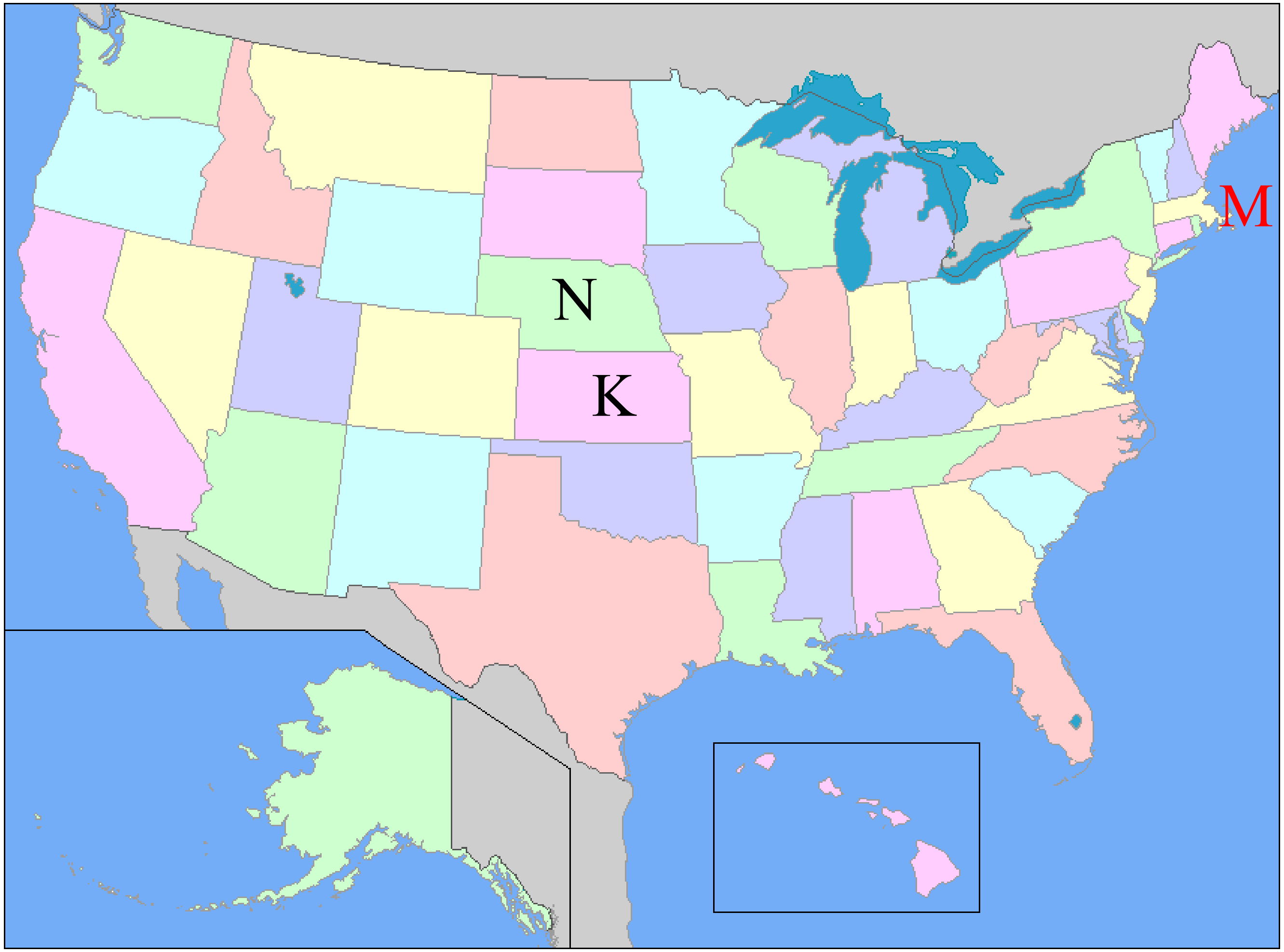
Financial and economic networks

Rosario Nunzio Mantegna
Central European University, Budapest, Hungary
Palermo University, Palermo, Italy

Milgram's 1967 "small world" experiment



Is it possible to deliver a message to a stock dealer in Boston starting from randomly extracted people in Nebraska and Kansas?

The small world effect: the Milgram's experiment (1967-1969)

1. **Starting points** of the chain of correspondence: **Omaha**, Nebraska and **Wichita**, Kansas. **End point** of the chain of correspondence: **Boston**, Massachusetts.

1. **Information** packets were **initially sent to "randomly" selected individuals in Omaha or Wichita**. They included letters, which detailed the study's purpose, and **basic information about a target contact person in Boston**. It additionally contained a roster on which they could write their own name, as well as business reply cards that were pre-addressed to Harvard.

1. Upon receiving the invitation to participate, the recipient was asked whether he or she personally knew the contact person described in the letter. If so, the person was to forward the letter directly to that person.

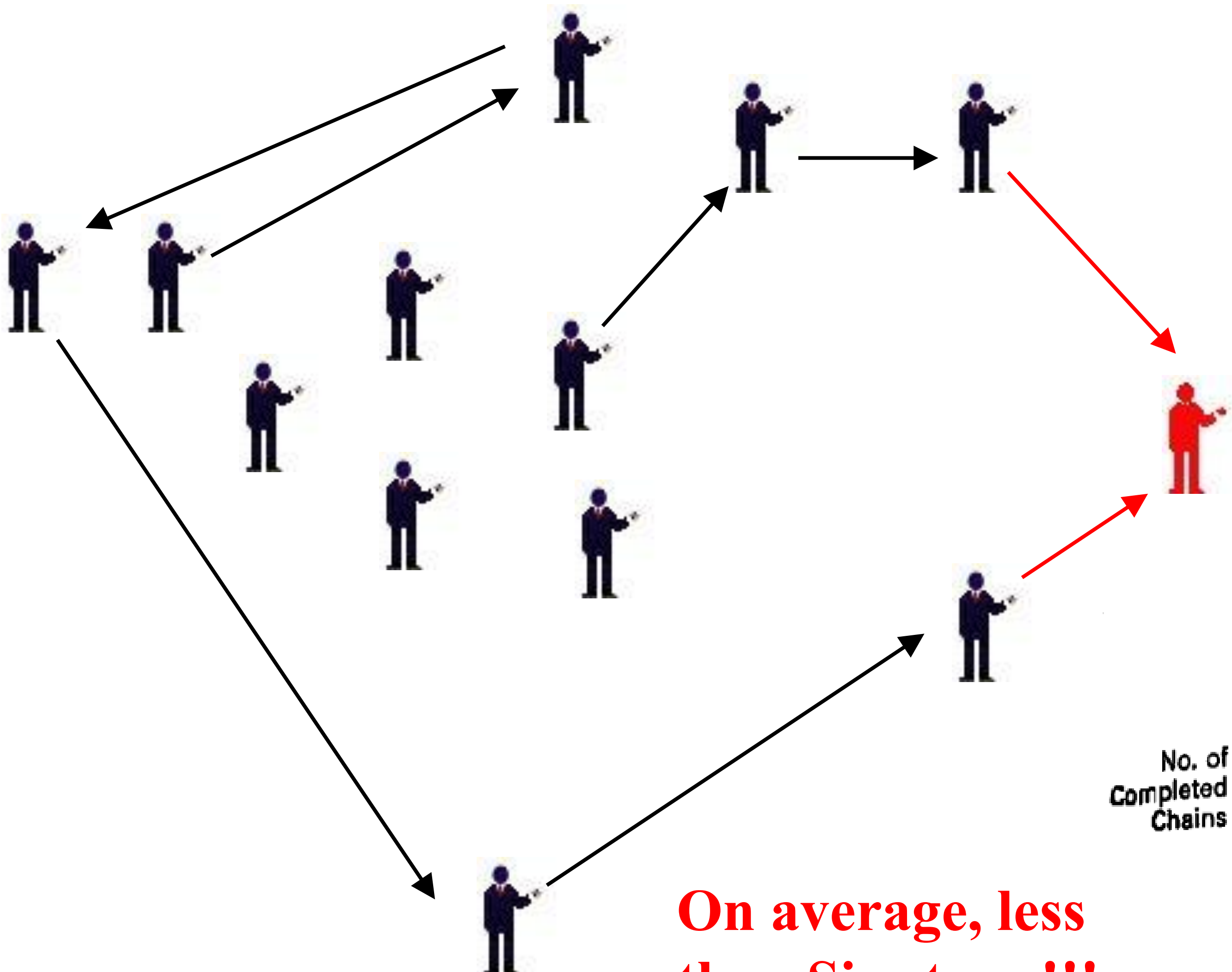
2. In the more likely case that the person did not personally know the target, then **the person was to think of a friend or relative they know personally that is more likely to know the target**. They were then directed to sign their name on the roster and forward the packet to that person. A postcard was also mailed to the researchers at Harvard so that they could track the chain's progression toward the target.

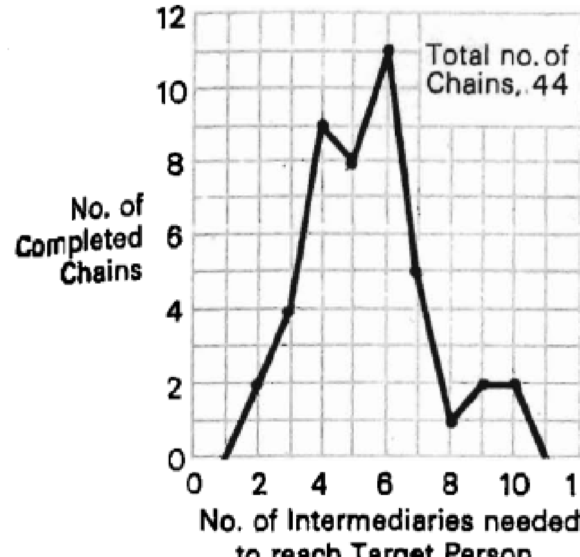
1. **When** and if **the package** eventually **reached the contact person in Boston**, the researchers could examine the roster to **count the number of times it had been forwarded from person to person**. Additionally, for packages that never reached the destination, the incoming postcards helped identify the break point in the chain.

The Small-World Problem

By Stanley Milgram

Six degrees of separation





On average, less than Six steps !!!

Network studies have been an area of research in Social Sciences since the fifties of the last century

Robust Action and the Rise of the Medici, 1400-1434

John F. Padgett and Christopher K. Ansell
University of Chicago

AJS Volume 98 Number 6 (May 1993): 1259–1319

American Journal of Sociology

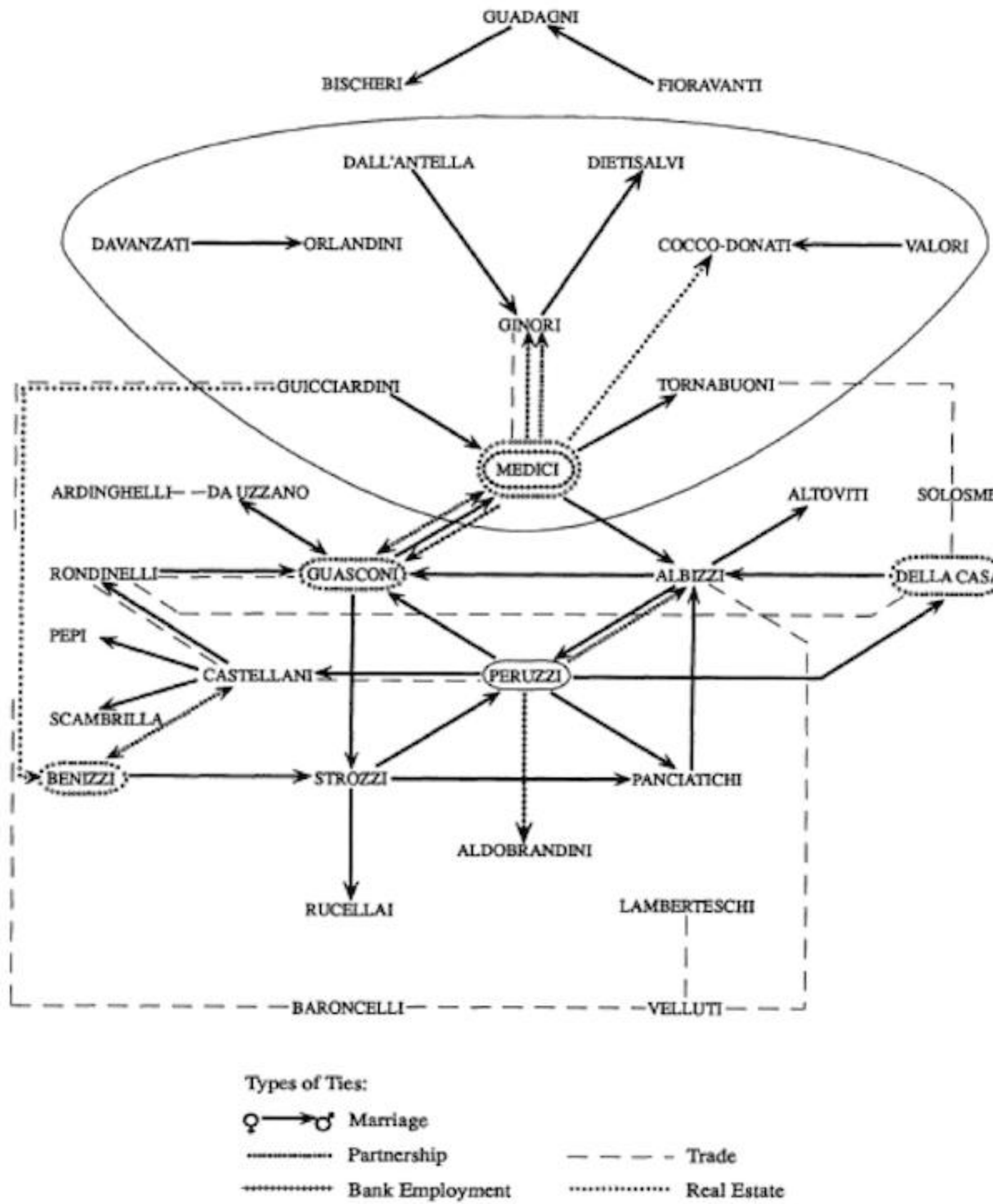


FIG. 2a.—Marriage and economic blockmodel structure (92 elite families)

Feb 4th, 2016

Mathematicians have modeled networks since 1950s. One prominent model of a class of networks is the random network

Erdös-Rényi model (1959-1961)

On random graphs I.

Dedicated to O. Varga, at the occasion of his 50^{th} birthday.

By P. ERDŐS and A. RÉNYI (Budapest).

d

Erdös-Rényi model (1959-1961)

On random graphs I.

Dedicated to O. Varga, at the occasion of his 50[th] birthday.

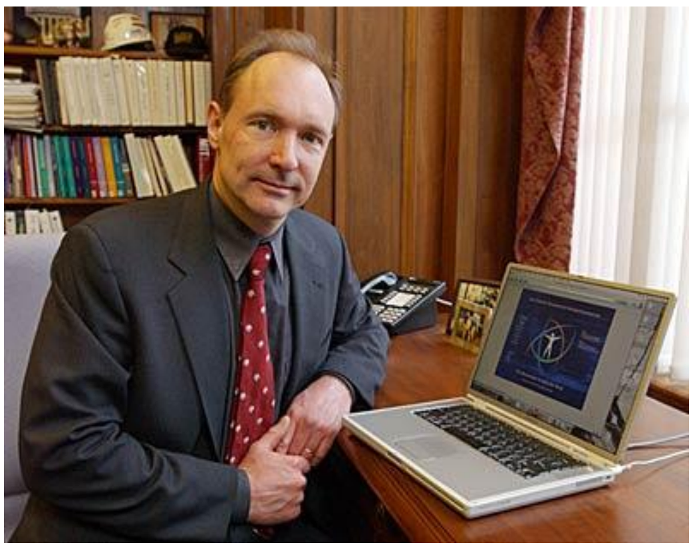
By P. ERDŐS and A. RÉNYI (Budapest).

Let us consider a "random graph" $\Gamma_{n,N}$ having n possible (labelled) vertices and N edges; in other words, let us choose at random (with equal probabilities) one of the $\binom{\binom{n}{2}}{N}$ possible graphs which can be formed from the n (labelled) vertices $P_1, P_2, \ldots, P_n$ by selecting N edges from the $\binom{n}{2}$ possible edges $\widehat{P_iP_j}$ $(1 \leqq i < j \leqq n)$. Thus the effective number of vertices of $\Gamma_{n,N}$ may be less than n, as some points P_i may be not connected in $\Gamma_{n,N}$ with any other point P_j; we shall call such points P_i *isolated* points. We consider the isolated points also as belonging to $\Gamma_{n,N}$. $\Gamma_{n,N}$ is called completely connected if it effectively contains all points $P_1, P_2, \ldots, P_n$ (i, e. if it has no isolated points) and is connected in the ordinary sense. In the present paper we consider asymptotic statistical properties of random graphs for $n \to +\infty$. We shall deal with the following questions:

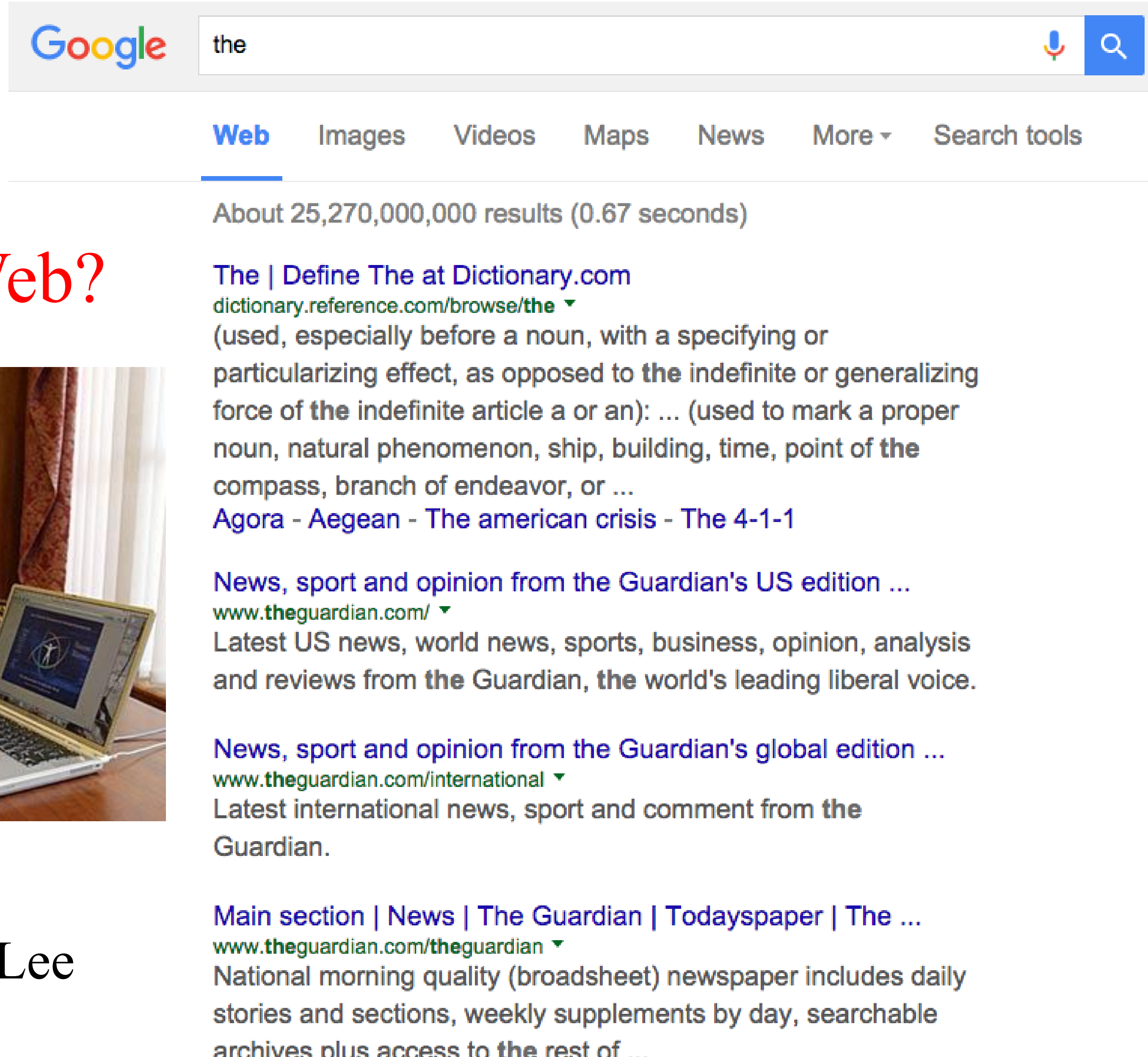
1. What is the probability of $\Gamma_{n,N}$ being completely connected?
2. What is the probability that the greatest connected component (sub-graph) of $\Gamma_{n,N}$ should have effectively $n-k$ points? $(k=0, 1, \ldots)$.
3. What is the probability that $\Gamma_{n,N}$ should consist of exactly $k+1$ connected components? $(k=0, 1, \ldots)$.
4. If the edges of a graph with n vertices are chosen successively so that after each step every edge which has not yet been chosen has the same probability to be chosen as the next, and if we continue this process until the graph becomes completely connected, what is the probability that the number of necessary steps ν will be equal to a given number l?

Why networks become so interesting at the end of the 1990s?

How big is the World Wide Web?



Sir Tim Berners-Lee



Google

the

Web Images Videos Maps News More Search tools

About 25,270,000,000 results (0.67 seconds)

The | Define The at Dictionary.com
dictionary.reference.com/browse/the
(used, especially before a noun, with a specifying or particularizing effect, as opposed to the indefinite or generalizing force of the indefinite article a or an): ... (used to mark a proper noun, natural phenomenon, ship, building, time, point of the compass, branch of endeavor, or ...
Agora - Aegean - The american crisis - The 4-1-1

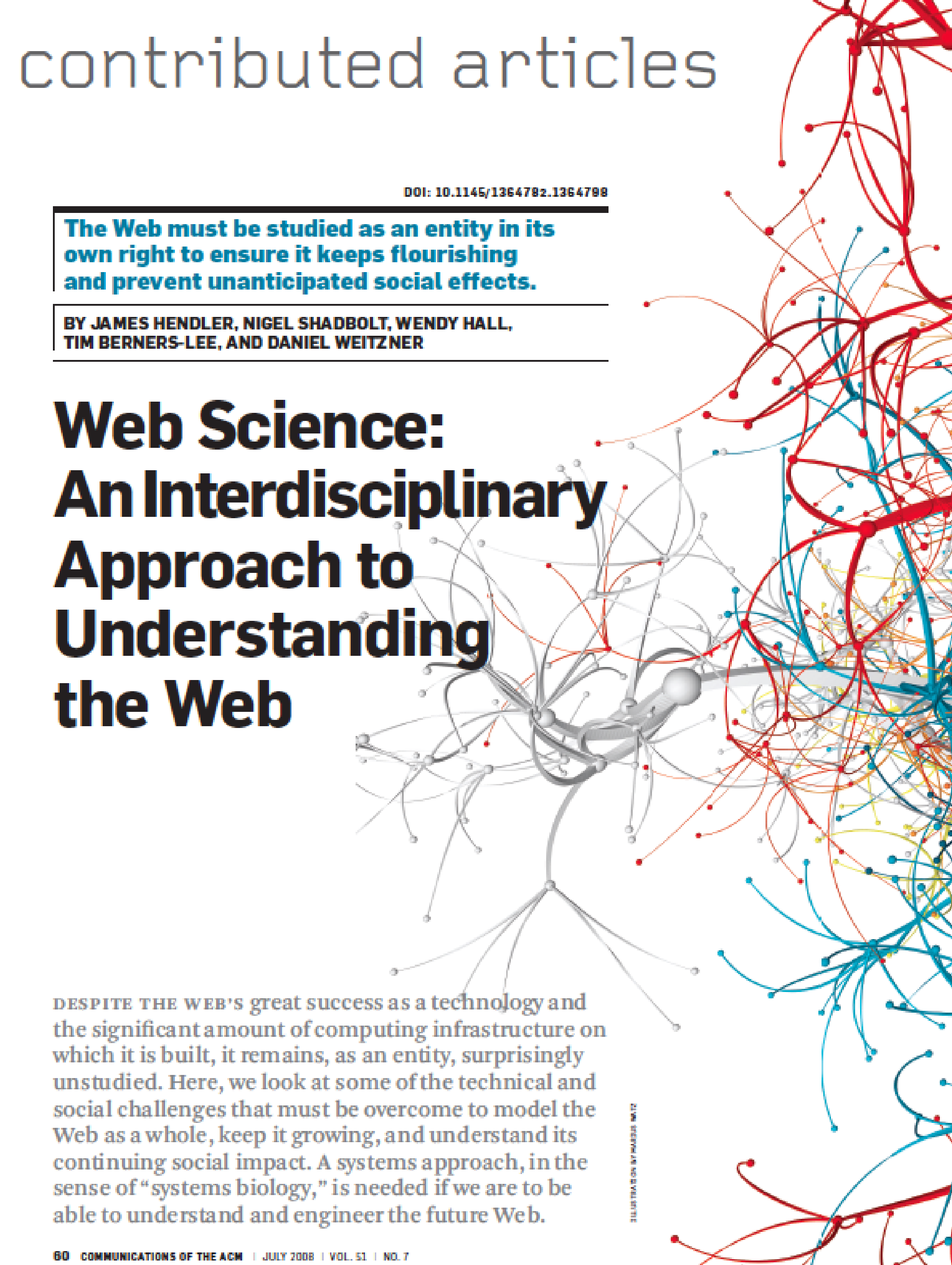
News, sport and opinion from the Guardian's US edition ...
www.theguardian.com/
Latest US news, world news, sports, business, opinion, analysis and reviews from the Guardian, the world's leading liberal voice.

News, sport and opinion from the Guardian's global edition ...
www.theguardian.com/international
Latest international news, sport and comment from the Guardian.

Main section | News | The Guardian | Todayspaper | The ...
www.theguardian.com/theguardian
National morning quality (broadsheet) newspaper includes daily stories and sections, weekly supplements by day, searchable archives plus access to the rest of ...

The WWW is a dynamic entity in continuous development. Tim Berners-Lee and collaborators have proposed to consider it as an object of scientific study developing a "*web science*".

