INFO I368
Introduction to Network Science

Filippo Menczer

CNetS Center for Complex Networks and Systems Research

INDIANA UNIVERSITY BLOOMINGTON

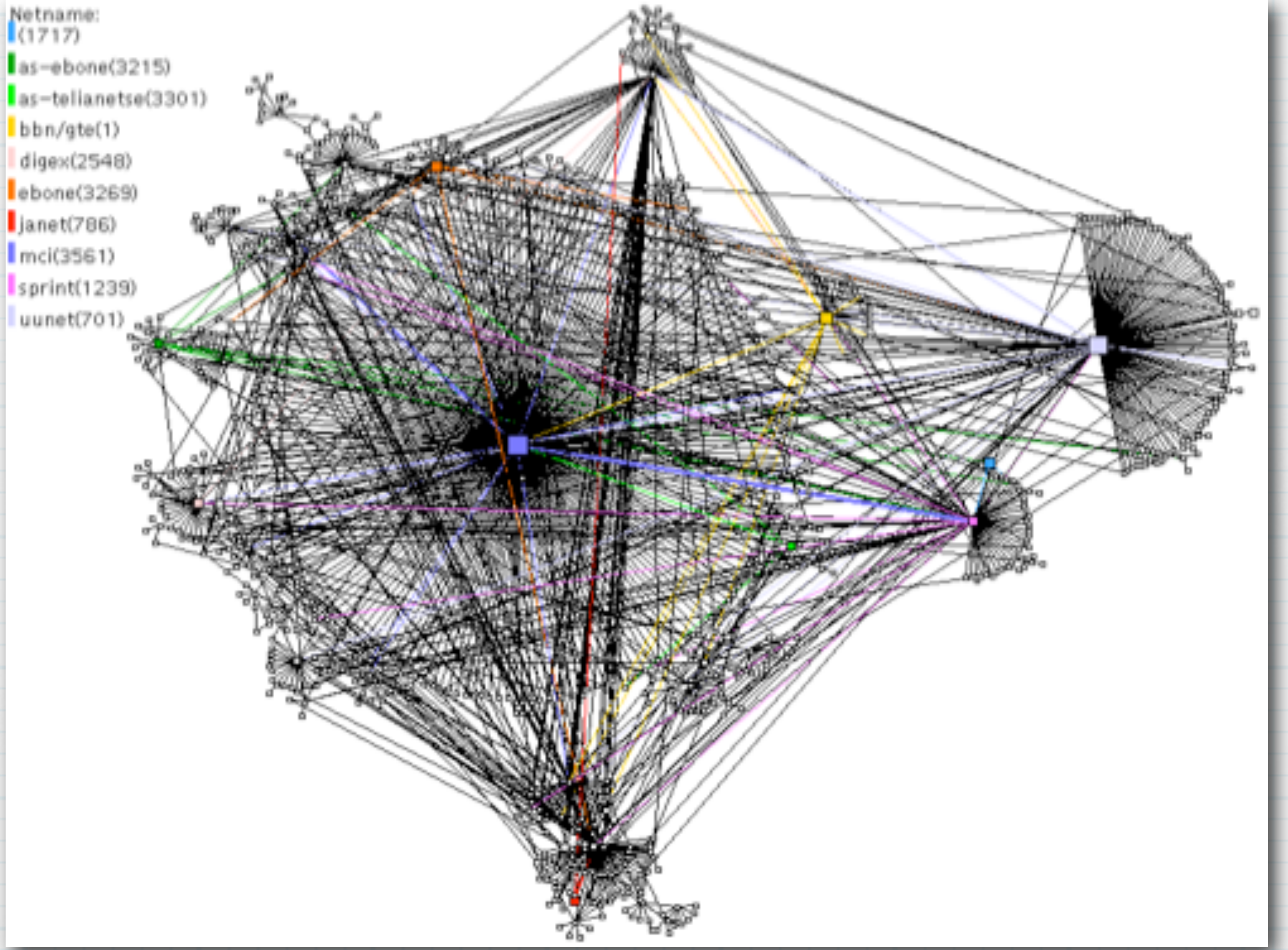
SCHOOL OF Informatics and Computing

- Midterm
- Review
  - Social vs random networks
- Quiz
  - Hubs
  - Universality

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Centrality

Degree as a measure of influence, importance, prestige...



© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

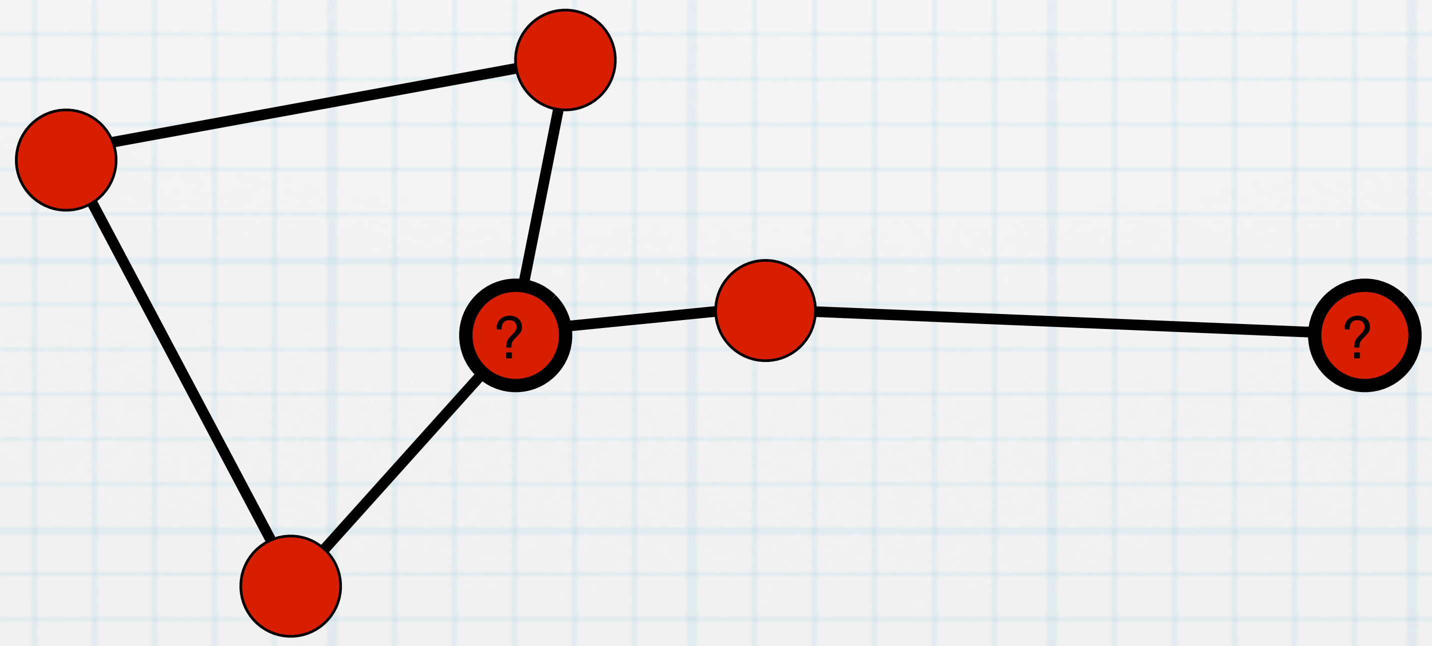
Centrality measures

* Degree centrality
* Closeness centrality
* Betweenness centrality
* PageRank
* Etc......

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Closeness Centrality

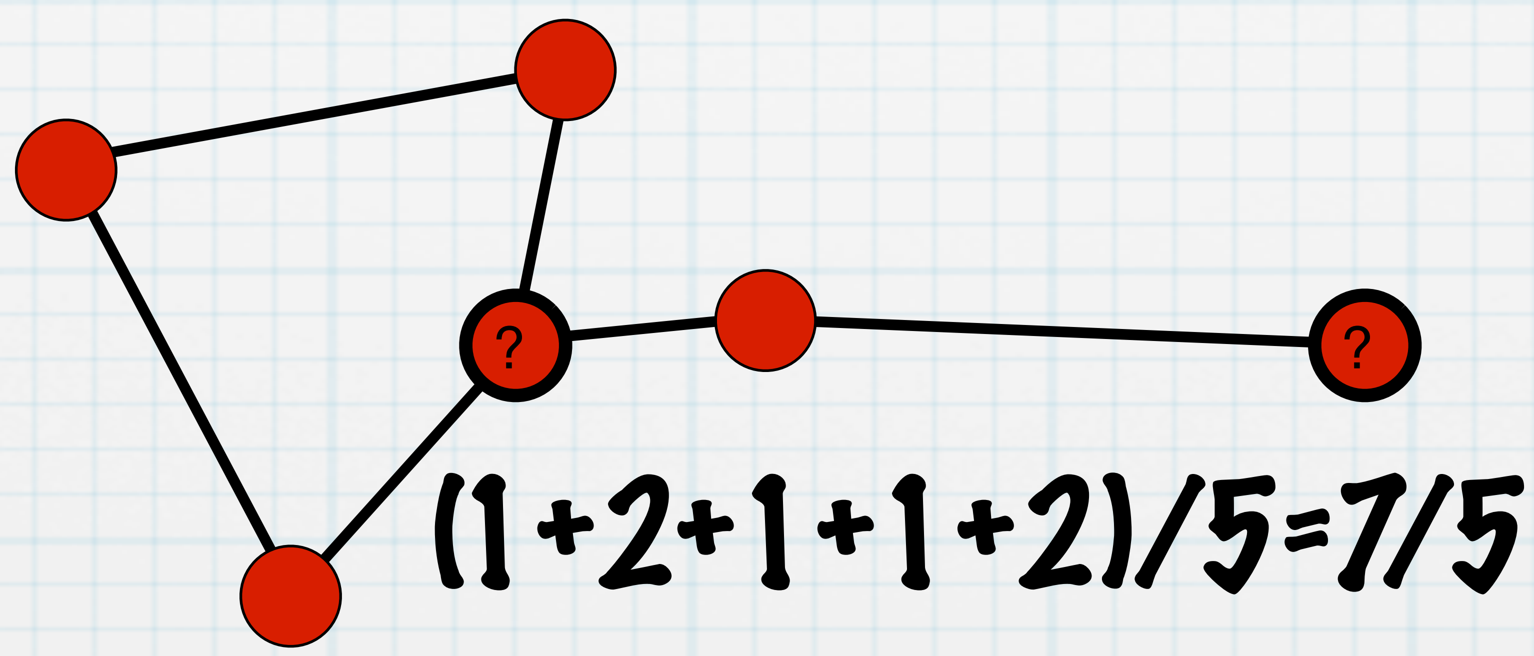
- Closeness centrality
  - Computes the average distance of each node to all other nodes in the system. Nodes are ranked accordingly to the minimum average distance.



© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Closeness Centrality

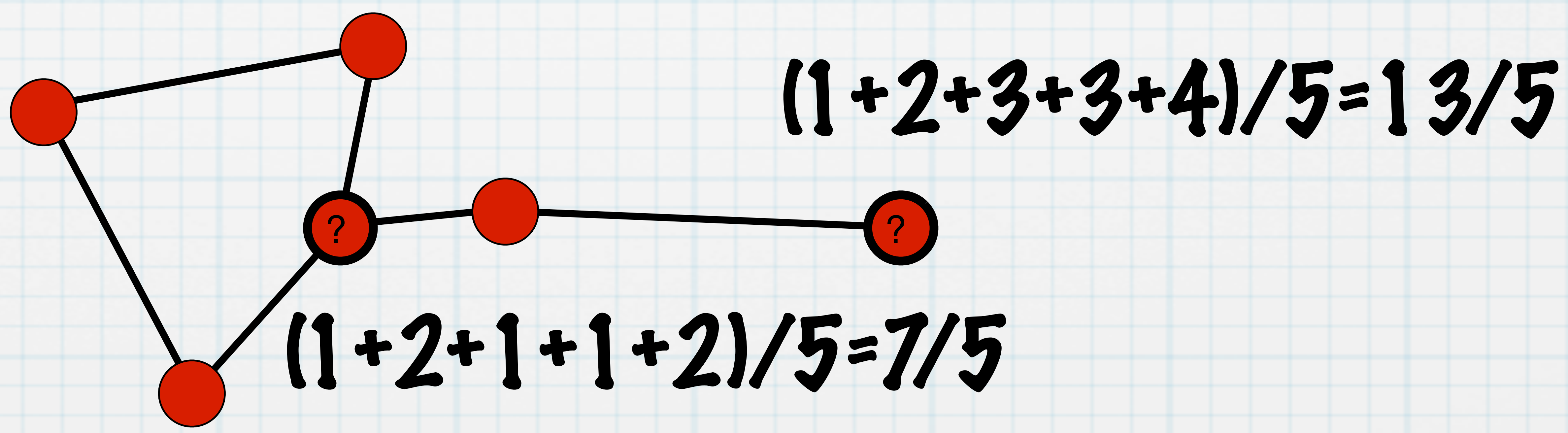
- Closeness centrality
  - Computes the average distance of each node to all other nodes in the system. Nodes are ranked accordingly to the minimum average distance.



© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Closeness Centrality

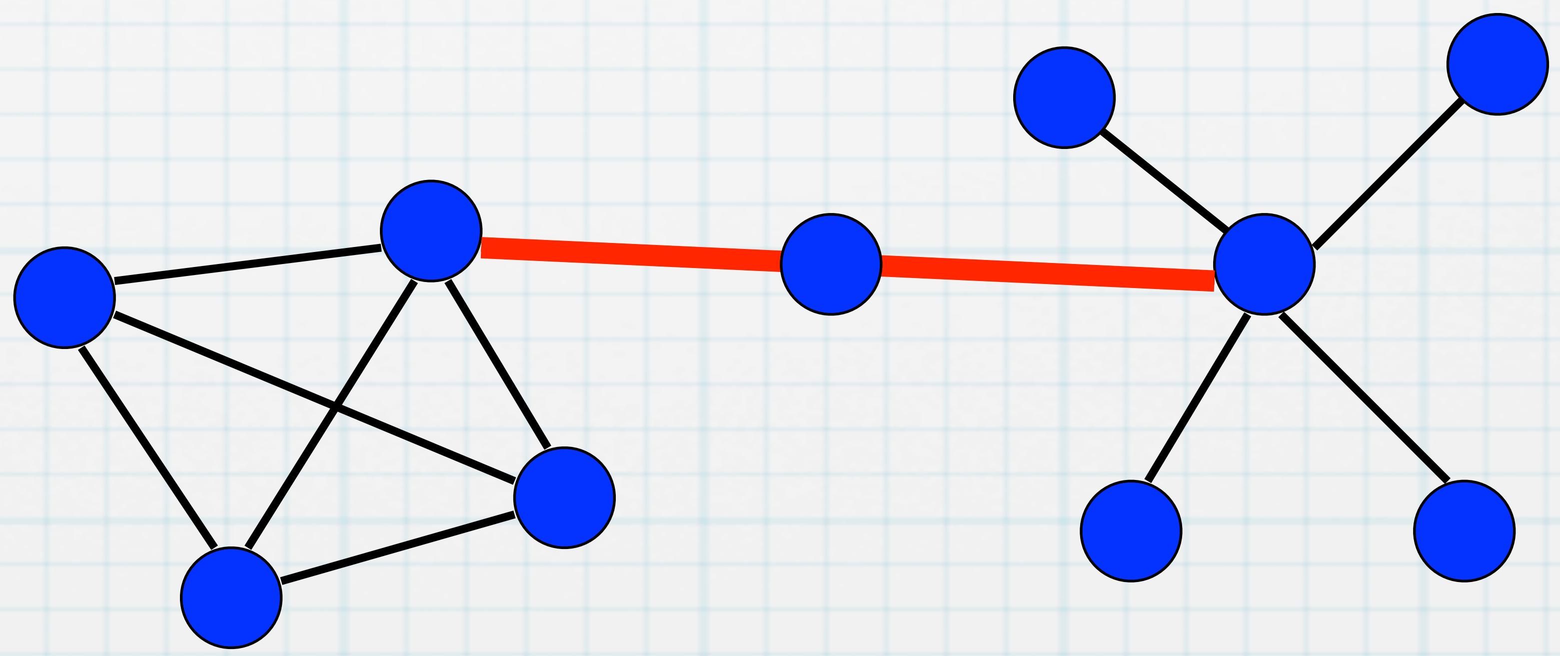
- Closeness centrality
  - Computes the average distance of each node to all other nodes in the system. Nodes are ranked accordingly to the minimum average distance.



© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering
INDIANA UNIVERSITY BLOOMINGTON

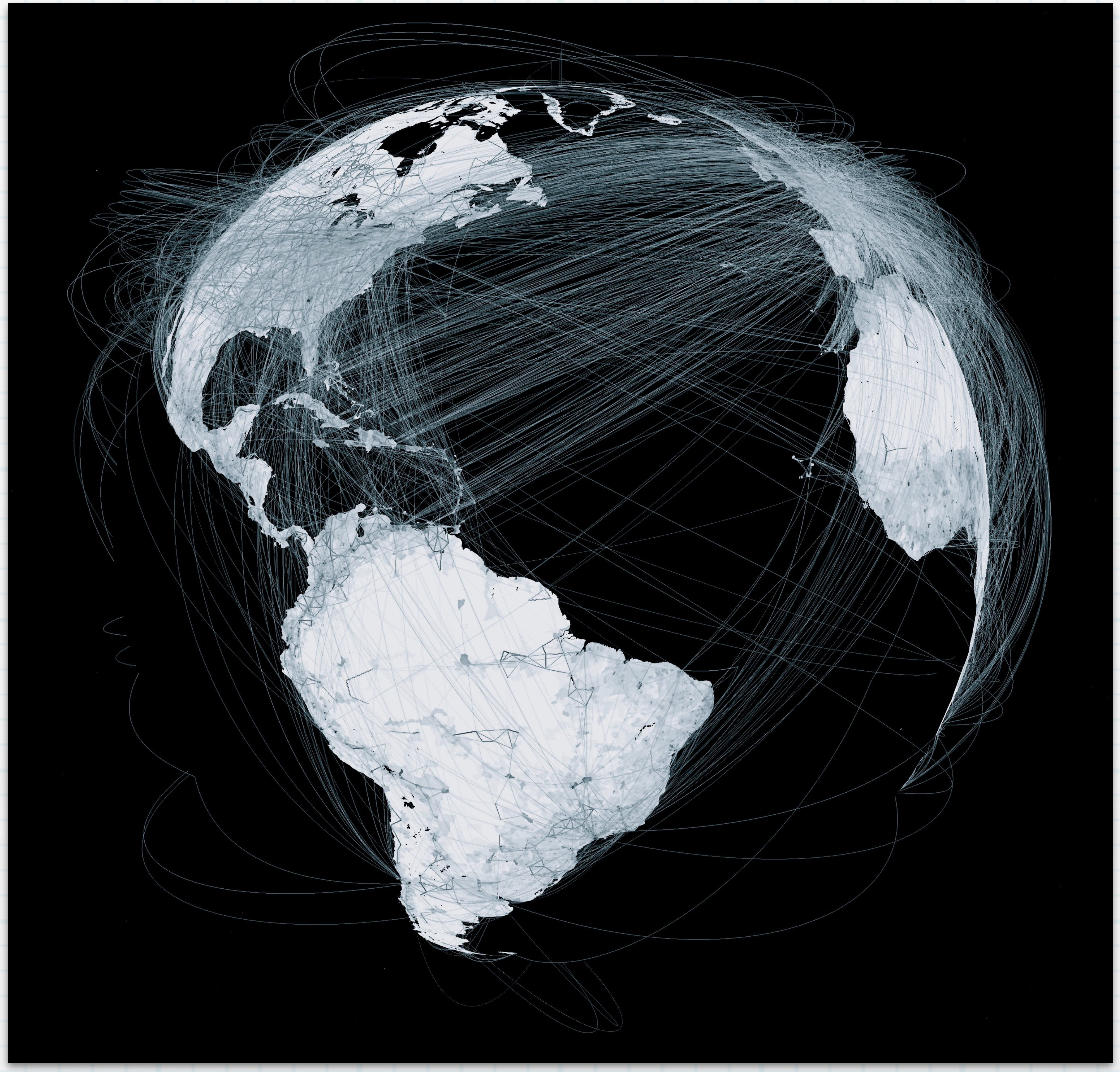
Betweenness Centrality

- Idea: flow of information through a node
- Problem: what if all we have is the structure of the network?
- **Proxy for betweenness centrality:** # of shortest paths traversing a vertex or edge if each individual sends a message to all other individuals



© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering
INDIANA UNIVERSITY BLOOMINGTON

Example: airport network



© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Betweenness Centrality

Rank	Degree	BC
1	Frankfurt	Frankfurt
2	Paris-CDG	Paris-CDG
3	Munich	Anchorage
4	Amsterdam	Tokyo-Narita
5	Atlanta	Port Moresby
6	London-Heathrow	London-Heathrow
7	Chicago (O'Hare)	Los Angeles
8	Duesseldorf	Singapore
9	Vienna	Vancouver
10	Brussels	Bangkok
11	Dallas Ft. Worth	Johannesburg
12	Houston	Toronto

