Very Short Introduction to Complex Networks

M. ÁNGELES SERRANO
Dpt. Física Fonamental
Universitat de Barcelona

Funding

Joint CRM-Imperial College School in Complex Systems
Barcelona, 8-11 April 2013

Complex system

No formal definition

- Large number of interacting units
- Self-organization: no centralized control or design
- Heterogeneity
- Nonlinear relationships
- Feedback loops
- Hierarchical, modular structure
- Coexistence of different spatial and time scales
- Open systems
- Out of equilibrium
- **Emergent collective phenomena** (not obvious from the properties of the individual components)

.
.
.

"The whole is more than the sum of its parts"

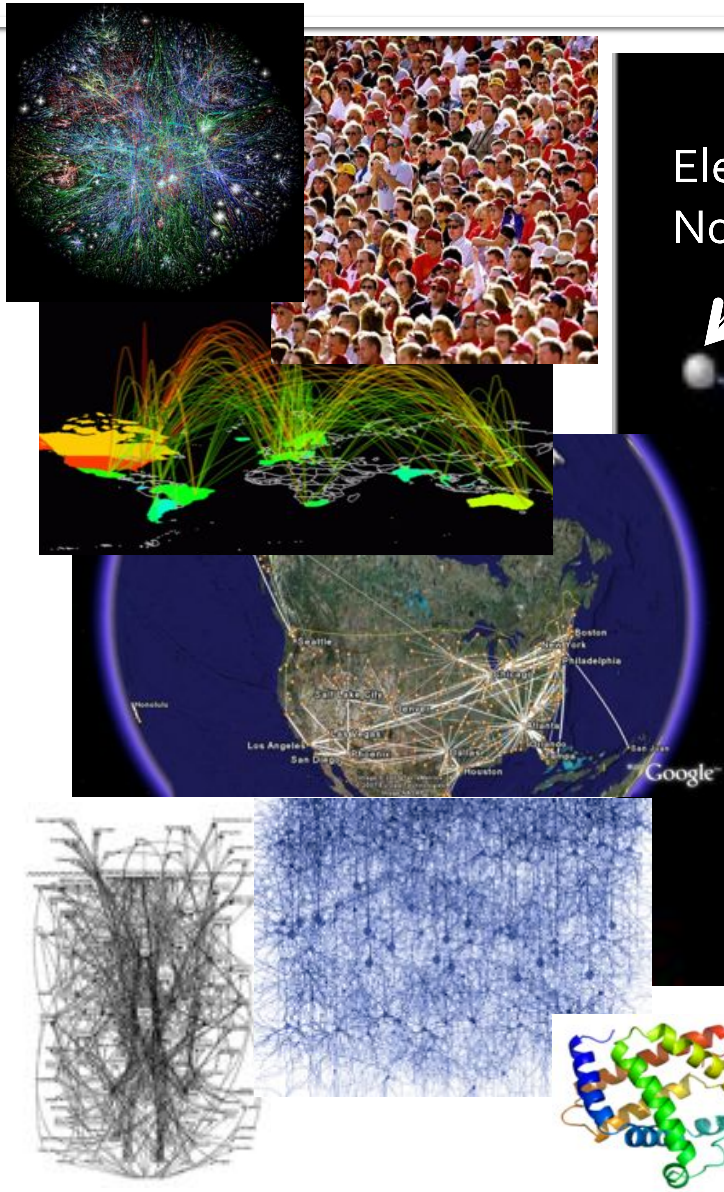
Lao Tse, *tao ching*, *VI BC*

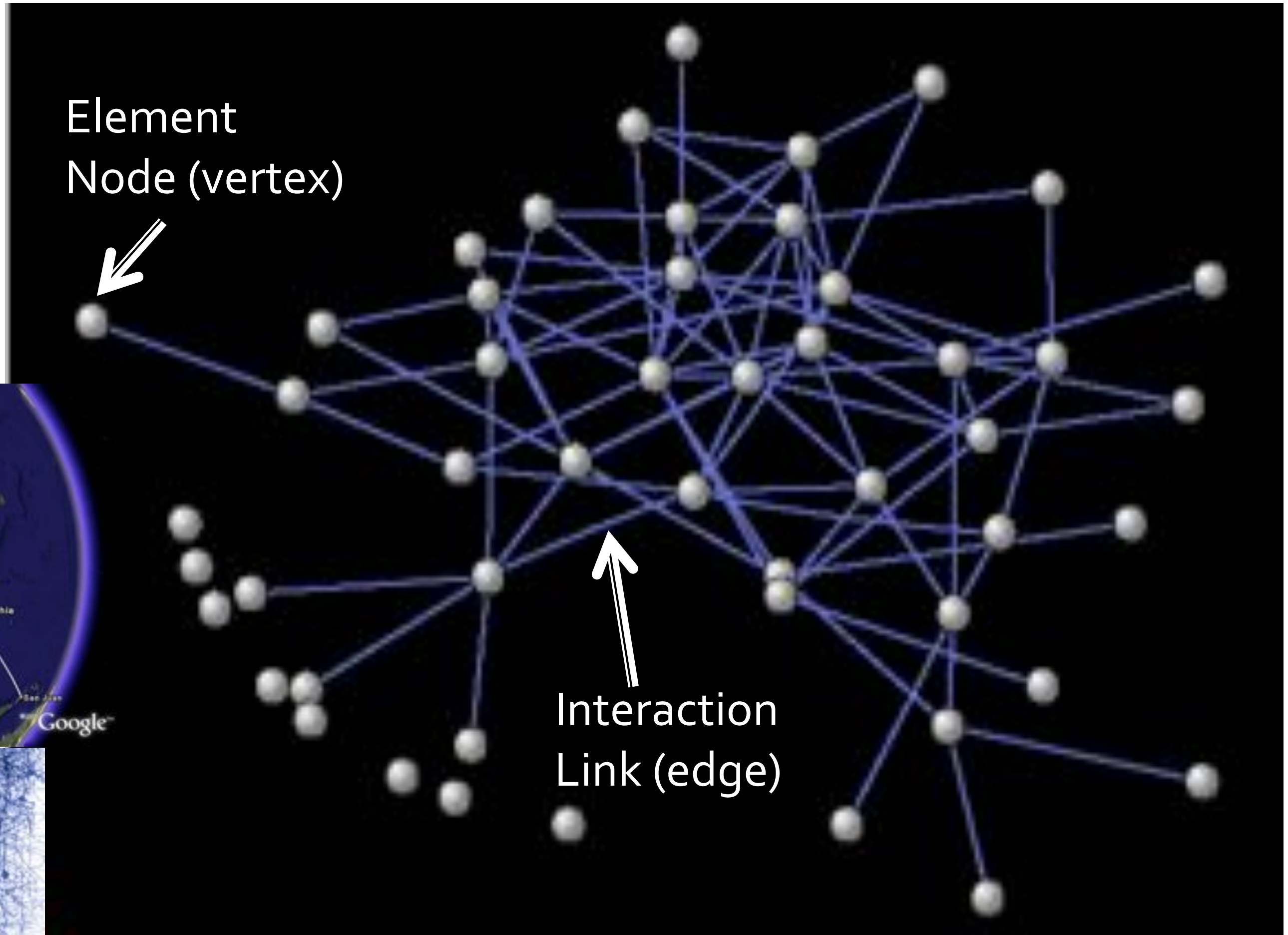
Aristotle, *Metaphysics*, *IV BC*

Complex network

Abstract representation of a discrete complex system

Graph





Between order and disorder/randomness

Complex networks

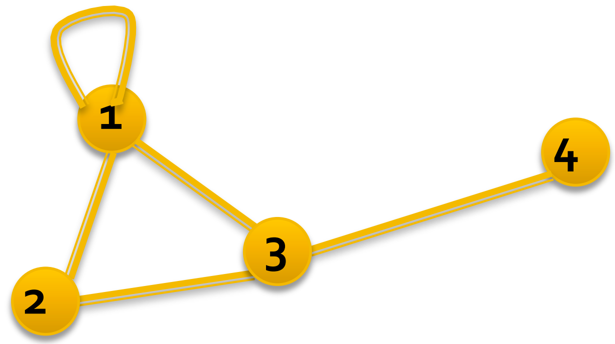
Adjacency matrix

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Edge list

1 1
2 1
3 1
2 3
3 4

Fundamental representation



Contains complete structural information,
but too detailed (large networks).....

Complex networks

- Connected components
- Small world property (topological distance)
- Degree distribution
- Correlations (two-point (ANND) and three-point (clustering))
- K-cores
- Centrality (measures the importance of nodes)
- Motifs (very small recurring functional units in complex networks)
- Modularity (communities)
- Hierarchy (structure on many scales)
- Backbones (minimal relevant subnetwork)
- Networks of networks: Multiplex and multilevel networks
- Self-similarity, underlying geometries

Weights - Directionality
Network models

DYNAMICS: dynamical processes on networks, coevolution of structure and processes, networks of dynamical units (oscillators…), adaptation and evolution of networks (interaction with environment), …

Basic statistics

N Number of nodes

E Number of edges

Complex networks are typically large in N

Complex networks are typically sparse (low E/N)

Node degree

Number of links per node

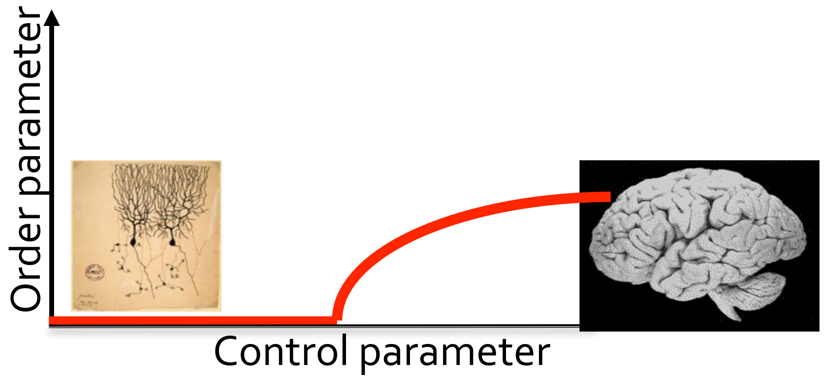
Notation: $k_i = 0, 1, 2, \ldots$

Average degree of a network $\langle k \rangle$

Connection between number of edges E and average degree

$$2E = N\langle k \rangle$$

Percolation

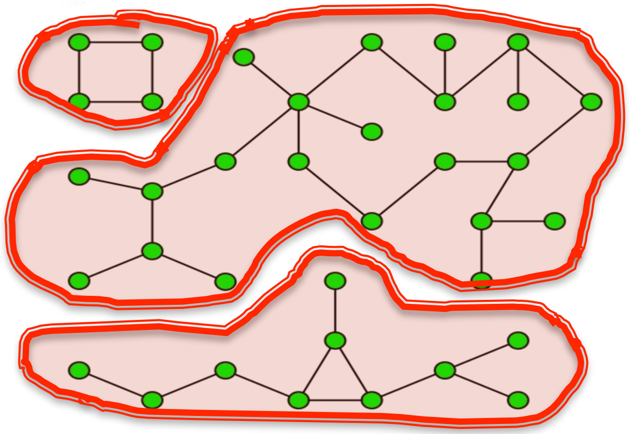


Giant connected component – complex networks are globally connected, what is essential to develop functionality

Percolation theory

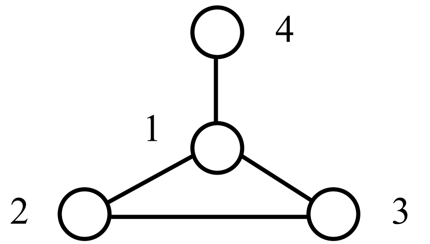
Mathematics: structure of connected clusters in random graphs

Physics: fluid flow in random media



Distances in networks

The distance between two nodes is measured as the number of links (unit length) in a shortest path connecting them



(a) shortest path length: Fewest number of steps between nodes i and j .