James Hendler, Nigel Shadbolt, Wendy Hall, Tim Berners-Lee, and Daniel Weitzner. Web science: An interdisciplinary approach to understanding the Web. Communications of the ACM, 51(7):60-69, 2008.



contributed articles

DOI: 10.1145/1364782.1364798

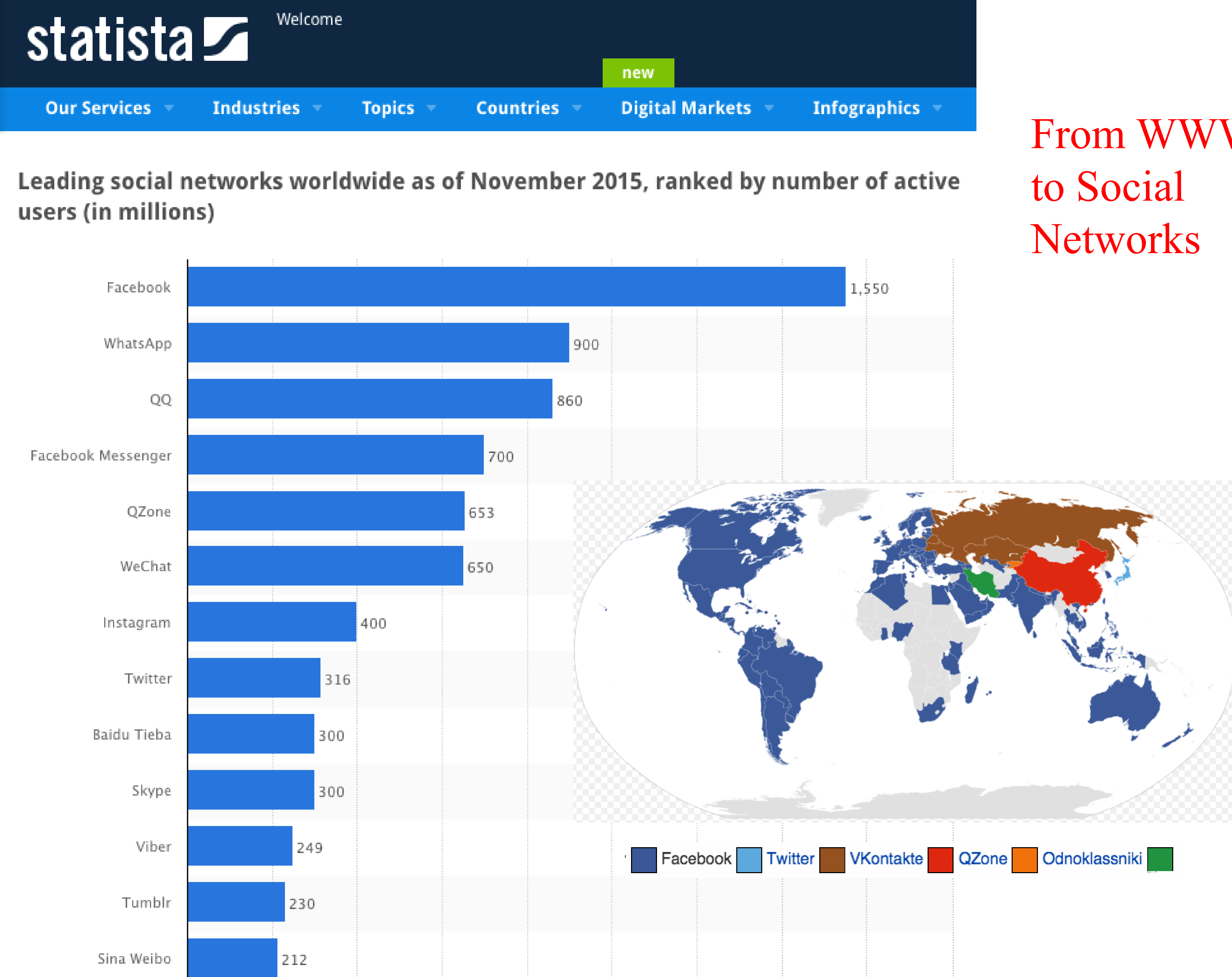
The Web must be studied as an entity in its own right to ensure it keeps flourishing and prevent unanticipated social effects.

BY JAMES HENDLER, NIGEL SHADBOLT, WENDY HALL, TIM BERNERS-LEE, AND DANIEL WEITZNER

Web Science: An Interdisciplinary Approach to Understanding the Web

DESPITE THE WEB'S great success as a technology and the significant amount of computing infrastructure on which it is built, it remains, as an entity, surprisingly unstudied. Here, we look at some of the technical and social challenges that must be overcome to model the Web as a whole, keep it growing, and understand its continuing social impact. A systems approach, in the sense of "systems biology," is needed if we are to be able to understand and engineer the future Web.

From WWW to Social Networks



Networks were and are investigated in several different disciplines.
It is therefore a genuinely multidisciplinary research topic.

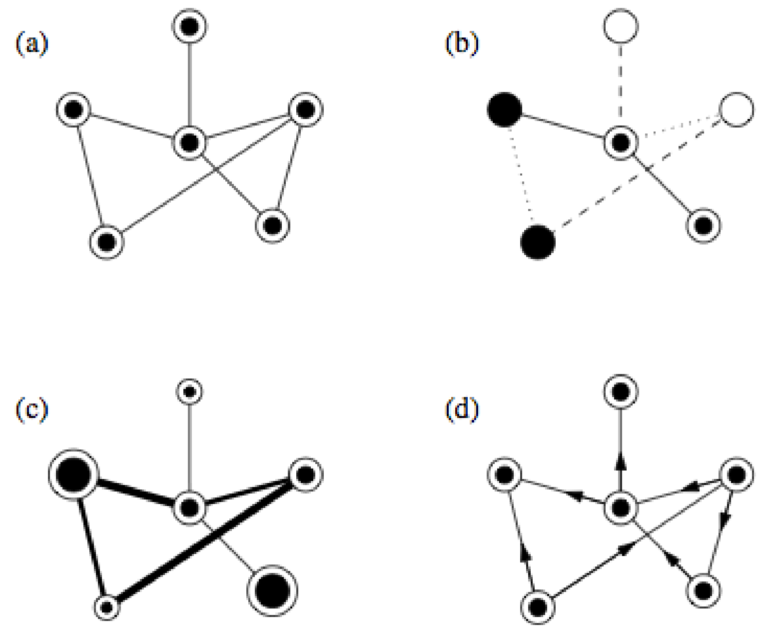
Networks are investigated in

Social sciences
Mathematics
Computer science
Statistical physics
Economics
Biology
Physiology

In its more general definition a *network* is "a collection of interconnected things" (Oxford English Dictionary).

In network modeling and network science "*network*" is often interchanged with the term "*graph*".

Types of networks



Examples of various types of networks: (a) an undirected network with only a single type of vertex and a single type of edge; (b) a network with a number of discrete vertex and edge types; (c) a network with varying vertex and edge weights; (d) a directed network in which each edge has a direction.

Newman, Mark EJ. "The structure and function of complex networks." SIAM review 45, no. 2 (2003): 167-256.

A graph $G=(V, E)$ is a mathematical object consisting of a set V of vertices (also called nodes) and a set E of edges (also called links).

Edges are defined in terms of unordered pairs $\{u,v\}$ of distinct vertices u,v belonging to the set V.

The number of vertices N_v is sometimes called the *order* of the graph

The number of edges N_E is sometimes called the *size* of the graph

A graph $H=(V_H,E_H)$ is a subgraph of $G=(V_G,E_G)$ if
$V_H \square V_G$ and $E_H \square E_G$

A graph $H'=(V_{H'},E_{H'})$ is an induced subgraph of $G'=(V_{G'},E_{G'})$ if
$V_{H'} \square V_{G'}$ is a pre-specified set of vertices and $E_{H'} \square E_{G'}$ are the edges observed among them

A graph where the relationship between vertex u and vertex v presents a directionality is called a directed graph or digraph.

Directed graphs have directed edges also called arcs

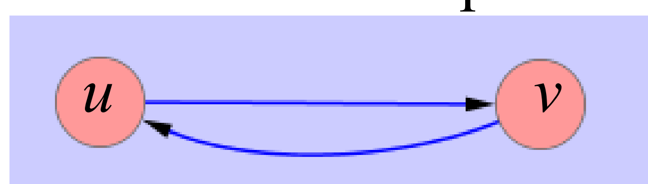
In a directed graph the direct edge $\{u,v\}$ is therefore different for $\{v,u\}$. Conventionally, the formalism $\{u,v\}$ reads tail u to head v

$\{u,v\}$

$\{v,u\}$

When we observe more than one type of directed edge between two vertices we are in the presence of a multi-digraphs

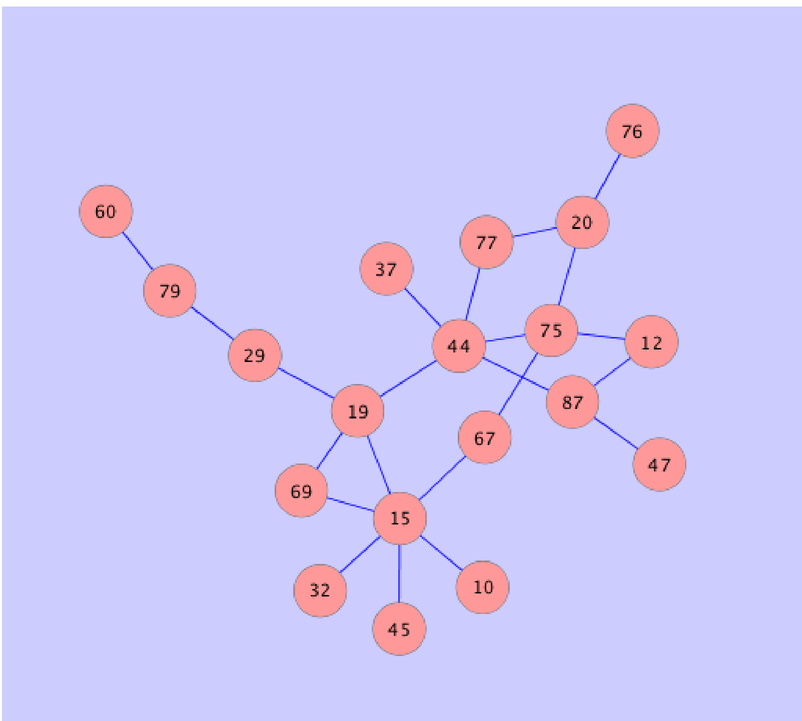
Note that in a simple digraph we might have up to 2 directed edges between two vertices. When both are present we say that the two arcs are mutual.



Two vertices $u, v \in V$ are adjacent if joined by an edge in E

Two edges $e_1, e_2 \in E$ are adjacent if connected by a vertex in V

The degree of a vertex is the number of incident edges on it

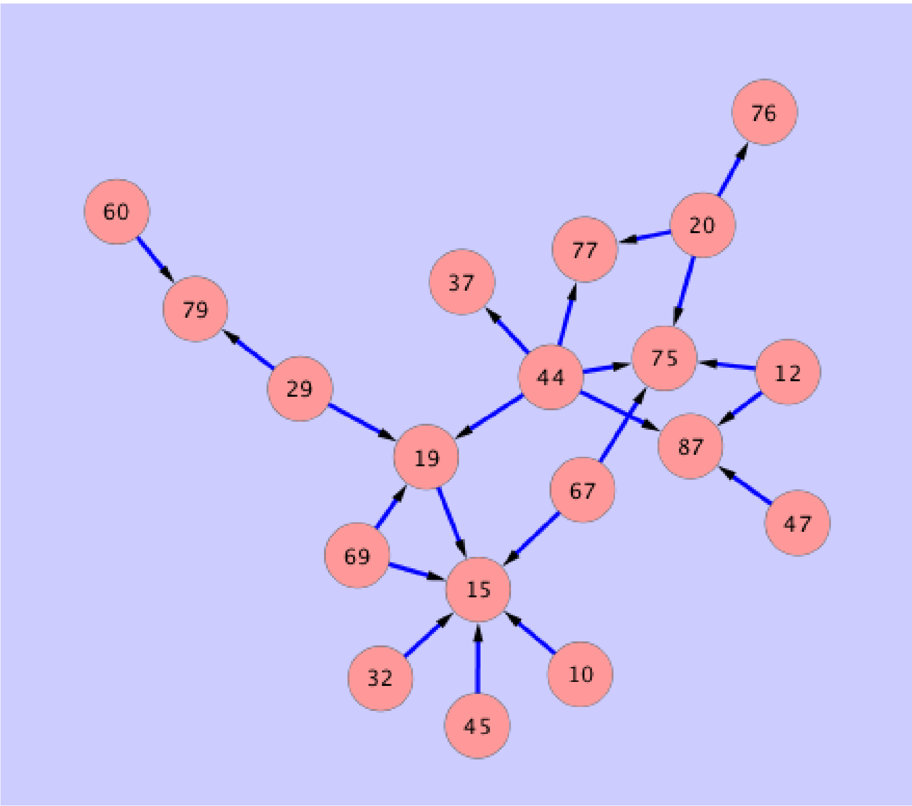


The degree sequence is obtained by arranging the degree of vertices in non-decreasing order

{1,1,1,1,1,1,1,2,2,2,2,2,2,3,3,4,4,5,6}

{10,32,37,45,47,60,76,12,29,67,69,
77,79,20,87,19,75,44,15}

In a digraph we have in-degree and out-degree



In the present example, the in-degree sequence is

{0,0,0,0,0,0,0,0,0,0,0,1,1,1,2,3,3,4,6}

{10,12,20,29,32,44,45,47,60,67,69, 37,76,79,77,19,87,75,15}

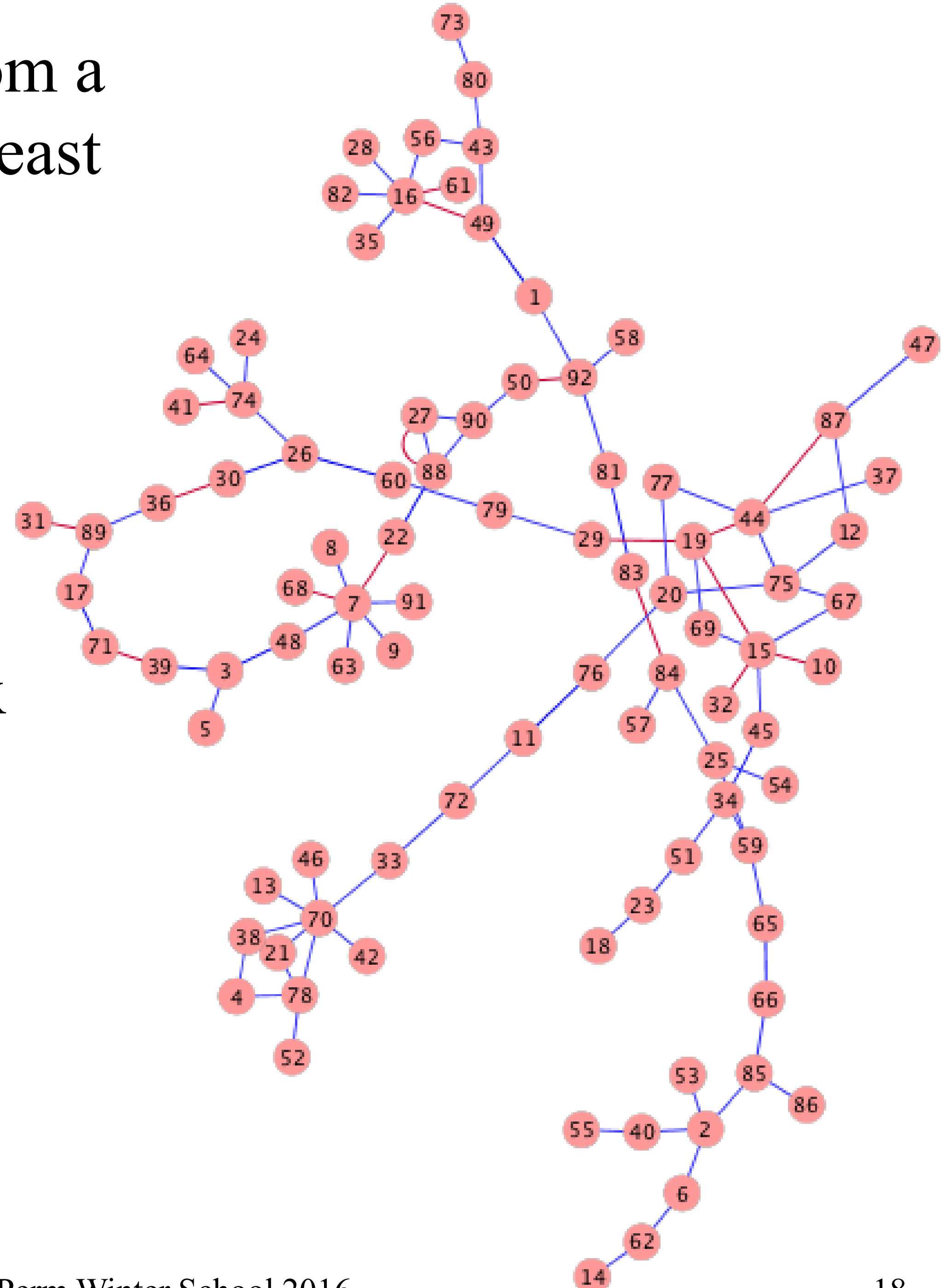
The out-degree sequence is

{0,0,0,0,0,0,1,1,1,1,1,1,1,2,2,2,2,3,5}

{15,37,75,76,77,87,10,19,32,45,47, 60,79,12,29,67,69,20,44}

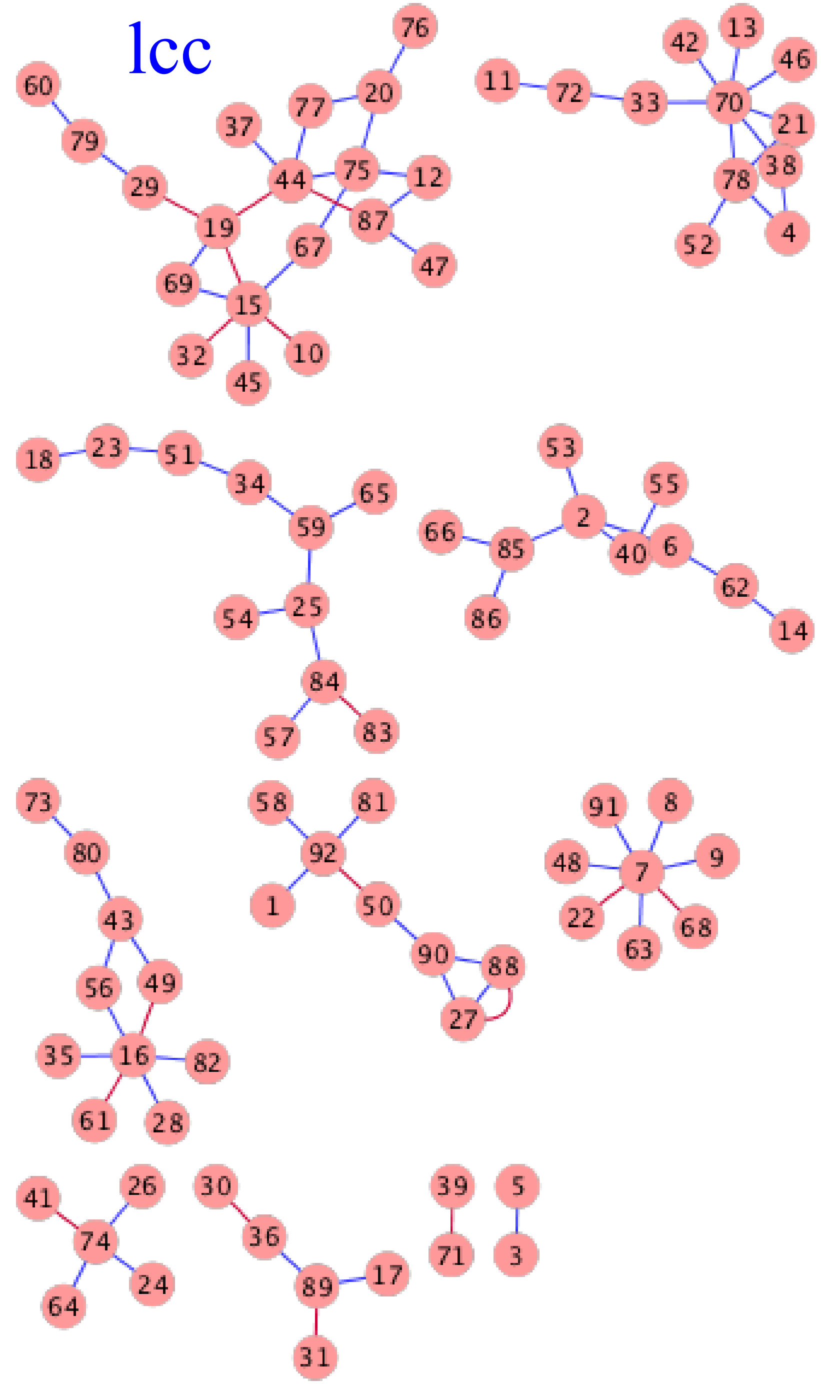
A vertex v is reachable from a vertex u if there exists at least a walk connecting u to v

A graph G is said to be connected if every vertex is reachable for every other one



A component of a graph is a maximally connected subgraph

The component with the largest number of vertices is called the largest connected component

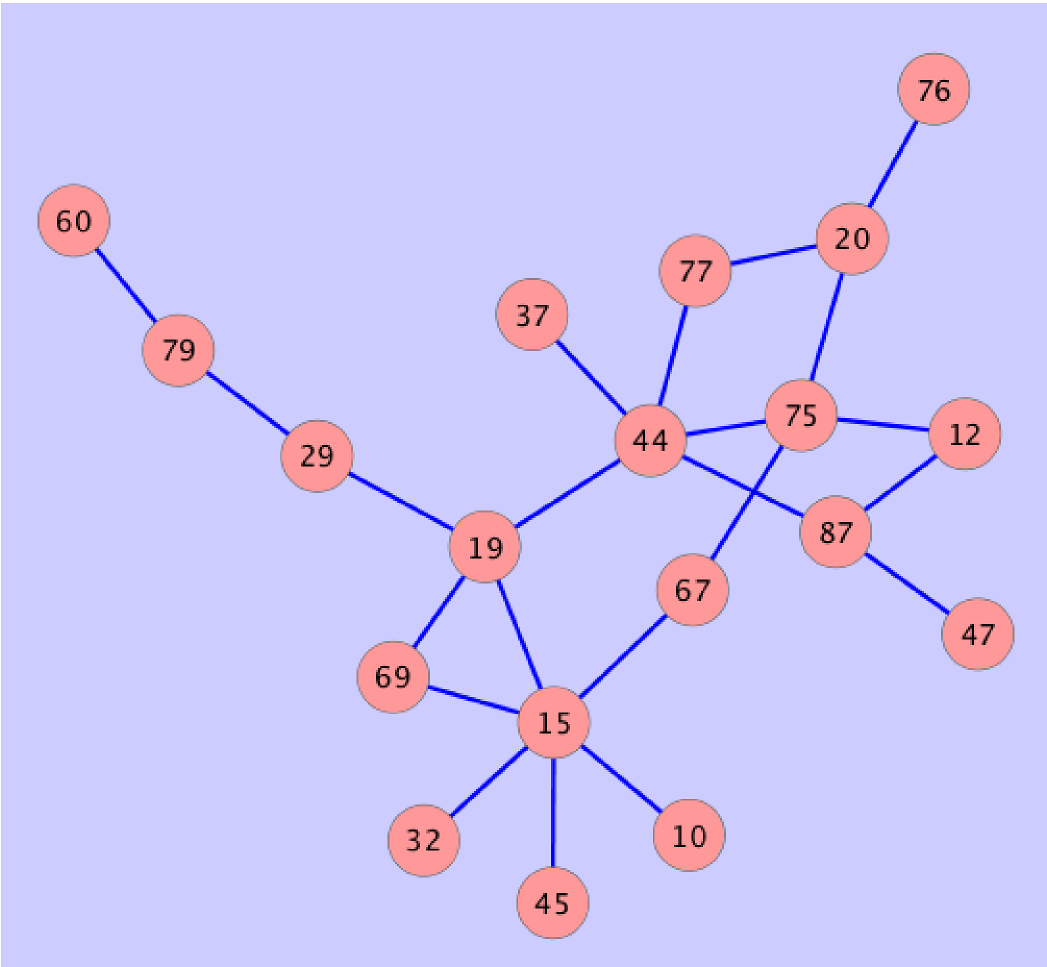


In the case of digraphs the concept of connectedness is specialized in two cases.

A digraph is weakly connected if the underlying undirected graph is connected

A digraph is strongly connected if every vertex v is reachable by every vertex u through a directed walk

A widely used notion of distance between two vertices of a graph is defined as the length of the shortest path(s) between the two vertices



Number of nodes 19

Average path length 3.053

Diameter 7
(see the paths from 60 to 76)

The **diameter** of a network is the maximum length of shortest paths

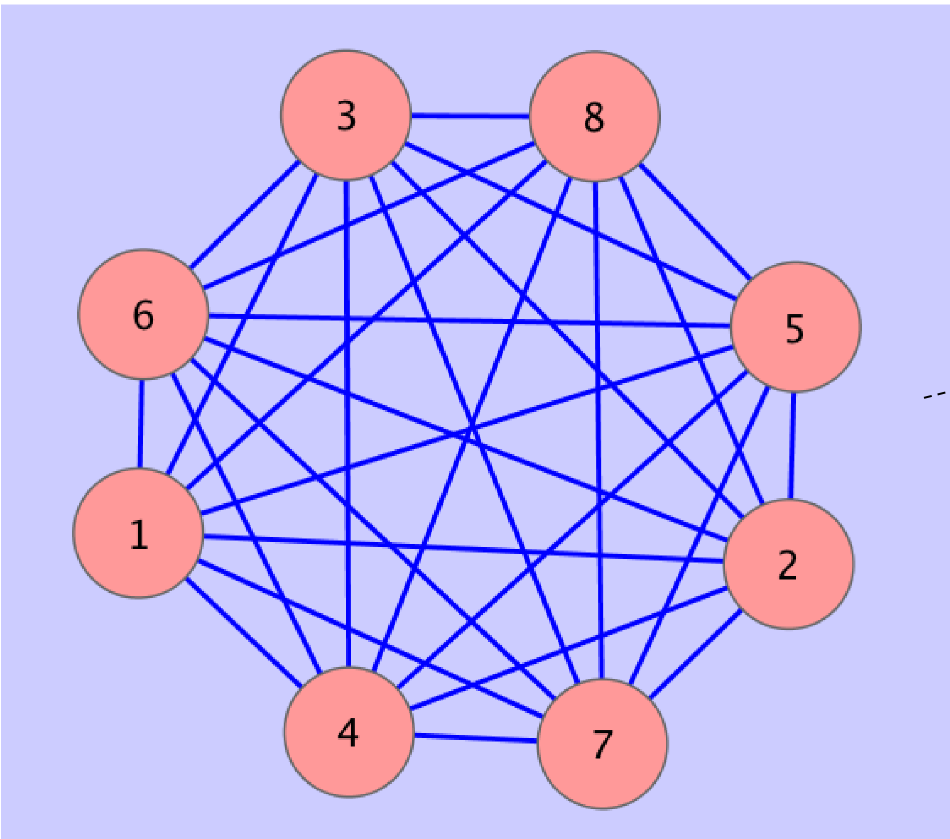
When an edge of a graph has associated a numerical weight the graph is called a weighted graph.

When edges are weighted the length of a walk (trail, path, etc) is defined as the sum of the values of the edges composing the walk. The distance is always defined as the weighted length of the shortest path.

The notion of degree is generalized to take into account the weights of the edges. The generalization is called the strength of the vertex and it is the sum of the weights of all incident edges.

