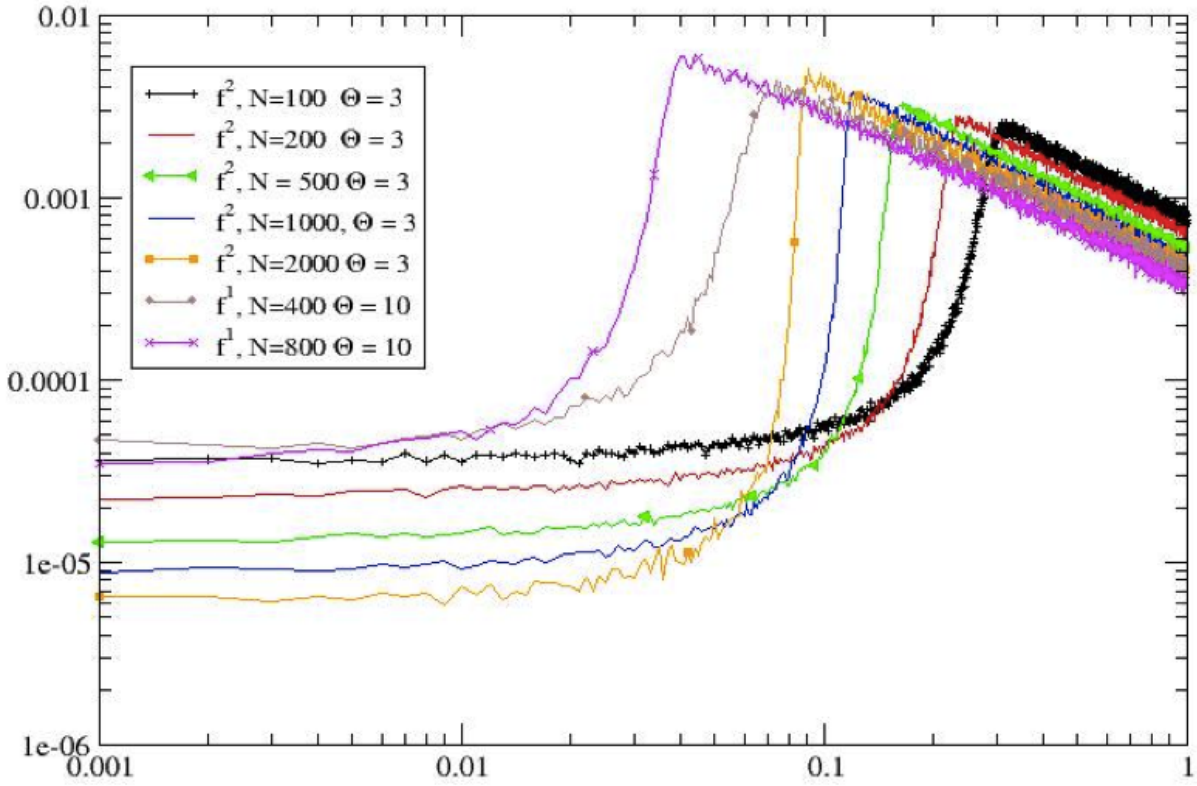
The linking probabilities that we used are