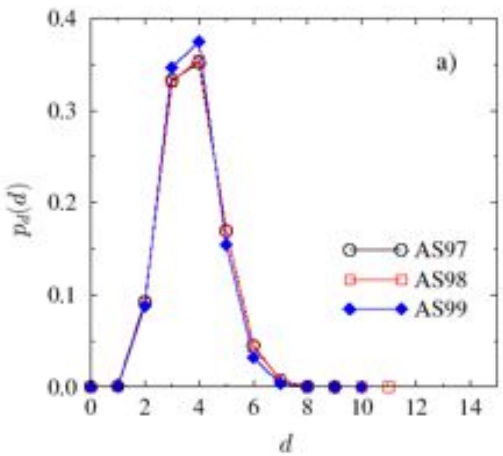
(b) average path length: Average shortest path length in the whole network.

(c) Network diameter: Maximum shortest path length in the network.

ASPL and D computationally expensive

$d_{12} = d_{13} = d_{14} = d_{23} = 1$

$d_{24} = d_{34} = 2$



Topological distances in networks are short and picked around a typical value (small-world property)

Small world property

Small-world property – networks are compact

each node in every pair is very close to each other

long range connections (shortcuts)

Collective dynamics of 'small-world' networks, D. J. Watts and S. H. Strogatz, Nature 393, 440-442(1998)

high clustering + small average path length

Six degrees of separation in social systems

Conjectured in 1929 by Frigyes Karinthy

Works in the 60's by Michael Gurevich, U.S. Kochen and de Sola Pool's

Famous experiments by S. Milgram in 1967

Repeated in 2001, 2003 , right now (Internet-based: emails,...)

de Sola Pool, I., Kochen, M. (1978–1979)."Contacts and Influence." Social Networks 1(1): 5–51

S. Milgram, "The Small World Problem", Psychology Today, 1967, Vol. 2, 60–67

The maximum distance (diameter) grows slower than any polynomial

$$l(N) \sim \log N$$

The number of nodes grows exponentially with the diameter of the network

$$N \approx e^{\bar{d}}$$

Degree distribution

Each node i has a degree k_i

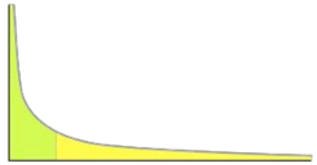
(a) Degree distribution: the probability of a node having k neighbors $P(k)$

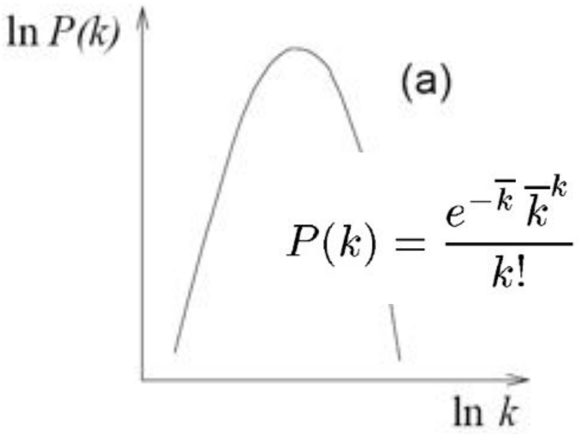
(b) Cumulative degree distribution: the probability of a node having at most degree k $P(k_i \leq k)$

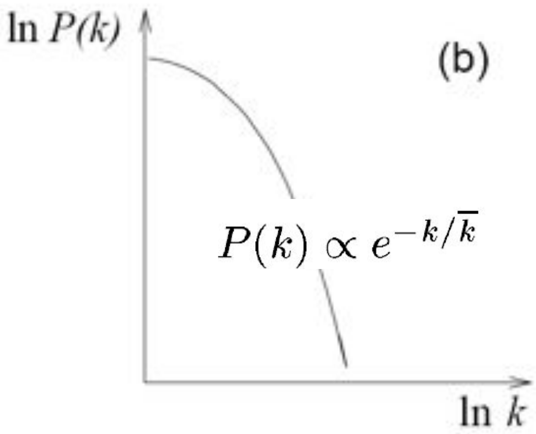
(c) Complementary cumulative degree distribution: probability of a node having a degree larger than k $P_c(k) = P(k_i > k)$

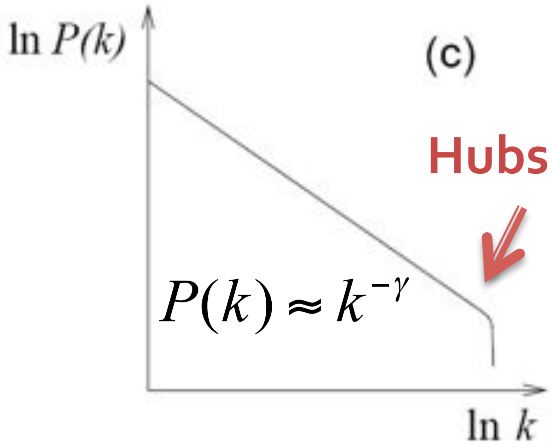
Degree distribution

Scale-free– power law degree distribution









Poisson

Classical random graphs

Exponential

Growing random graph

Scale-free

Barabasi-Albert model

Sometimes just fat-tailed or heterogeneous degree distribution

Degree distribution

Scale-free degree distributions - infinite size

Exponent typically between 2 and 3 for real networks

Normalization factor (γ must be greater than 1)

$$A = \sum k_i^{-\gamma}$$

Riemann zeta function with argument s=γ

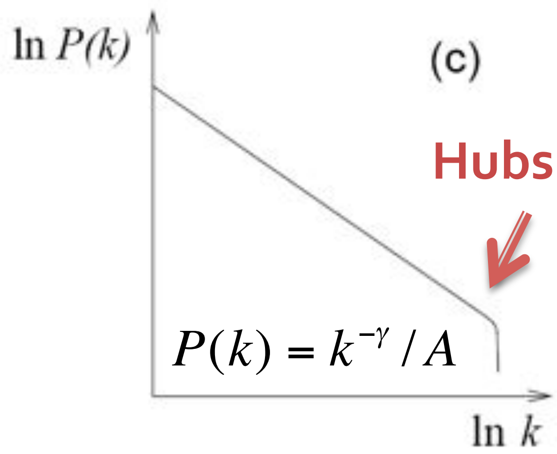
$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \cdots$$

If k is taken as a continuous variable $P(k) = \frac{\gamma - 1}{k_{\min}} \left(\frac{k}{k_{\min}} \right)^{-\gamma}$

First moment

$$\overline{k} = \sum_{k=0}^{\infty} k P(k)$$

In infinite networks, all moments of order greater or equal to γ-1 diverge



Correlations

M. A. Serrano, M. Boguñá, R. Pastor-Satorras and A. Vespignani. Correlations in complex networks, in Large Scale Structure and Dynamics of Complex Networks, G. Caldarelli and A. Vespignani editors, World Scientific (2007), pp. 35-66.

Two-point degree correlations:

P(k,k'), probability that a link selected at random connects two nodes of degrees k and k'

Three-point degree correlations:

P(k,k',k''), prob. that a randomly chosen triangle connects three vertices of degrees k, k', and k''

From a local perspective, P(k'|k) and P(k'',k'|k)

probability that a node of degree k is connected to a node of degree k'

prob. that a node of degree k is simultaneously connected to two nodes of degree k' and k''

In particular, P(k'|k) obeys the detailed balance condition

(closure of the network, physical conservation of edges)

$$kP(k' \mid k)P(k) = k'P(k \mid k')P(k') = \langle k \rangle P(k, k')$$

Correlations

Proof of the detailed balance condition

$$kP(k' \,|\, k)P(k) = k'P(k \,|\, k')P(k') = \langle k \rangle P(k, k')$$

$E_{kk'}$ symmetric matrix counting number of edges between degree classes k and k' (2 times for the diagonal)

$$\sum_{k'} E_{kk'} = kN_k,$$

$$\sum_{k,k'} E_{kk'} = \langle k \rangle N = 2E,$$

$$P(k' \,|\, k) = \frac{E_{k'k}}{kN_k}$$

$$P(k, k') = \frac{E_{kk'}}{\langle k \rangle N}$$

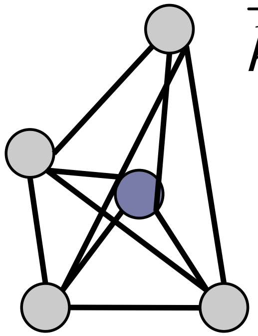
$$P(k' \,|\, k) = \frac{\langle k \rangle P(k, k')}{kP(k)}$$

Symmetry of P(k,k') → **DBC**

M. A. Serrano, M. Boguñá, R. Pastor-Satorras and A. Vespignani. Correlations in complex networks, in Large Scale Structure and Dynamics of Complex Networks, G. Caldarelli and A. Vespignani editors, World Scientific (2007), pp. 35-66.

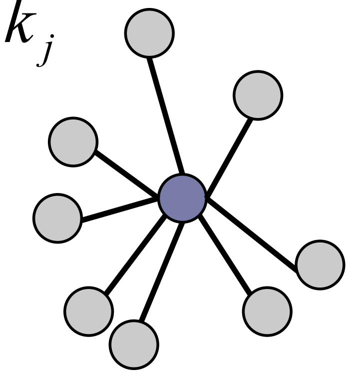
Average nearest neighbors degree

For empirical measurements: uniparametric projections



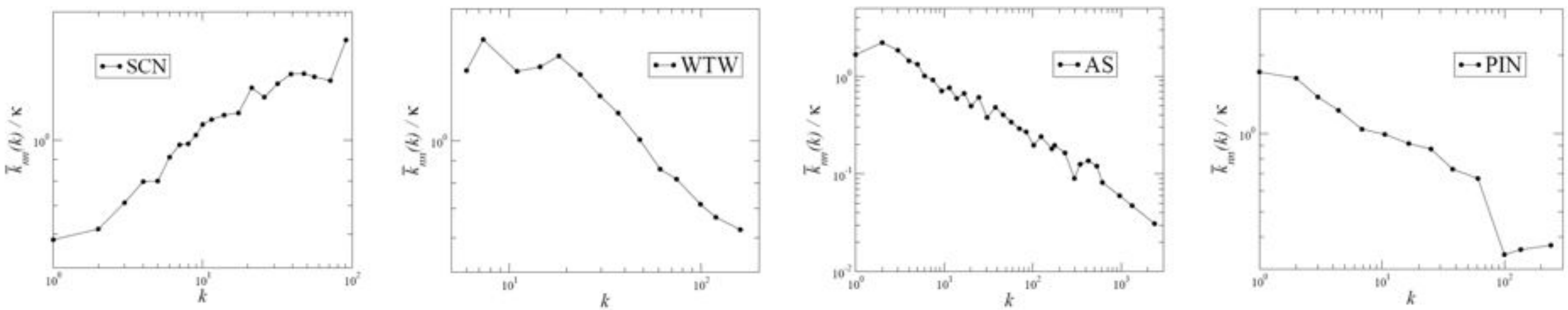
$$\bar{k}_{nn}(k) = \sum_{k'} k' P(k'|k) = \frac{1}{N_k} \sum_{i \in \nu(k)} \frac{1}{k_i} \sum_j a_{ij} k_j$$

It is a measure of the tendency of nodes
to connect to peers in terms of degree



Assortative mixing

Disassortative mixing

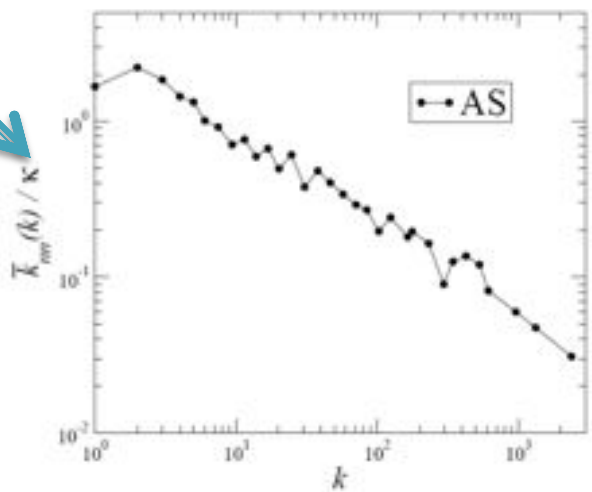


Average nearest neighbors degree

Empirical measurements

$$\bar{k}_{nn}(k) = \sum_{k'} k' P(k' | k)$$

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



Correction that makes comparable the ANND functions of different real networks

The function ANND/κ has average 1 independently of the network and the pattern of correlations and so it can be used to compare two point correlations in different networks