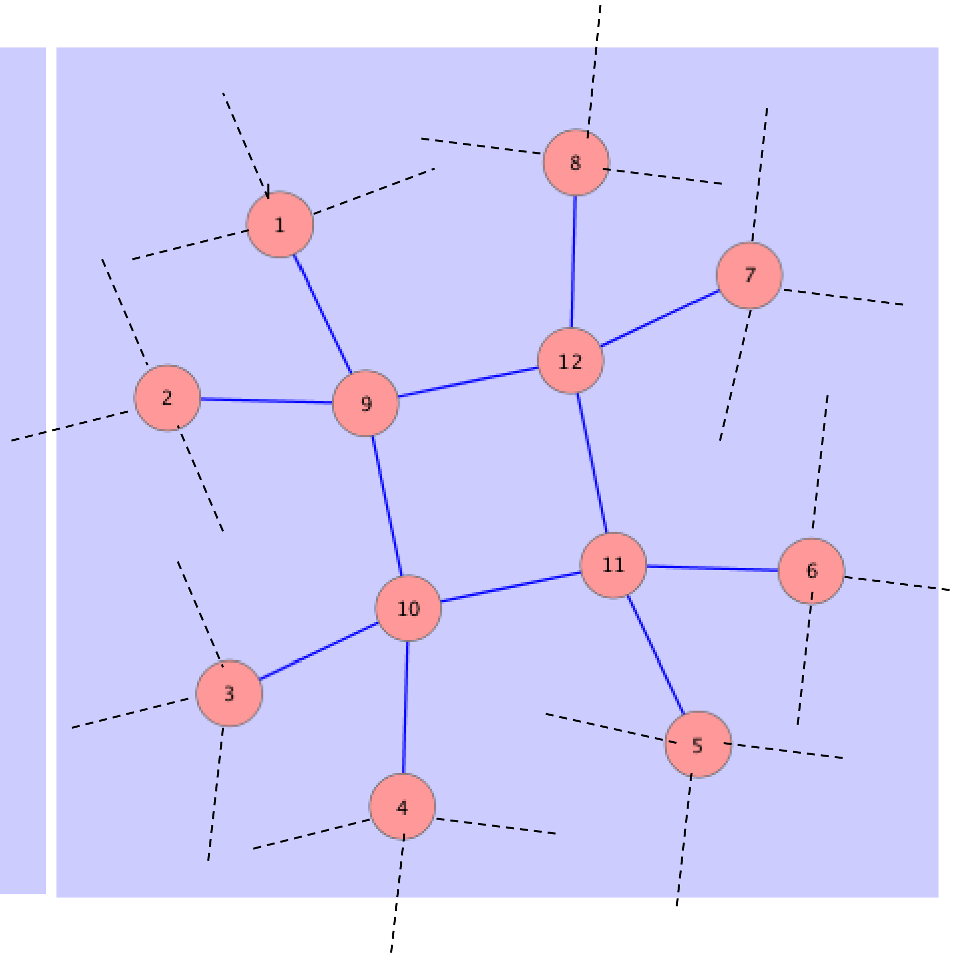
Some reference graphs

A complete graph is a graph where every vertex is linked to every other vertex.

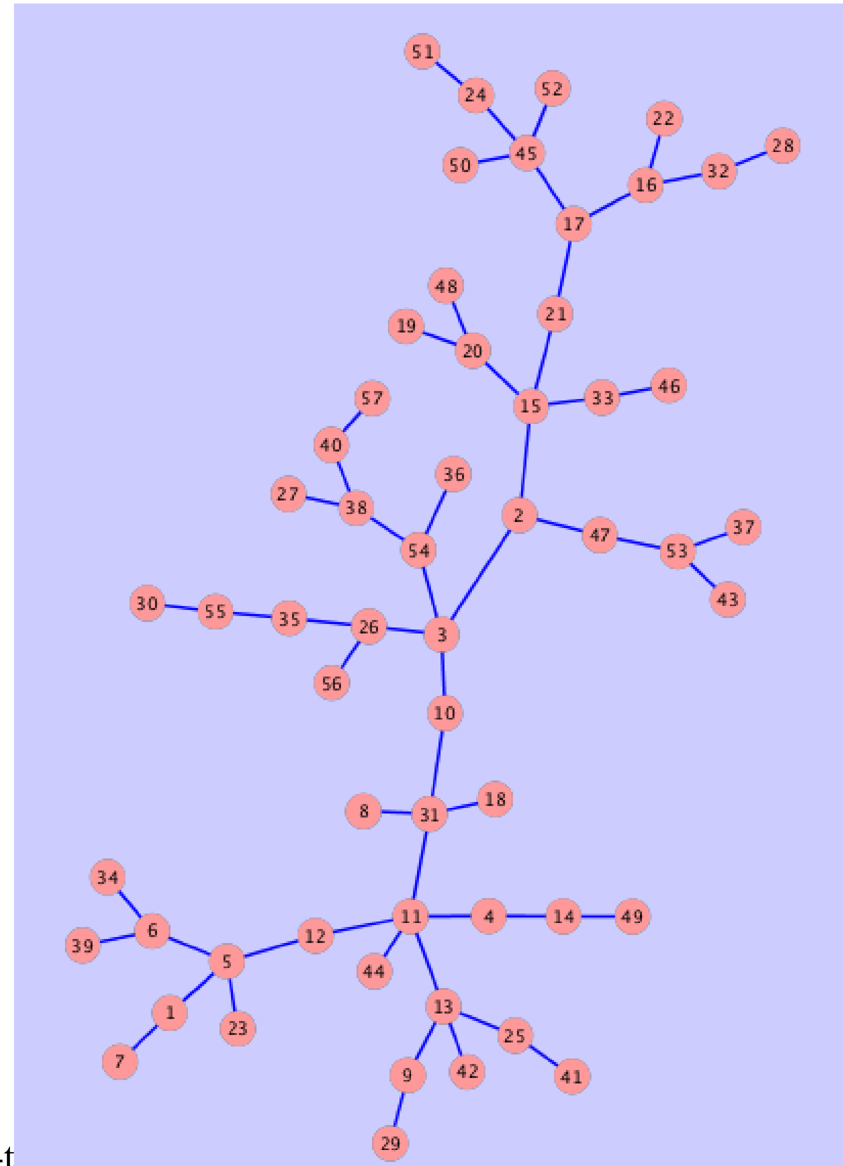


A complete subgraph is called a clique

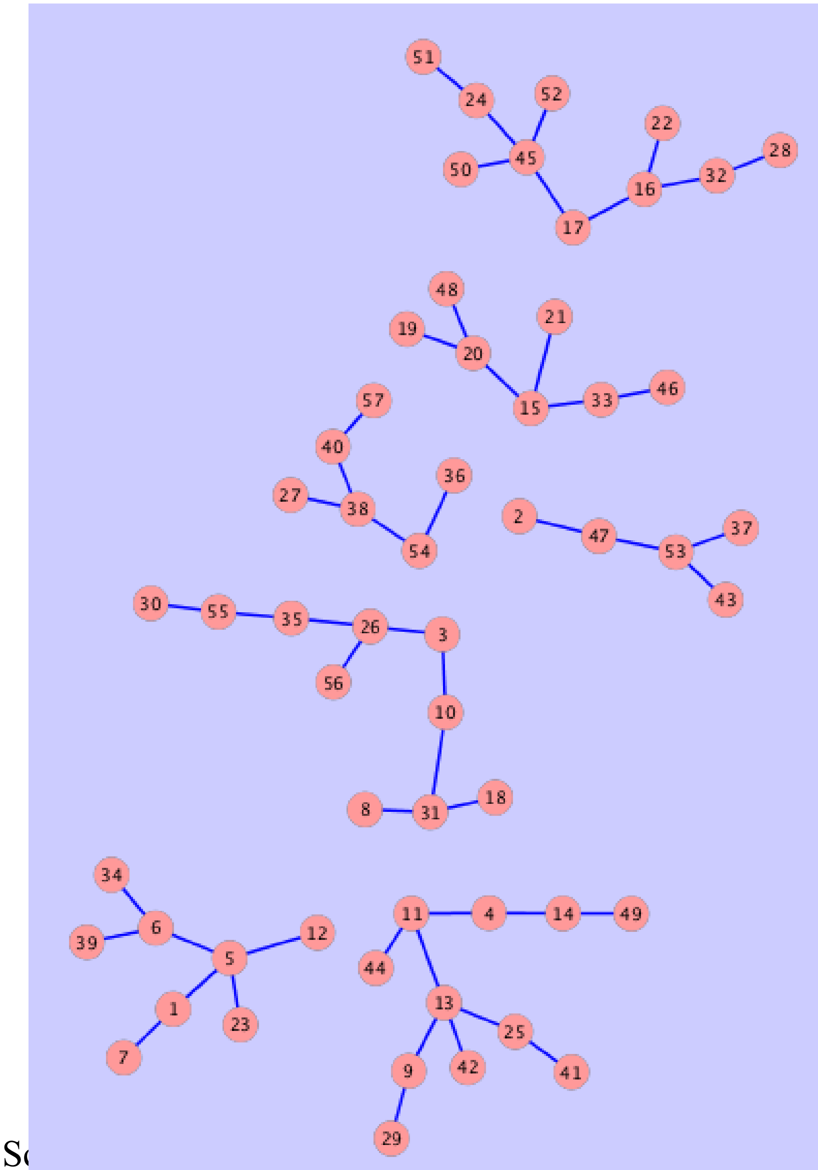
A d-regular graph is a graph where all vertices have degree d



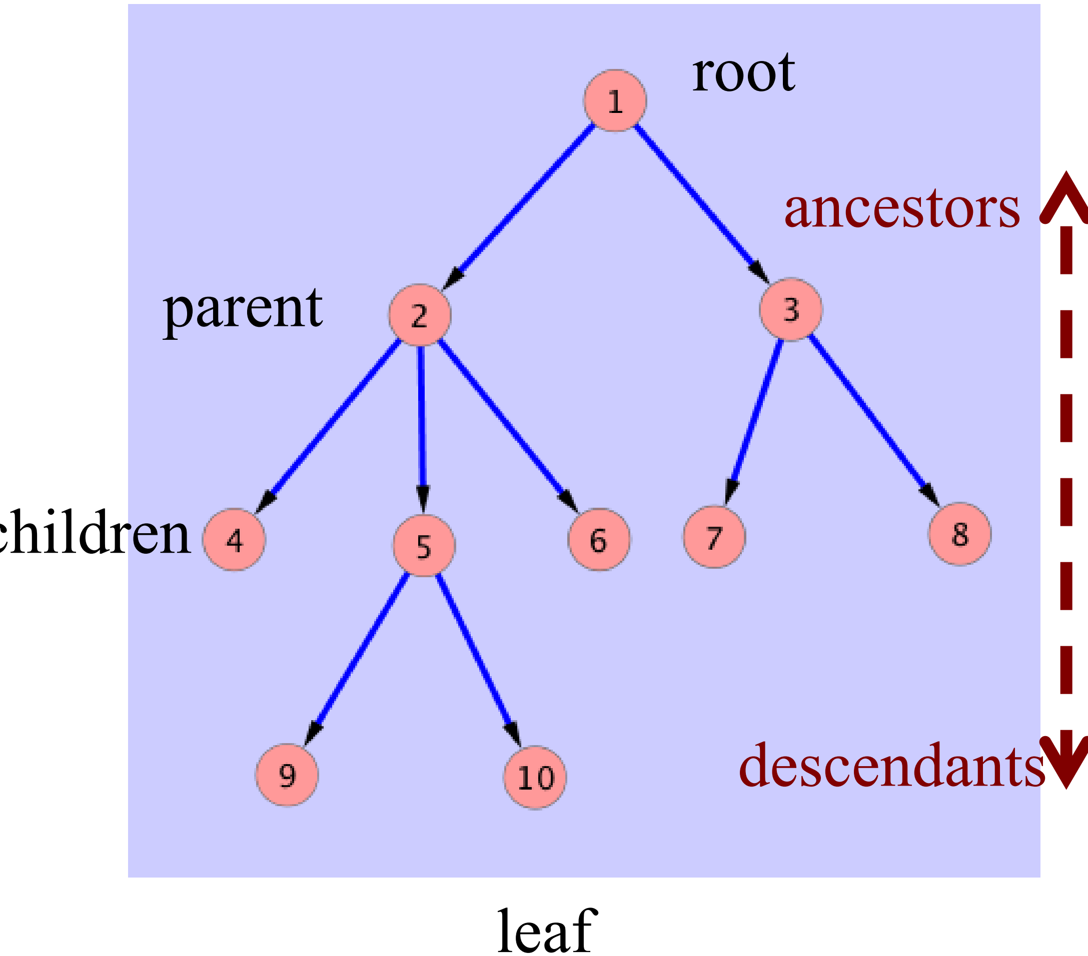
A connected graph with no cycles is called a tree



The disjoint union of trees is called a forest

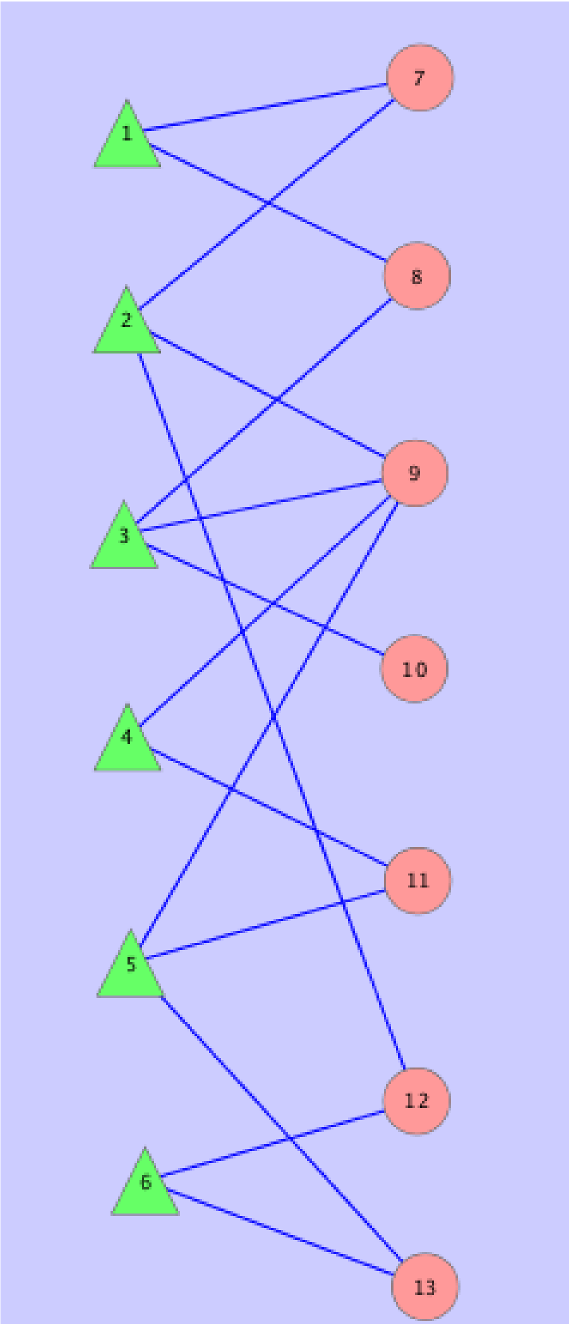


A digraph whose underlying graph is a tree is called a directed tree



Directed tree may show a root, i.e. a unique vertex from which there is a directed path to any other vertex of the graph.

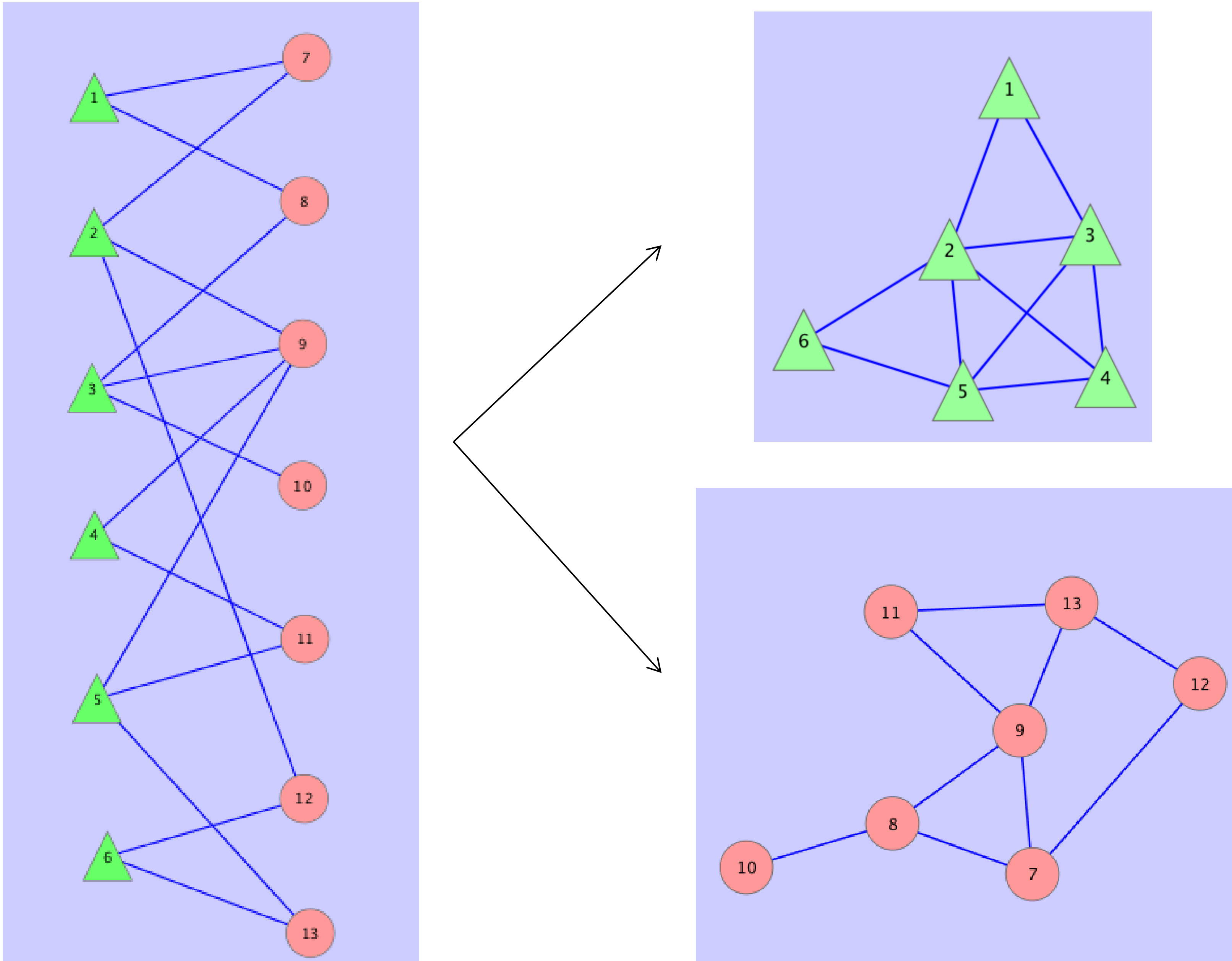
When a root is present in a directed tree the tree is called a rooted tree



Bipartite graphs

A bipartite graph is a graph where the set of vertices can be partitioned in two disjoint sets and edges are present only between a pair of vertices of different nature (for example, actors and movies, genes and organisms, students and courses, etc)

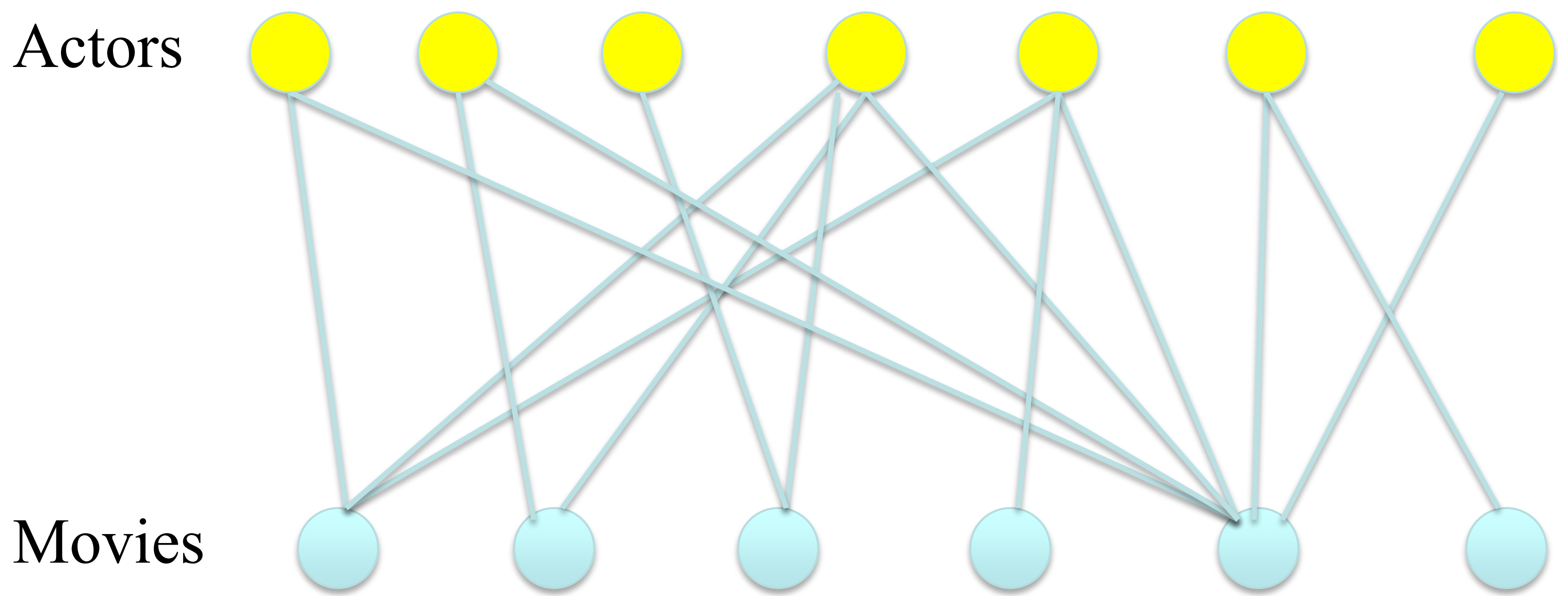
From a bipartite graph it is quite common to extract projected graphs.



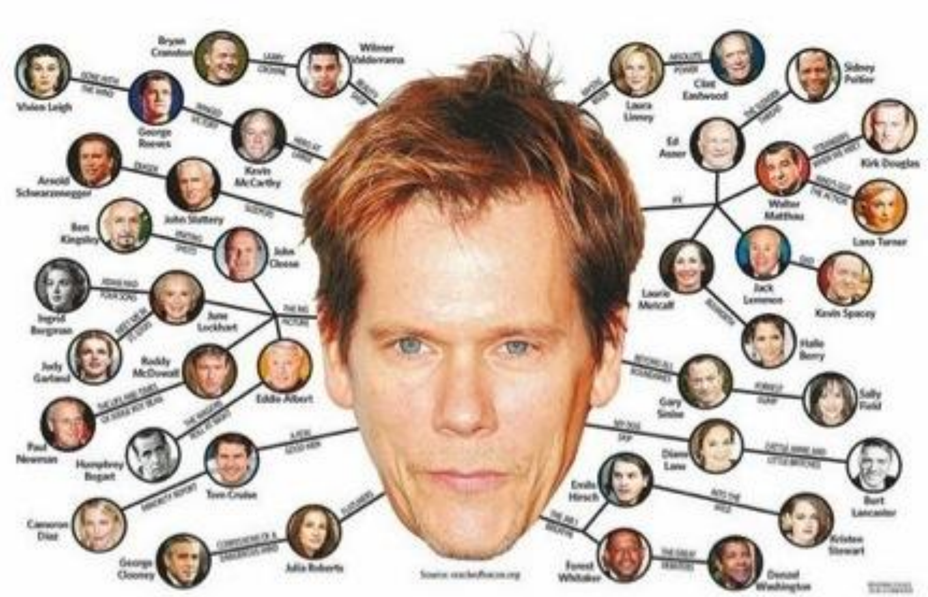
A classic example is the bipartite network of Movies and Actors

An example is the set of all world movies produced during the period 1990-2008 which are present in the International Movie Data Base.

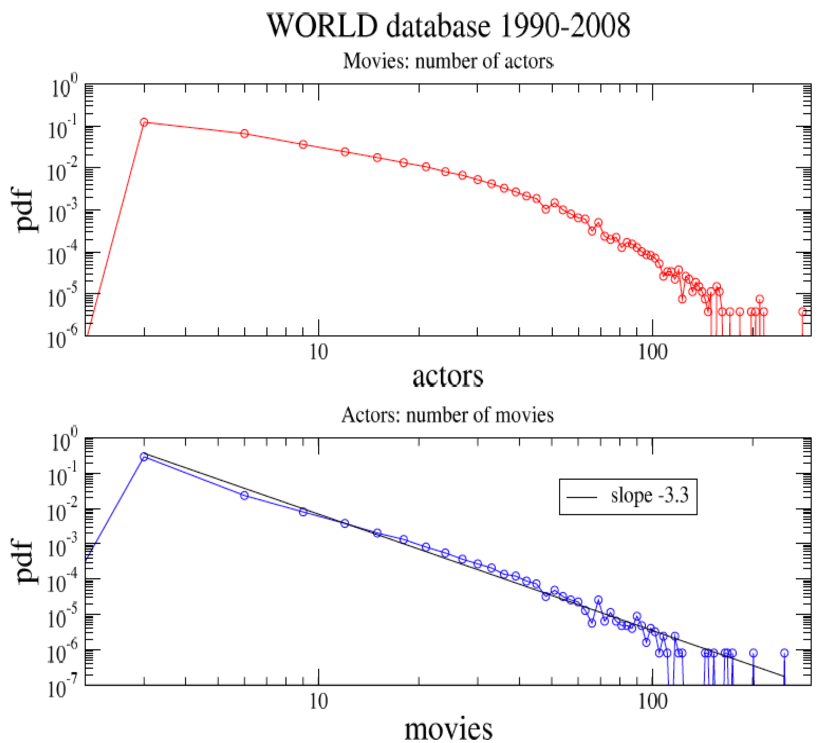
This set of data comprised 89605 movies realized in 158 countries. 412,143 different actors played in these movies.