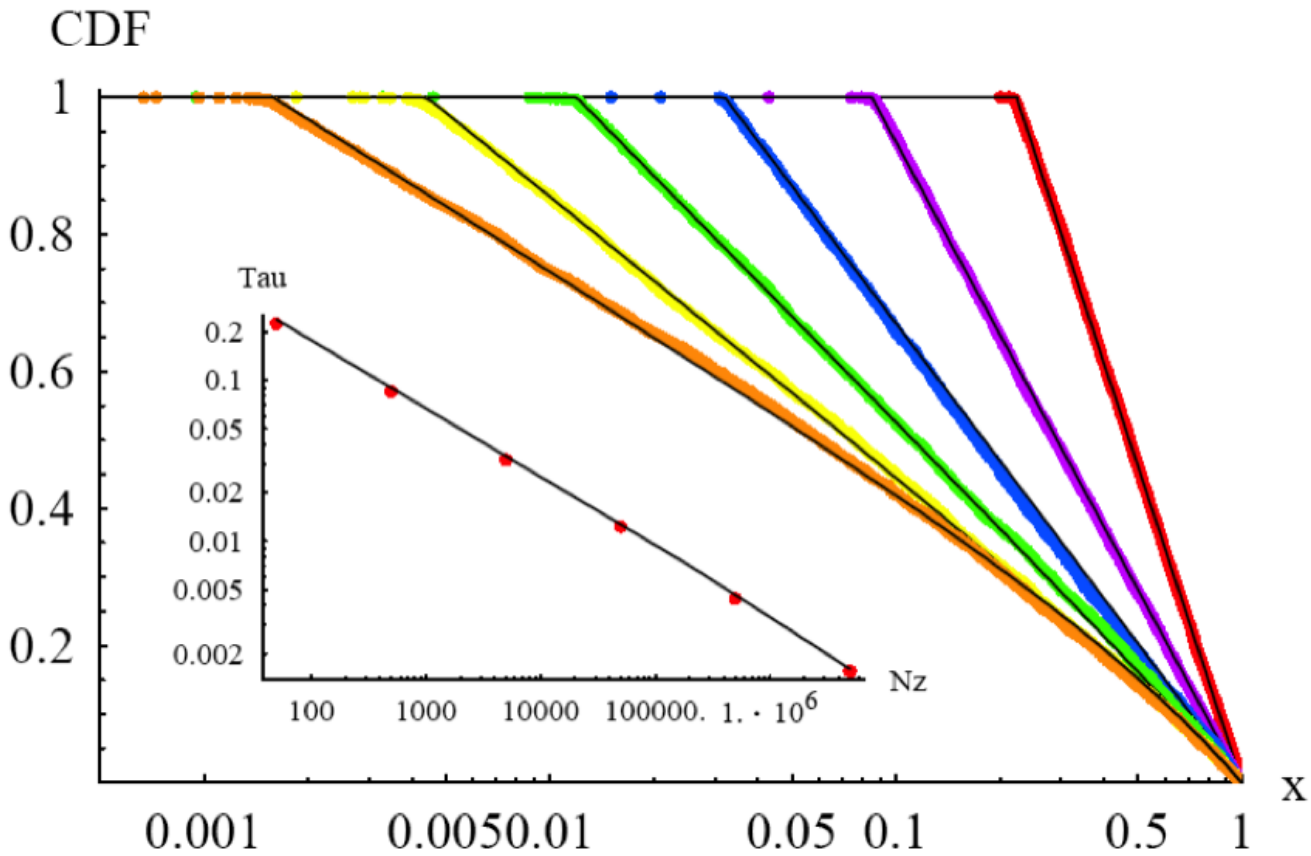
- $f_1(x_i, x_j) \propto z(x_i + x_j)$
- $f_2(x_i, x_j) \propto z x_i x_j$
- $f_3(x_i, x_j) \propto z x_i x_j 1 + z x_i x_j$

The fitness refreshment rule we used for the neighbors are

- $x_i^{new} = \eta$
- $x_i^{new} = \frac{1}{k_i} \left[(k_i - 1)\, x^{old} + \eta \right]$

where $(\eta \in [0, 1])$


0.01
0.001
0.0001
1e-05
1e-06
0.001
0.01
0.1
1
f², N=100 Θ = 3
f², N=200 Θ = 3
f², N = 500 Θ = 3
f², N=1000, Θ = 3
f², N=2000 Θ = 3
f¹, N=400 Θ = 10
f¹, N=800 Θ = 10


CDF
1
0.8
0.6
0.4
0.2
Tau
0.2
0.1
0.05
0.02
0.01
0.005
0.002
100
1000
10000
100000.
1. · 10^6
Nz
X
0.001
0.005
0.01
0.05
0.1
0.5
1