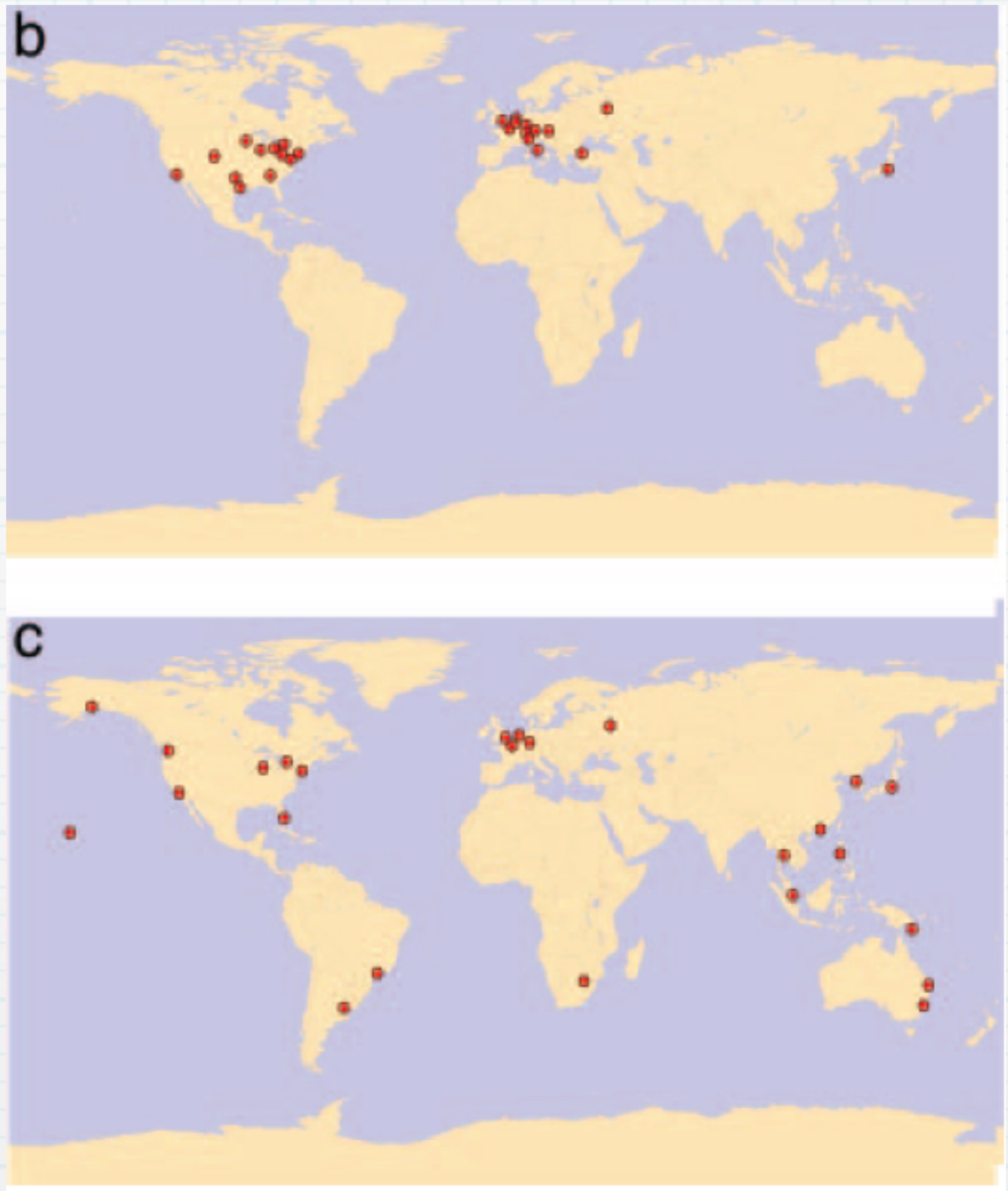
© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Betweenness Centrality

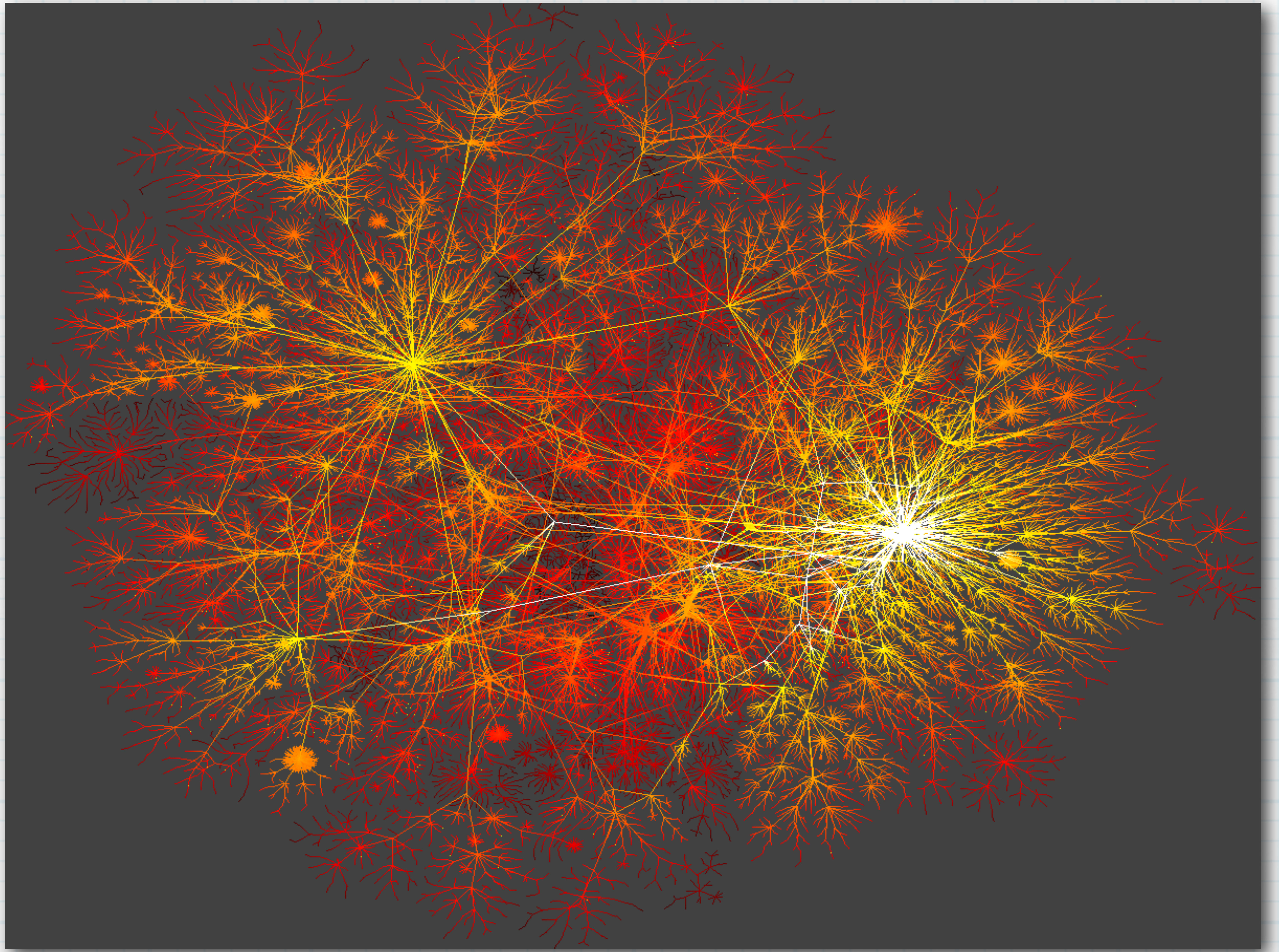
Degree ranking

Betweenness ranking



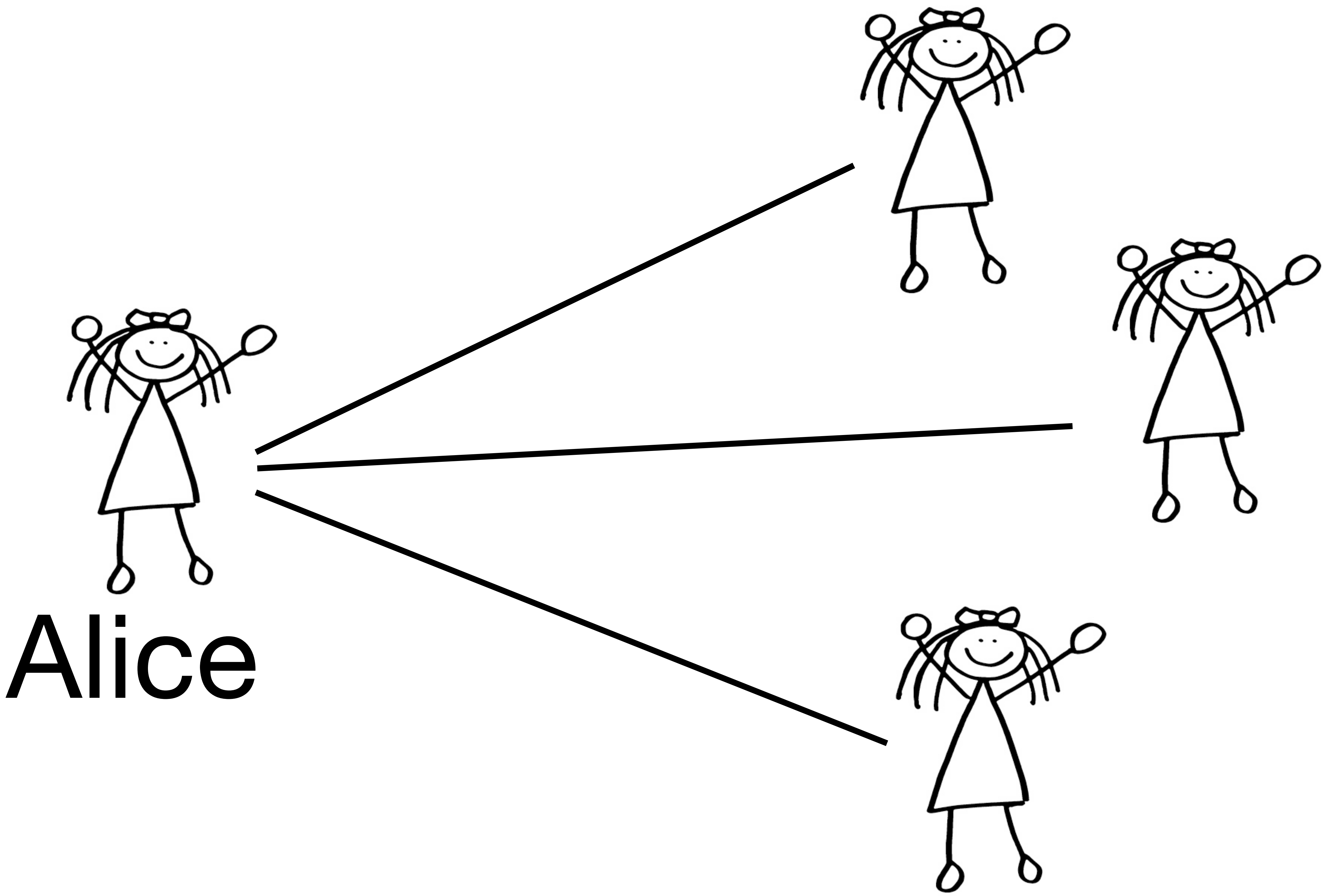
© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