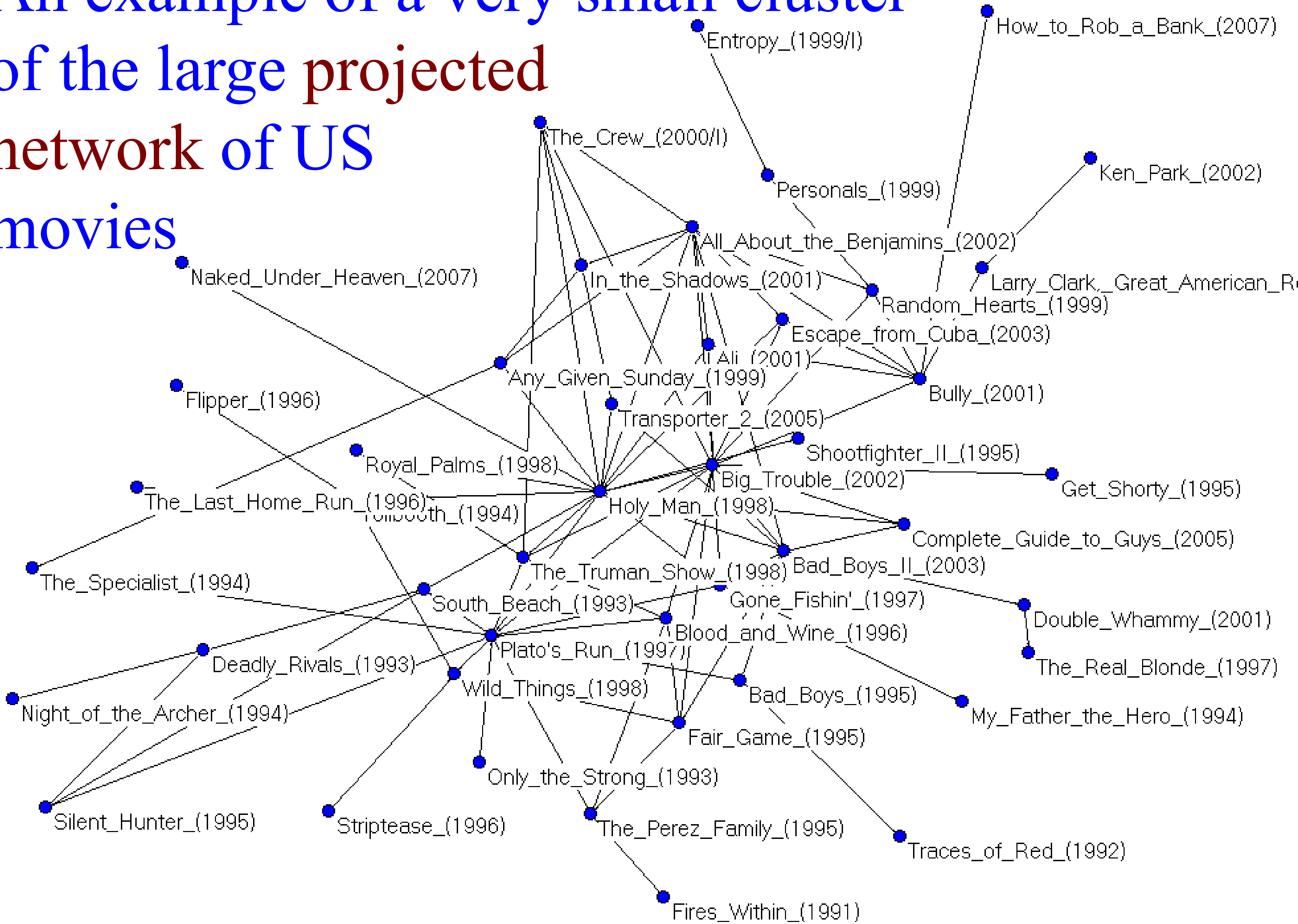
One typical characteristic of complex systems and complex network is to be heterogeneous



Kevin Bacon

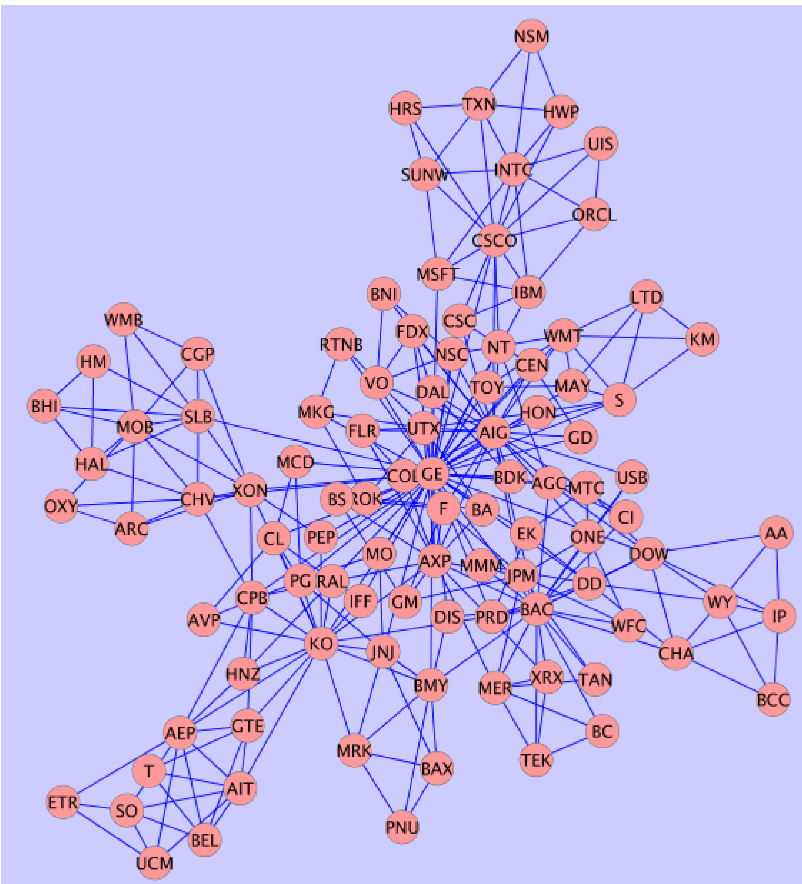


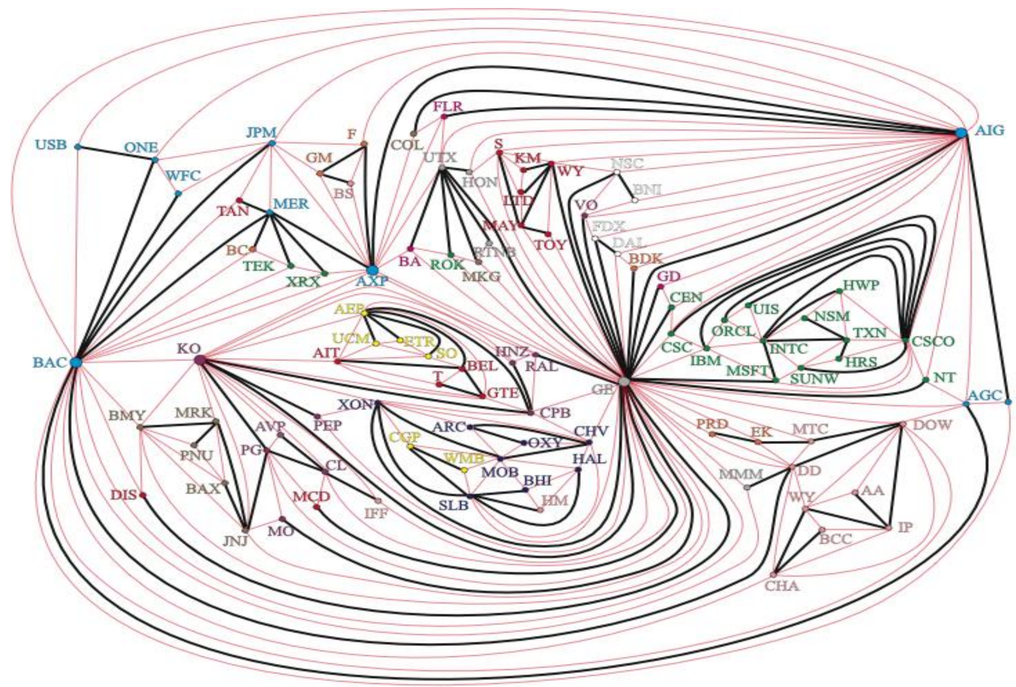
An example of a very small cluster of the large projected network of US movies



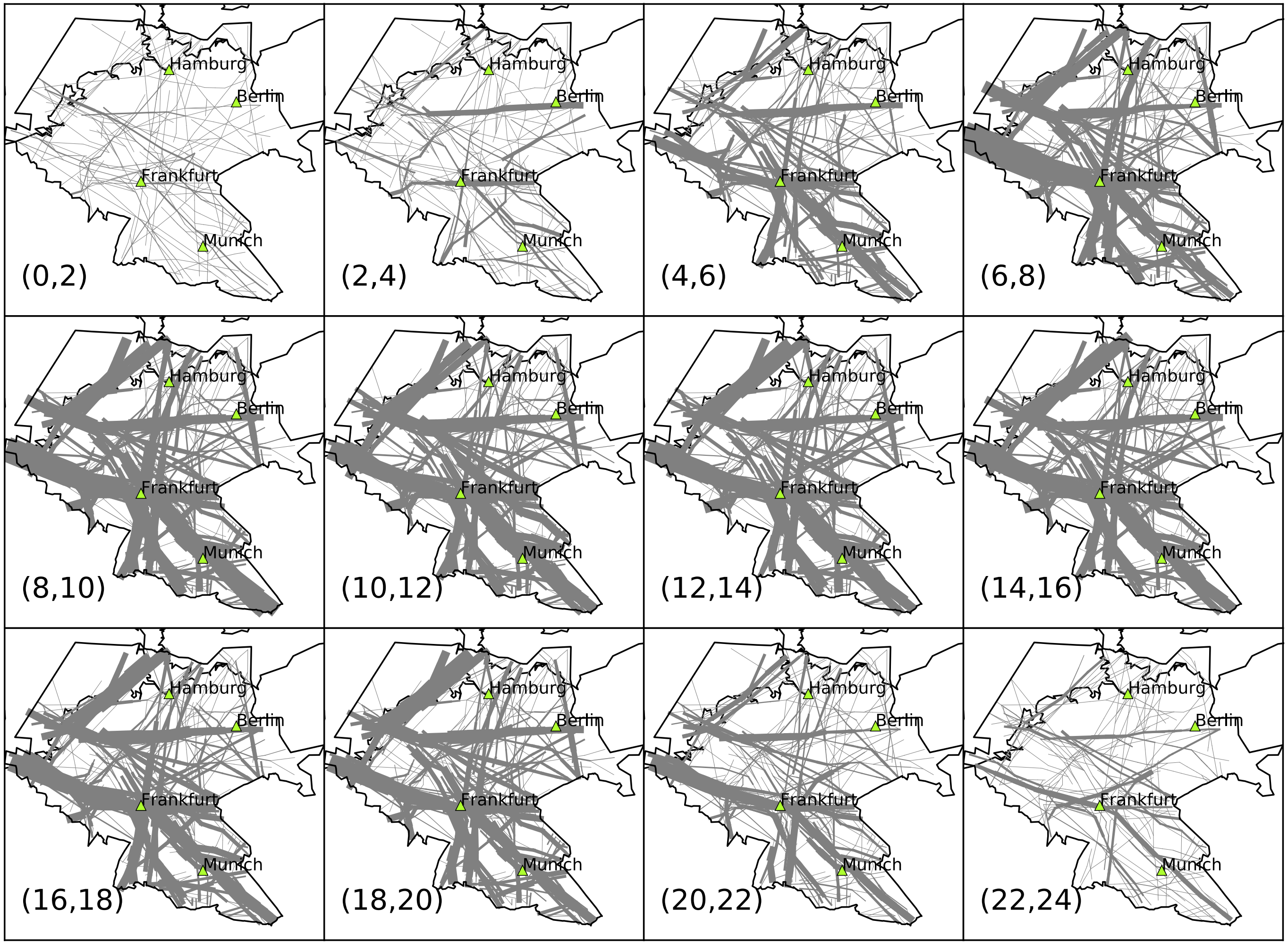
Planar graphs

A graph is planar if its edges can be embedded on a surface of genus 0, i.e. a surface like a plane or a sphere, without intersections of the edges.





Planar graphs are naturally observed in the presence of geographical constraints



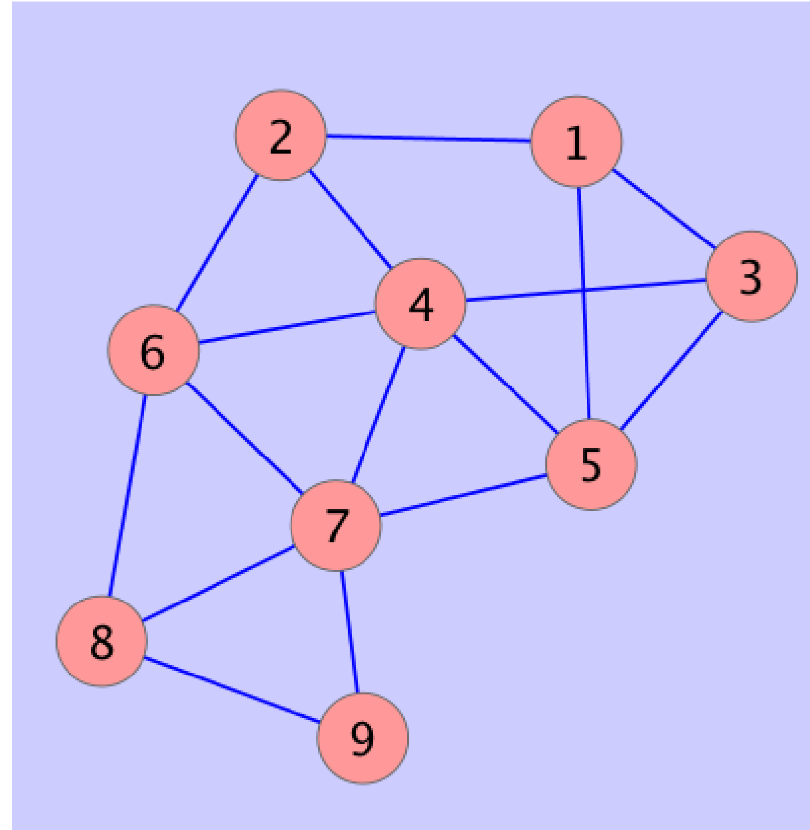
The network of navigation points of the German airspace during different time interval of the day

Adjacency matrix A_{ij}

	1	2	3	4	5	6	7	8	9
1	0	1	1	0	1	0	0	0	0
2	1	0	0	1	0	1	0	0	0
3	1	0	0	1	1	0	0	0	0
4	0	1	1	0	1	1	1	0	0
5	1	0	1	1	0	0	1	0	0
6	0	1	0	1	0	0	1	1	0
7	0	0	0	1	1	1	0	1	1
8	0	0	0	0	0	1	1	0	1
9	0	0	0	0	0	0	1	1	0

3
3
3
5
4
4
5
3
2

degree



$$A_{ij} = \begin{cases} 1, & \text{if } \{i, j\} \in E, \\ 0 & \text{otherwise,} \end{cases}$$

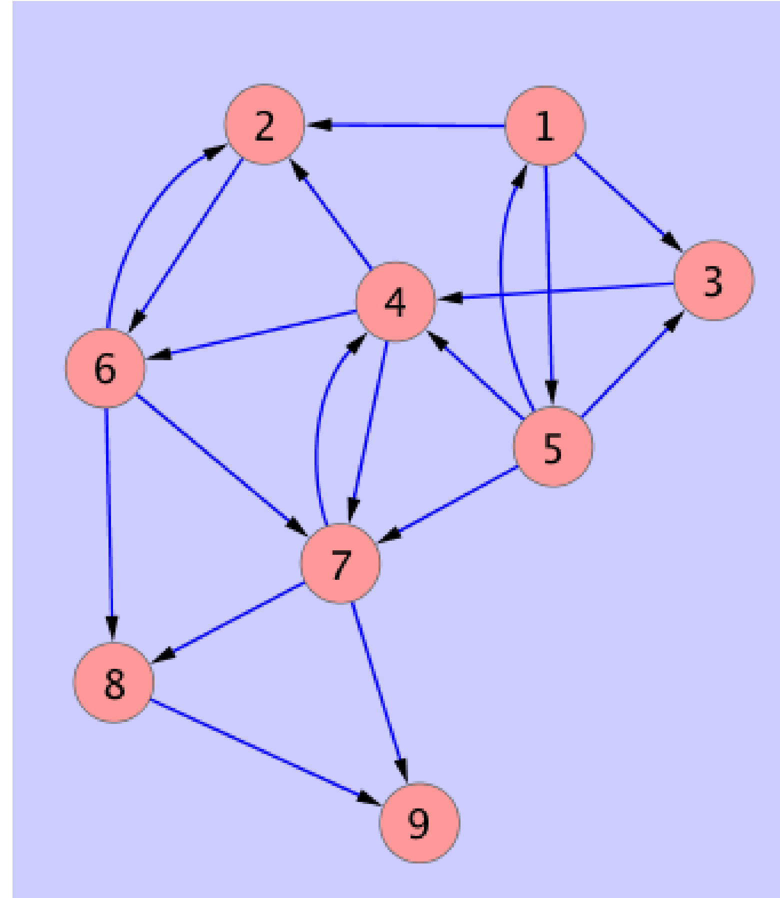
For digraphs the matrix A_{ij} is not symmetric

	1	2	3	4	5	6	7	8	9	
1	0	1	1	0	1	0	0	0	0	3
2	0	0	0	0	0	1	0	0	0	1
3	0	0	0	1	0	0	0	0	0	1
4	0	1	0	0	0	1	1	0	0	3
5	1	0	1	1	0	0	1	0	0	4
6	0	1	0	0	0	0	1	1	0	3
7	0	0	0	1	0	0	0	1	1	3
8	0	0	0	0	0	0	0	0	1	1
9	0	0	0	0	0	0	0	0	0	0

1	3	2	3	1	2	3	2	2

in-degree

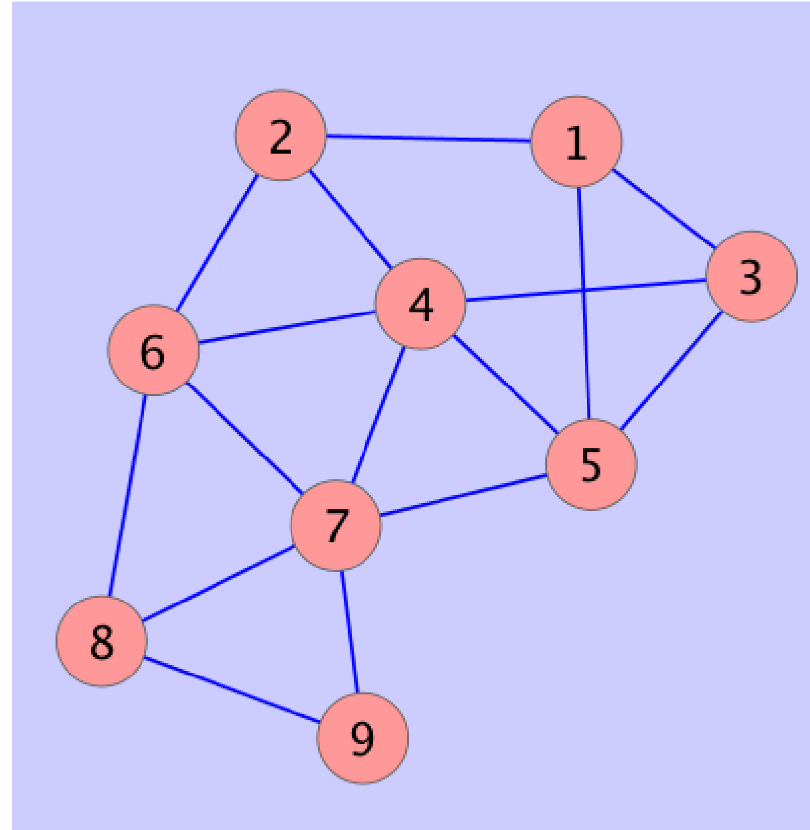
out-degree



$$A_{ij} = \begin{cases} 1, & \text{if } \{i, j\} \in E, \\ 0 & \text{otherwise,} \end{cases}$$

Other properties of the adjacency matrix A_{ij}

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	0	1	1	0	1	0	0	0	0
2	1	0	0	1	0	1	0	0	0
3	1	0	0	1	1	0	0	0	0
4	0	1	1	0	1	1	1	0	0
5	1	0	1	1	0	0	1	0	0
6	0	1	0	1	0	0	1	1	0
7	0	0	0	1	1	1	0	1	1
8	0	0	0	0	0	1	1	0	1
9	0	0	0	0	0	0	1	1	0



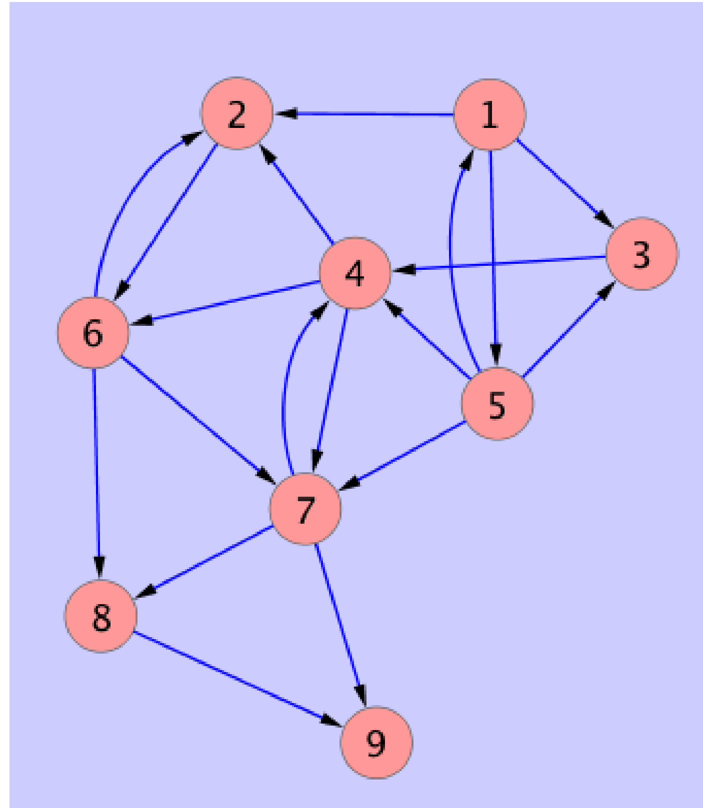
The r-th power of A_{ij} has elements A^r_{ij} providing the number of walks of length r between vertices i and j

Data structures for a graph

A_{ij}

	1	2	3	4	5	6	7	8	9
1	0	1	1	0	1	0	0	0	0
2	0	0	0	0	0	1	0	0	0
3	0	0	0	1	0	0	0	0	0
4	0	1	0	0	0	1	1	0	0
5	1	0	1	1	0	0	1	0	0
6	0	1	0	0	0	0	1	1	0
7	0	0	0	1	0	0	0	1	1
8	0	0	0	0	0	0	0	0	1
9	0	0	0	0	0	0	0	0	0

$$O\left(N_V^2\right)$$

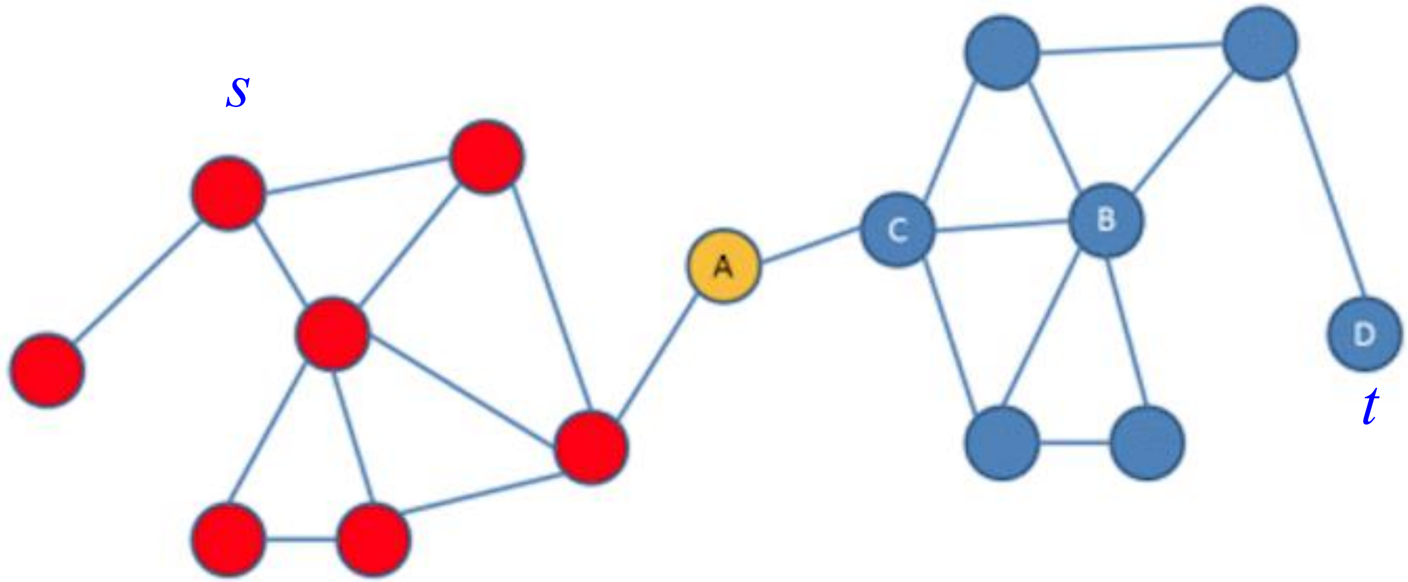


$$O\left(N_v + N_E\right)$$

Edge list

1	2
1	3
1	5
2	6
3	4
4	2
4	6
5	3
5	4
5	7
4	7
6	7
7	8
7	9
6	8
8	9
5	1
6	2
7	4

Another vertex centrality measure: the vertex betweenness

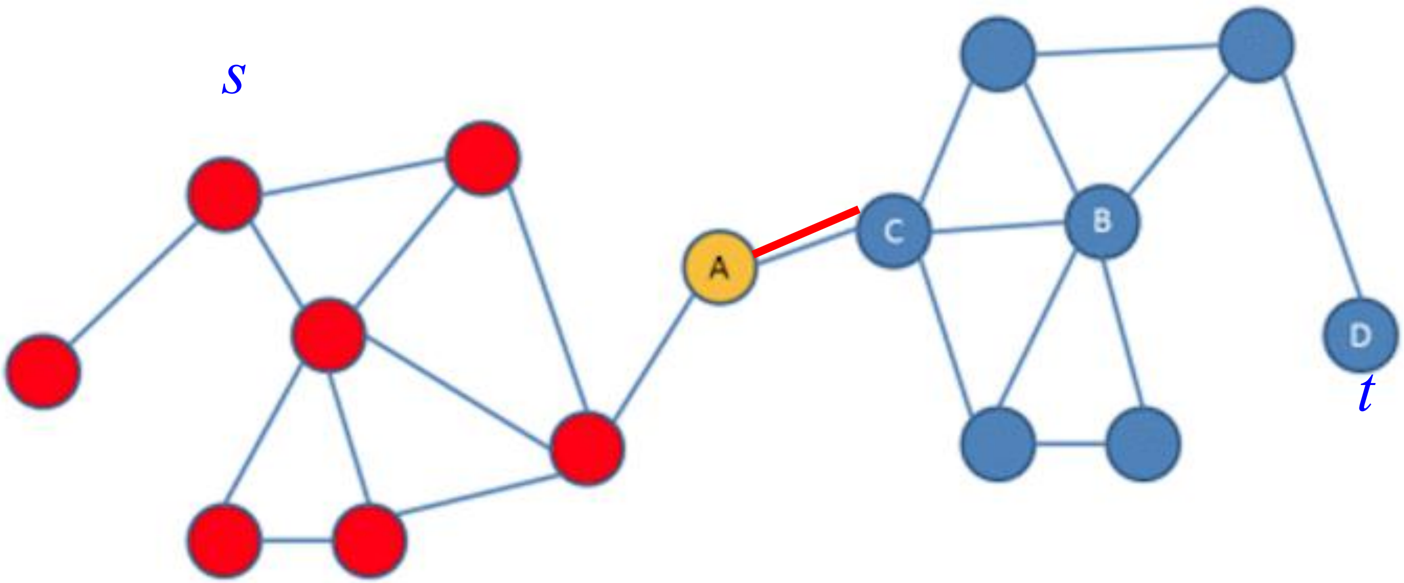


The vertex **betweenness** is a centrality measure defined as

$$C_B(v) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

The edge betweenness



The **edge betweenness** is defined as:

$$C_B(e) = \sum_{s \neq t \neq v} \frac{\sigma_{st}(e)}{\sigma_{st}}$$

where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(e)$ is the number of those paths that pass through edge e.