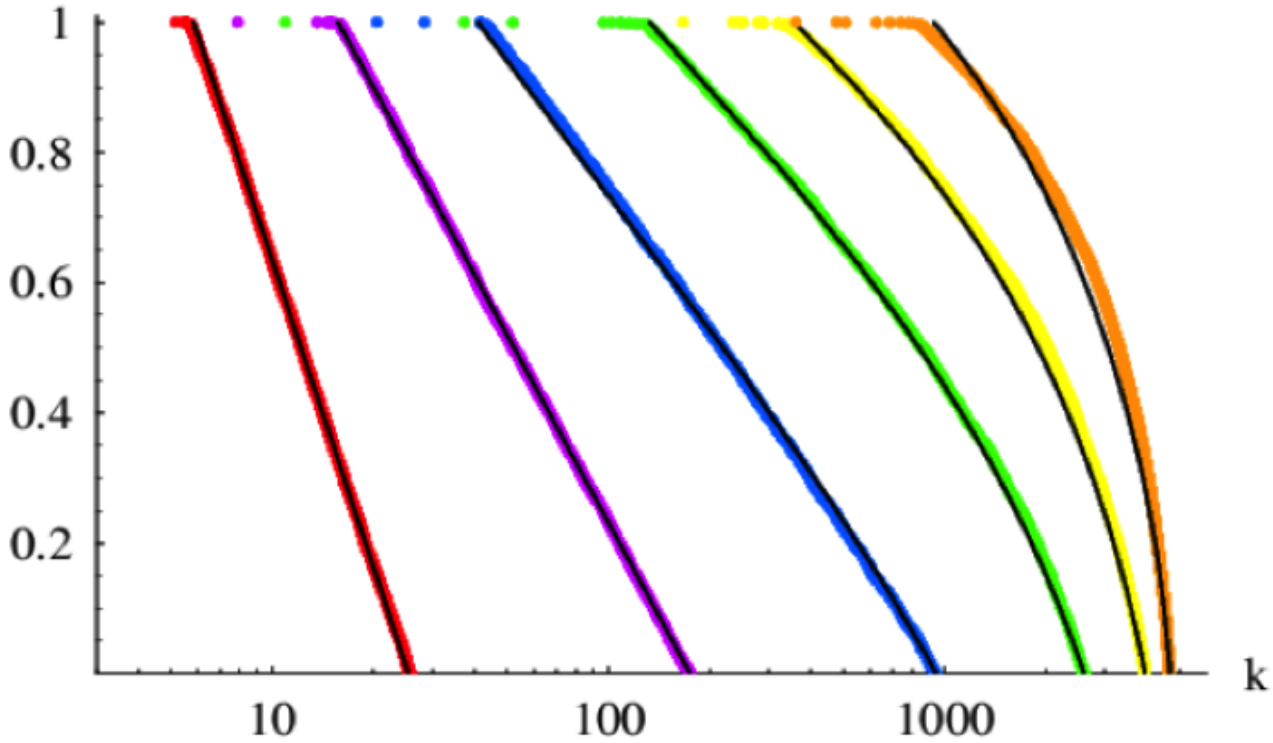
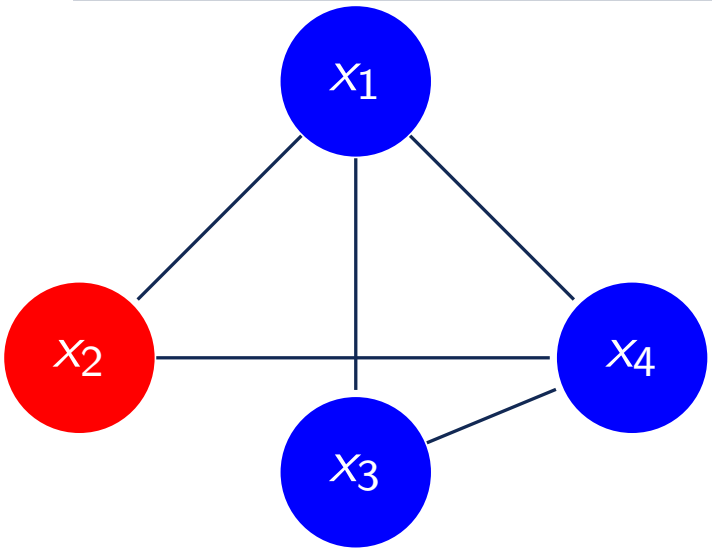