Proof from detailed balance:

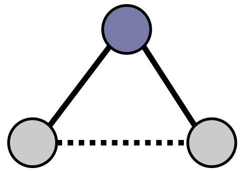
$$kP(k' \mid k)P(k) = k'P(k \mid k')P(k')$$

$$\langle k^2 \rangle = \sum_{k'} k'P(k') \sum_{k} kP(k \mid k')$$

Boguna, M; Pastor-Satorras, R; Vespignani, A
PHYSICAL REVIEW LETTERS 90, 28701 (2003)

Clustering

It is a measure of the likelihood that two neighbors of a given vertex are neighbors themselves



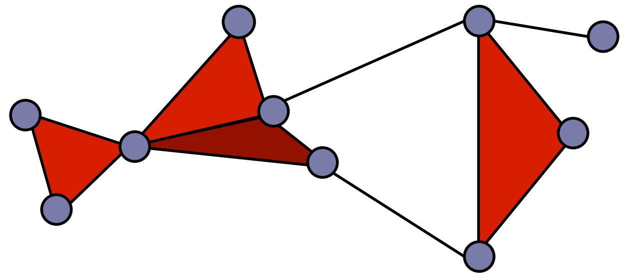
- Local measures

$$c_i = \frac{2T_i}{k_i(k_i - 1)}$$

It is a measure of the number of triangles in a network

- Global measures

$$C_\Delta = \frac{3 \times (\text{number of triangles})}{(\text{number of connected triples})}$$



$$C = \frac{\sum_i c_i}{N}$$

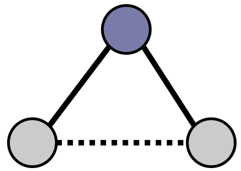
Clustering coefficient

Projections of the three-point distribution function P(k,k′,k″)

Clustering

- **Degree dependent Clustering coefficient**

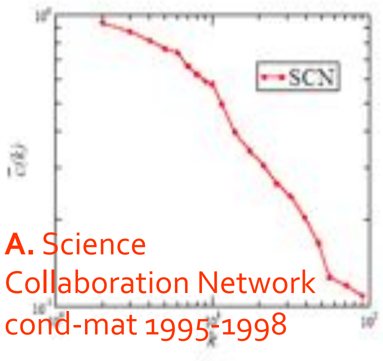
average of local clustering over degree classes



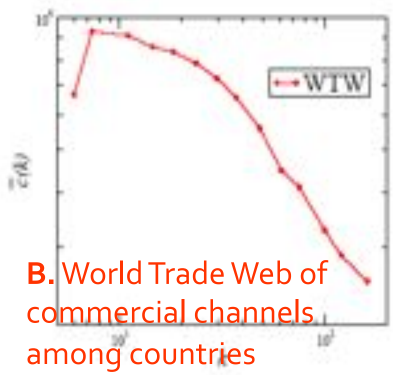
$$\bar{c}(k) = \frac{1}{N_k} \sum_{i \in \Upsilon(k)} c_i = \frac{1}{k(k-1)N_k} \sum_{i \in \Upsilon(k)} 2T_i$$

$$\bar{c} = \sum_k P(k)\bar{c}(k)$$

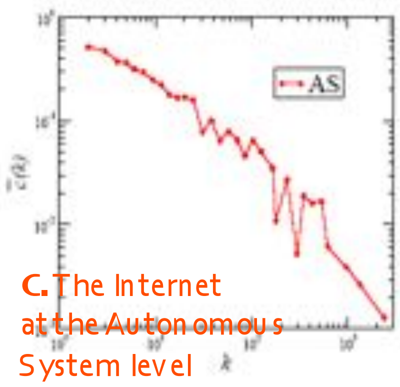
$$\bar{c}(k) = \sum_{k'k''} P(k'k''|k) r^k_{k'k''} = \frac{1}{N_k} \sum_{i \in v(k)} \frac{1}{k_i(k_i-1)} \sum_{jl} a_{ij} a_{il} a_{jl}$$



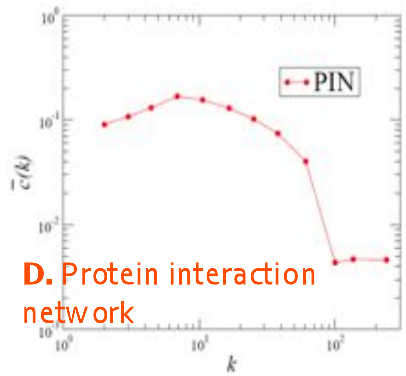
A. Science Collaboration Network cond-mat 1995-1998



B. World Trade Web of commercial channels among countries



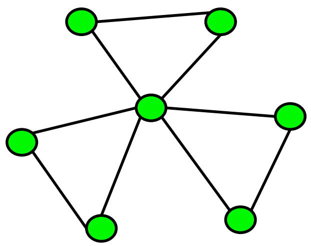
C. The Internet at the Autonomous System level



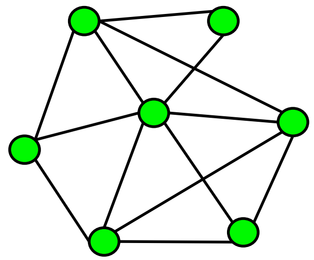
D. Protein interaction network

Clustering

Also important:
Weak vs Strong clustering $\bar{c}(k) < \dfrac{1}{k-1}$



Triangles are disjoint

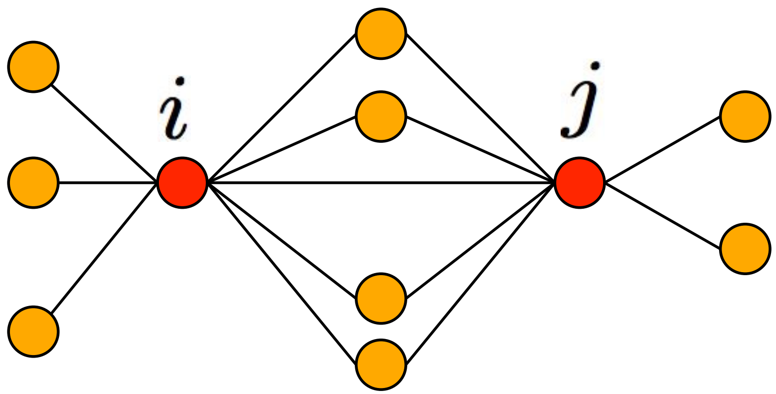


Triangles coalesce

The locally tree like assumption is a good approximation for networks with weak clustering, and enables analytical approaches like the generating function formalism to estimate percolation thresholds, sizes of giant components…

Edge multiplicity m_{ij}

Number of triangles passing through edge i-j



M. A. Serrano and M. Boguñá, Clustering in complex networks. I. General formalism, Physical Review E 74, 056114 (2006)

K-cores

A network is organized as a set of successively enclosed k-cores (like a Russian nesting doll)

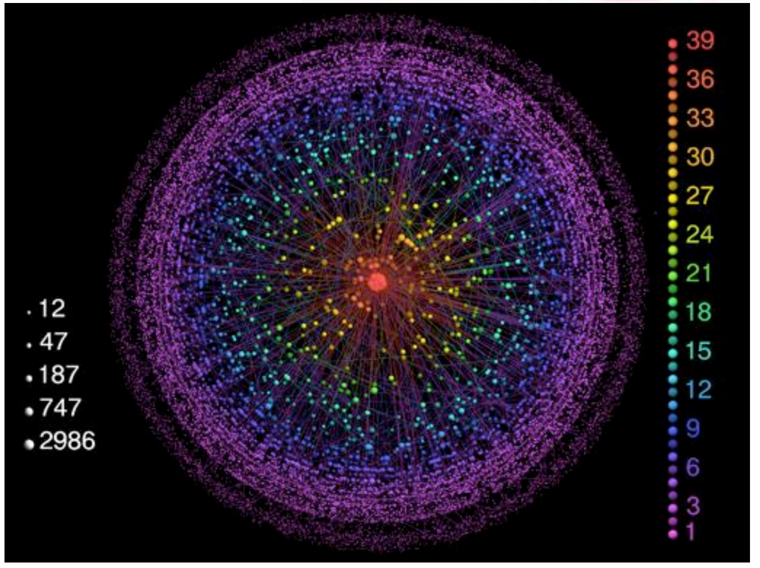
k-core: subgraph in which every node has at least k connections with other neighbors in the subgraph

The k-cores are obtained in a recursive way:

i) remove from the original graph all nodes (and their connections) with degree less than k

ii) remove from the remaining graph all nodes (and their connections) with degree less than k

iii) repeat ii) until no further removal is possible

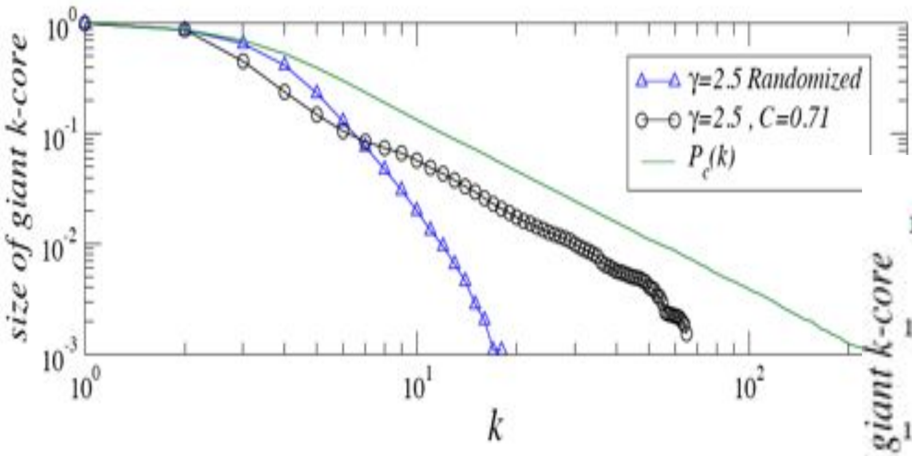


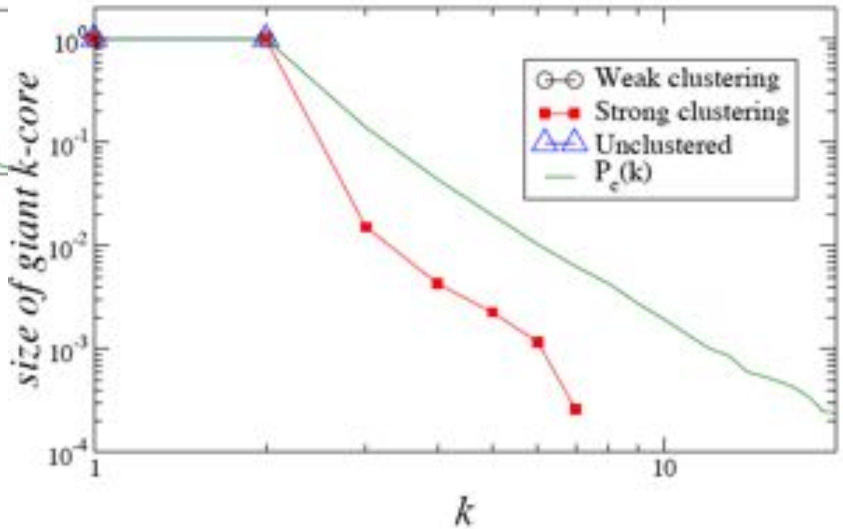
LANET-VI visualization

K-cores

The k-core organization of a network gives idea of its sparsity/ tree-likeness

Stronger clustering produces a deeper hierarchy of k-cores

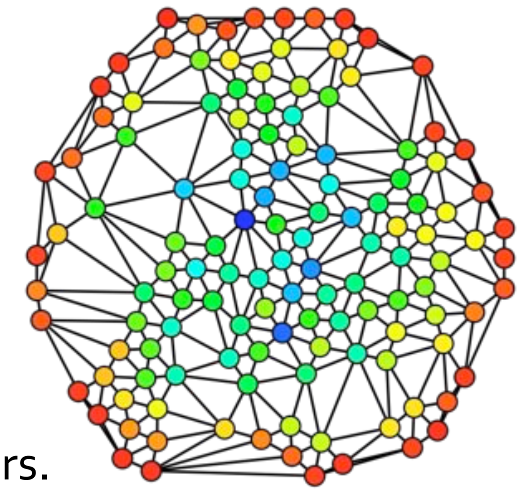




Centrality

Structural measures of a node's 'importance'