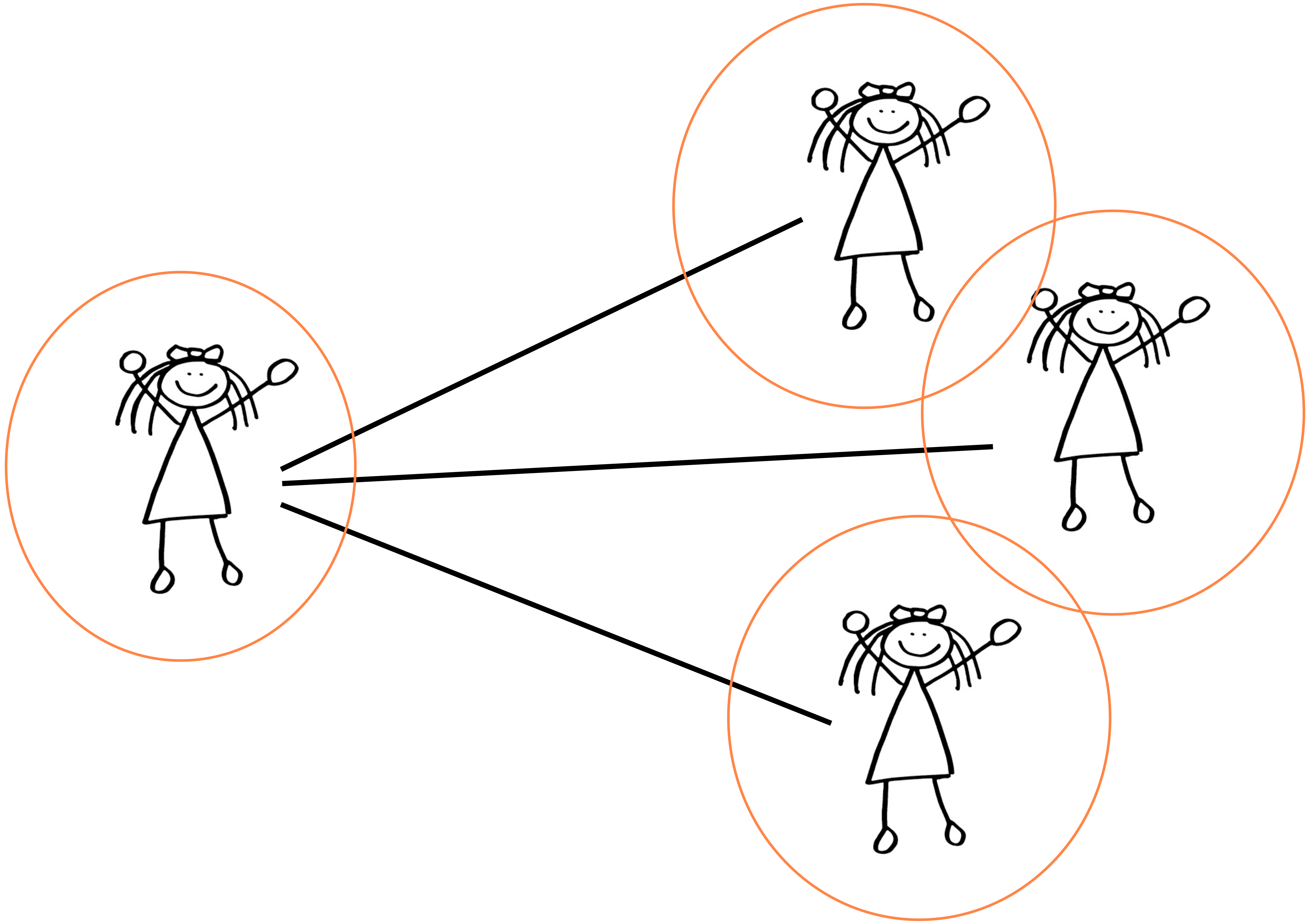
Back to degree centrality...



© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Imagine selecting someone at random, then looking at their friends





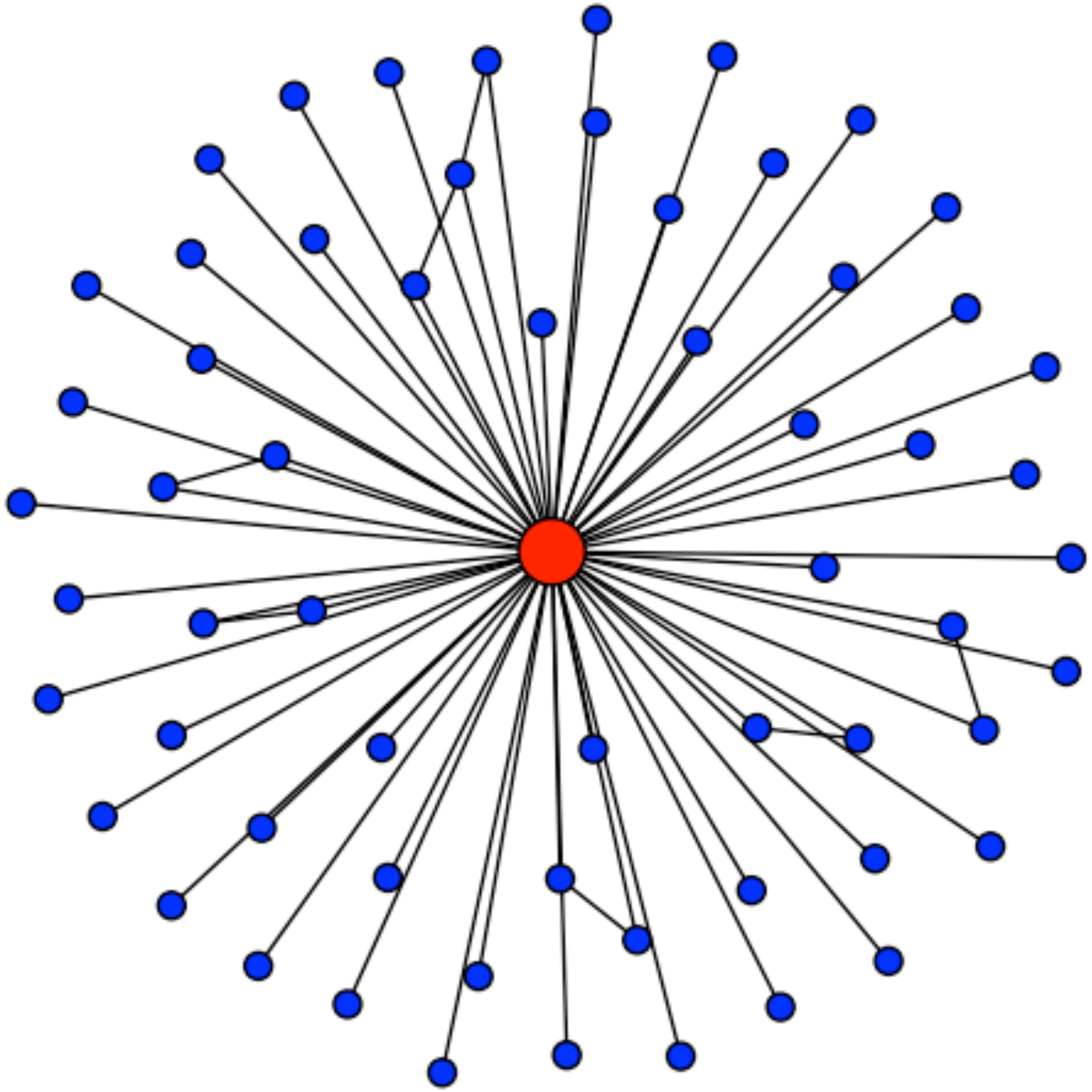
Q: Who will be more popular, Alice or her friends (on average)?

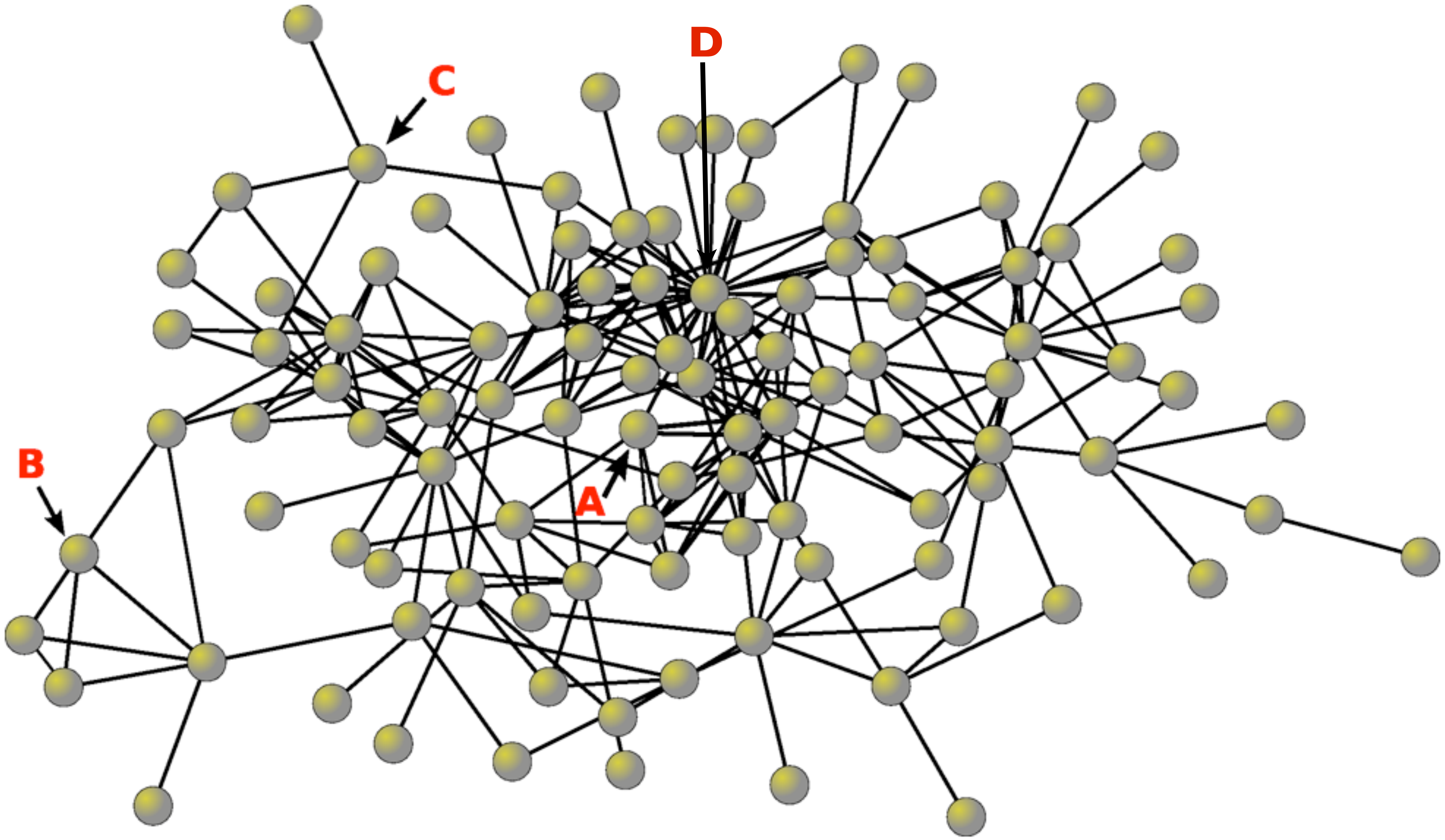
1. Alice
2. Friend
3. Same

The "Friendship Paradox"

curiosity.com/paths/how-popular-are-you-number-hub-ep-28-head-squeeze-head-squeeze/

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering





“Hubs”

The plan

* We discussed two properties of social networks: short paths and high clustering
  * We saw how random graphs have short paths, but not high clustering
  * We studied the small-world model that generates networks with both properties
* Next let's study another important property shared by social networks and many other real-world networks: the "scale-free" property