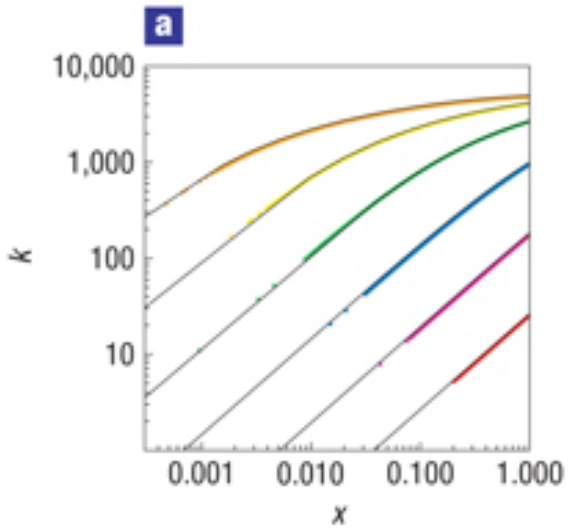
FIG. 3: Cumulative degree distribution. Again the scale is log-linear, therefore straight lines are logarithm corresponding to $P(k) \propto k^{-1}$. The isolated points are the simulation results

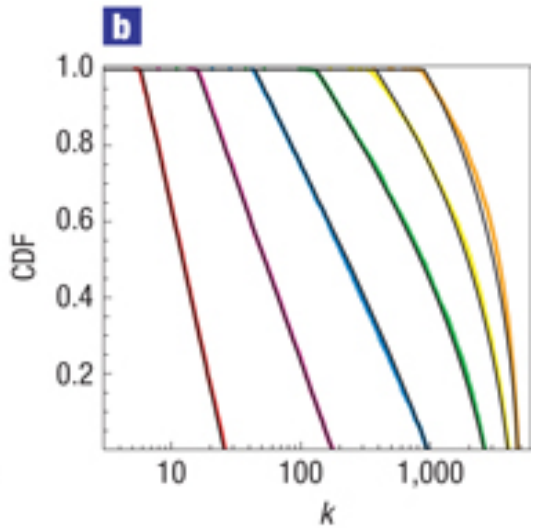
Fitness Model

Unexpected power laws

- $\rho(x) = e^{-x}$, $f(x, y) = \theta(x + y - z) \rightarrow P(k) \propto k^{-2}$.
- Self-Organized Processes $\rightarrow P(k) \propto k^{-1}$







WTW Modelling

GDP

The GDP determines the property of the network

A fitness model based on GDP reproduces the data

$$x_i = \frac{w_i}{\sum_{j=1}^{N} w_j} \qquad f(x_i, x_j) = \frac{\delta x_i x_j}{1 + \delta x_i x_j}$$