Node betweenneess



- **Degree centrality**: k

- **Node/Edge betweenness**: fraction of all paths between any two vertices that pass through that node/edge

- **Eigenvector centrality**. The score of a node is proportional to the sum of the scores of all its neighbors. A node that is linked to many nodes with high score receives a high score itself. Eigenvector with the greatest eigenvalue (positive scores) of the adjacency matrix.

$$x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j = \frac{1}{\lambda} \sum_{j=1}^{N} a_{i,j} x_j \qquad \mathbf{Ax} = \lambda \mathbf{x}$$

Variants: Google's PageRank for web pages (and its precursor Hubs and Authorities). Probability distribution giving the likelihood that a person randomly clicking on links will arrive at any particular page.

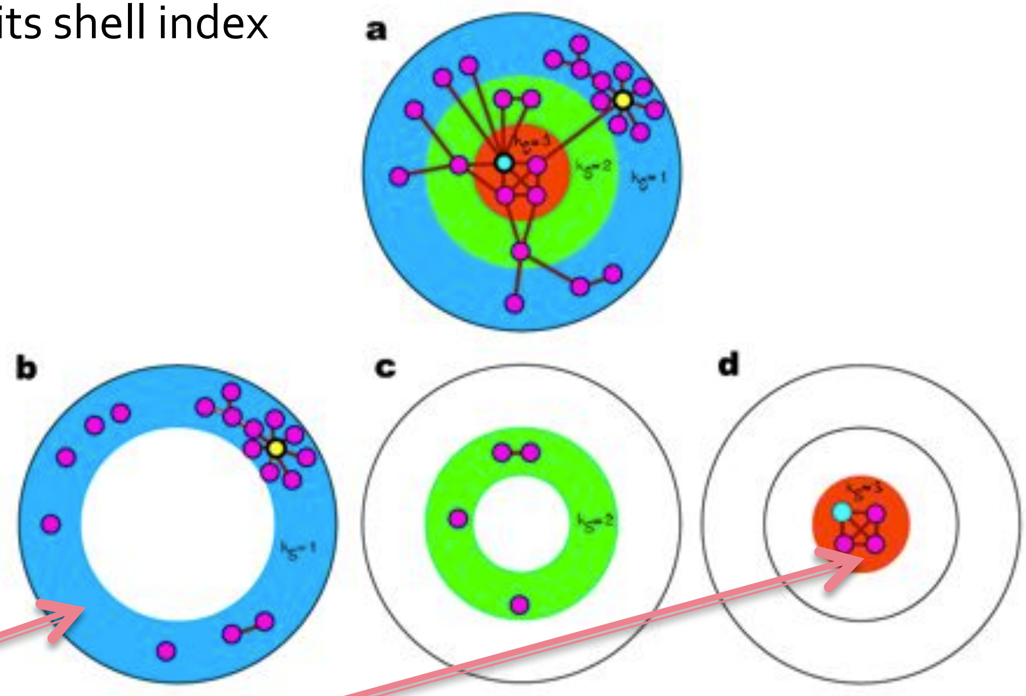
Centrality

Structural measures of a node's 'importance'

Kitsak et al., Nat. Phys. 6, 888 (2012)

- **k-shell decomposition**: the coreness of a node (the higher k-core it belongs to) is its shell index

In supercritical epidemic spreading originating in single nodes, **most efficient spreaders** seem to be those located **within the core** of the network as identified by the k-shell decomposition analysis



Periphery

Core

k-shell decomposition

Centrality

The interplay between structure and dynamics is what determines a node's 'importance'

- **Dynamical influence (DI):** measures how strongly a node's dynamical state affects collective behavior. It is defined for a system of N nodes with coupled linear dynamics $\dot{x} = Mx$.

$\mu_{\max}$ maximum eigenvalue of M with left and right eigenvectors c and e, respectively

$\mu_{max} < 0$ x(t) converges to a null vector that represents a stable fixed point solution

$\mu_{max} > 0$ indefinite growth from almost all initial conditions

$\mu_{max} = 0$ **Conserved quantity** $\phi_c = c \cdot x$ $\quad \frac{d\phi_c}{dt} = c \cdot \dot{x}(t) = [cM] \cdot x(t) = 0.$

$x(\infty) := \lim_{t \to \infty} x(t) = \frac{c \cdot x(0)}{c \cdot e} e,$ **c_i is the DI of element i on the dynamics**

The coefficient ci quantifies the extent to which the initial condition x(0) at node i affects the final state.

Masuda, N. & Kori, H., Phys. Rev. E 82, 056107 (2010).

K. Klemm, M. A. Serrano, V. M. Eguíluz, M. San Miguel, A measure of individual role in collective dynamics. Scientific Reports 2, 292 (2012).

Dynamical influence

Example: Epidemic spreading model SIR (and SIS)

In the SIR model, all nodes are initially in the susceptible state (S) except for some seeds in the infectious state (I). At each time step, infected nodes infect their susceptible neighbors with probability β and then enter the recovered state (R) with probability λ, where they cannot be infected again.

Other models of critical phenomena (Ising mode); diffusive processes (voter mode, exact);strongly non-linear systems (Kuramoto oscillators)

The linear regime corresponds to small perturbations to the stationary state with all nodes S, $M = \beta A^{\mathrm{T}} - I$

$x_j(t)$ is the probability of node j to be infected at time t

DI as before implies that β has to tuned to $\beta = \beta_c = 1/\alpha_{\max}$ and then DI is given by a right eigenvector of A for maximum eigenvalue α_{max} .

The expected outbreak size from an initial infection described by the probability vector x(0) is proportional to c x(0).

K. Klemm, M. A. Serrano, V. M. Eguíluz, M. San Miguel,
A measure of individual role in collective dynamics. Scientific Reports 2, 292 (2012).

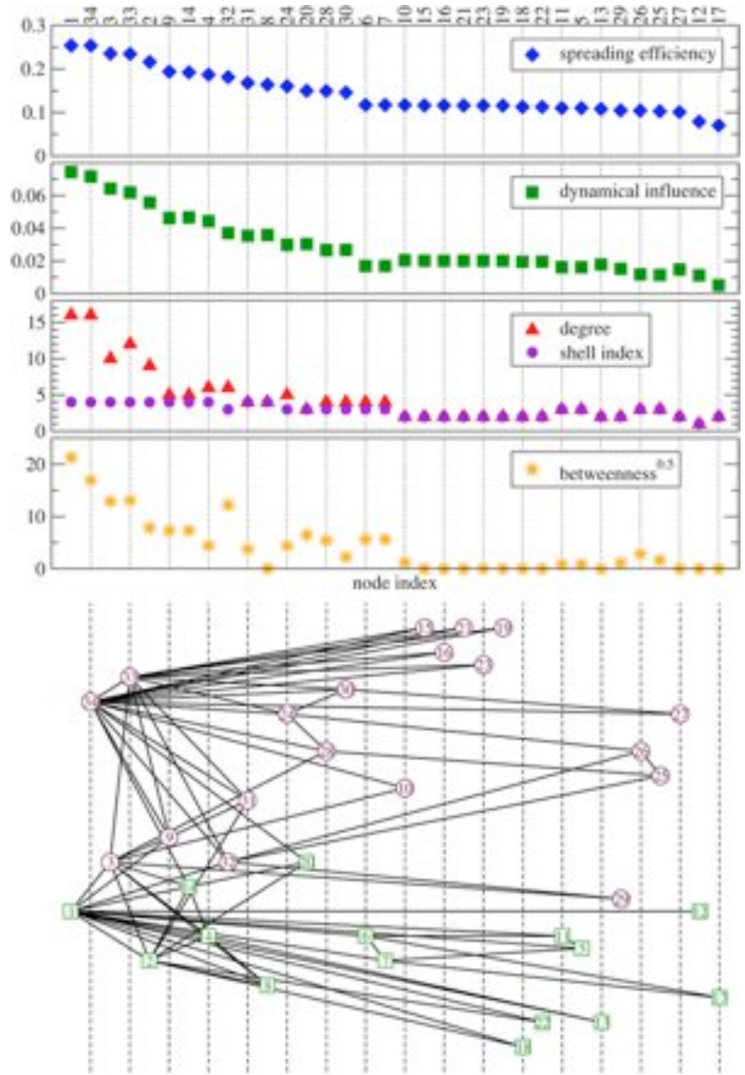
Dynamical influence

Example: Epidemic spreading model SIR

• DI is compared with spreading efficiency, the expected fraction of nodes reached by an epidemic outbreak initiated with node i infected (seed node), all others susceptible.

• c_i is a good predictor of SIR spreading efficiency at critical parameter value $\beta = \beta c$. (Predictive power is quantified by the rank order correlation).

• Dynamical influence c_i outperforms the predictions made by degree, shell index and betweenness centrality.



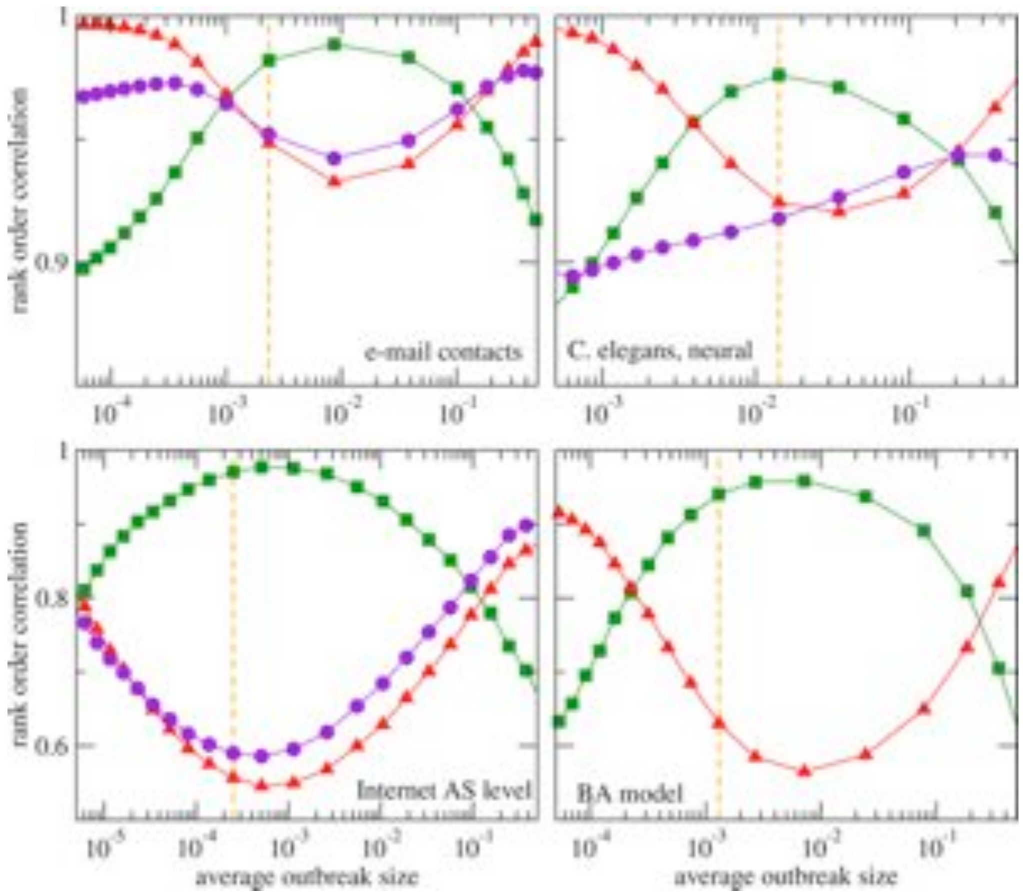
Dynamical influence

Example: Spreading model SIR (and SIS)

influence (squares), degree (triangles) and shell index (circles)