Robust Action and the Rise of the Medici, 1400–1434[1]

John F. Padgett and Christopher K. Ansell
University of Chicago

AJS Volume 98 Number 6 (May 1993): 1259–1319

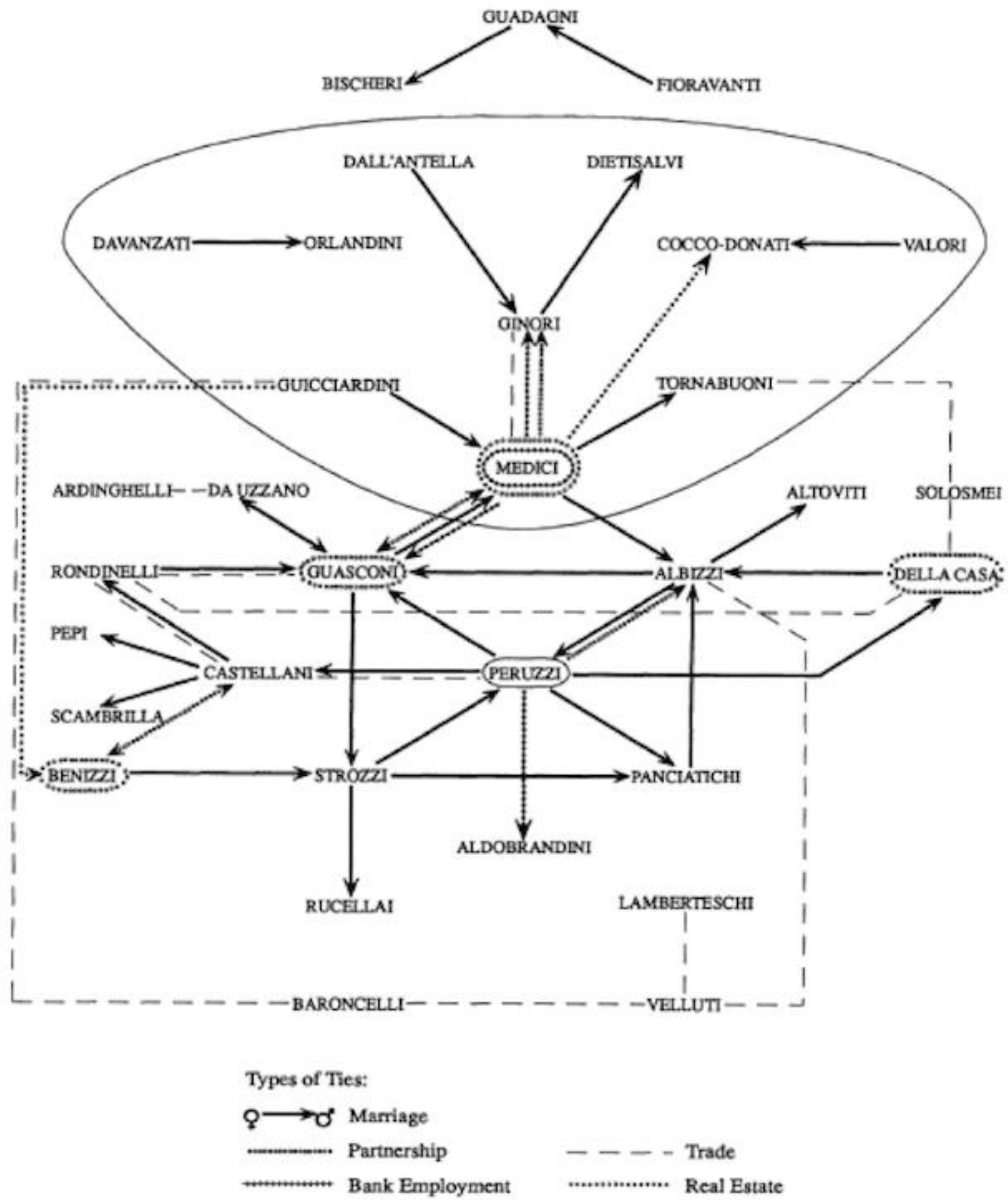


FIG. 2a.—Marriage and economic blockmodel structure (92 elite families)

Feb 4th, 2016

[31] Graph centralization can be measured with the network betweenness statistic (C_B) of Freeman (1979, p. 230), usually interpreted as intensity of concentration of resource or information flows. Among Medicean families, marriage relations were concentrated at the level C_B = .362. Among oligarch partisans, marriage C_B = .184. On the economic front, C_B = .429 among Mediceans, compared to C_B = .198 among oligarchs. (Economic ties, including personal loans, were pooled for the latter calculations. Personal loans were included because otherwise the density of intra-Medicean ties was too low. All data were binarized and symmetrized for these calculations, as required by the Freeman measure.) We thank an *AJS* referee for the suggestion of calculating these statistics.

Centrality betweenness:

Marriage relations:
Medici C_B=0.362
Other oligarch C_B=0.184

Economic ties (including personal loans):
Medici C_B=0.429
Other oligarch C_B=0.198

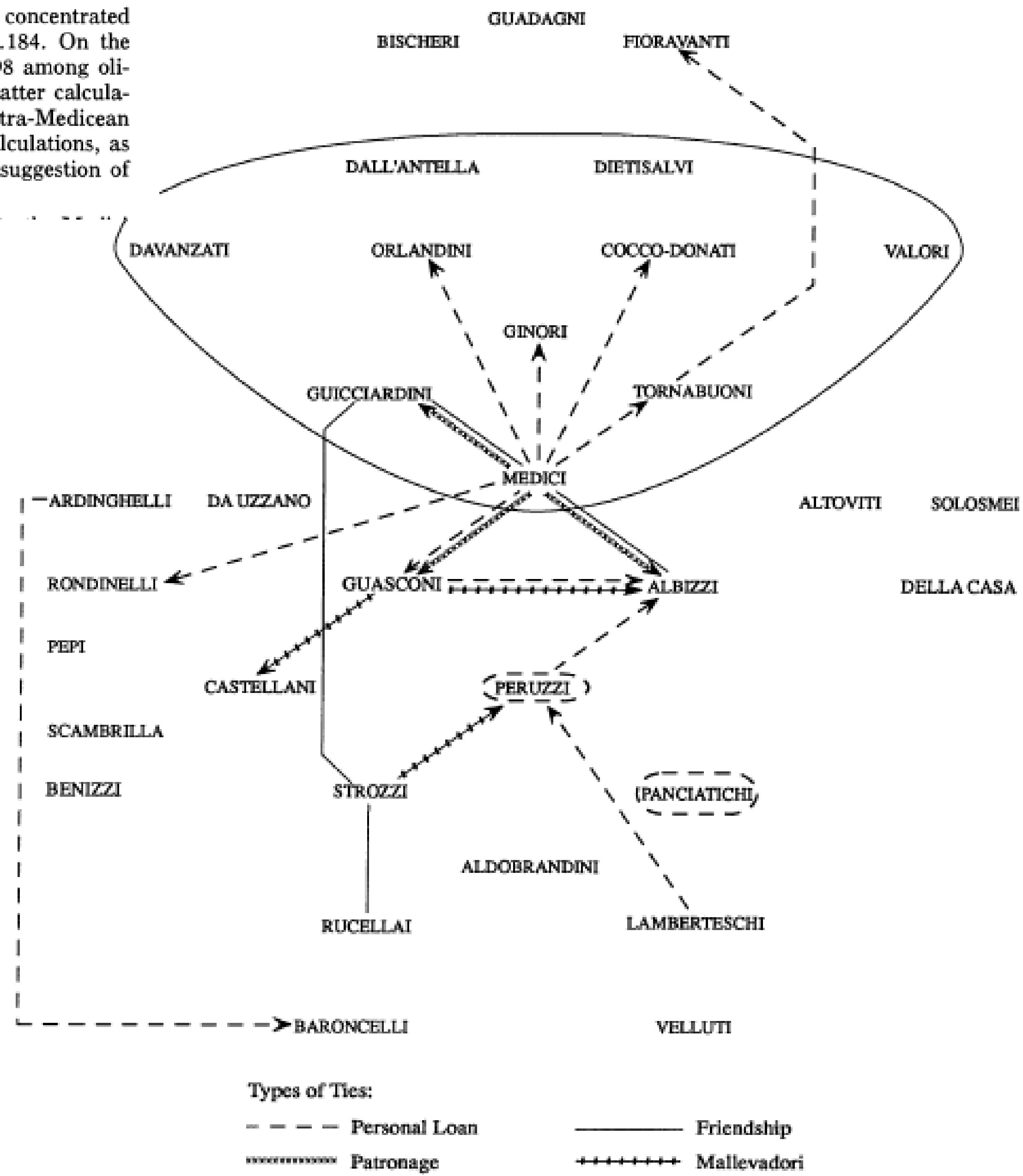
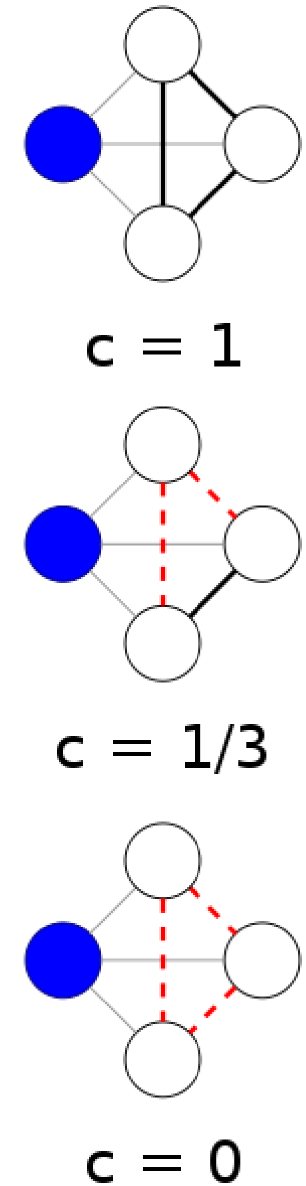


FIG. 2b.—"Political" and friendship blockmodel structure (92 elite families)

Other network indicators

Local clustering coefficient



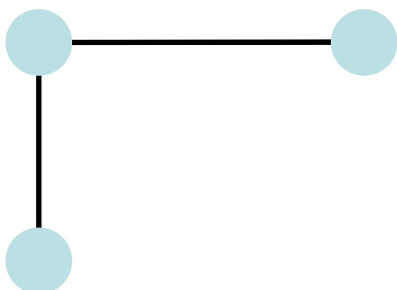
$$C_i = \frac{2\left|\{e_{jk}\}\right|}{k_i\left(k_i - 1\right)}$$

$\{e_{jk}\}$ is the number of edges observed between pairs of vertices *jk*, ... linked to *i*

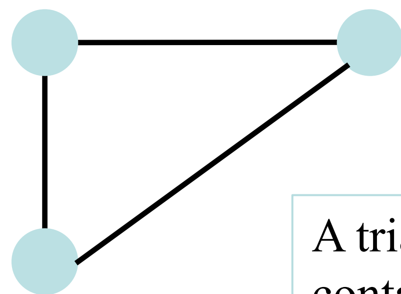
k_i is the degree of vertex *i*

Global clustering coefficient

Connected triple



Triangle

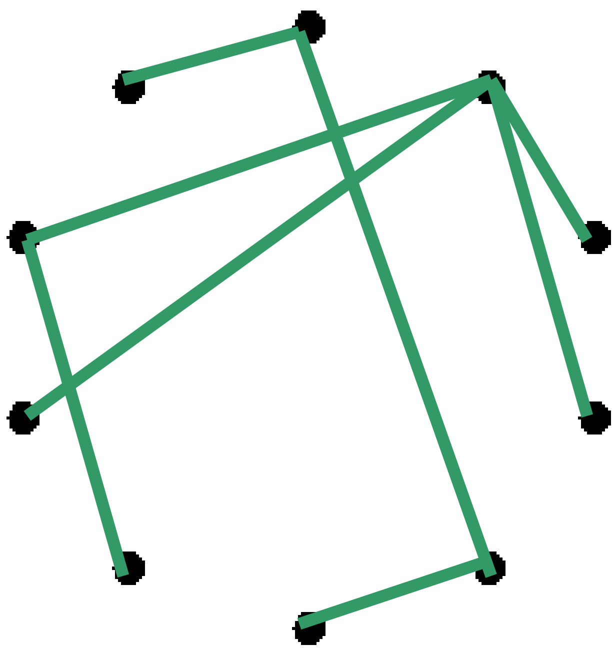


A triangle contains 3 triples

Fraction of path of length 2 in the network that are close

$$C = \frac{3 \times \text{ total number of triangles}}{\text{total number of connected triples}}$$

Models of graphs: Erdös-Rényi model (1960): Random network



Connect with probability p

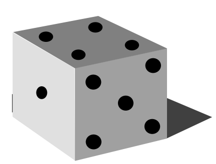
p=1/6 n=10

$<d>=z=$

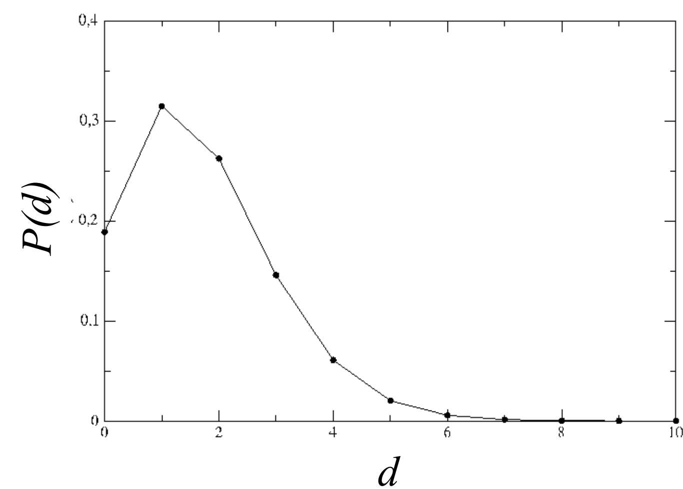
p(n-1) ~ 1.7

The tail of the degree distribution is decaying quickly

Poisson distribution for large n and z=cost

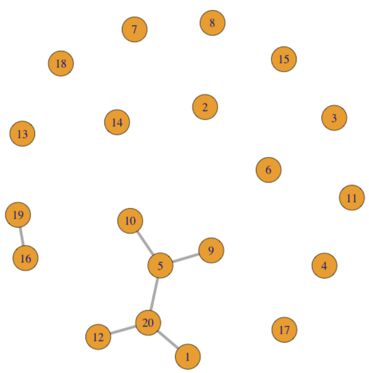


$$P(d) = \binom{n}{d} p^d (1-p)^{n-d} \cong \frac{(z)^d e^{-z}}{d!}$$

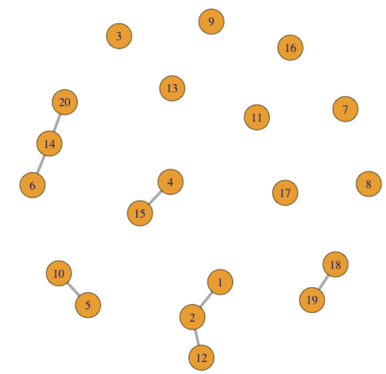


What is happening as a function of the average degree?

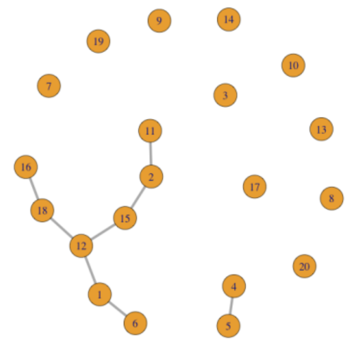
N_v=20



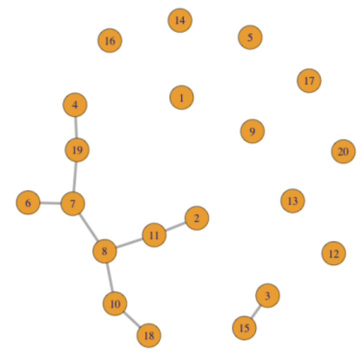
<d>=0.6
LCC 6



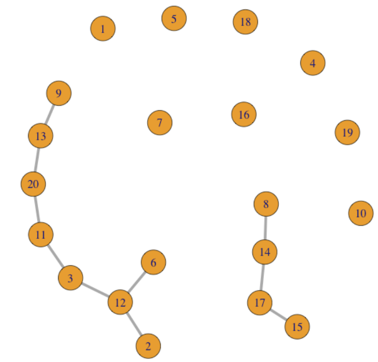
<d>=0.7
LCC 3



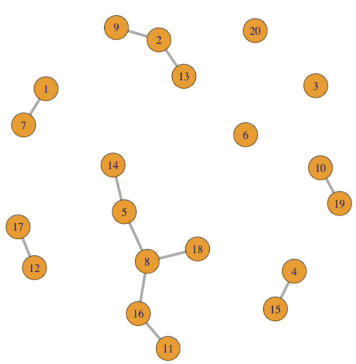
<d>=0.8
LCC 8



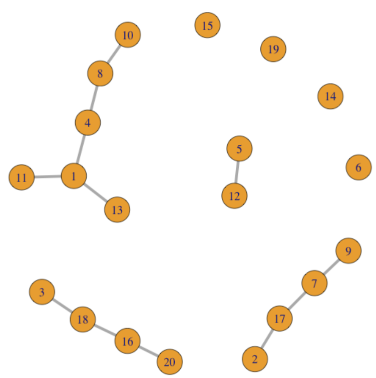
<d>=0.9
LCC 9



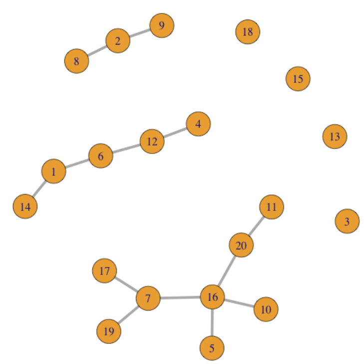
<d>=1.0
LCC 8



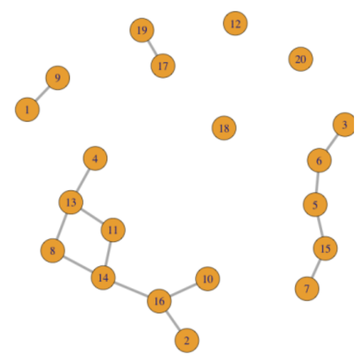
<d>=1.1
LCC 6



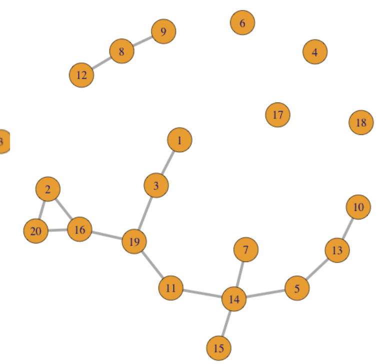
<d>=1.2
LCC 6



<d>=1.3
LCC 8



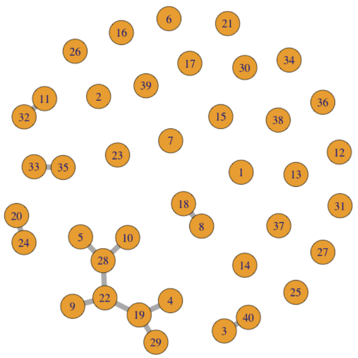
<d>=1.4
LCC 8



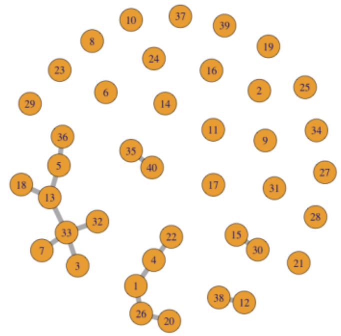
<d>=1.5
LCC 13

What is happening as a function of the average degree for large networks?

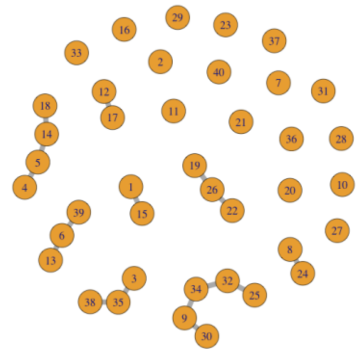
N_v=40



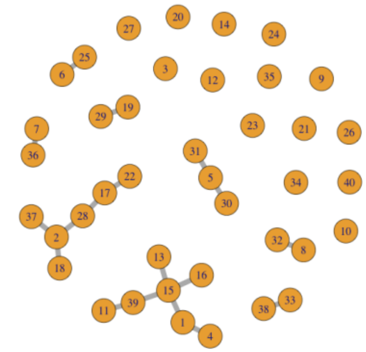
<d>=0.6
LCC 8



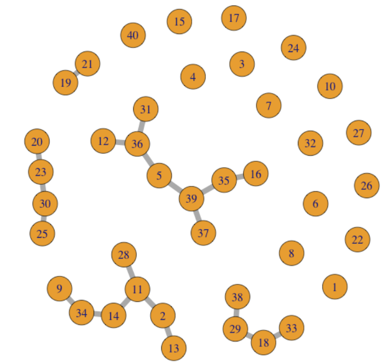
<d>=0.7
LCC 8



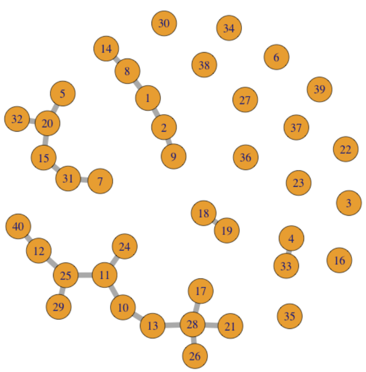
<d>=0.8
LCC 5



<d>=0.9
LCC 7

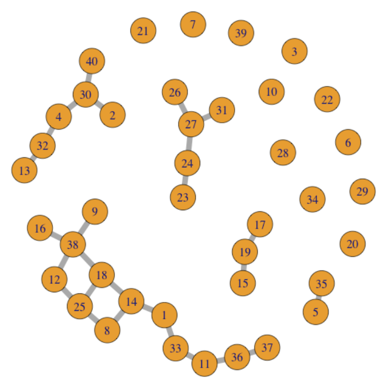


<d>=1.0
LCC 8

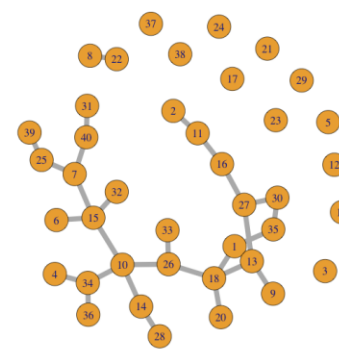


<d>=1.1
LCC 12

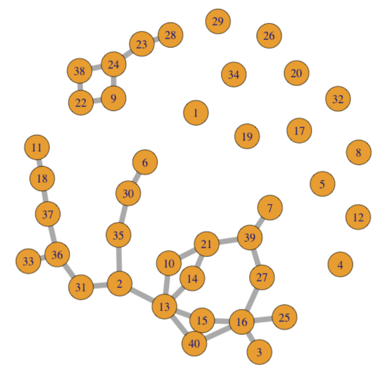
<d>=1.2
LCC 18



<d>=1.3
LCC 13



<d>=1.4
LCC 27



<d>=1.5
LCC 22

In fact for an Erdős-Rényi network we observe

THE STRUCTURE AND FUNCTION OF COMPLEX NETWORKS

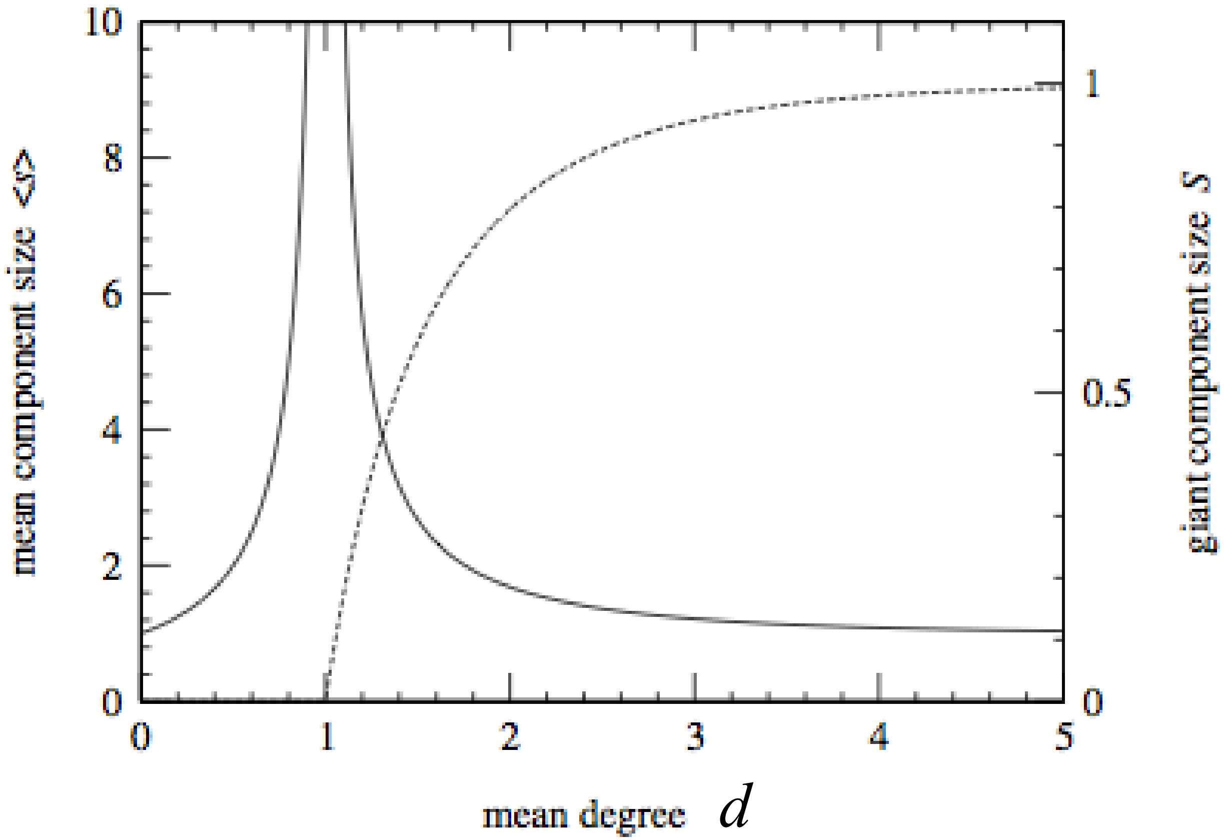
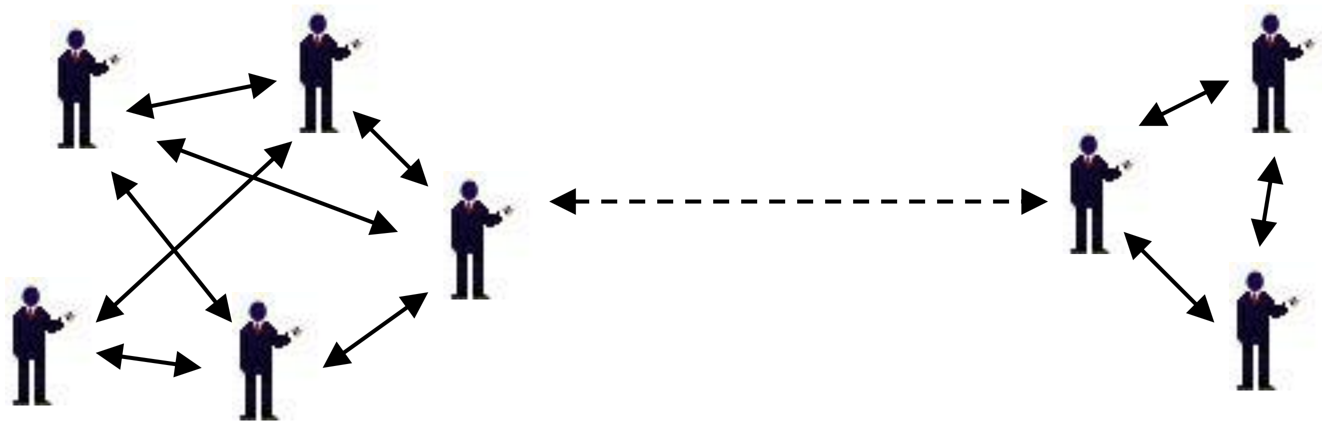
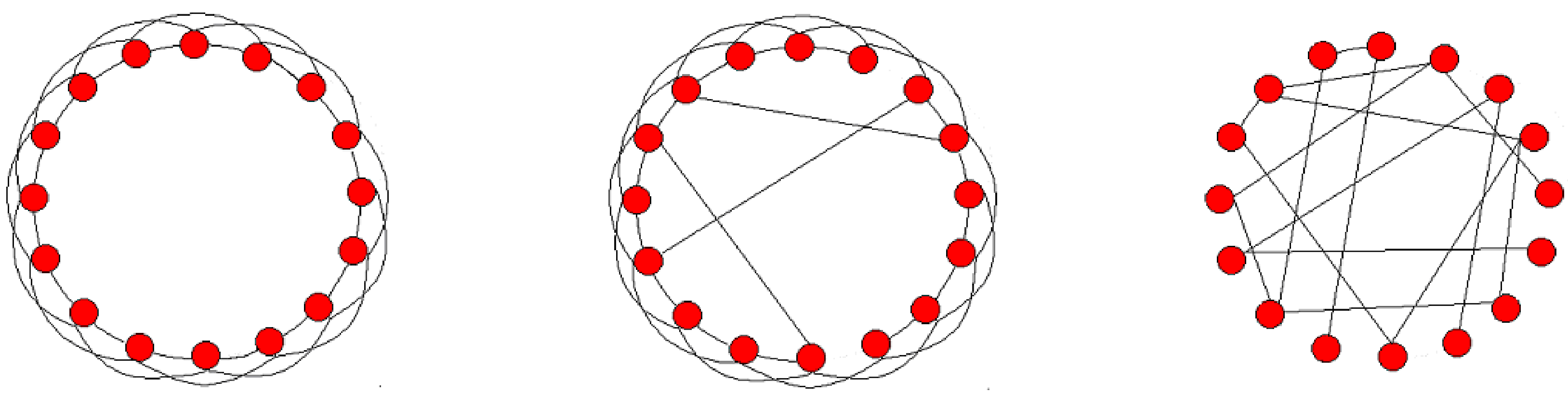


Fig. 4.1 *The mean component size (solid line), excluding the giant component if there is one, and the giant component size (dotted line), for the Poisson random graph,* (4.3) *and* (4.4).