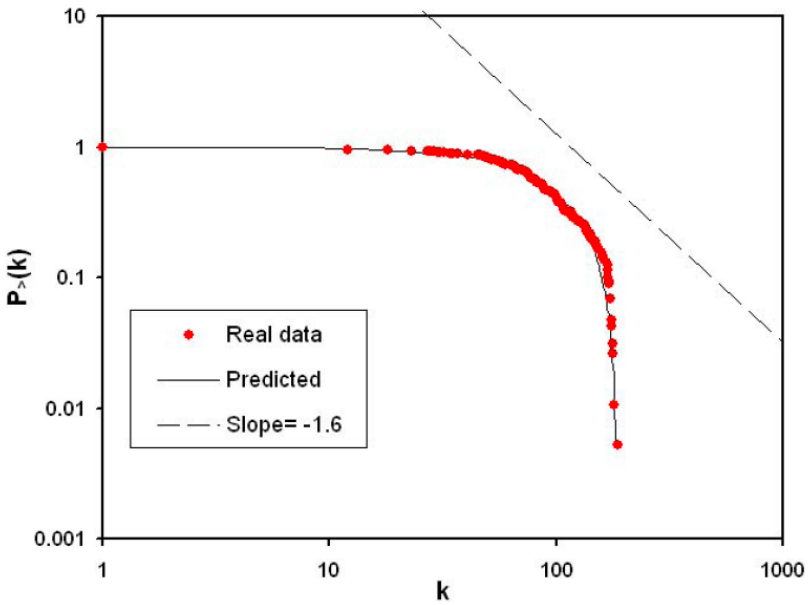
where w_i is the GDP of country i and $\delta > 0$ is the only free parameter of the model

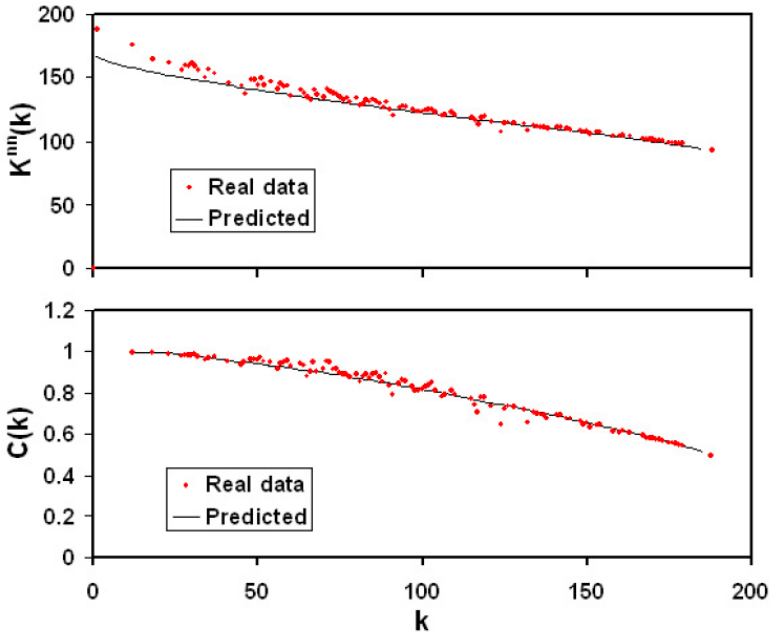
G. Caldarelli, A. Capocci, P. De Los Rios and M.-A, Muñoz, *PRL* **89**, 258702 (2002)

WTW Modelling 2

GDP

The Pareto's shape of GDP propagates in the WTW





D. Garlaschelli and M.I. Loffredo, *PRL* **93**, 188701 (2004)

D. Garlaschelli, T. D. Matteo, T. Aste, G. Caldarelli, and M. I. Loffredo, *EPJB* **57**, 159 (2007)

IMT SCHOOL FOR ADVANCED STUDIES LUCCA

References

- Bender, E. A. & Canfield, E. R. "The asymptotic number of labeled graphs with given degree sequences" *J. Comb. Theory, Ser. A* **24**, 296?307 (1978).
- Bollobás, B. *Random Graphs*. (Academic Press, London, 1985).
- Erdős, P., Renyi A. "On random graphs" *Publicationes Mathematicae Debrecen*, **6**, 290-297, (1959)
- Gilbert, E.N. "Random graphs", *Annals of Mathematical Statistics* **30** 1141-1144 (1959).
- Molloy, M. & Reed, B. A "Critical Point for Random Graphs with a Given Degree Sequence" *Random Struct. Algorithms* **6**, 161?180 (1995).
- Stauffer, D. Classical percolation. Lect. Notes Phys. 762, 1?19 (2009).