In the subcritical regime, spreading is sparse and typically confined to the neighborhood of the seed node i

in the supercritical regime, the epidemics rarely fails to spread to the whole system



Good predictor of critical spreading capabilities (also for other critical phenomena like the Ising model)

High-quality proxy for driving efficiency, uncovering which are the best target nodes in real networks to be forced in order to drive the system towards specific states

Networks and hidden metric spaces

Three key properties

scale-free degree distribution+small world property+high clustering

The natural degree distribution of a random geometric graph embedded in hyperbolic space is a scale-free with characteristic degree 3…

The small-world property explains why hyperbolic hidden metric spaces underlying complex networks offer a natural embbedding…

Clustering gives the clue to connect complex networks to underlying hidden metric spaces because of the triangle inequality…

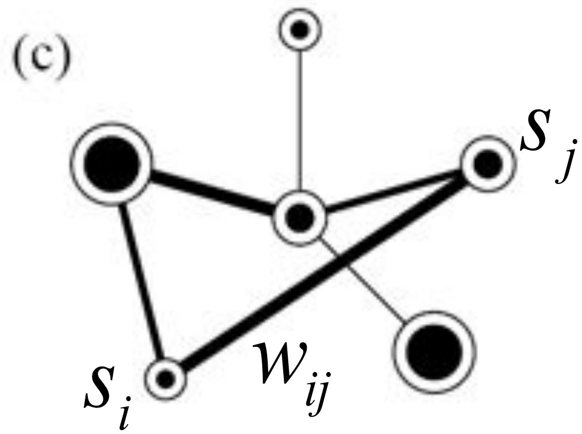
Weighted networks

Links are not binary (just present or absent) but are characterized by different intensities, w

The sum of all weights incident on a node is its strength, s

$$s_i = \sum_j w_{ij}$$

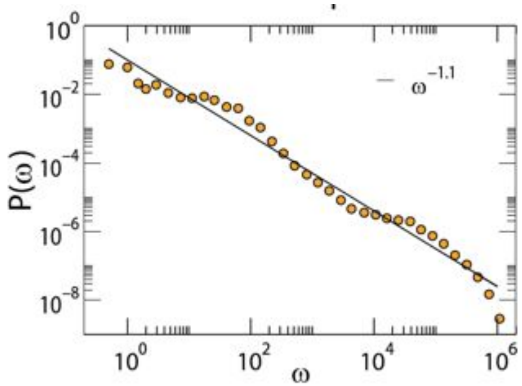
The degree distribution has to be complemented with the distribution of weights associated to links and the distribution of strengths associated to nodes

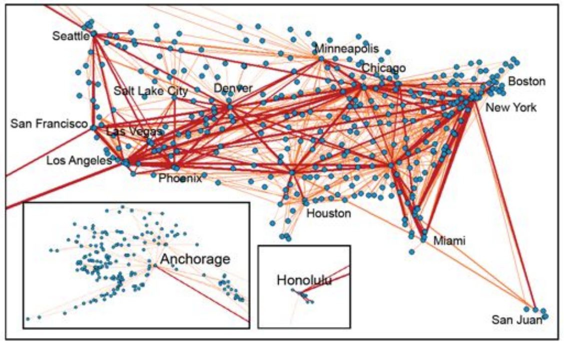


$$P(k) \rightarrow P(k), P(w), P(s)$$

Weighted networks

In many networks, P(w) is broadly distributed spanning several orders of magnitude





M. A. Serrano, M. Boguñá, A. Vespignani. Extracting the multiscale backbone of complex weighted networks. PNAS 106, 6483-6488 (2009).

Relationship between the strength of a node and its degree

$$s(k) \sim k^{\beta}$$

Weights are locally correlated and non-trivially coupled to topology (β>1)

$$s_i = \langle w \rangle \Sigma_j \, a_{ij} = \langle w \rangle k_i$$

Absence of correlations between weights and degrees (β=1); weights do not add information

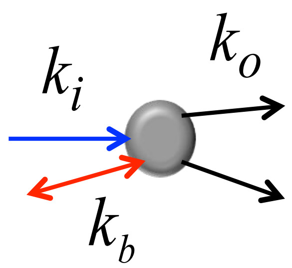
A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani. PNAS 101, 3747-3752 (2009).

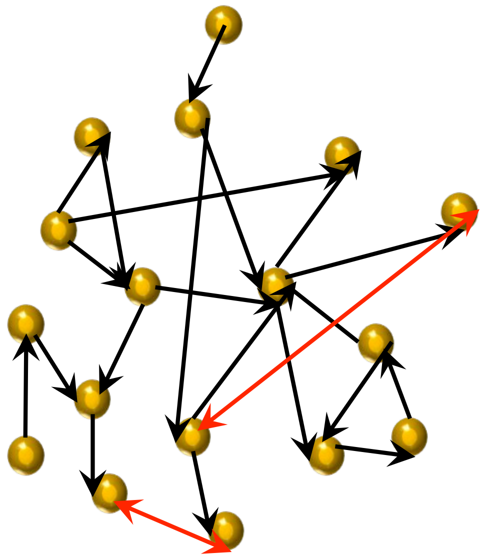
Directed networks

Links have an associated directionality, fro i to j, from j to i, or are bidirectional

$$P(k) \rightarrow P(k_i, k_o, k_b)$$

$k \rightarrow k_{in}$ and k_{out} and k_b





Many times, bidirectional links are decoupled, and just $P(k_i)$ and $P(k_o)$ are considered (usually, one distribution is fat-tailed and the other is homogeneous…)

$$k_b P(\mathbf{k}) P_b(\mathbf{k}'|\mathbf{k}) = k_b' P(\mathbf{k}') P_b(\mathbf{k}|\mathbf{k}')$$

detailed balance conditions

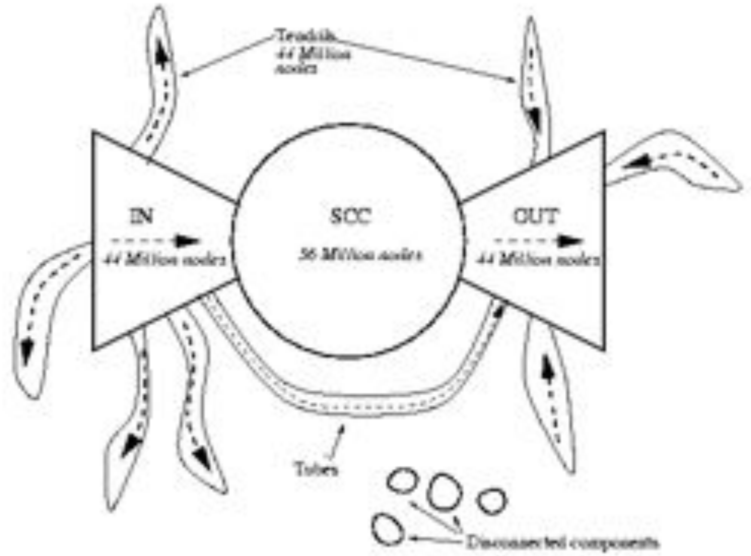
$$k_o P(\mathbf{k}) P_o(\mathbf{k}'|\mathbf{k}) = k_i' P(\mathbf{k}') P_i(\mathbf{k}|\mathbf{k}')$$

M. Boguñá and M. A. Serrano. Generalized percolation in random directed networks, Physical Review E 72, 016106 (2005).

Directed networks

Global connectivity structure – nodes – BOW TIE

- SCC (strongly connected component):
  - can reach all nodes from any other by following *directed* edges
- IN
  - can reach SCC from any node in 'IN' component by following directed edges
- OUT
  - can reach any node in 'OUT' component from SCC
- Tendrils and tubes
  - connect to IN and/or OUT components but not SCC
- Disconnected
  - isolated components



The bow-tie structure of the Web graph
A. Broder *et al.*, Computer Networks, 33:309–320, 2000.

Systems characterized by transport phenomena (of matter, energy, information....

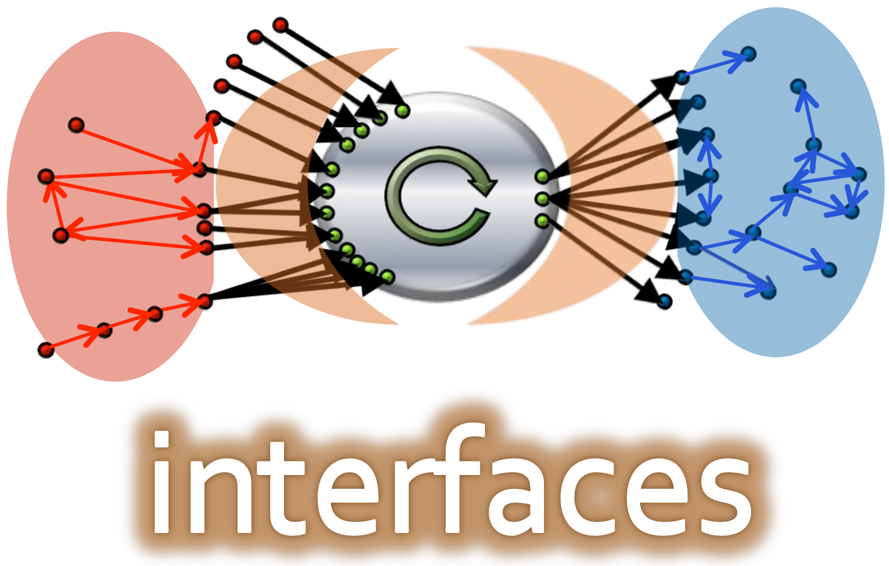
Directed networks

Global connectivity structure – edges – L BOW TIE

- The connected components are computed in terms of edges instead of nodes.

- The IN and OUT peripheral components are connected to the core SCC through interfaces, so that the L-Bow tie structure is characterized by five main components instead of three.

- The specific conformation of the interfaces could bring new light into the discussion of how structure is interwoven with functionality in transport networks.