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

- Review
  - Centrality
  - Hubs
- Quiz
  - Scale-free property

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

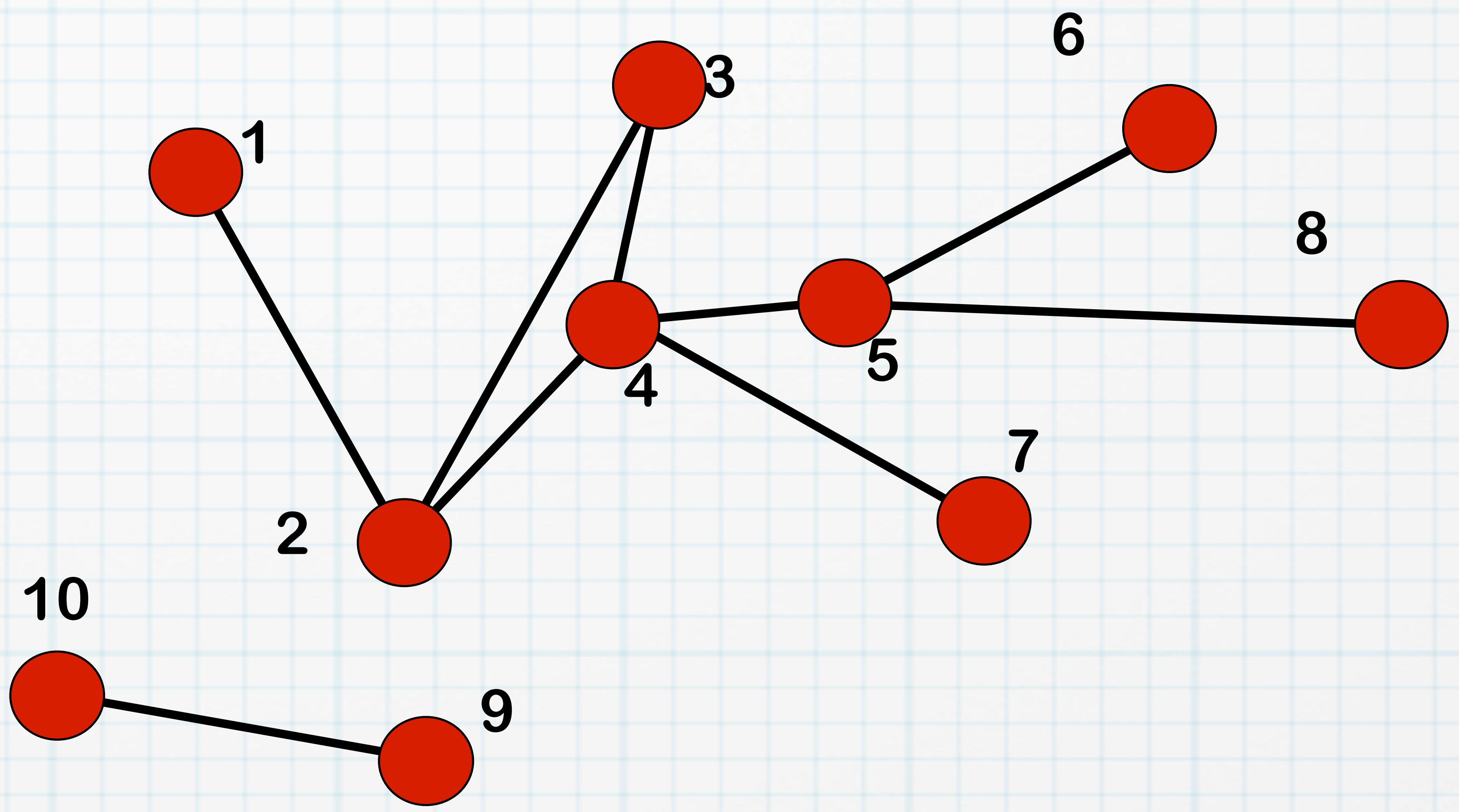
A bit of statistics review

* We have a sample of values $x_1, ..., x_N$
  * Average (a.k.a. mean): typical value
    $\langle x \rangle = (x_1 + x_1 + ... + x_N)/N = \Sigma_i x_i /N$
  * Standard deviation: fluctuations around typical value
    $\sigma_x = \sqrt{\Sigma_i (x_i - \langle x \rangle)^2/N}$

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Back to networks

* We can apply these statistical properties to describe an entire network
* Average degree
  * $\langle k \rangle = \sum_i k_i / N = \sum_i \sum_j A_{ij} / N$
    $= \sum_{ij} A_{ij} / N = 2E/N$



© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

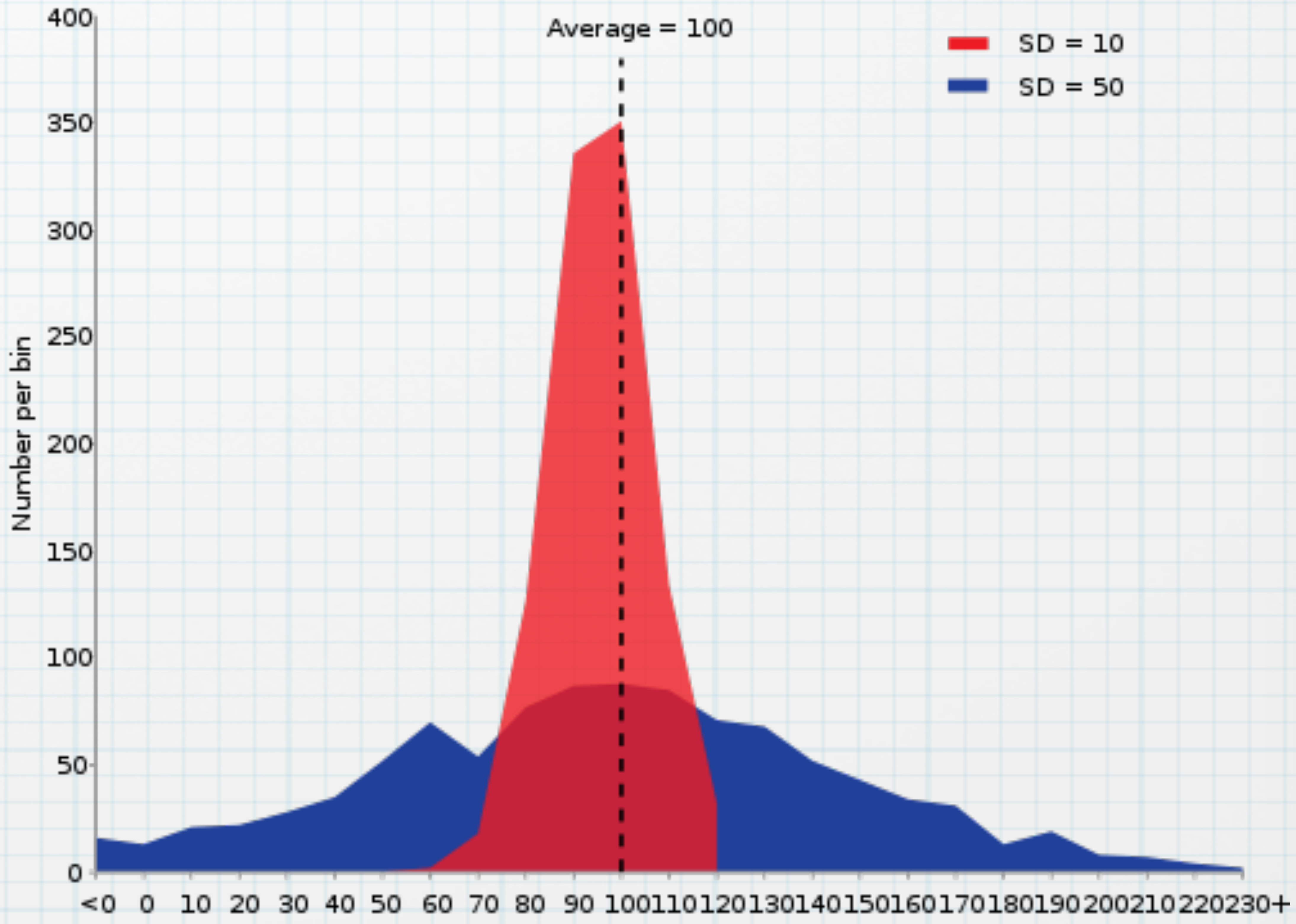
A bit of statistics review

histograms

* We have a sample of values $x_1, \ldots, x_N$
  * Distribution of x: probability that a randomly chosen value is x

    $P(x) = (\#\text{ values } x) / N$

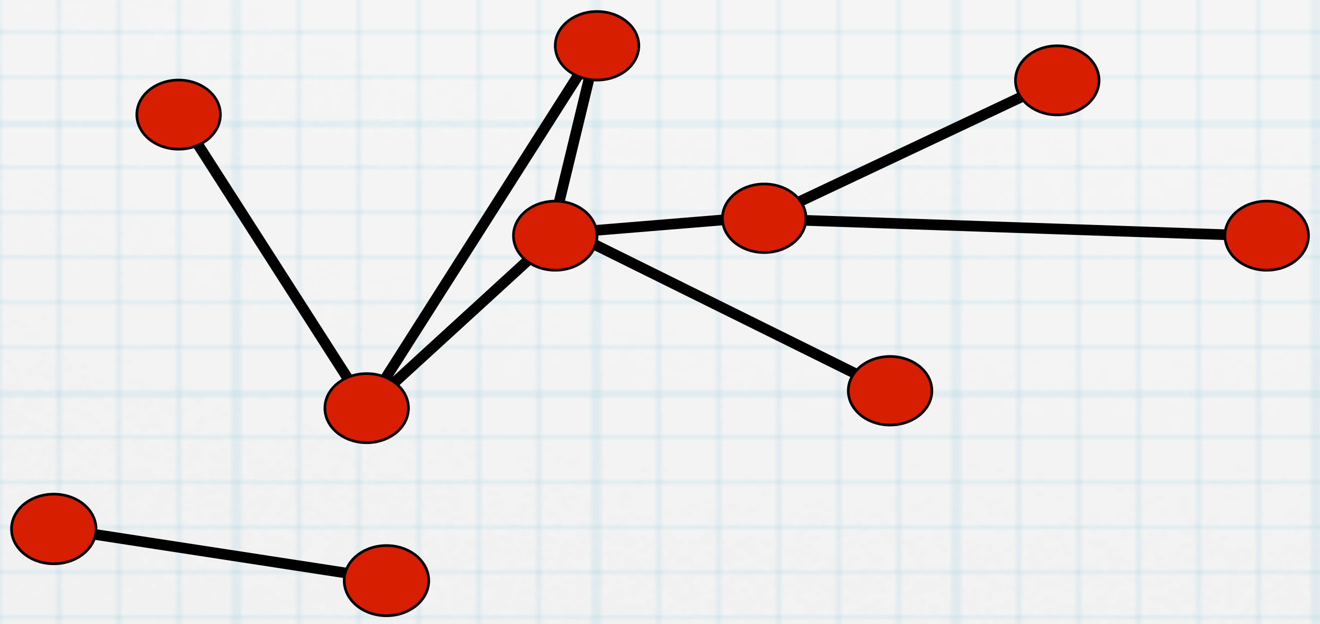
    $\sum_i P(x_i) = 1$ always!
* If continuous rather than discrete values: binning, PDF