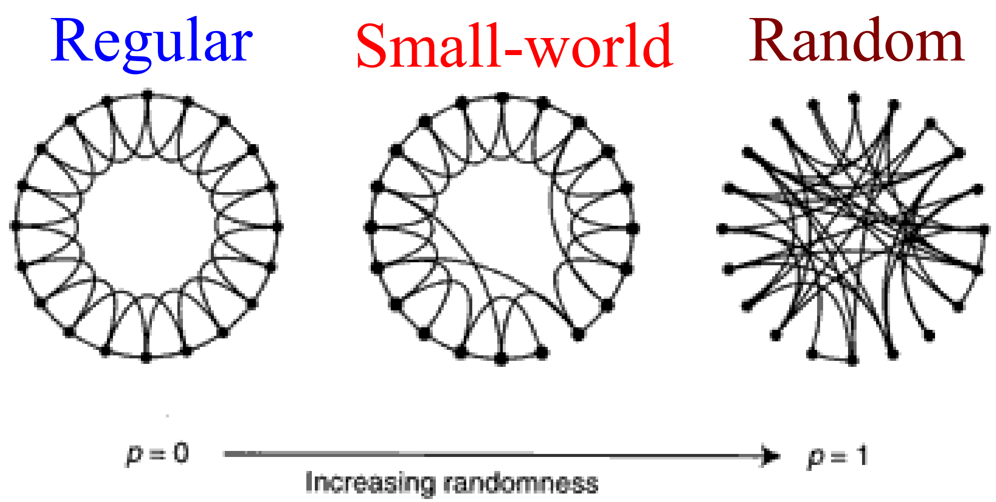
Newman, Mark EJ. "The structure and function of complex networks." SIAM review 45, no. 2 (2003): 167-256.

Small world and weak links

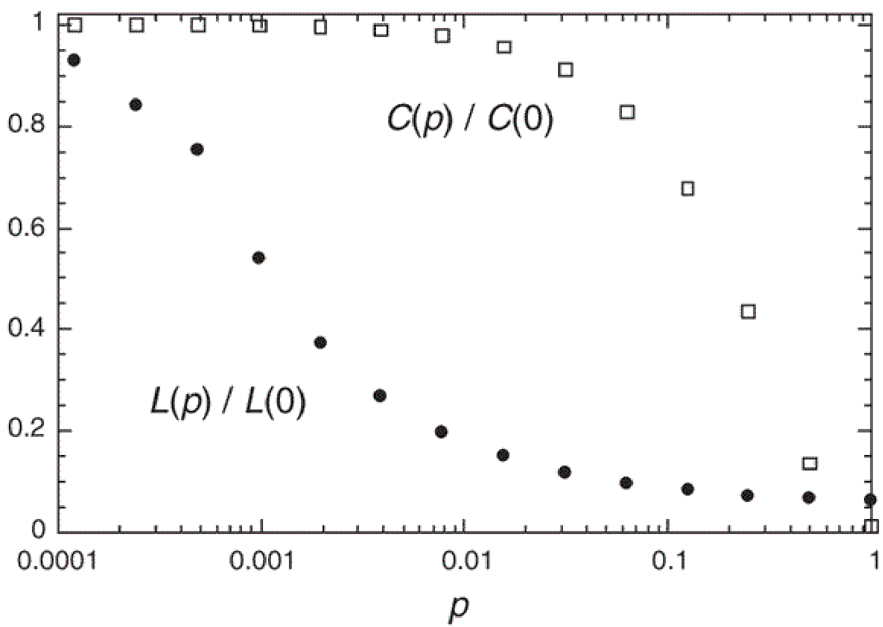




Watts-Strogatz model



Regular



Random

C(p) : average clustering coeff.
L(p) : average path length

C(0) and L(0) refer to the regular lattice

Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of ‘small-world’networks." nature 393, no. 6684 (1998): 440-442.

Networks are clustered [large C] but have a small characteristic path length [small L].

Network	C	C_{rand}	L	N
WWW	0.1078	0.00023	3.1	153127
Internet	0.18-0.3	0.001	3.7-3.76	3015-6209
Actor	0.79	0.00027	3.65	225226
Coauthorship	0.43	0.00018	5.9	52909
Metabolic	0.32	0.026	2.9	282
Foodweb	0.22	0.06	2.43	134
C. elegance	0.28	0.05	2.65	282

Degree distribution of several real networks

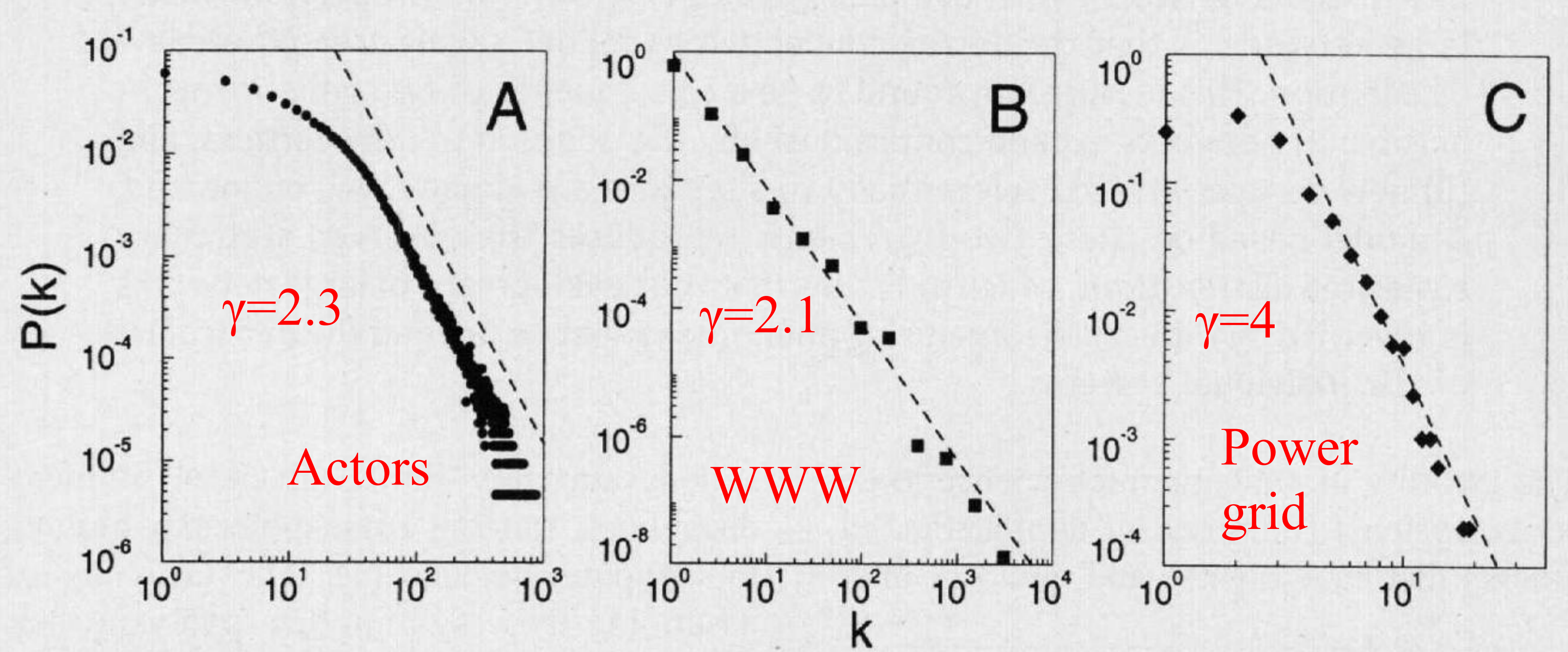
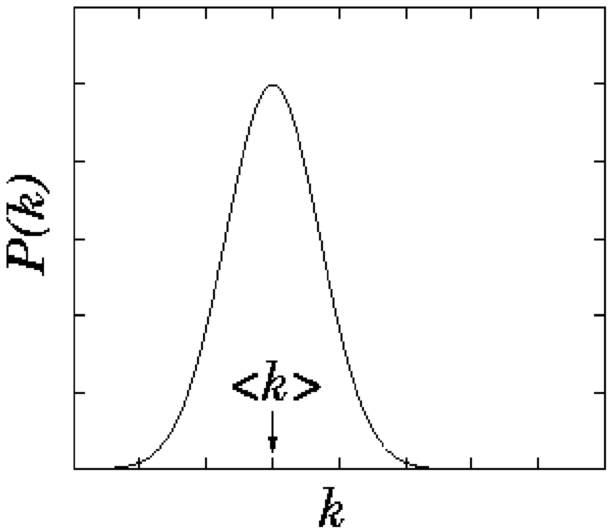


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with N = 212,250 vertices and average connectivity $\langle k \rangle$ = 28.78. (**B**) WWW, N = 325,729, $\langle k \rangle$ = 5.46 (6). (**C**) Power grid data, N = 4941, $\langle k \rangle$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

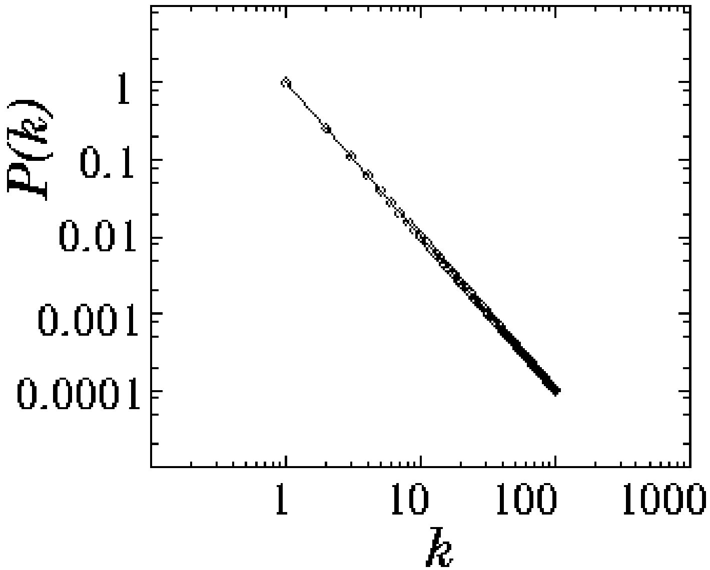
Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." Science 286, no. 5439 (1999): 509-512.

Degree distribution

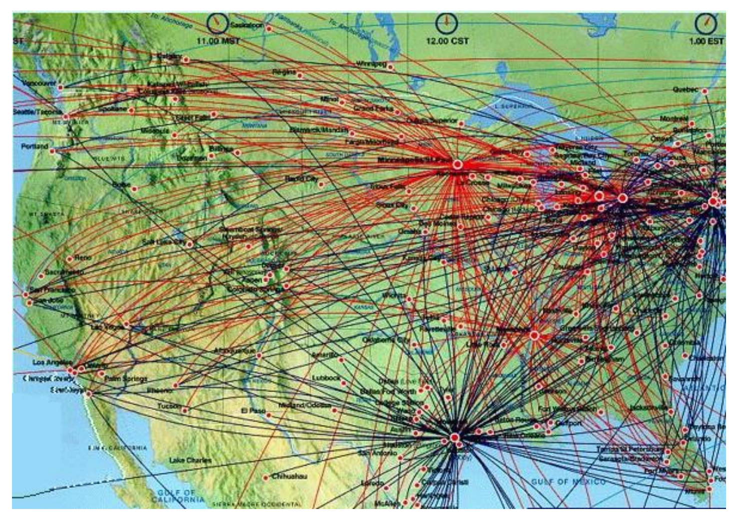
Poisson distribution



Power-law distribution



Poisson Network



Scale-free Network

Scale free networks are sparse networks

The probability that any node on the network is highly connected to many others is *very low*.

The probability that a very large number of nodes are connected loosely or not at all is *very high*.

Complex networks are typically sparse networks

Barabási and Albert (1999)

The Barabási-Albert model of preferential attachment

At each time step a new vertex is added to the network.

The new vertex forms m connections with existing vertices.

The probability that a new edge attaches to a vertex with degree d is

$$\frac{dp(d)}{2 \times m}$$

where $2m$ is the mean degree of the network.

Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." Science 286, no. 5439 (1999): 509-512.

One can write a master equation describing the dynamics of $p_n(d)$, which is the degree distribution when the network has n vertices

$$(n+1)\,p_{n+1}(d) - np_n(d) = \begin{cases} \frac{1}{2}(d-1)\,p_n(d-1) - \frac{1}{2}dp_n(d) & \text{for } d > m \\ 1 - \frac{1}{2}mp_n(m) & \text{for } d = m \end{cases}$$

where the mean number of vertices of degree d that gain an edge when a single vertex with m edges is added is

$$m \times \frac{dp(d)}{2m} = \frac{1}{2}dp(d)$$

Looking for stationary solutions

$$p_{n+1}(d) = p_n(d) = p(d)$$

Under the stationary assumption the above master equation implies

$$p(d)=\begin{cases}\frac{1}{2}(d-1)p(d-1)-\frac{1}{2}dp(d) & \text{for } d>m\\ 1-\frac{1}{2}mp(m) & \text{for } d=m\end{cases}$$

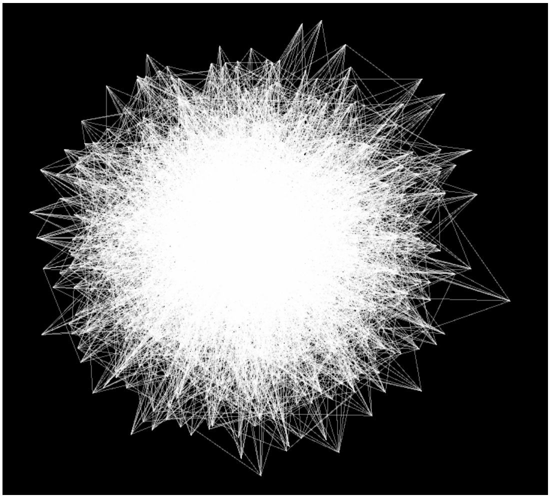
By solving this equations recursively we have

$$p(m)=\frac{2}{m+2} \qquad p(d)=p(d-1)\frac{k-1}{k+2}$$

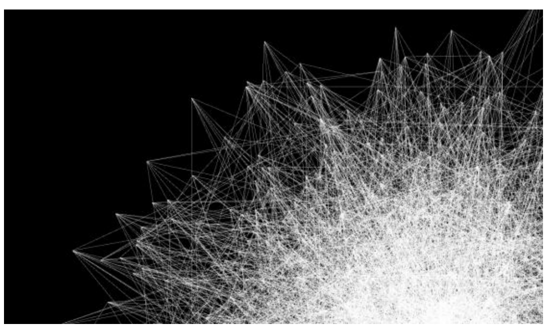
Which are equivalent to

$$p(d)=\frac{(d-1)(d-2)\cdots(m+1)m}{(d+2)(d+1)d\cdots(m+3)}p(m)=\frac{2m(m+1)}{(d+2)(d+1)d}\cong\frac{2m(m+1)}{d^3}$$

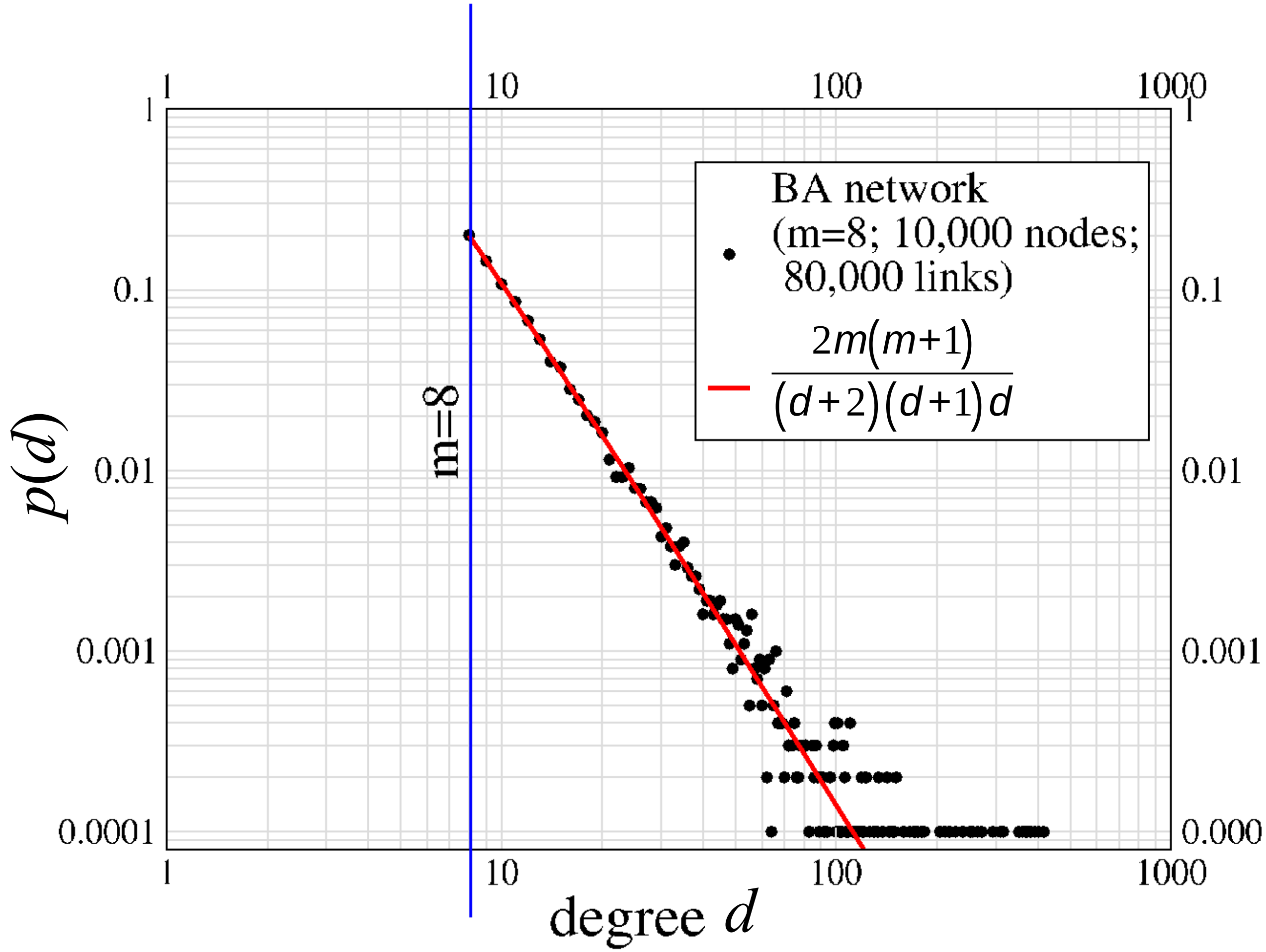
A numerical simulation of the model



Degree and age of nodes are correlated



Younger nodes have smaller degree



In network theory the most basic model of a random graph presenting a given degree distribution is the *configuration model*.

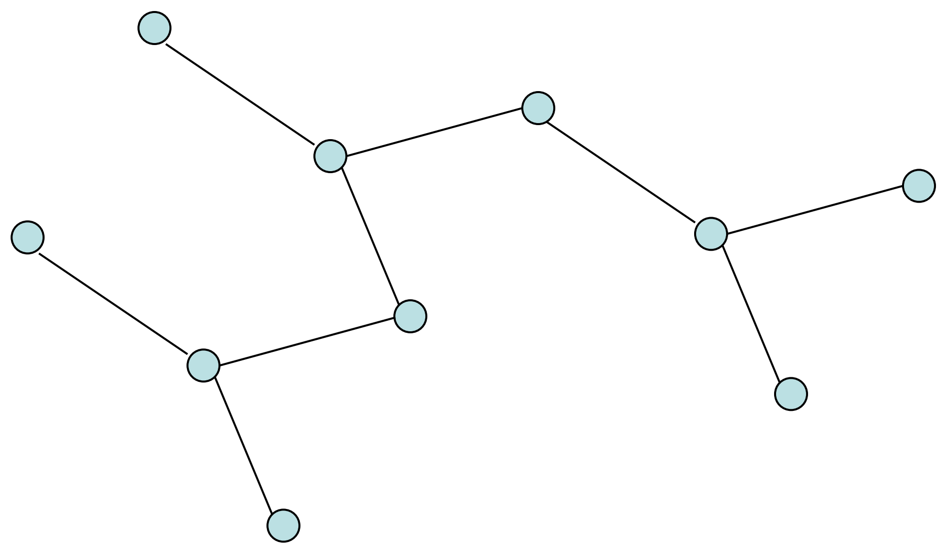
It should be realized that the configuration model does not fully characterize a network but only describes the most basic characteristics of it. In fact, different networks characterized by quite different local structures can be characterized by the same configuration model.

A scientific question widely investigated has been: By assuming a given configuration model, when is a giant component observed to be present in the considered network?

Emergence of a giant component in a network described by a given configuration model

A heuristic argument describing quite well the setting of a giant component[¶]

A giant component does not set up in the system until a negligible number of cycles (closed paths) are observed in the network. Therefore, when the giant component is not yet formed, the paths observed in the network are indistinguishable from paths observed on trees.



[¶]Cohen, Reuven, Keren Erez, Daniel Ben-Avraham, and Shlomo Havlin. "Resilience of the Internet to random breakdowns." Physical review letters 85, no. 21 (2000): 4626.

The argument of Cohen et al is as follows: Let us estimate the probability that 2 occupied nodes are connected in a component with *s* nodes. This probability is proportional to $\left(\frac{s}{n}\right)^2$

Therefore the fraction of links that are belonging to cycles is of the order of $\sum_i \left(\frac{s_i}{n}\right)^2$

Where s_i is the size of the component i in the network

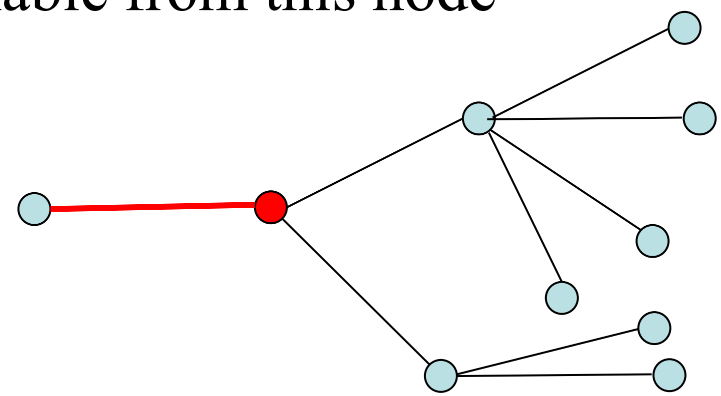
We can write the inequality $\sum_i \left(\frac{s_i}{n}\right)^2 \le \sum_i \frac{s_i S}{n^2}$

where S is the size of the giant component.

Since $\sum_i s_i = n$ The fraction of links that lie on cycles is of the order of no more than S/n

¶Cohen, R., K. Erez, D. Ben-Avraham, and S. Havlin. "Resilience of the Internet to random breakdowns." Physical review letters 85, no. 21 (2000): 4626.