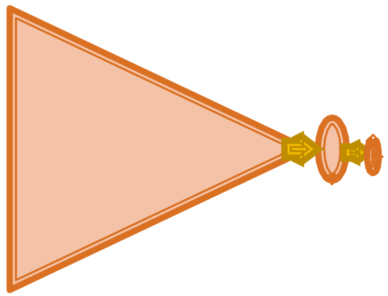


Phys. Rev. E 76, 056121 (2007)

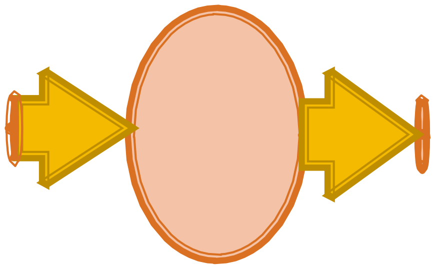
Directed networks

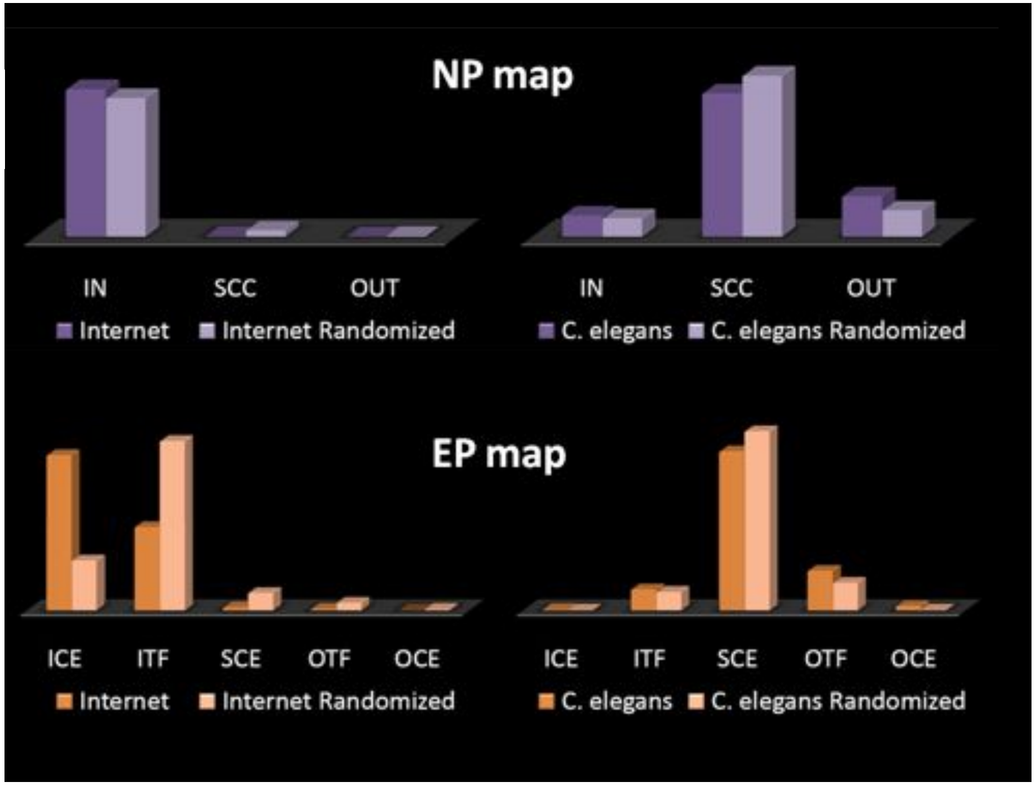
Global connectivity structure

Internet customer-provider AS relationships
(connections are payment for services provided)



Synaptic neuronal structure of C. Elegans
(connections are chemical synapses)





M. A. Serrano and P. De Los Rios. Structural efficiency of percolated landscapes in flow networks. PLoS ONE 3(11): e3654 (2008).

Bipartite networks

Example: metabolic networks

(*a*)

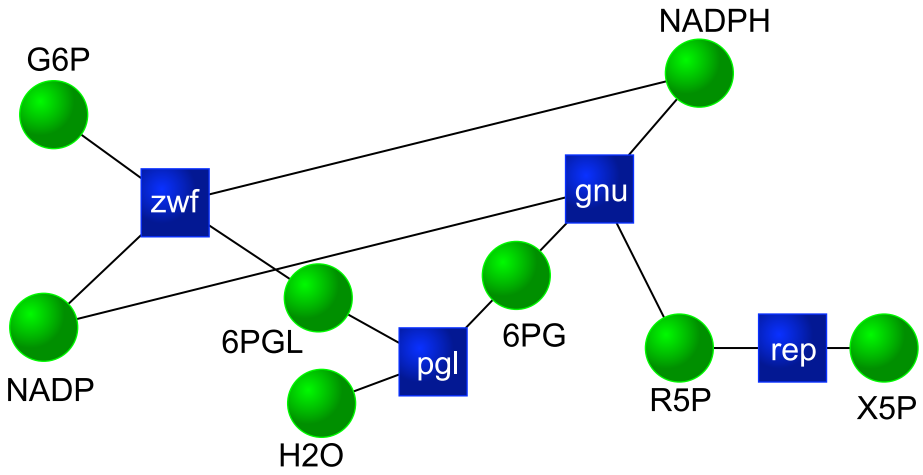
$$1 \text{ glucose 6-phosphate (G6P)} + 1 \text{ NADP}^+ \xrightarrow{zwf} 1 \text{ 6-phosphoglucono } \delta\text{-lactone (6PGL)} + 1 \text{ NADPH}$$

$$1 \text{ 6-phosphoglucono } \delta\text{-lactone} + 1 \text{ H}_2\text{O} \xrightarrow{pgl} 1 \text{ 6-phosphogluconate (6PG)}$$

$$1 \text{ 6-phosphogluconate} + 1 \text{ NADP}^+ \xrightarrow{gnd} 1 \text{ ribulose 5-phosphate (R5P)} + 1 \text{ NADPH}$$

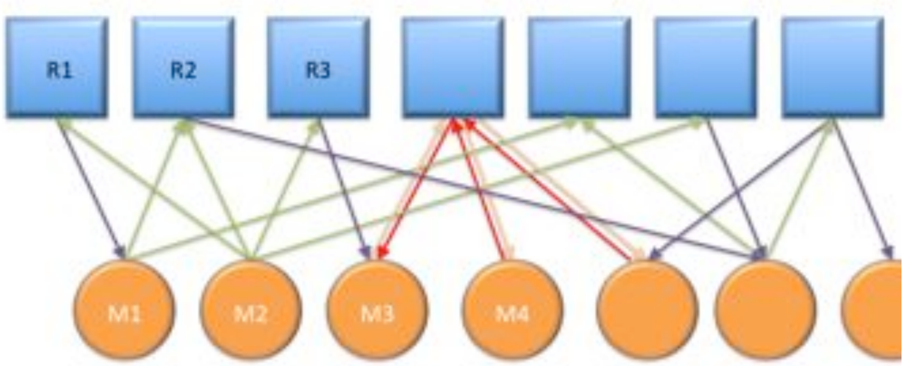
$$1 \text{ ribulose 5-phosphate} \xleftrightarrow{rpe} 1 \text{ xylulose 5-phosphate (X5P)}$$

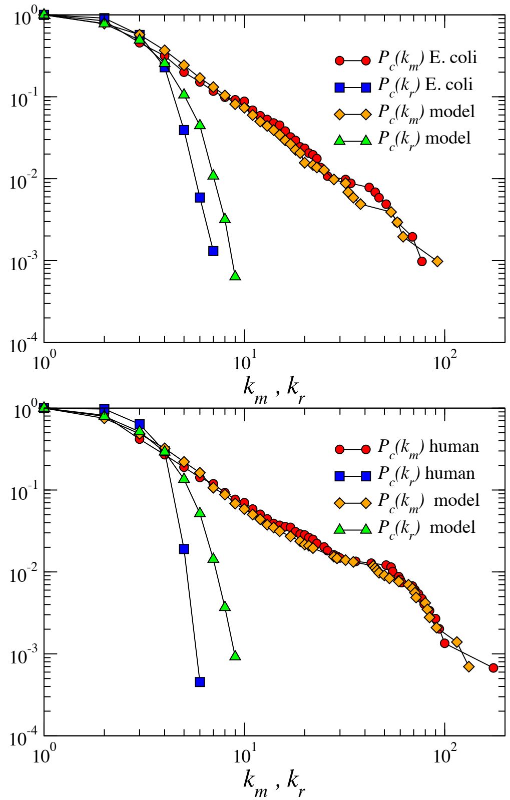
four coupled equations in the pentose-phosphate pathway of E. coli



Bipartite networks

Bipartite networks: two types of nodes without connections in the same class



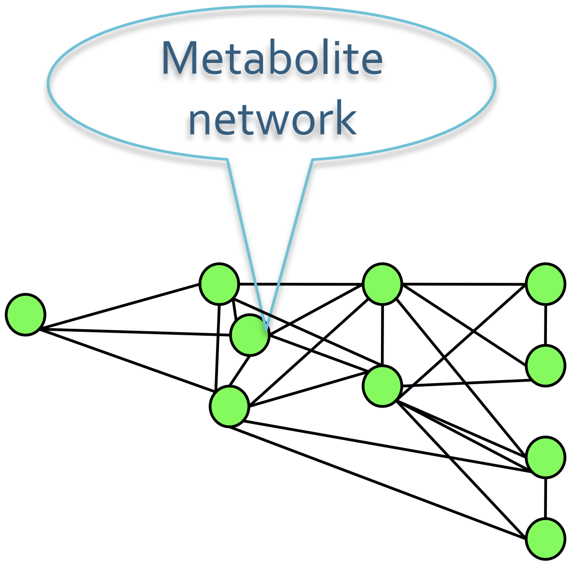


Real bipartite networks are typically characterized by a scale-free degree distribution for nodes in one class and a bounded/homogeneous distribution for nodes in the other

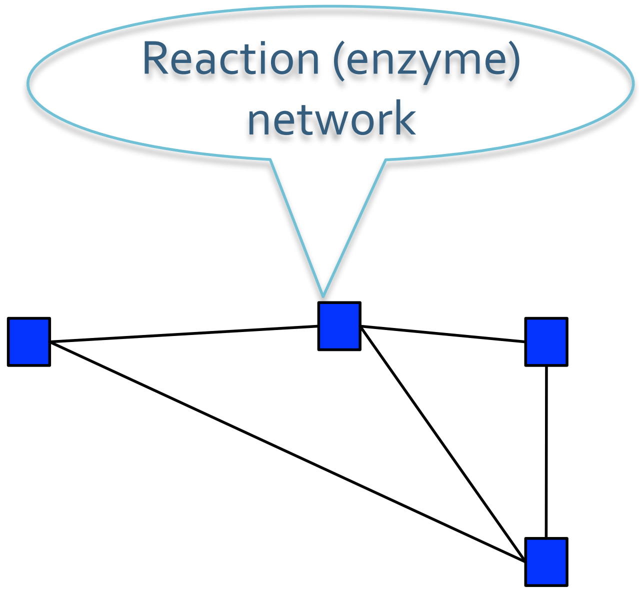
M. A. Serrano, M. Boguñá, F. Sagués, Uncovering the hidden geometry behind metabolic networks. Molecular BioSystems, 8, 843-850 (2012).