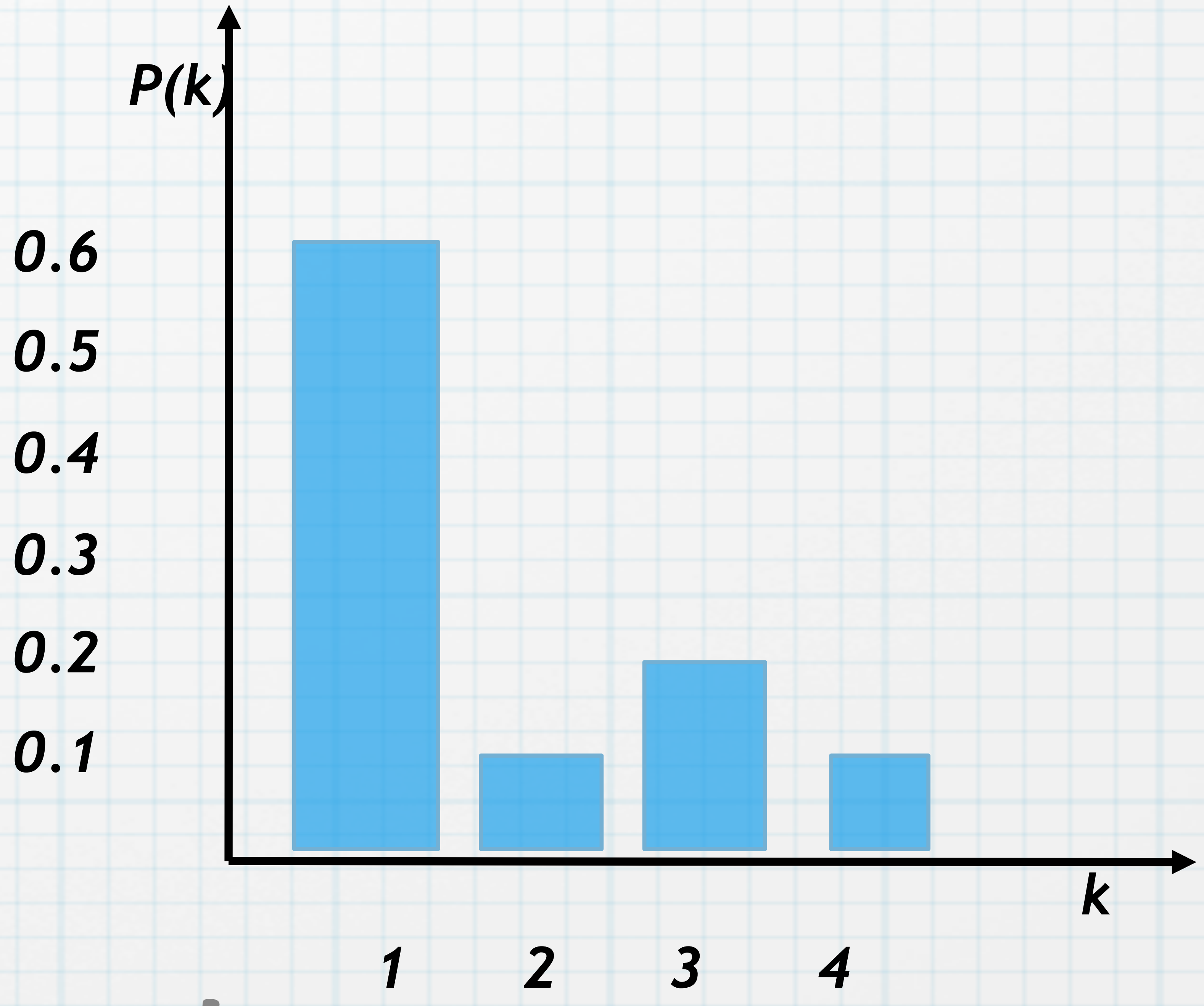


© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Networks stats

* Degree distribution P(k): probability that a randomly chosen vertex has degree k
    * N_k = # nodes with degree k
    * $P(k) = N_k / N$ → plot





* This will be an important way to describe networks

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

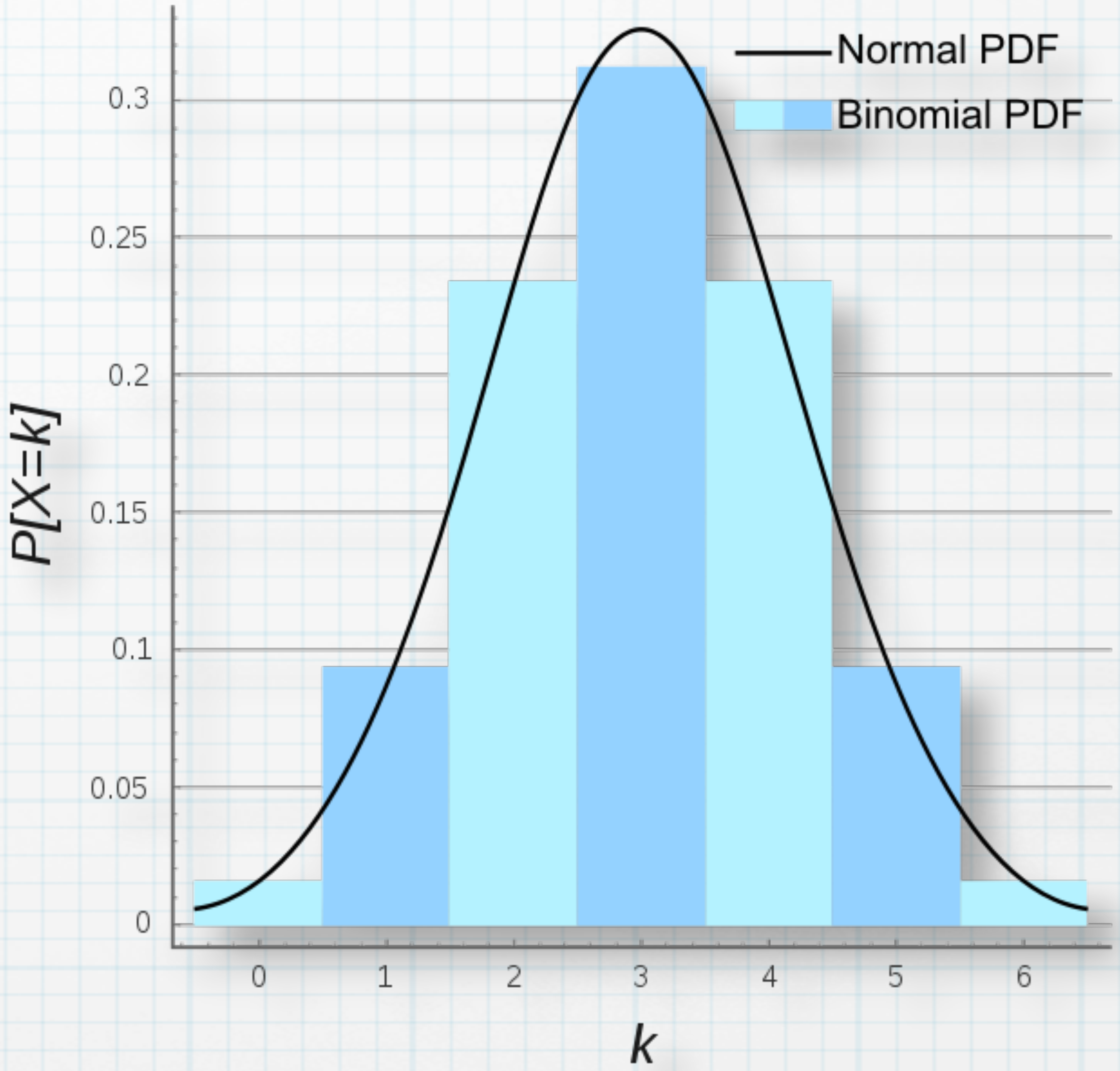
Degree distribution in random graphs

Degree distribution: Binomial

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

In the limit of large N and constant <k> ≈ pN, this is well approximated by the Poisson distribution:

$$P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$



If <k> is not too small, this is in turn approximated by the Normal (Gaussian) distribution with mean <k> and variance <k>: the average <k> is a "typical value" and most values are close to it, making it a good descriptor.

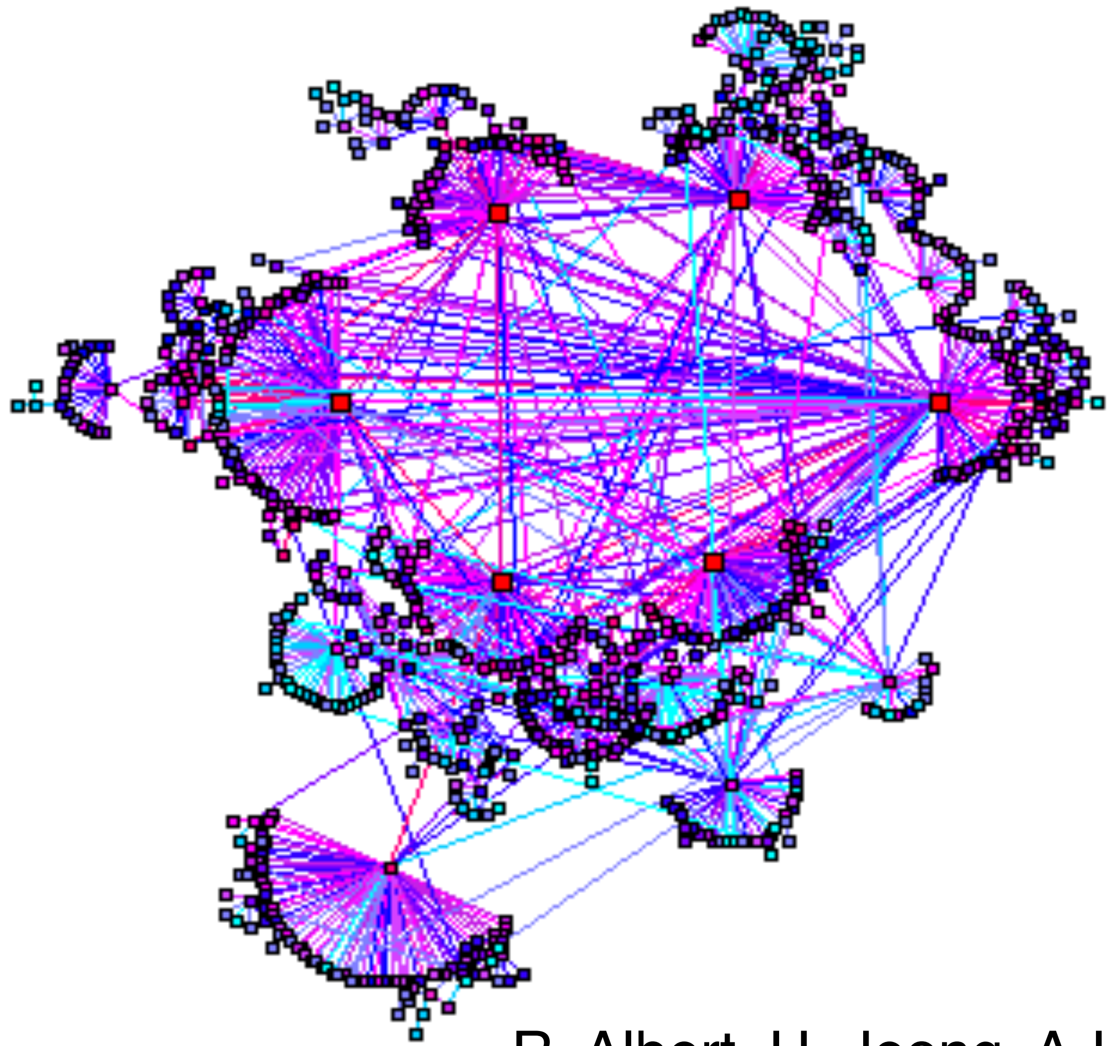
© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