Let ϕ be the number of nodes that can be reached in the network by selecting randomly a link and then one node of the link and exploring all nodes reachable from this node



By assuming that the giant component is still not present, i.e. the number of cycles is negligible

$$\phi = 1 + \sum_{d=1}^{\infty} \frac{P(d)\,d}{\langle d \rangle}(d-1)\,\phi$$

limit number of nodes reached from one selected

Distribution of degree of node found selecting a link at random and selecting one of the nodes

The previous equation states that: $$\phi = 1 + \frac{\left[\langle d^2 \rangle - \langle d \rangle\right]\phi}{\langle d \rangle}$$

or $$\phi = \frac{1}{2 - \langle d^2 \rangle / \langle d \rangle}$$

The quantity ϕ is finite and positive when

$$2 - \frac{\langle d^2 \rangle}{\langle d \rangle} > 0$$ which means $$\langle d^2 \rangle - 2\langle d \rangle < 0$$

The giant component will set in when the above condition is not verified $$\langle d^2 \rangle - 2\langle d \rangle > 0$$

Therefore a good approximation for the threshold setting the presence of a giant component is

$$\langle d^2 \rangle = 2\langle d \rangle$$

Let us examine some popular models of network

a) Poisson degree distribution (Erdős-Rényi model)

$$P(d) \approx \frac{(np)^d e^{-np}}{d!}$$ with

$$\langle d^2 \rangle = \langle d \rangle + \langle d \rangle^2$$

Therefore the condition for the onset of a giant components is

$$\langle d^2 \rangle - 2\langle d \rangle = \langle d \rangle + \langle d \rangle^2 - 2\langle d \rangle > 0$$

or

$$\langle d \rangle > 1$$

In fact for an Erdős-Rényi network we observe

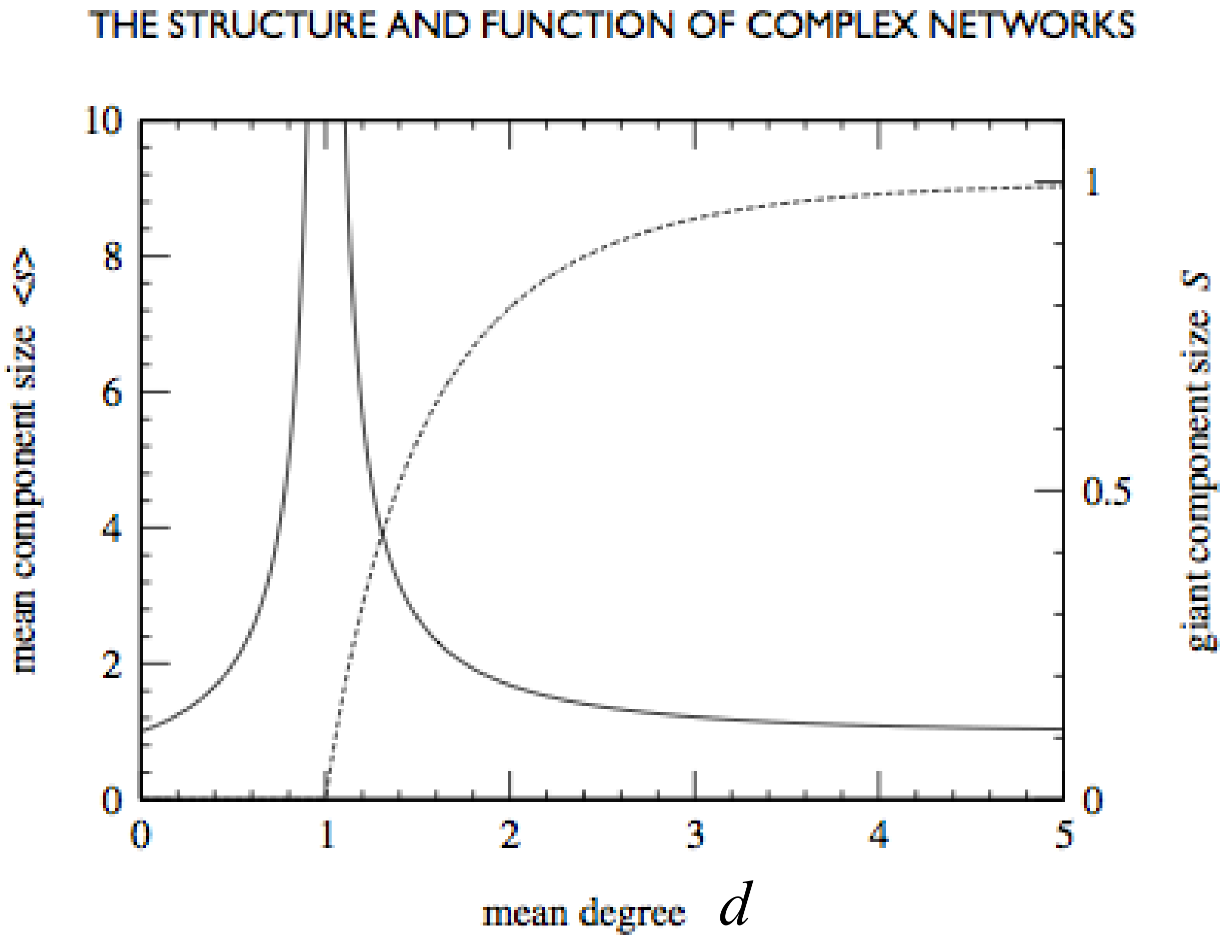


Fig. 4.1 *The mean component size (solid line), excluding the giant component if there is one, and the giant component size (dotted line), for the Poisson random graph,* (4.3) *and* (4.4).

Newman, Mark EJ. "The structure and function of complex networks." SIAM review 45, no. 2 (2003): 167-256.

b) Scale free networks (Example: Barabási-Albert model)

$$P(d) = cd^{-g} \quad \text{with} \quad g = 3$$

$\langle d^2 \rangle$ diverges

Therefore there is a giant component regardless of the details of the distribution

Let us consider the problem of the diffusion of something (an innovation, a rumor, a disease, etc) on a network

By defining $\lambda=\nu/\delta$ where ν is the transmission rate of the "infection" and δ is the "recovery rate".

The by assuming a Susceptible, Infected, Susceptible model one can estimate the role of the topology of the network in network diffusion.

Under the assumptions of the model and some simplifying assumption the result obtained for the spreading of the "epidemics" for different net

$$r » q = \frac{1}{l\left(e^{1/l} - 1\right)}$$

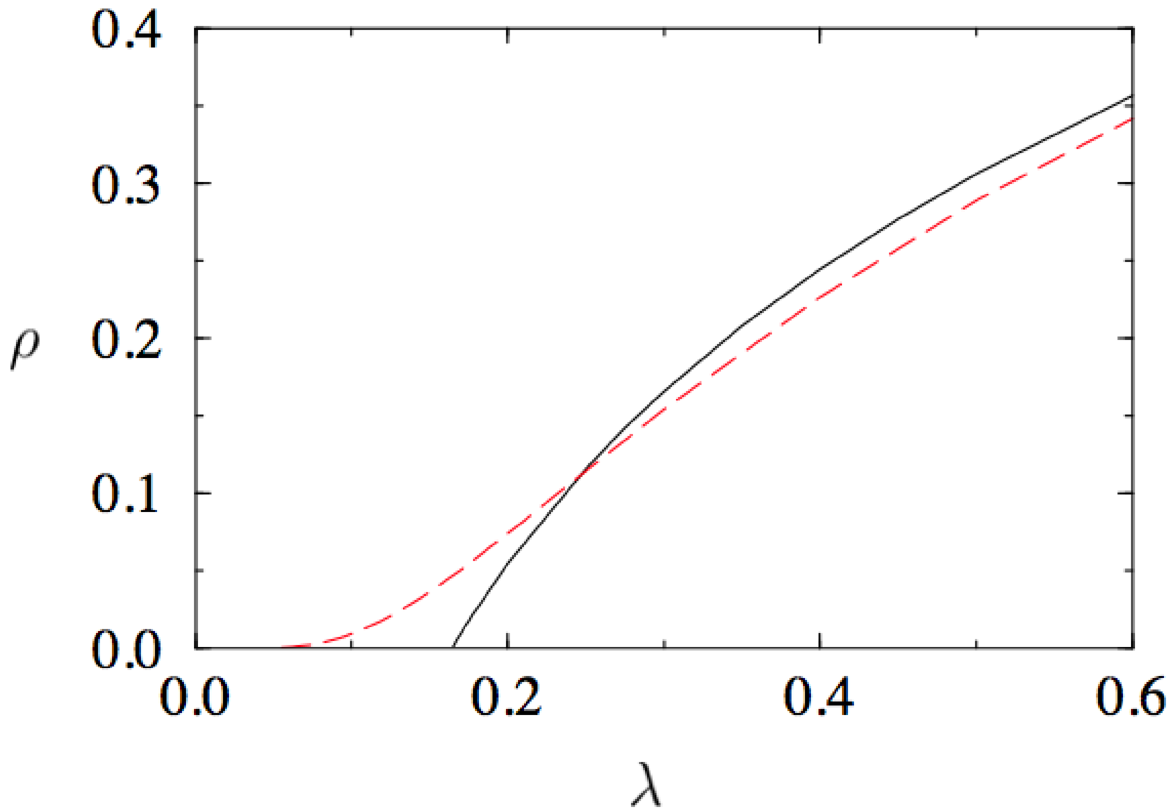


FIG. 1. Density of infected nodes ρ as a function of λ in the WS network (full line) and the BA network (dashed line).

Pastor-Satorras, Romualdo, and Alessandro Vespignani. "Epidemic dynamics and endemic states in complex networks." Physical Review E 63.6 (2001): 066117.

Networked markets in economics: Empirical evidence of the role of social networks in the labor market

In an early study of the textile industry Myers and Shultz (1951) shown that 62% of interviewed workers found their job through a social contact, 23% by direct application, and 15% through an employment agency, advertisement or other means.

Similar results were obtained by Rees and Shultz (1970) and by Granovetter (1995).

Summarizing the results of 24 studies, Truman Bewley (1999) estimated that 30 to 60 percent of jobs were found through friends or relatives.

TABLE 1
Education and Methods of Job Search

Category	1	2	3	4	5	6	7	8	Full	All
	did nothing	pub. agency	priv. agency	curr. emplr	other emplr	friend acquai.	ads	other activ.	searchers	PSID
Unemployed										
Sample frequencies	11.5	21.8	9.5	5.7	29.4	15.5	33.5	31.0	5.8	100
Years ≤ 8	17.9	20.5	18.2	24.2	21.0	23.3	16.9	14.4	17.0	16.7
8 < Years < 11	41.8	25.2	25.5	30.3	28.7	28.9	27.2	30.6	31.6	17.9
Years = 12	26.9	32.2	25.5	30.3	28.7	27.8	32.3	28.9	30.8	31.0
Years = 12 + nonac.	7.5	18.7	25.5	15.2	17.0	12.2	15.4	17.8	14.6	18.7
13 < Years < 15	4.5	9.5	5.5	0.0	3.5	6.7	6.7	5.0	4.3	9.4
BA + adv.	1.5	6.2	0.0	0.0	1.2	1.1	1.5	3.3	1.7	6.2
Searching on-the-job										
Sample frequency	60.5	6.3	3.0	2.0	9.4	8.5	16.0	17.0	8.1	100
Years ≤ 8	7.1	9.8	16.7	6.3	13.2	11.6	13.2	6.6	8.3	16.7
8 < Years < 11	17.2	17.7	12.5	12.5	13.2	14.5	11.6	8.0	14.3	17.9
Years = 12	33.7	41.2	25.0	31.3	34.2	29.0	38.0	38.7	34.9	31.0
Years = 12 + nonac.	27.2	25.5	25.0	43.8	22.4	21.7	18.6	22.6	25.3	18.7
13 < Years < 15	9.4	3.9	16.7	6.3	15.8	13.0	14.0	18.3	11.9	9.4
BA + adv.	5.3	2.0	4.2	0.0	1.32	10.1	4.7	5.8	5.2	6.2

Notes: The categories are whether: 1. did nothing; 2. searched with a public employment agency; 3. searched with a private employment agency; 4. checked with the current employer; 5. checked with other employer; 6. checked with friend or relative; 7. placed or answered ads; or, 8. engaged in other activity. The results are summarized in the following table. The entries in the lines labelled "sample frequencies" are not mutually exclusive—some respondents may be engaged in more than one method—and thus do not add up to 100. The entries for educational attainment sum up to 100 in each column. The column labelled "Full" gives the educational attainments for the respective subsample of unemployed and those searching on the job in the 1993 sample of the PSID. The column labelled "All" gives the educational attainments for the entire 1993 sample of the PSID.

The Panel Study of Income Dynamics - PSID - is the longest running longitudinal household survey in the world.

From: Ioannides, Y. M., and L. Datcher Loury. "Job information networks, neighborhood effects, and inequality." Journal of economic literature (2004): 1056-1093.

TABLE 2
Urban Size and Methods of Job Search

Category	1 did nothing	2 pub. agency	3 priv. agency	4 curr. emplr	5 other emplr	6 friend acquai.	7 ads	8 other activ.	Full searchers	All PSID
Unemployed										
Sample frequencies	11.5	21.8	9.5	5.7	29.4	15.5	33.5	31.0	5.8	100
> 500,000	38.20	27.17	15.60	35.10	13.17	51.11	22.21	23.69	23.86	16.02
[100000, 500000)	12.05	19.73	10.95	3.20	34.59	14.48	19.07	21.66	20.99	24.15
[50000, 100000)	1.69	7.28	15.34	21.16	7.12	8.37	15.61	6.86	9.39	11.76
[25000, 50000)	17.14	7.88	30.43	26.91	7.53	9.87	13.58	16.94	14.36	13.44
[10000, 25000)	3.31	14.57	22.81	5.88	17.94	7.39	20.86	18.36	14.30	15.74
10,000 >	27.62	23.38	4.87	5.21	17.14	8.78	8.11	12.16	15.85	17.38
Employed										
Sample frequency	60.5	6.3	3.0	2.0	9.4	8.5	16.0	17.0	8.1	100
> 500,000	16.19	6.36	7.85	.00	12.28	22.43	14.03	10.17	14.79	16.02
[100000, 500000)	30.77	22.63	22.92	19.27	27.47	22.24	28.45	27.48	30.09	24.15
[50000, 100000)	11.67	13.15	41.89	29.72	25.93	17.68	18.38	18.94	13.43	11.76
[25000, 50000)	12.21	13.29	.00	21.76	12.43	17.49	18.04	10.90	12.90	13.44
[10000, 25000)	13.68	5.98	2.18	1.33	6.73	5.81	10.42	16.27	13.08	15.74
10,000 >	12.10	38.60	25.16	27.93	15.17	14.34	10.68	16.25	13.44	17.38

Notes: The categories are whether: 1. did nothing; 2. searched with a public employment agency; 3. searched with a private employment agency; 4. checked with the current employer; 5. checked with the other employer; 6. checked with friend or relative; 7. placed or answered ads; or, 8. engaged in other activity.
The entries in the lines labelled "sample frequency" are not mutually exclusive—some respondents may be engaged in more than one methods—and thus do not add up to the number in column "Full". The column labelled "Full" gives the relative geographical distribution of the two respective categories, unemployed and employed looking for job, for the entire 1993 sample of the PSID. "All" gives the geographical distribution of the entire 1993 sample of the PSID. All calculations are weighted by means of the latest weight in PSID.
The geographical categories are defined in terms of the size of the largest city in the county of a household's residence. The categories are: SMSA with largest city 500,000 or more; SMSA with largest city between 100,000 and 499,000; SMSA with largest city 50,000 to 99,999; non SMSA with largest city 25,000 to 49,999; non-SMSA with largest city 10,000 to 24,999; non SMSA with largest city less than 10,000.

Strong and weak ties

Granovetter's work has been extremely influential in this area of research.

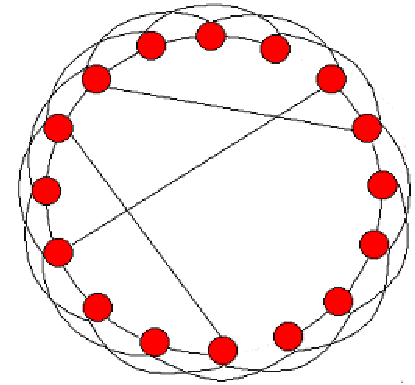
Granovetter obtained a proxy of the strength of a ties by considering the number of times that the two social actors have interacted during a year (strong= at least twice a week, medium= less than twice a week but more than once a year, and weak=once a year or less).

In his pioneering study, of the 54 people who found job through a social contact, 16.7% found their job through a strong tie, 55.6% through a medium tie and 27.8% through a weak one.

The Strength of Weak Ties[1]

Mark S. Granovetter
Johns Hopkins University

1360 *AJS* Volume 78 Number 6



In a random sample of recent professional, technical, and managerial job changers living in a Boston suburb, I asked those who found a new job through contacts how often they *saw* the contact around the time that he passed on job information to them. I will use this as a measure of tie strength.[15] A natural a priori idea is that those with whom one has strong ties are more motivated to help with job information. Opposed to this greater motivation are the structural arguments I have been making: those to whom we are weakly tied are more likely to move in circles different from our own and will thus have access to information different from that which we receive.

I have used the following categories for frequency of contact: often = at least twice a week; occasionally = more than once a year but less than twice a week; rarely = once a year or less. Of those finding a job through contacts, 16.7% reported that they saw their contact often at the time, 55.6% said occasionally, and 27.8% rarely ($N = 54$).[16] The skew is clearly to the weak end of the continuum, suggesting the primacy of structure over motivation.

A stylized time dependent labor market model: Calvò-Armengol and Jackson

In the basic setting the model presents the following characteristics:

- job information arrives directly or through neighbors;
- information about new job opening arrives randomly to the agents of the network;
- if the worker is unemployed she takes the job;
- if the worker is employed she select randomly an unemployed neighbor and transfer the information;
- there is only one type of link (no difference between strong and weak ties).

The system operates over time and final state at time *t*-1 is the starting state at the beginning of the next period

Calvo-Armengol, Antoni, and Matthew O. Jackson. "The effects of social networks on employment and inequality." American economic review (2004): 426-454.

The model therefore allows tracking the time dynamics of unemployment in the system. One can therefore deduce long-run steady state distribution of employment and the role of network structure in it. It can also be able to model the empirical observation of the duration dependence (i.e. the fact that workers who have been unemployed for longer times are less likely to find a work than workers who are just recently unemployed).

Model description:

- all jobs are identical and there is a single wage level;

- n agents are connected by an undirected network represented by a n x n symmetric adjacency matrix g which has entries 0 and 1.

- time evolves in discrete periods indexed by t=1,2, ... ,

- the n-dimensional vector s_t describes the employment status of agents at time t. At the end of the t period s_{it}=1 indicates employment and s_{it}=0 unemployment of agent i.

- each agent directly hears about a job with probability

$$a \in [0,1]$$

- the arrival of job information is independent across agents. When agent i receives job information and is unemployed he or she becomes employed. If he or she is already employed he or she passes the information to a random unemployed agent (if present).

The probability that agent i learns about a job and this job is taken by agent j is

$$p_{ij}(s_{t-1}) = \begin{cases} a & \text{if } s_{i,t-1} = 0 \text{ and } i=j \\ \dfrac{a}{\sum_{k:s_{k,t-1}=0} g_{ik}} & \text{if } s_{i,t-1} = 1,\ s_{j,t-1} = 0, \text{ and } g_{ij}=1 \\ 0 & \text{otherwise} \end{cases}$$

At the end of the period some agents lose their jobs with probability

$$b \in [0,1]$$

Long run steady state probability

a) Isolated agent

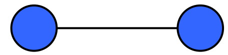
Let us indicates the long run steady state probability of an agent to be employed as μ

For the isolated agent

$$m=(1-b)\left(m+a(1-m)\right)$$

where $(1-b)\mu$ is the probability of being employed and not loosing the job and $a(1-b)(1-\mu)$ is the probability of being unemployed, getting a job and not loosing it.

Therefore $$m=\frac{1}{1+\frac{b}{(1-b)a}}$$

The case of a dyad 

The steady state distribution can be obtained in terms of the probabilities μ_0, μ_1 and μ_2. Where μ_0 is the probability that no agent of the dyad is employed, μ_1 is the probability that one agent of the dyad is employed and μ_2 is the probability that both agents are employed.

The relations among these variables are given by the following equations

$$\begin{pmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} (1-a+ab)^2 & (1-a)^2 b(1-b)+b^2 & b^2 \\ 2a(1-b)(1-a+ab) & (1-b)\left((1-a)^2(1-2b)+2b\right) & 2b(1-b) \\ a^2(1-b)^2 & \left(1-(1-a)^2\right)(1-b)^2 & (1-b)^2 \end{pmatrix} \begin{pmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \end{pmatrix}$$

Let us consider in detail the relation

$$m_2 = a^2 (1- b)^2 m_0 + \left(1- (1- a)^2\right)(1- b)^2 m_1 + (1- b)^2 m_2$$

The steady state probability μ_2 of a dyad showing employment for both agents is given by

$$a^2 (1- b)^2 m_0$$

which is the probability that both agents were unemployed, get employed and both stayed employed, plus

$$\left(1- (1- a)^2\right)(1- b)^2 m_1$$

which is the probability that one agent was employed and the unemployed receives an offer of employment directly or through the already employed and both stay employed, plus

$$(1- b)^2 m_2$$

probability that both agents were employed, and stayed employed

The solution of the linear systems can be worked out but it is quite cumbersome.

$$\begin{pmatrix} m_0 \\ m_1 \\ m_2 \end{pmatrix} = \begin{pmatrix} b^2\left(1+(1-b)(1-a)^2\right)/X \\ 2a(1-b)\left(1+(1-b)(1-a)\right)/X \\ a^2(1-b)^2\left((1-a)(3-a)(1-b)+1\right)/X \end{pmatrix}$$

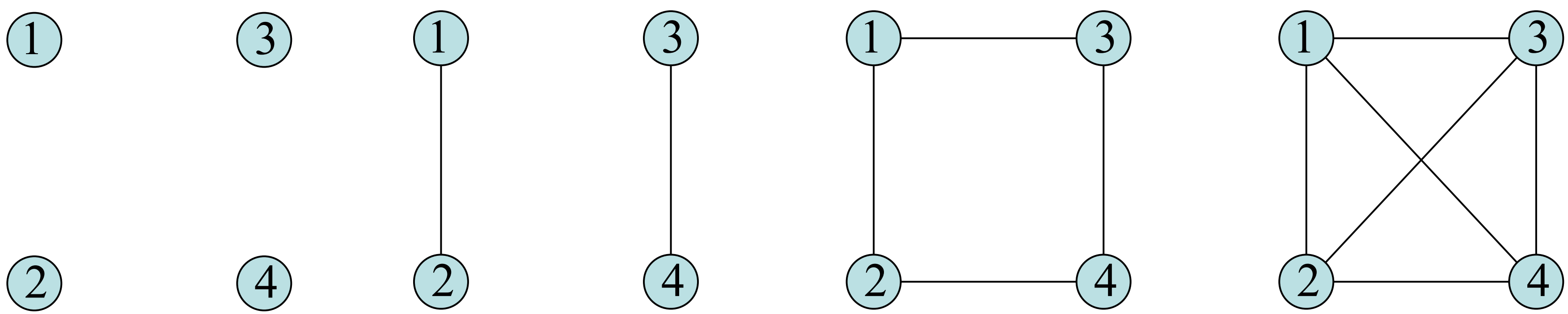
where

$$X = b^2\left(1+(1-b)(1-a)^2\right)+2a(1-b)\left(1+(1-b)(1-a)\right)$$
$$+a^2(1-b)^2\left((1-a)(3-a)(1-b)+1\right)$$

Therefore the network type has an impact on the average employment rate and on the correlation of employment of pairs.

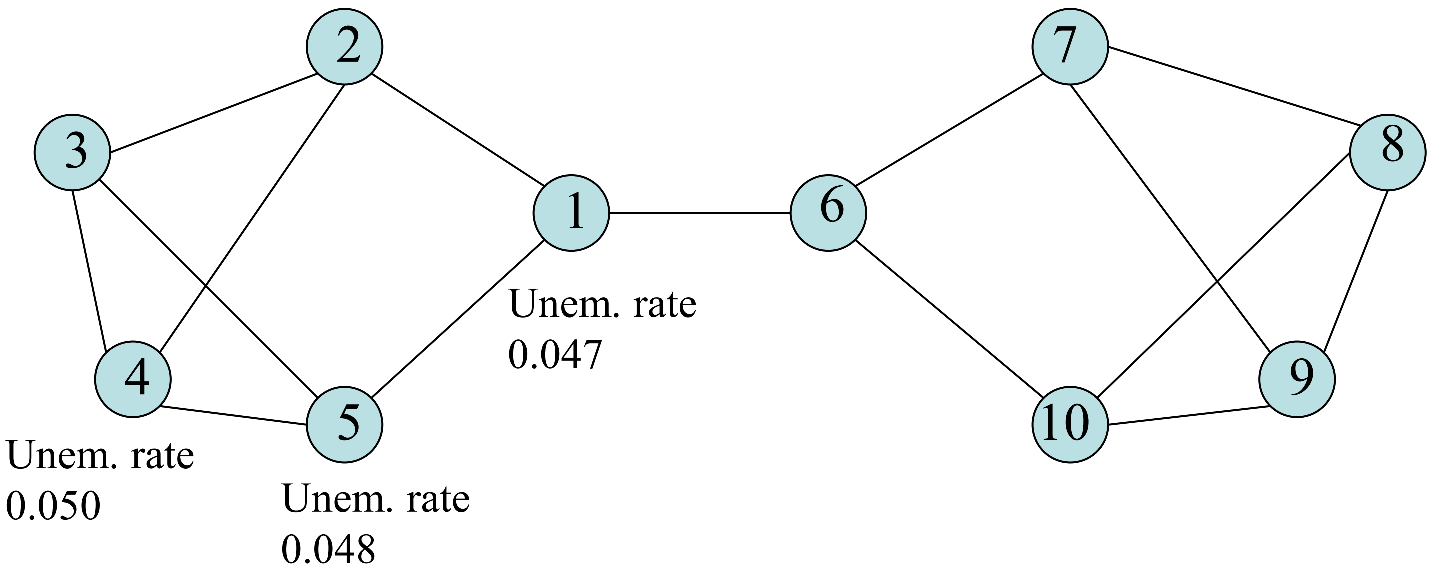
For other different network shapes it is difficult to analytically estimate the steady state of the unemployment rate but such information can be obtained by numerical simulations

For example for the following cases Calvò-Armengol and Jackson obtain when a=0.100 and b=0.015



probability 1 unemployment 0.132	probability 1 unemployment 0.083 corr(1,2)=0.041	probability 1 unemployment 0.063 corr(1,2)=0.025 corr(1,4)=0.019	probability 1 unemployment 0.050 corr(1,2)=0.025 corr(1,4)=0.025

By investigating a more structured network Calvò-Armengol and Jackson found



The degree of each node is 3

The parameters of the model for this simulation are a=0.100 and b=0.015

A system highly studied in econophysics is the international trade network (also called world trade web).

Nodes are countries

Edges are directed and weighted according to the amount in dollars of imported or exported goods

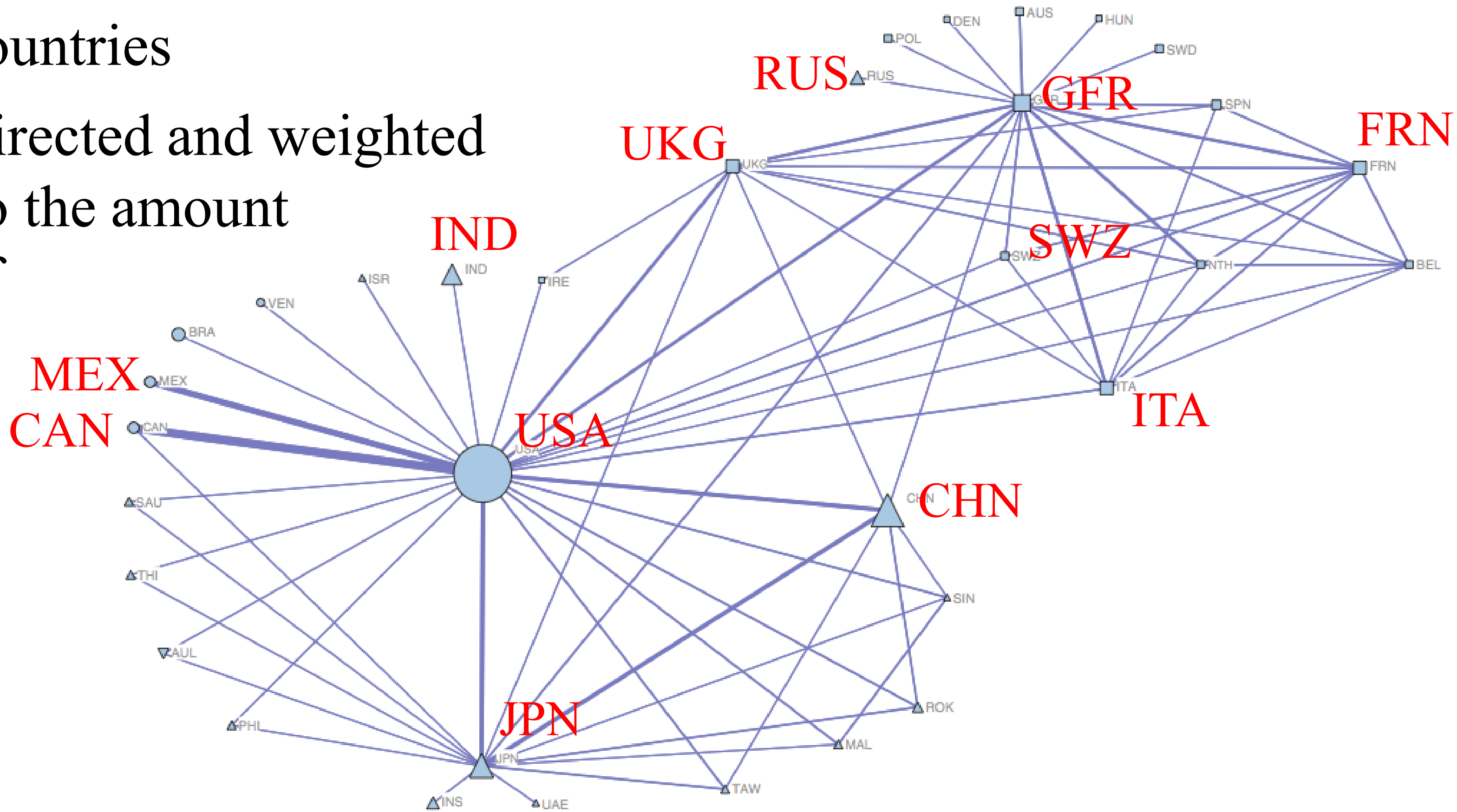


Fig. 5 A partial visualization of the original weighted ITN (W). Thickness of links is proportional to their weight. Only the largest 1% of links are shown. Node sizes are proportional to country's GDP. *Node shapes* represent the continent which the country belongs to (*Circles*: America; *Empty Squares*: Europe; *Upright Triangles*: Asia; *Crossed Squares*: Africa; *Reversed Triangles*: Pacific)

Fagiolo, G., 2010. The international-trade network: gravity equations and topological properties. Journal of Economic Interaction and Coordination, 5(1), pp.1-25.