Bipartite networks

One-mode projections (weights can be added – number of neighbors shared)



Two metabolites connected if they share a reaction



Two reactions/enzymes connected if they have common metabolites

Network models

Set of rules to interconnect a number of elements in order to reproduce specific features (but always SWs)

- **Equilibrium network models**

  The number of nodes is fixed to N
  - Classical random graphs, Erdös and Rényi model
  - Watts-Strogatz model
  - Configuration model

- **Non-equilibrium network models**

  The number of nodes N grows
  - Classical random growing grahs
  - Preferential attachment, Barabasi-Albert model

Equilibrium network models: ER

Classical random graphs, Erdös and Rényi

- **(1) Each pair of nodes in the network are connected with probability p**

  Grand canonical statistical ensamble of graphs with equal probability

  (standard statistical definition: temperature and chemical potential constant)

- **(2) A given number of links E connect randomly chosen pairs of vertices.** Self-connections and multiple connections may happen unless a specific restriction to avoid them is added

  Microcanonical statistical ensamble of graphs with equal probability

  (standard statistical definition: energy and number of particles constant)

In the large graph limit (thermodynamic limit) both constructions are equivalent

Equilibrium network models: ER

Classical random graphs, Erdös and Rényi

Degree distribution: Binomial,
for large N (p small, sparse limit) it takes the Poisson form

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

$$P(k) = e^{-\bar{k}} \bar{k}^k / k!$$

$$\bar{k} = p(N-1)$$

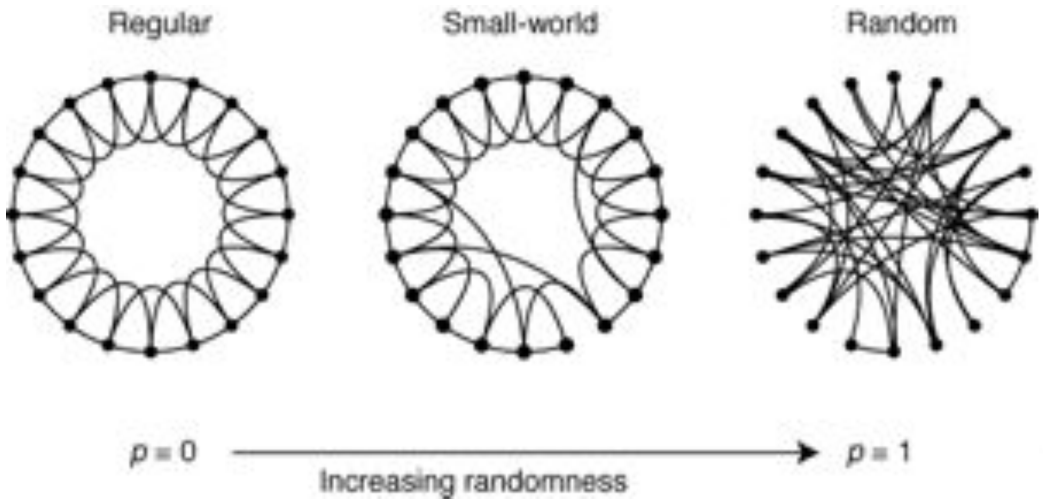
$$E = pN(N-1)/2$$

Uncorrelated

$$P(k'|k) = \frac{k' P(k')}{\langle k \rangle} \Rightarrow \bar{k}_{nn}^{unc}(k) = \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{<k>^2 - <k>}{<k>} = <k> - 1$$

$$C = \frac{(\text{number of triangles})}{(\text{number of connected triples})} = \frac{\binom{n}{3} p^3}{\binom{n}{3} p^2} = p = \frac{\langle k \rangle}{n-1} \Rightarrow$$ Locally tree-like

Equilibrium network models: WS



random network

Low clustering

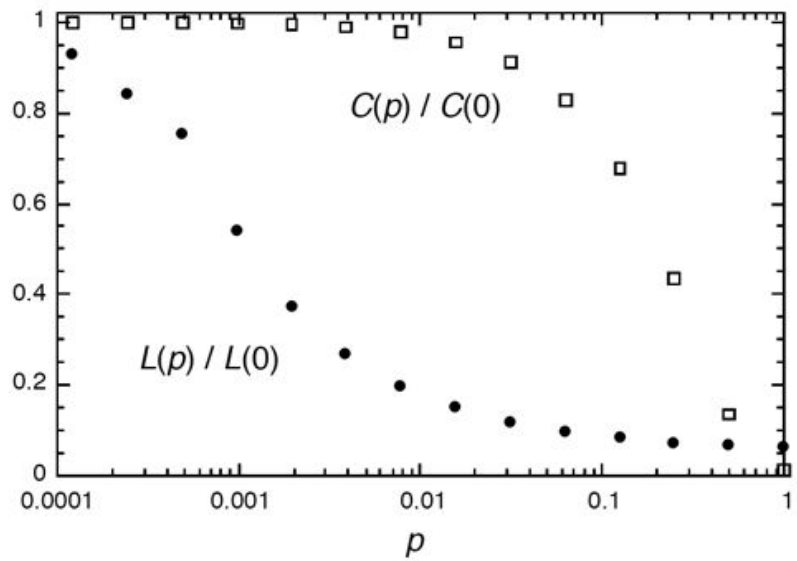
Small-world property

regular ring lattice

High clustering

High average path length

small worlds even with small concentrations of shortcuts



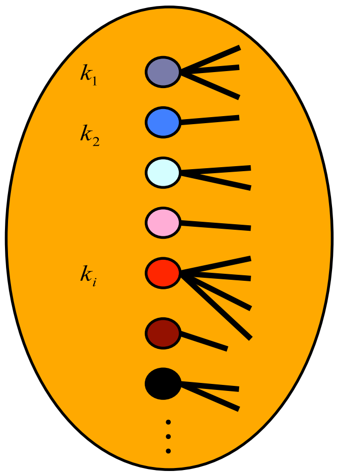
Collective dynamics of 'small-world' networks, D. J. Watts and S. H. Strogatz, Nature 393, 440-442 (1998)

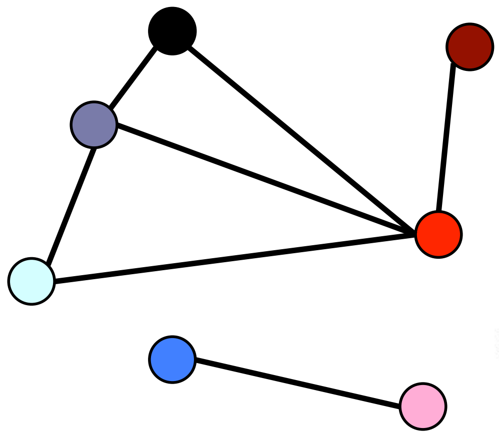
Equilibrium network models: CM

The **configuration model** is usually used as a null model for real complex networks. It produces maximally random networks with a preassigned degree sequence (uncorrelated except for structural unavoidable correlations necessary to close the networks)

$N \quad P(k) \qquad k_1 \quad k_2 \quad \cdots \quad k_i \quad \cdots \quad k_N$

ANND and clustering (except for structural correlations)





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

$$\bar{c}(k) = C = \frac{1}{N}\frac{(\langle k^2 \rangle - \langle k \rangle)^2}{\langle k \rangle^3}, \ k > 1$$

For weighted networks

M. A. Serrano and M. Boguñá. Weighted Configuration Model, AIP Conference Proceedings 776 (1), 101-107 (2005).

M. Molloy and B. Reed, Random Struct. Algorithms 6, 161 (1995)
M. Molloy and B. Reed, Combinatorics, Probab. Comput. 7, 295 (1998)

Structural correlations

Scale-free degree distributions - finite size

Natural cut-off : The expected maximum value of the degree (random variable) given the degree distrution and that the network is of finite size N

Definition: It is calculated taking into account that the number of nodes with degrees greater than the cut-off has to be of the order of 1

$$N \int_{k_{cut}}^{\infty} dk\, P(k) \sim 1$$

$$k_{\max} = \min\left(N-1, k_0 N^{1/(\gamma-1)}\right)$$

Power law degree distributions are observable only in large networks!!!

The same scaling is found from extreme value theory: if we draw N observations (random variables) from the degree distribution (many times), the maximum value of the sample will also be a random variable, the natural cut-off is the average value of the maximum (increasing with N)

Boguna, M; Pastor-Satorras, R; Vespignani, A;
EUROPEAN PHYSICAL JOURNAL B 38, 205-209 (2004)

Structural correlations

Structural cut-off for the degree distribution (for uncorrelated networks):

In real networks, the degrees are not drawn independetly from P(k), but must satisfy some topological constraints due to network structure and closeness

$$k_s(N) \sim (\langle k \rangle N)^{1/2}$$

The structural cut-off is calculated by imposing that the probabiliy that a link extist between two nodes of degrees k and k' is less or equal to 1 (physical constraint)

$$r_{kk'} = \frac{kk'}{\langle k \rangle N}$$

Boguna, M; Pastor-Satorras, R; Vespignani, A;
EUROPEAN PHYSICAL JOURNAL B 38, 205-209 (2004)

Structural correlations

$$k_c(N) \sim N^{1/(\gamma-1)} \qquad k_s(N) \sim (\langle k \rangle N)^{1/2}$$

The structural cut-off vs the natural cut-off

• The structural cut-off coincides with the natural cut-off when $\gamma=3$ (Barabasi-Albert networks)

• If $\gamma>3$, the structural cut-off diverges faster than the natural one and the later dominates

• If $\gamma<3$, the natural cut-off diverges faster than the structural one and the later dominates. If the actual cut-off is imposed to be larger than the structural cut-off, then some negative two-point correlations (or multiple connections) are necessary to fulfill network closeness contraints

Structural correlations!!!

Boguna, M; Pastor-Satorras, R; Vespignani, A;
EUROPEAN PHYSICAL JOURNAL B 38, 205-209 (2004)

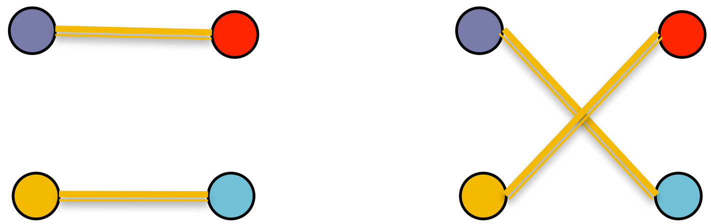
Equilibrium network models: Rewiring

Alternatively, one could apply a random rewiring process to links in order to destroy correlations (again, structural correlations necessary to close the network are not avoidable). If one wants to preserve the degree distribution:

i) select two different links in the network

ii) exchange one end of one of the two links for one of the ends of the second link (avoiding multiple and self-connections)

iii) repeat at least E times



Different possibilities for weighted networks

M. A. Serrano. Rich-club vs rich-multipolarization phenomena in weighted networks. Physical Review E 78, 026101 (2008).

Growing network models: BA

- **At each time step, add one vertex to the network and attach it to a randomly chosen vertex** Growing exponential network with exponential degree distribution
- **At each time step, add one vertex to the network and attach it to m randomly chosen vertices with probability proportional to their degree** Preferential attachment, BA model, power-law degree distributions

Growth alone or preferential attachment alone is not sufficient to produce scale-free degree distributions

A little bit of history about preferential attachment

First use to explain power-law distributions by Yule in 1925.

The modern master equation method was applied by H. A. Simon in 1955 in the course of studies of the sizes of cities and other phenomena.

It was first applied to the growth of networks by D. de Solla Price in 1976, who was interested in the networks of citation between scientific papers.

Growing network models: BA

- **At each time step, add one vertex to the network and attach it to m randomly chosen vertices with probability proportional to their degree**
  power-law degree distribution with exponent -3

$$\frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{2m}$$

<k>

Probability that a new edge attaches to a vertex of degree k

$$m \times kp_k/2m = \tfrac{1}{2}kp_k$$

Probability that m new edges attach to (m different) vertices of degree k

$$(n+1)p_{k,n+1} - np_{k,n} = \tfrac{1}{2}(k-1)p_{k-1,n} - \tfrac{1}{2}kp_{k,n},$$

for $k > m$, or

$$(n+1)p_{m,n+1} - np_{m,n} = 1 - \tfrac{1}{2}mp_{m,n},$$

for $k = m$, and there are no vertices with $k < m$.