WORLD WIDE WEB

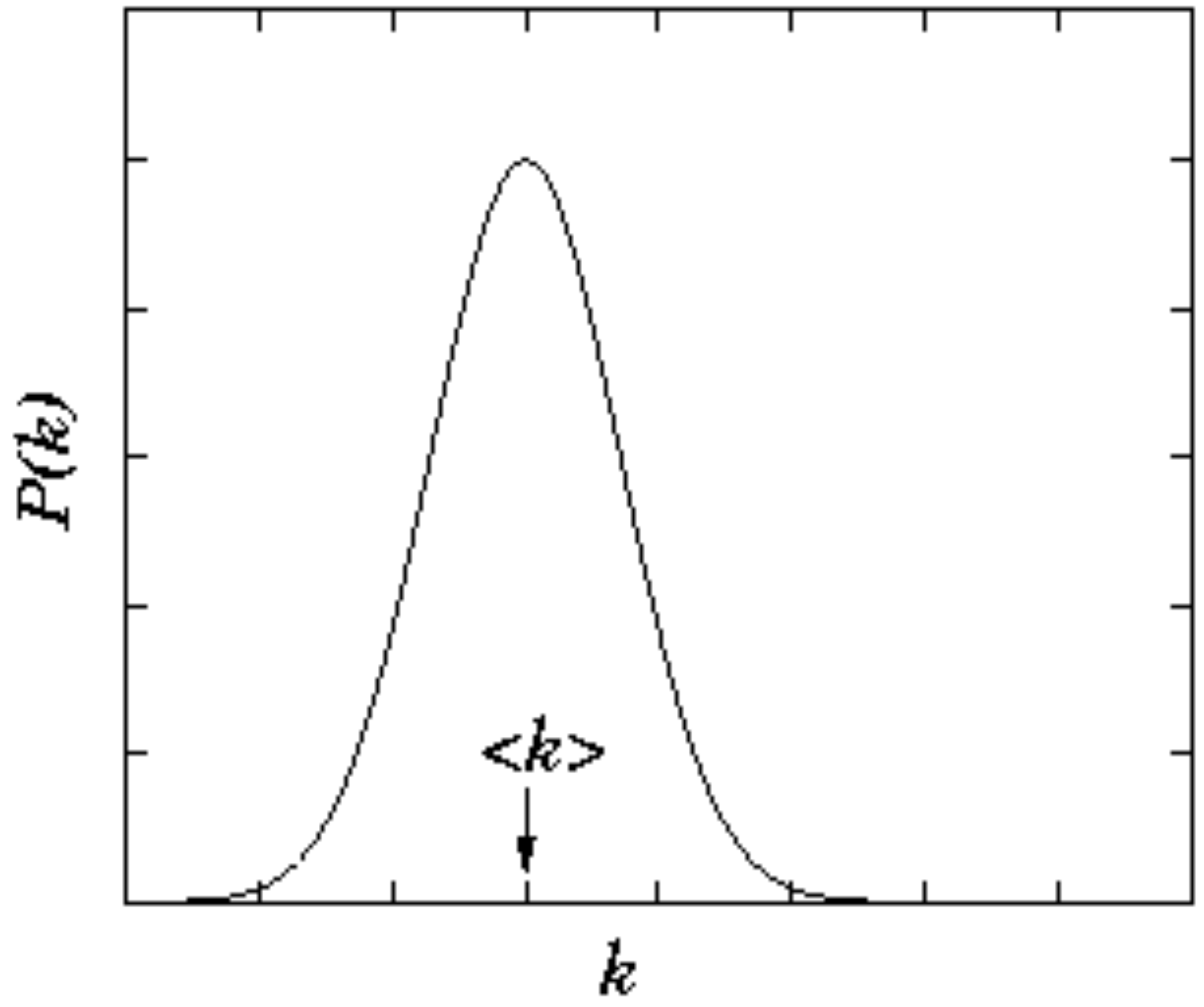
Nodes: **WWW page**
Links: **URL links**

Over 3 billion documents

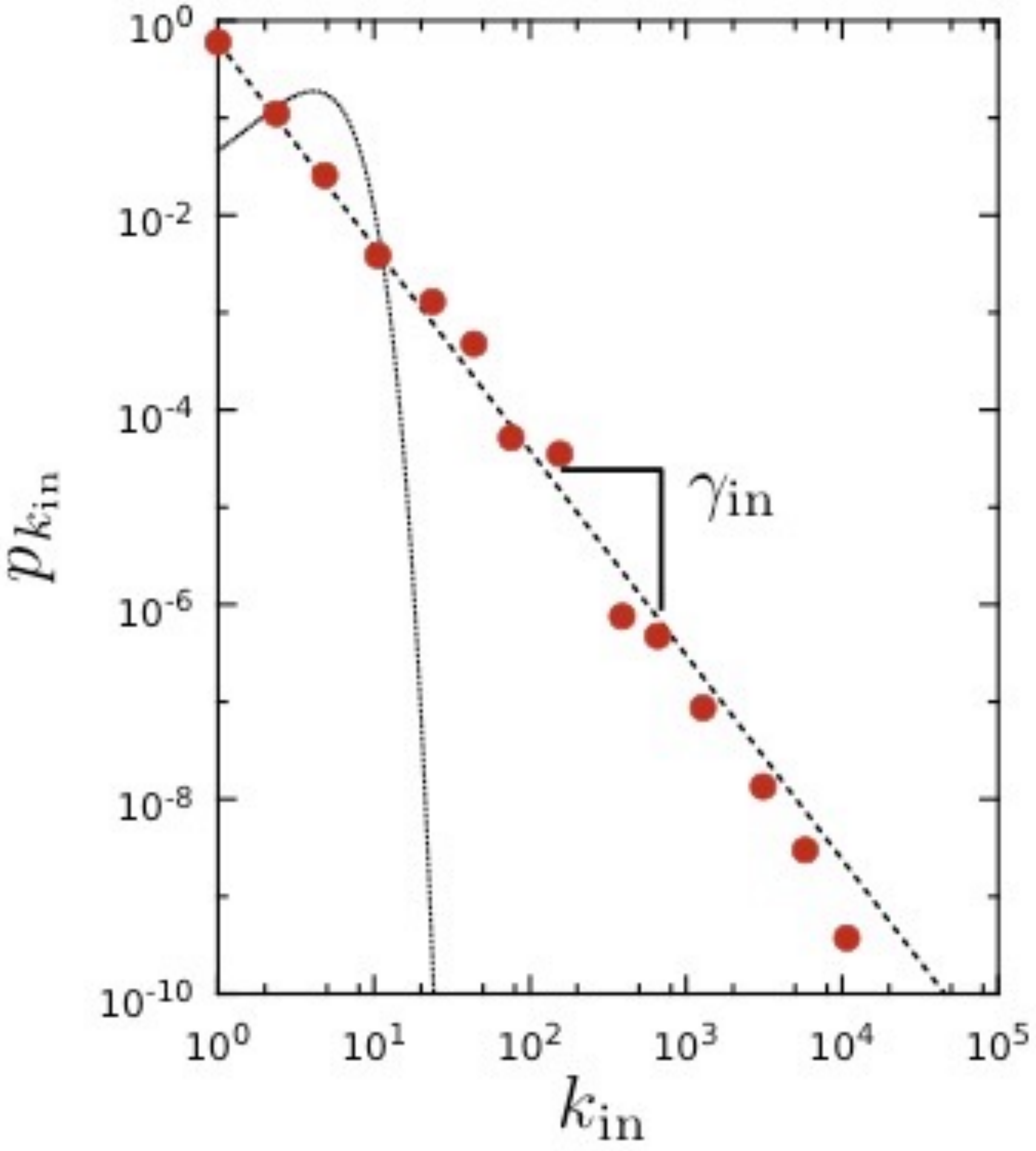
ROBOT: collects all URLs found in a page and follows them recursively