Most of the earlier studies have focused on the topological properties of the international trade network.

Serrano, M.Á. and Boguñá, M., 2003. Topology of the world trade web. Physical Review E, 68(1), p.015101.

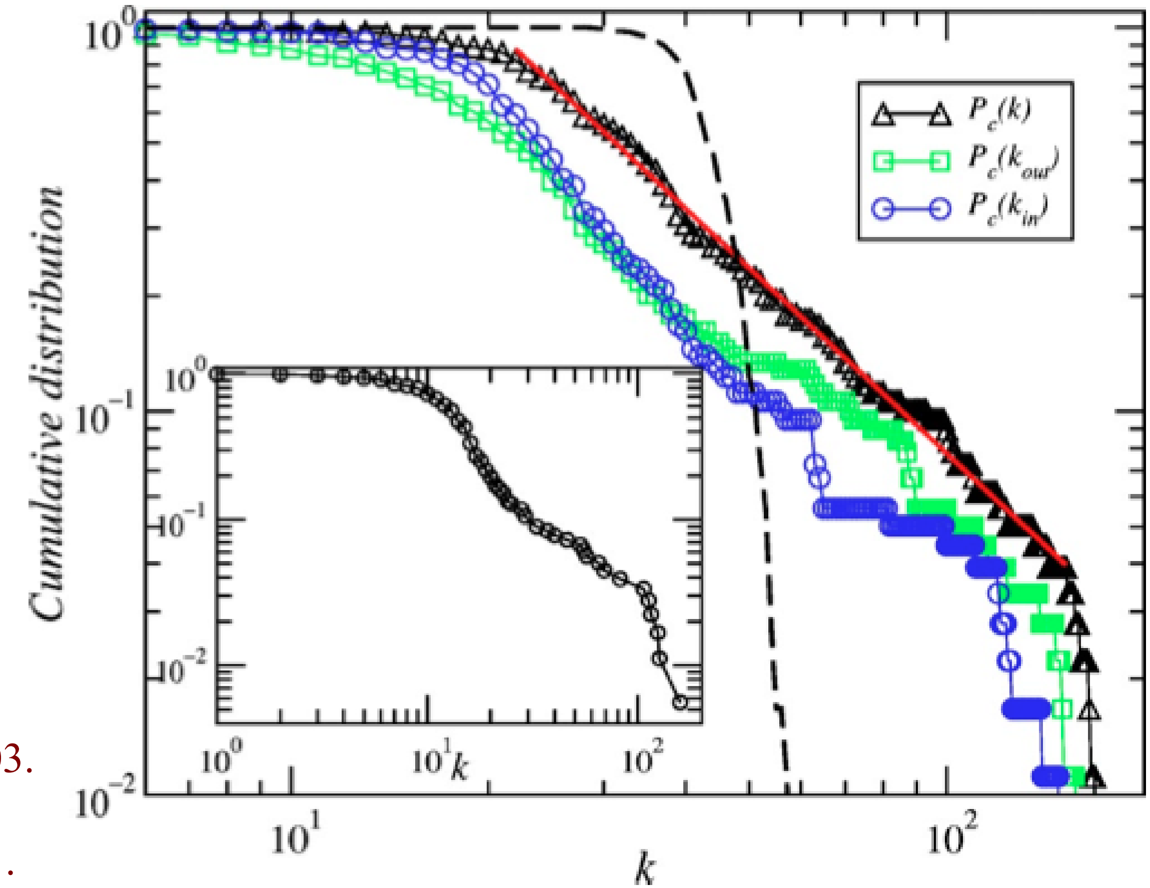


FIG. 1. Cumulative in- and out-degree distributions, $P_c(k_{in})$ and $P_c(k_{out})$, and undirected $P_c(k)$ corresponding to the import and export world trade web. The solid line is a power law fit of the form $P_c(k) \sim k^{\gamma-1}$ with $\gamma = 2.6 \pm 0.1$. The cumulative distribution of the equivalent random network with the same average degree is also shown. Inset: cumulative distribution for the WTW with reciprocal edges.

Local clustering coefficient

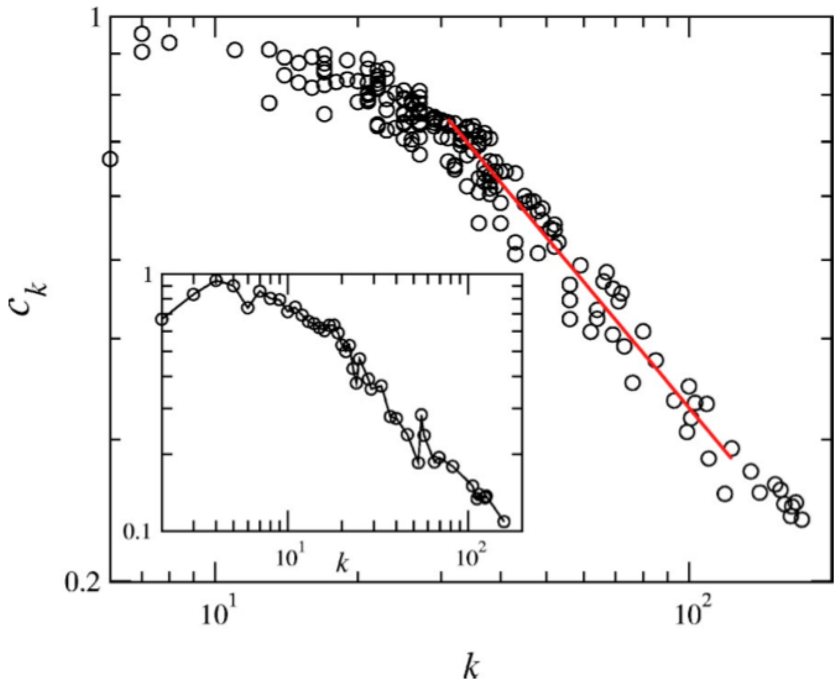


FIG. 2. Clustering coefficient of single countries as a function of their degree for the undirected version of the WTW. The solid line is a power law fit of form $c_k \sim k^{-\omega}$ with $\omega = 0.7 \pm 0.05$. Inset: average clustering coefficient, as a function of the degree, for the WTW with reciprocal edges.

Average nearest neighbors degree

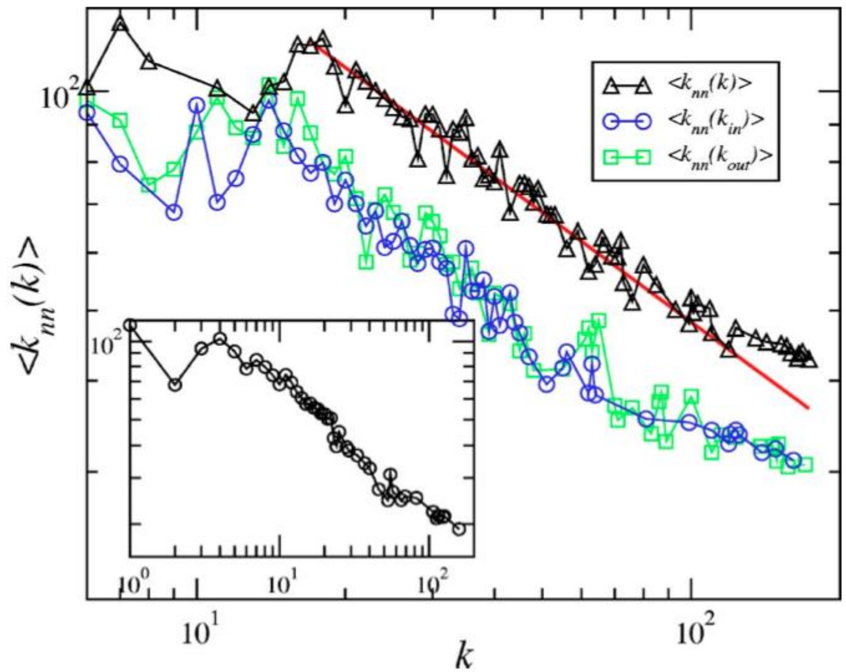
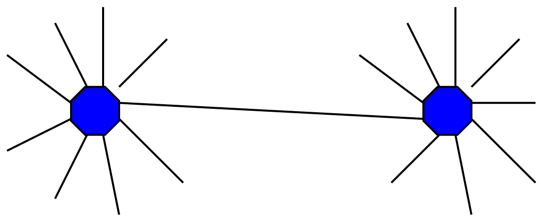


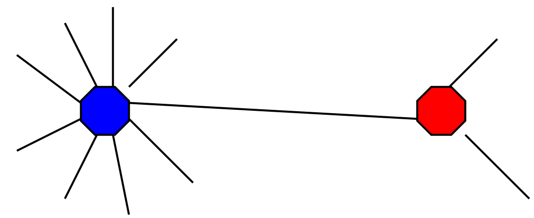
FIG. 3. Average in (out, undirected) nearest neighbors degree as a function of the in (out, undirected) degree of the vertex. The solid line is a fit of the form $\langle k_{nn}(k) \rangle \sim k^{-\nu_k}$ with $\nu_k = 0.5 \pm 0.05$. Inset: the same for the WTW with reciprocal edges.

Serrano, M.Á. and Boguñá, M., 2003. Topology of the world trade web. Physical Review E, 68(1), p.015101.

One way to measure the "assortativity" of a network is through the average nearest neighbor degree (or in the weighted case strength)



Assortative networks



Disassortative networks

$$\langle k_i^{nn} \rangle = \frac{\sum_j \left(p_{ij} + p_{ji}\right) k_j}{k_i} \qquad p_{ij} = \frac{k_i k_j}{1 + k_i k_j}$$

The international trade network is disassortative. What about comparing this stylized fact with a "null" model?

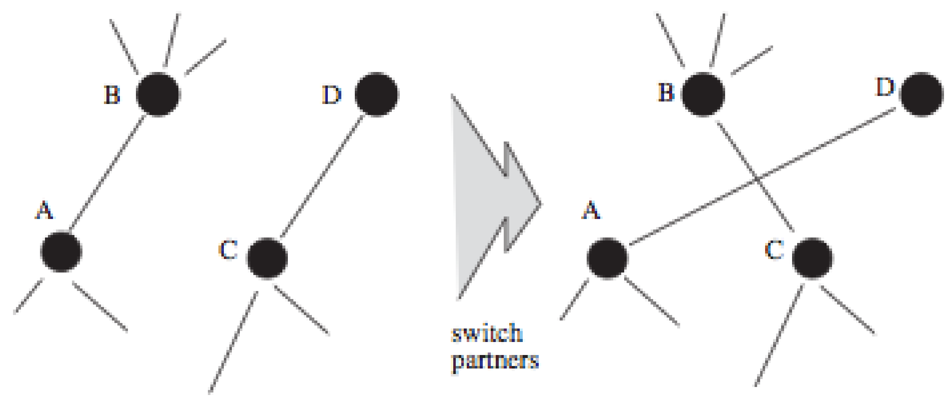
The first step is to clearly identify the null model.

In the simplest case the unweighted network is considered and a good random "null" model is the so-called "configuration model". The configuration model is just a network model characterized by the same degree sequence as the one investigated but with links randomly defined.

A configuration model is usually numerically investigated.

In the numerical approach random realizations of the configuration model are obtained through the repeated application of the rewiring procedure

S. Maslov et al. / Physica A 333 (2004) 529–540



Maslov, S., Sneppen, K. and Zaliznyak, A., 2004. Detection of topological patterns in complex networks: correlation profile of the internet. Physica A333, pp.529-540.

There are also analytical approaches based on the maximization of an Entropy. This approach estimates the set of graphs G such that the P(**G**) subject to some specified constraints $\{C_a\}$ maximizes the entropy

$$S = -\sum_{\mathbf{G}} P(\mathbf{G}) \ln P(\mathbf{G})$$

The formal solution to the entropy maximization problem can be written in terms of the so-called Hamiltonian H(**G**) providing the "energy" or fitness associated to a given graph G

$$H(\mathbf{G}) = \sum_{a} \theta_a C_a(\mathbf{G})$$

The maximum entropy is obtained when

$$P(\mathbf{G}) = \frac{e^{-H(\mathbf{G})}}{Z} \quad \text{with} \quad Z = \sum_{\mathbf{G}} e^{-H(\mathbf{G})}$$

Squartini, T., Fagiolo, G. and Garlaschelli, D., 2011. Randomizing world trade. I. A binary network analysis. Physical Review E, 84(4), p.046117.

The analytical approach provides results such that the ensemble average of each constraint C_a is equal to the empirically observed value whereas the numerical rewiring procedure guarantees that the observed value in the null model is exactly equal to the empirically observed one. The results of the two approaches converge for large networks.

One plus of the analytical approach is that it works both for unweighted (called by the authors "binary") and weighted networks. The rewiring procedure is not unique and less well defined in weighted networks. In both cases the degree sequence is usually not controlled.

Squartini, T., Fagiolo, G. and Garlaschelli, D., 2011. Randomizing world trade. I. A binary network analysis. Physical Review E, 84(4), p.046117.

Squartini, T., Fagiolo, G. and Garlaschelli, D., 2011. Randomizing world trade. II. A weighted network analysis. Physical Review E, 84(4), p.046118.

Comparison with a maximum entropy null model in the International Trade Network

unweighted	weighted

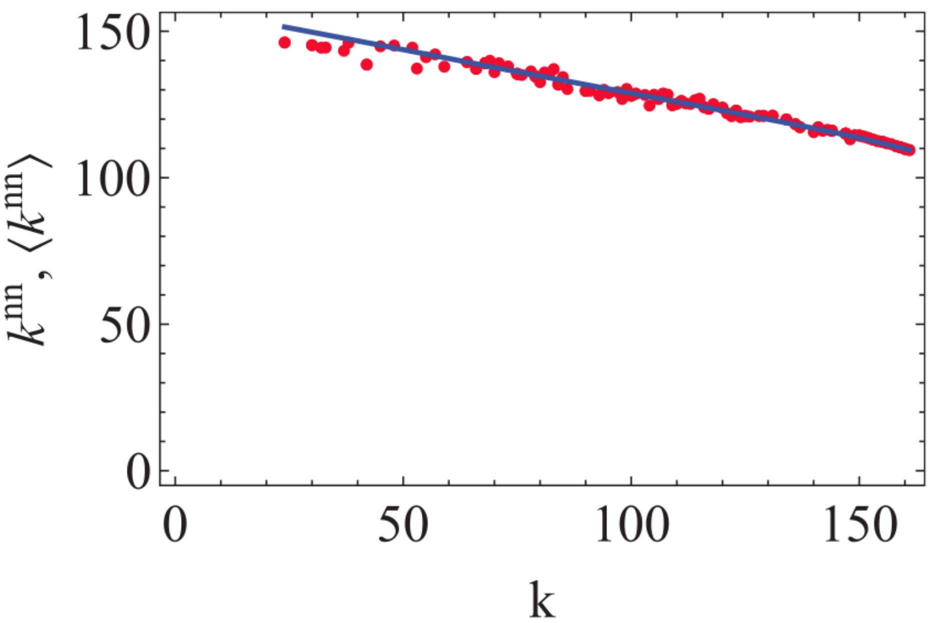


FIG. 1. (Color online) Average nearest-neighbor degree k_i^{nn} versus degree k_i in the 2002 snapshot of the real binary undirected ITN (red points), and corresponding average over the maximum-entropy ensemble with specified degrees (blue solid curve).

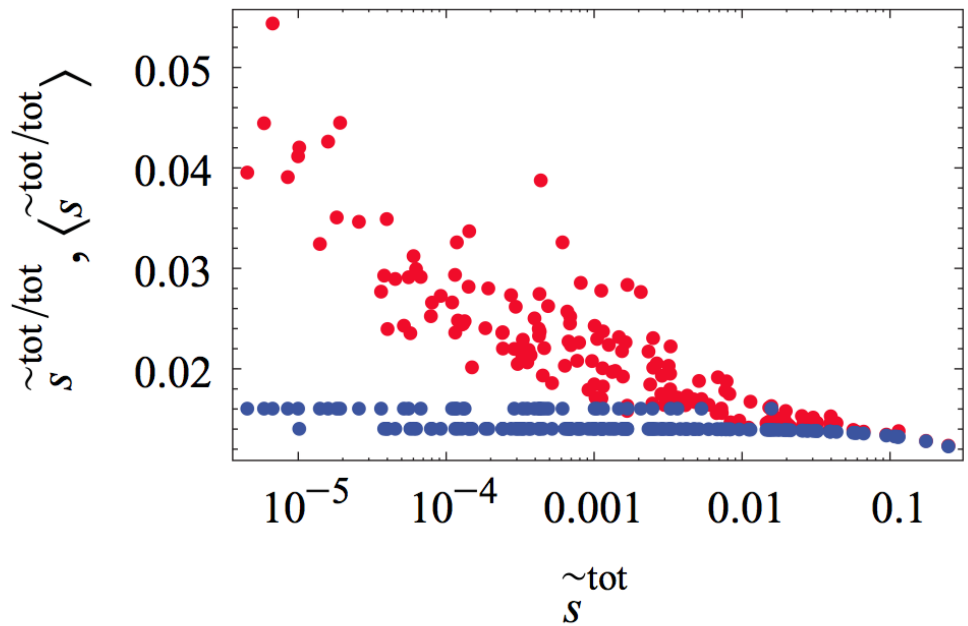


FIG. 9. (Color online) Total average nearest-neighbor strength $\tilde{s}_i^{tot/tot}$ versus total strength $\tilde{s}_i^{tot}$ in the 2002 snapshot of the real weighted directed ITN (red, upper points), and corresponding average over the null model with specified out-strengths and in-strengths (blue, lower points).

Squartini, T., Fagiolo, G. and Garlaschelli, D., 2011. Randomizing world trade. I. A binary network analysis. Physical Review E, 84(4), p.046117.

Squartini, T., Fagiolo, G. and Garlaschelli, D., 2011. Randomizing world trade. II. A weighted network analysis. Physical Review E, 84(4), p.046118.

An investigation performed yearly over a long period of time

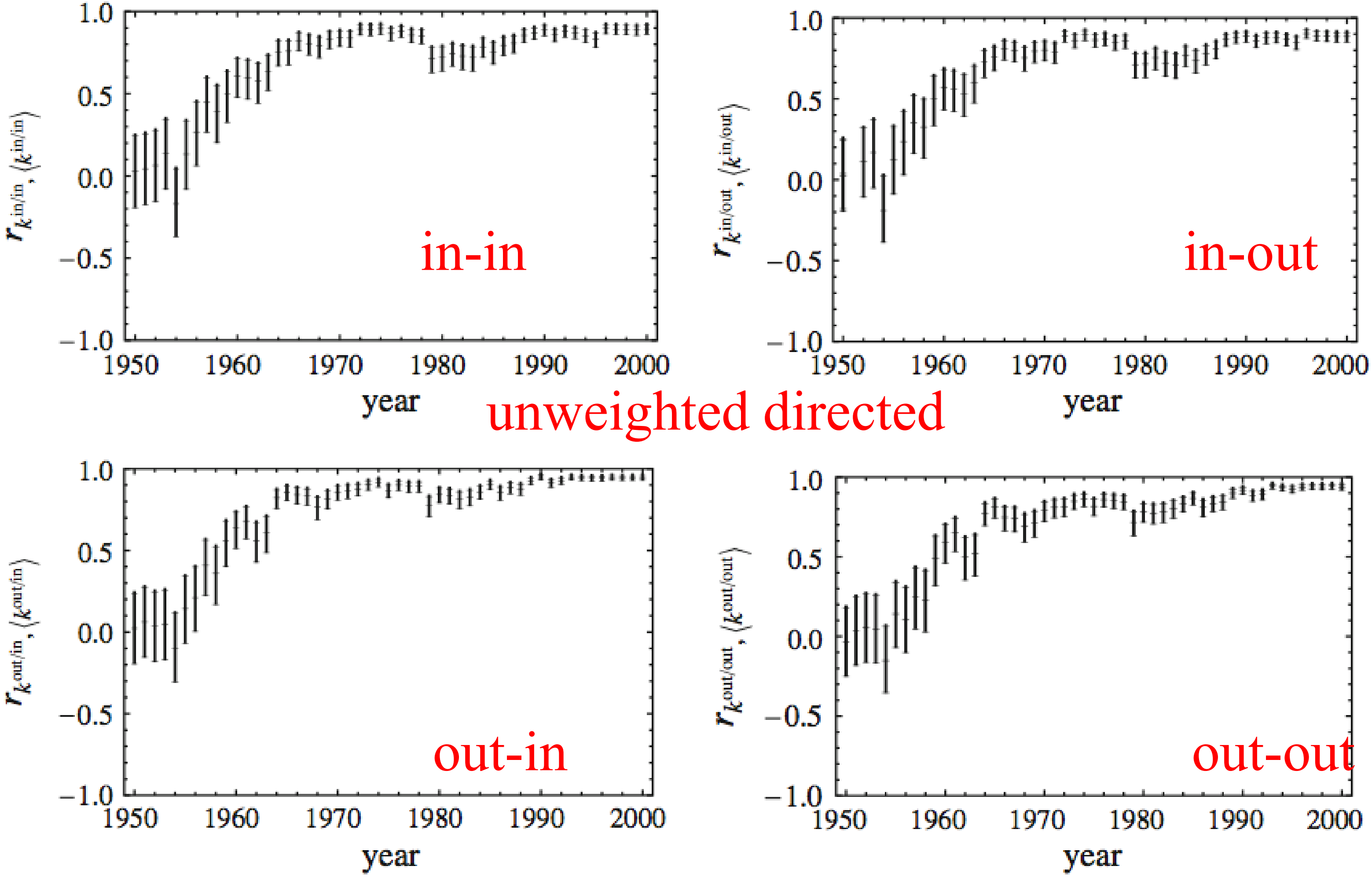


Fig. 3 The binary-directed WTW: Pearson correlation coefficient between observed and null-model node ANND. *Top-left*: IN-IN ANND. *Top-right*: IN-OUT ANND. *Bottom-left*: OUT-IN ANND. *Bottom-right*: OUT-OUT ANND

After 1965 the null model explain quite well the assortativity of the network

A different conclusion is reached for the weighted ITN

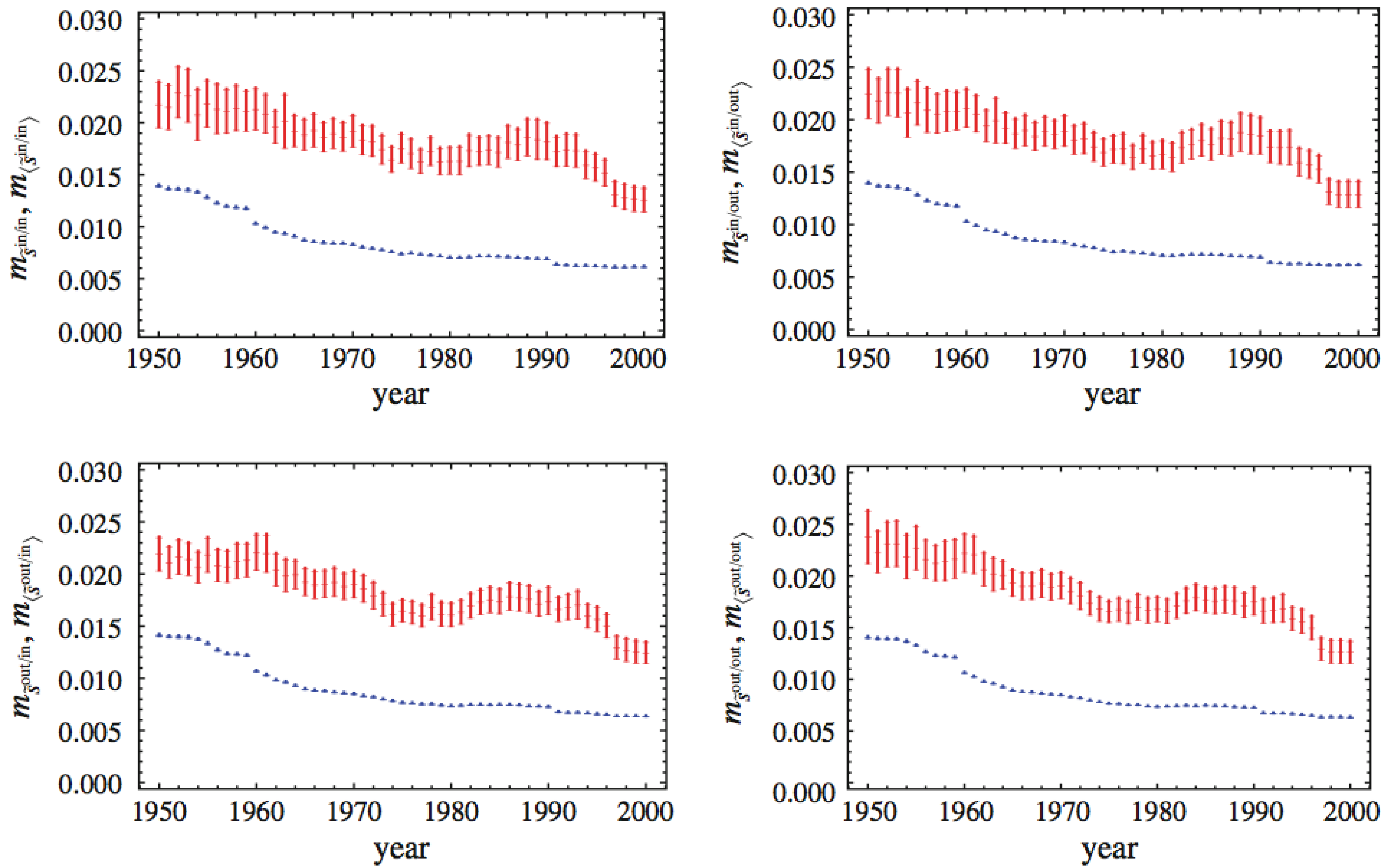


Fig. 10 The weighted-directed WTW: average nearest neighbor strengths and 95 % confidence bands. *Red*: observed quantities. *Blue*: null-model fit. *Top-left*: IN-IN ANNS. *Top-right*: IN-OUT ANNS. *Bottom-left*: OUT-IN ANNS. *Bottom-right*: OUT-OUT ANNS (color figure online)

Fagiolo, G., Squartini, T. and Garlaschelli, D., 2013. Null models of economic networks: the case of the world trade web. Journal of Economic Interaction and Coordination, 8(1), pp.75-107.

Is information about preferential trading and/or indirect trading present in the weighted international trade network?

The comparison with the null hypothesis is suggesting a positive answer.

Several filtering methods have been devised to detect the links that are highly informative in characterizing the system and/or in detecting clustering of nodes (called communities in network science) and over-expression of small sub-graph (called triads or motifs).