Growing network models: BA

- **At each time step, add one vertex to the network and attach it to m randomly chosen vertices with probability proportional to their degree**
  power-law degree distribution with exponent -3

$$p_{k,n+1} = p_{k,n} = p_k$$ Stationary solutions

$$p_k = \frac{(k-1)(k-2)\ldots m}{(k+2)(k+1)\ldots(m+3)} p_m = \frac{2m(m+1)}{(k+2)(k+1)k}$$

In the limit of large degrees $p_k \sim k^{-3}$

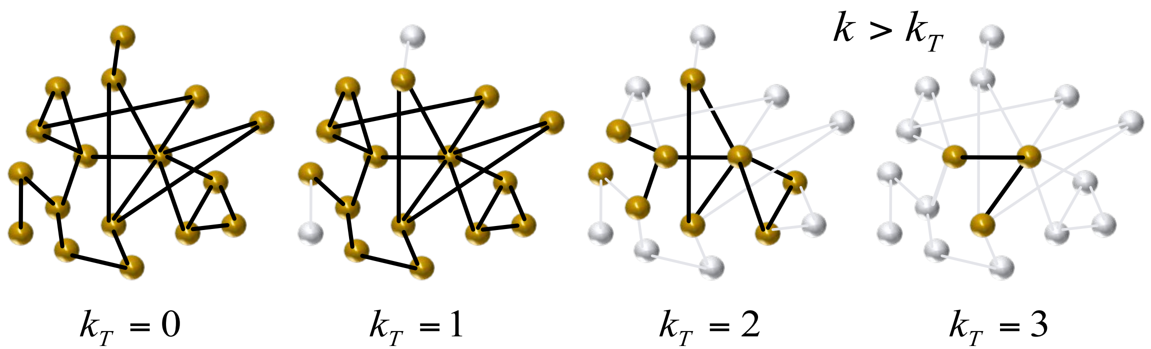
Other features

- Correlation between age and degree
- Very weak two point disassortative degree correlations that vanish in the thermodynamic limit
- Clustering vanishes in the thermodynamic limit

Importance of null models

Rich club phenomenon

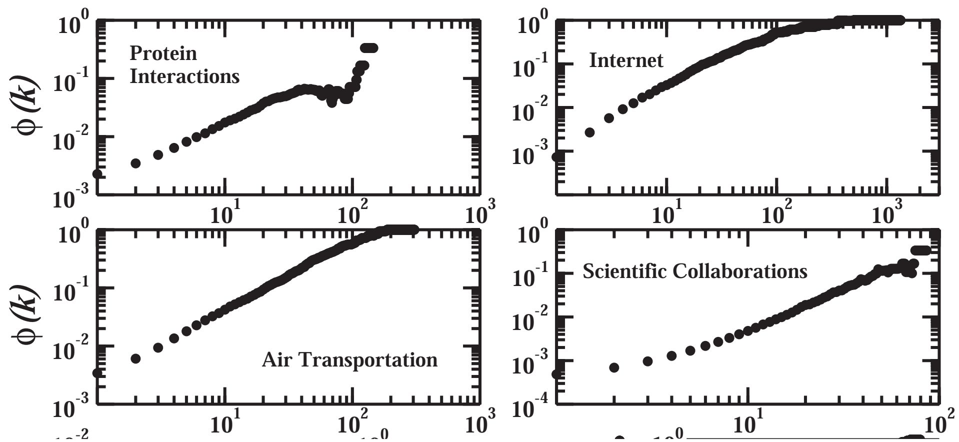
Zhou, S. & Mondragon, R. J. The rich-club phenomenon in the Internet topology. *IEEE Commun. Lett.* **8,** 180–182 (2004).



$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k}-1)}$$

[1]This nested hierarchy of subgraphs turns out to have self-similarity properties for some real scale-free networks such as the Internet at the autonomous system level; see [7].

M. A. Serrano, D. Krioukov, and M. Boguñá, Phys. Rev. Lett. **100**, 078701 (2008).

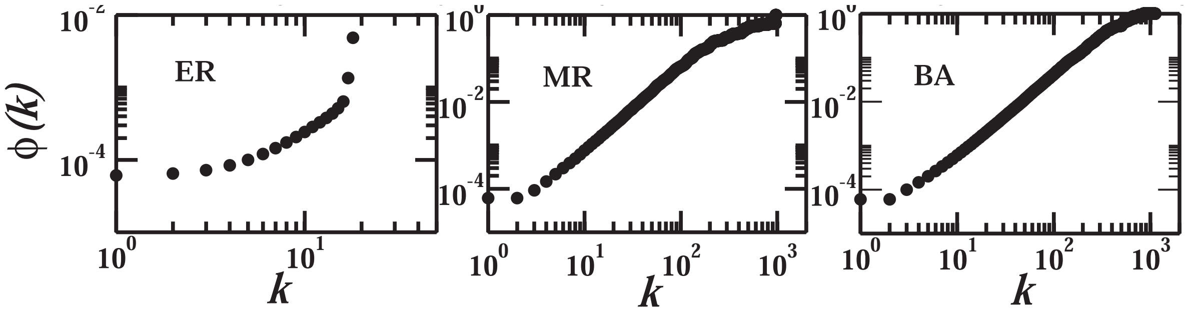


Original interpretation: the growing behavior denotes the presence of rich clubs

Importance of null models

Rich club phenomenon

The same for uncorrelated networks



$$\phi(k) = \frac{N\langle k\rangle \int_k^{k_{\max}} dk' \int_k^{k_{\max}} dk'' P(k',k'')}{\left[N\int_k^{k_{\max}} dk' P(k')\right]\left[N\int_k^{k_{\max}} dk' P(k') - 1\right]}$$

$$P_{\mathrm{unc}}(k,k') = kk'P(k)P(k')/\langle k\rangle^2$$

$$\boxed{\phi_{\mathrm{unc}}(k) = \frac{1}{N\langle k\rangle}\left[\frac{\int_k^{k_{\max}} dk' k' P(k')}{\int_k^{k_{\max}} dk' P(k')}\right]^2 \underset{k,k_{\max}\to\infty}{\sim} \frac{k^2}{\langle k\rangle N}}$$

V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani. Detecting rich-club ordering in complex networks, Nature Physics 2, 110-115 (2006).
M. A. Serrano. Rich-club vs rich-multipolarization phenomena in weighted networks. Physical Review E 78, 026101 (2008).

Importance of null models

Rich club phenomenon

Compare $\phi(k) = \dfrac{2E_{>k}}{N_{>k}(N_{>k}-1)}$ with $\phi_{\text{expected}}(k)$

Structural correlations have to be discounted!!!

$$\rho_{\text{unc}}(k) = \phi(k)/\phi_{\text{unc}}(k)$$

$$\phi_{\text{unc}}(k) = \frac{1}{N\langle k\rangle}\left[\frac{\int_k^{k_{\max}} \mathrm{d}k'\, k' P(k')}{\int_k^{k_{\max}} \mathrm{d}k'\, P(k')}\right]^2 \underset{k,\, k_{\max}\to\infty}{\sim} \frac{k^2}{\langle k\rangle N}$$

Not feasible when structural correlations are present $P(k) \approx k^{-\gamma}$

$$\rho_{\text{ran}}(k) = \phi(k)/\phi_{\text{ran}}(k)$$

Maximally random network (MRN) with the same P(k)

Configuration model or rewiring

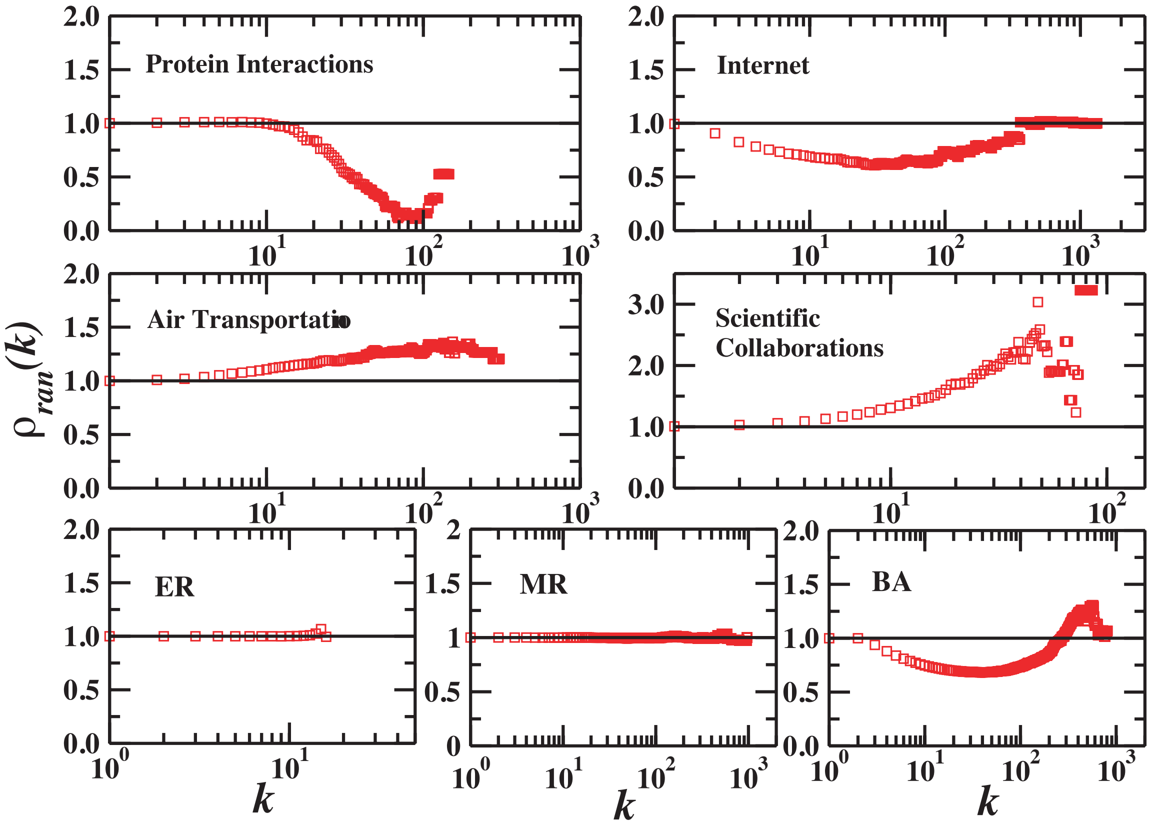
$\rho_{\text{ran}}(k) > 1$ ⇒ RC effect

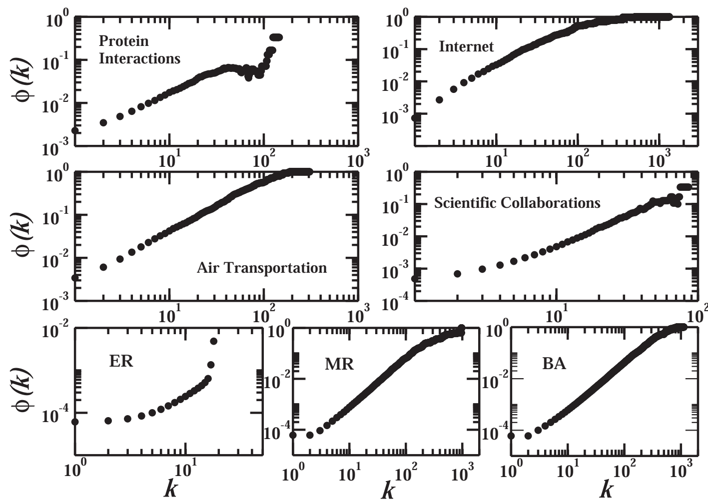
V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani. Detecting rich-club ordering in complex networks, Nature Physics 2, 110-115 (2006).

M. A. Serrano. Rich-club vs rich-multipolarization phenomena in weighted networks. Physical Review E 78, 026101 (2008).

Importance of null models

Rich club phenomenon





V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani. Detecting rich-club ordering in complex networks, Nature Physics 2, 110-115 (2006).
M. A. Serrano. Rich-club vs rich-multipolarization phenomena in weighted networks. Physical Review E 78, 026101 (2008).

References

Reviews

[R1] *Statistical mechanics of complex networks,* R. Albert, and A.-L. Barabási *, Rev. Mod. Phys.* **74**, 47-97 (2002)

[R2] *The structure and function of complex networks*, M. E. J. Newman, *SIAM Review* **45**, 167–256, 2003

[R3] *Critical phenomena in complex networks*, S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008)

Books

[B1] *Evolution of Networks: From Biological Nets to the Internet and WWW*, S. N. Dorogovtsev and J. F. F. Mendes, Oxford University Press (2003)

[B2] *Networks: An Introduction*, M. E. J. Newman, Oxford University Press (2010)

[B3] *Dynamical Processes on Complex Networks*, A. Barrat, M. Barthélémy, and A. Vespignani, Cambridge University Press (2012)

Popular science

[P1] A.-L. Barabási. *Linked: How Everything Is Connected to Everything Else and What It Means*

[P2] D. J. Wats. *Small Worlds: The Dynamics of Networks between Order and Randomness*

[P3] R. Solé. *Redes Complejas. Del genoma a Internet*