Expected

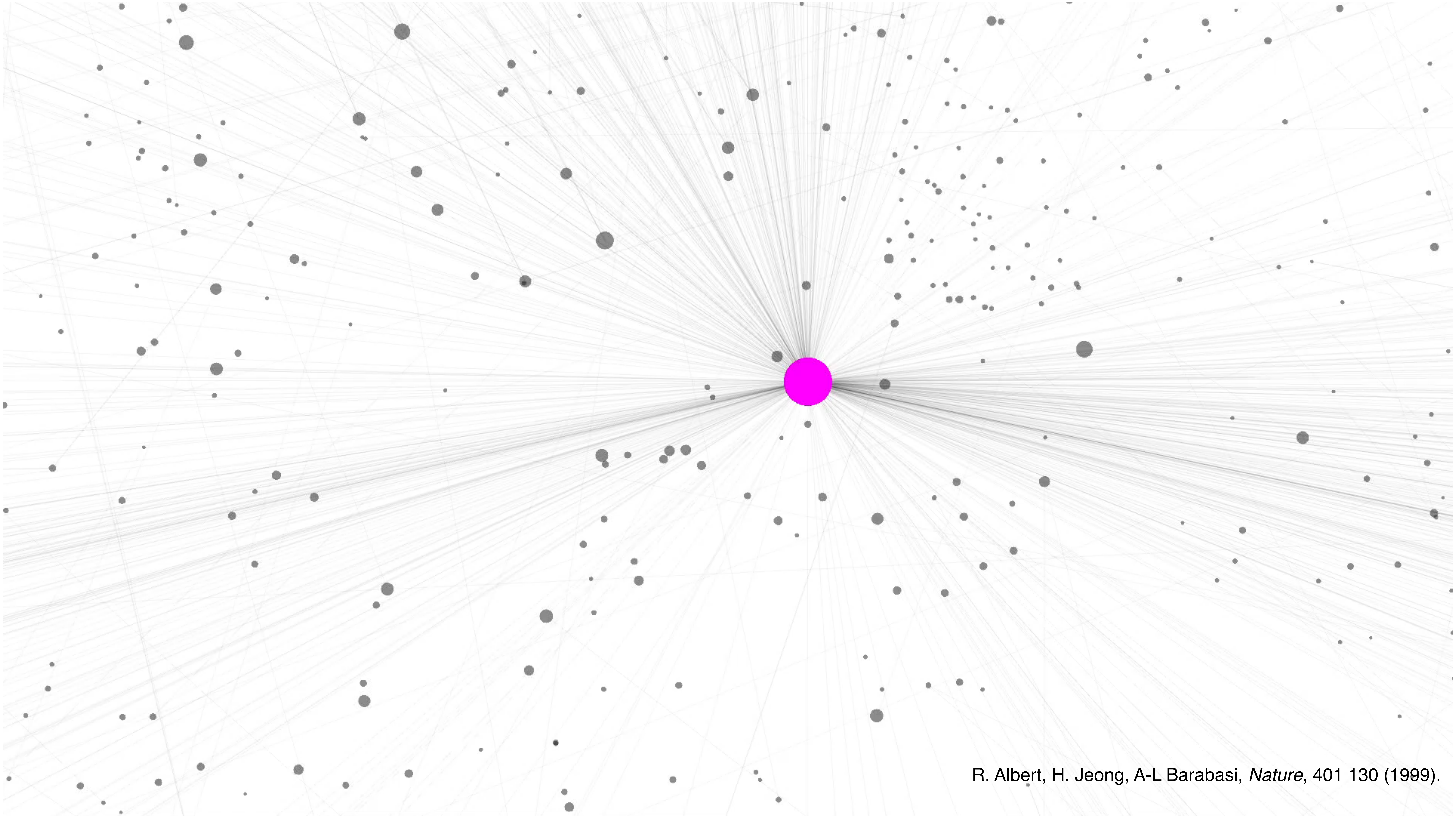


Actual

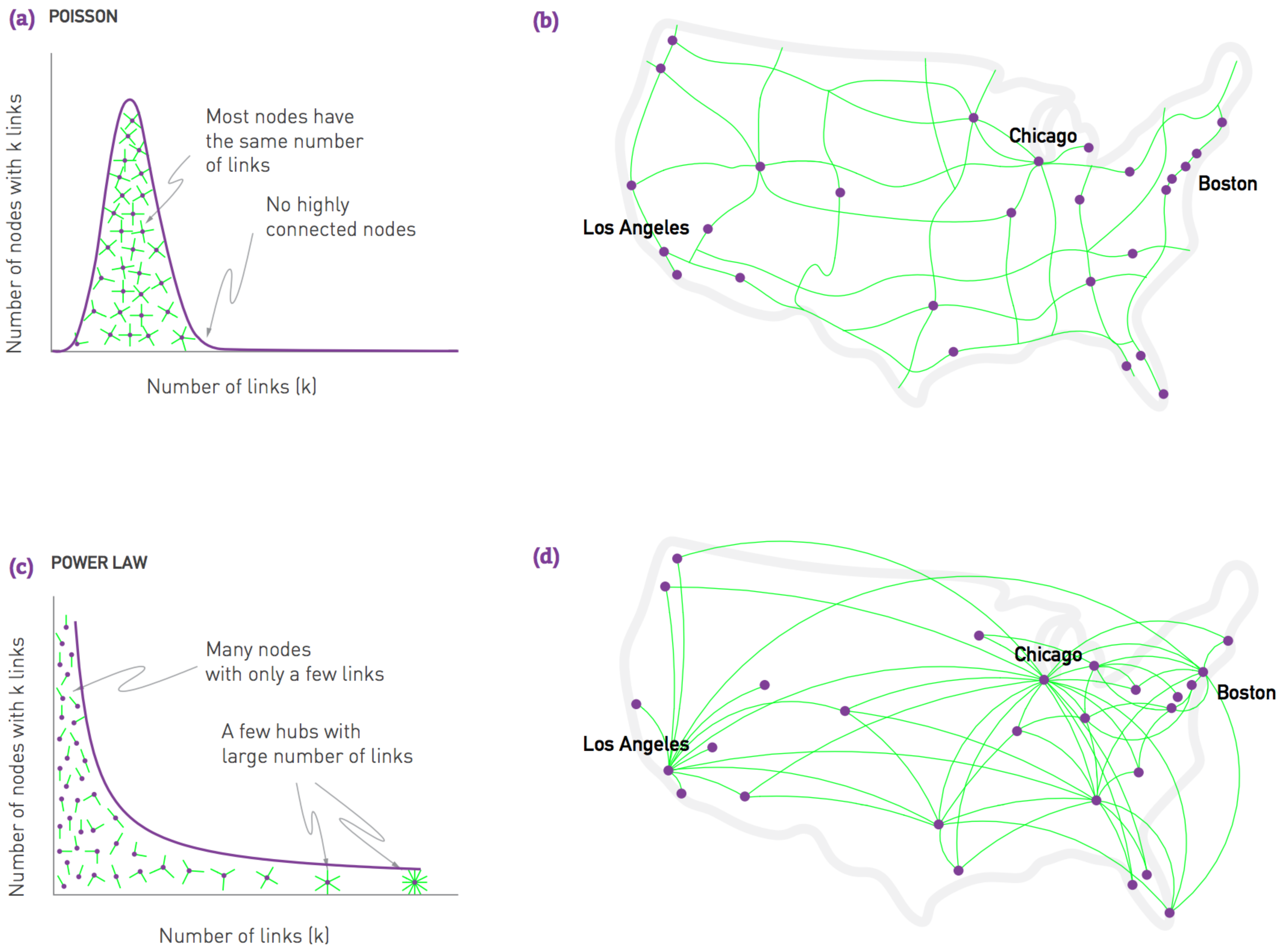


R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).

WORLD WIDE WEB

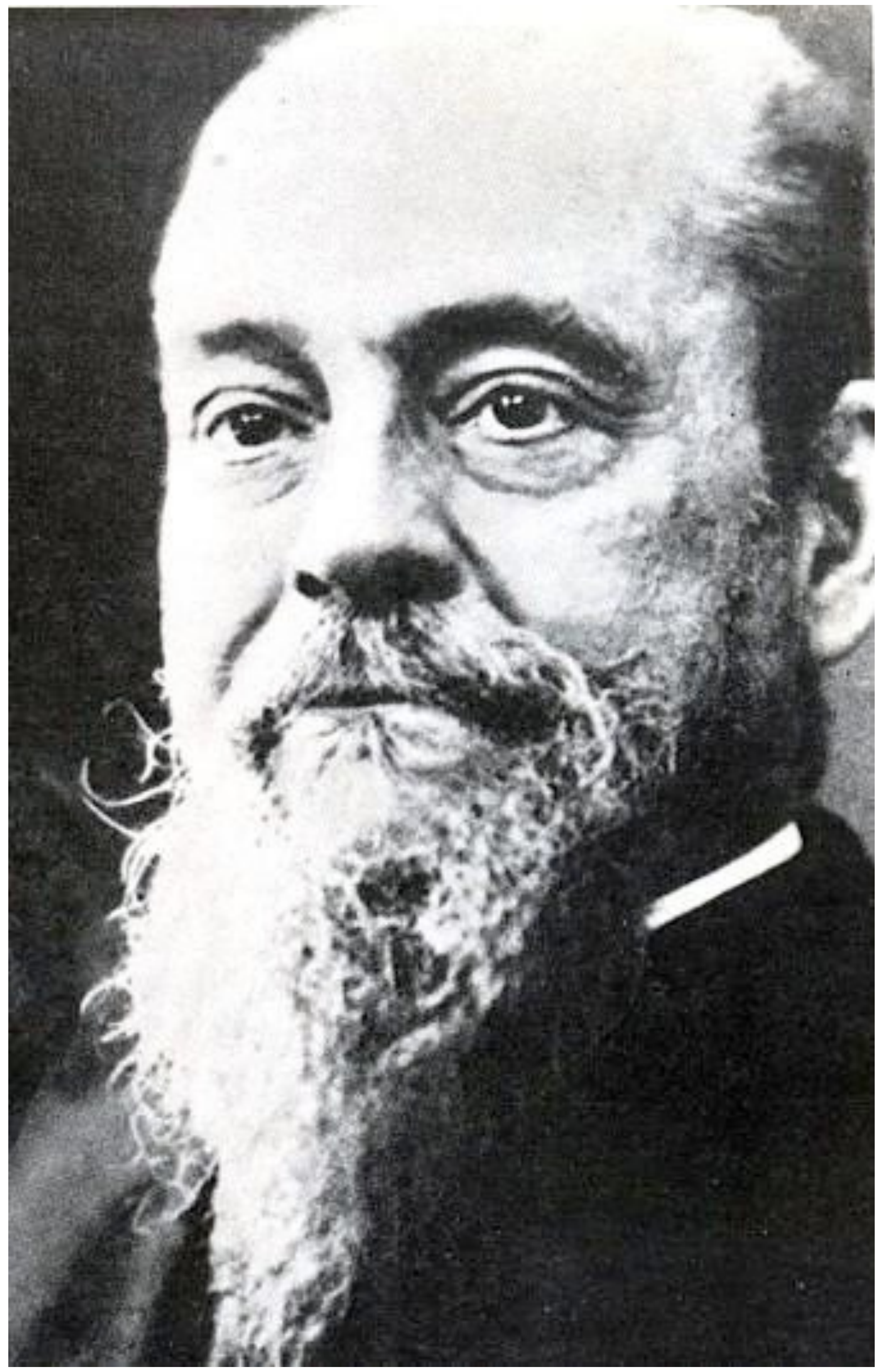


R. Albert, H. Jeong, A-L Barabasi, *Nature*, 401 130 (1999).

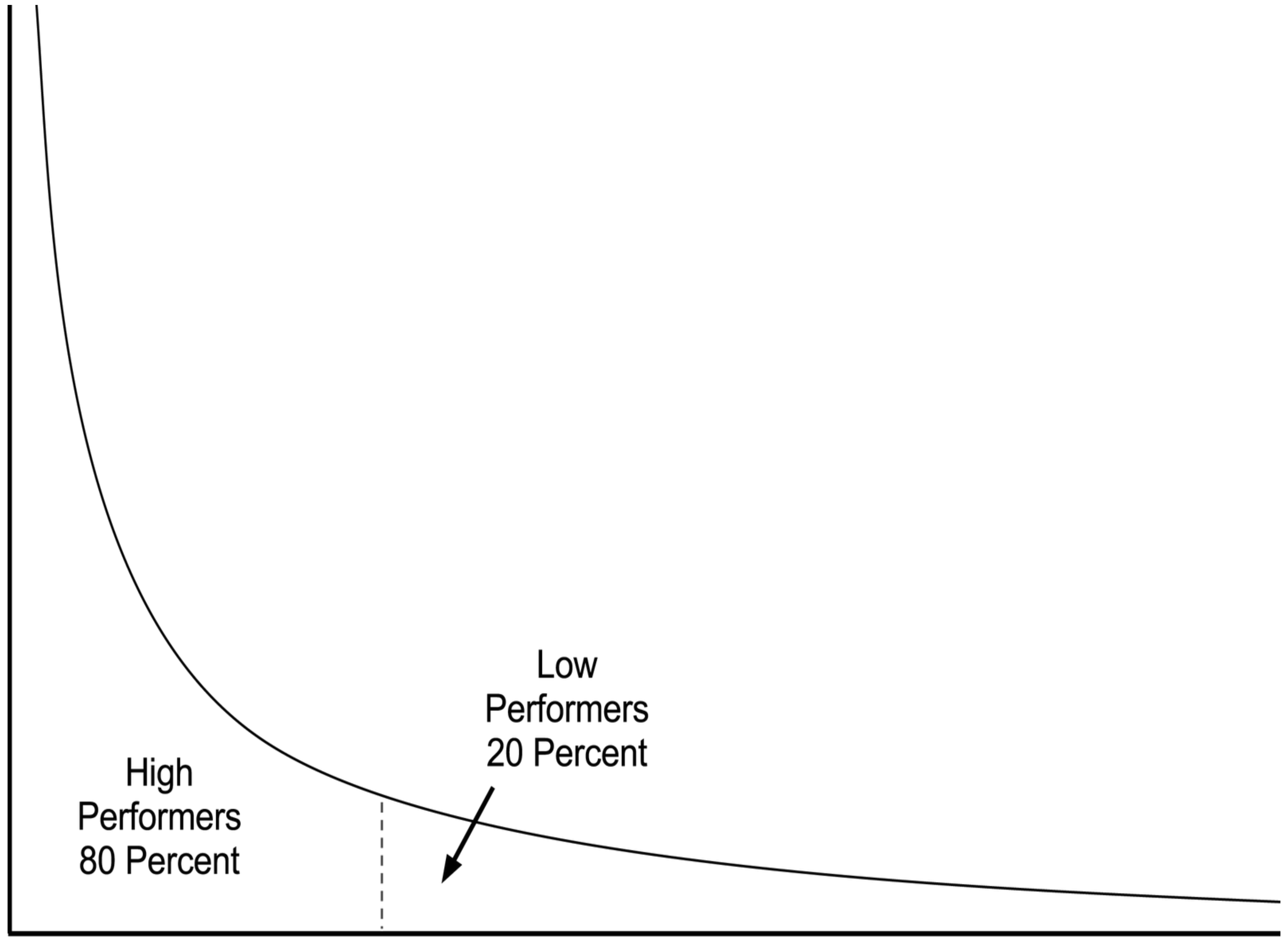


(a) POISSON
Number of nodes with k links
Most nodes have the same number of links
No highly connected nodes
Number of links (k)
(b)
Chicago
Boston
Los Angeles
(c) POWER LAW
Number of nodes with k links
Many nodes with only a few links
A few hubs with large number of links
Number of links (k)
(d)
Chicago
Boston
Los Angeles

80/20 RULE

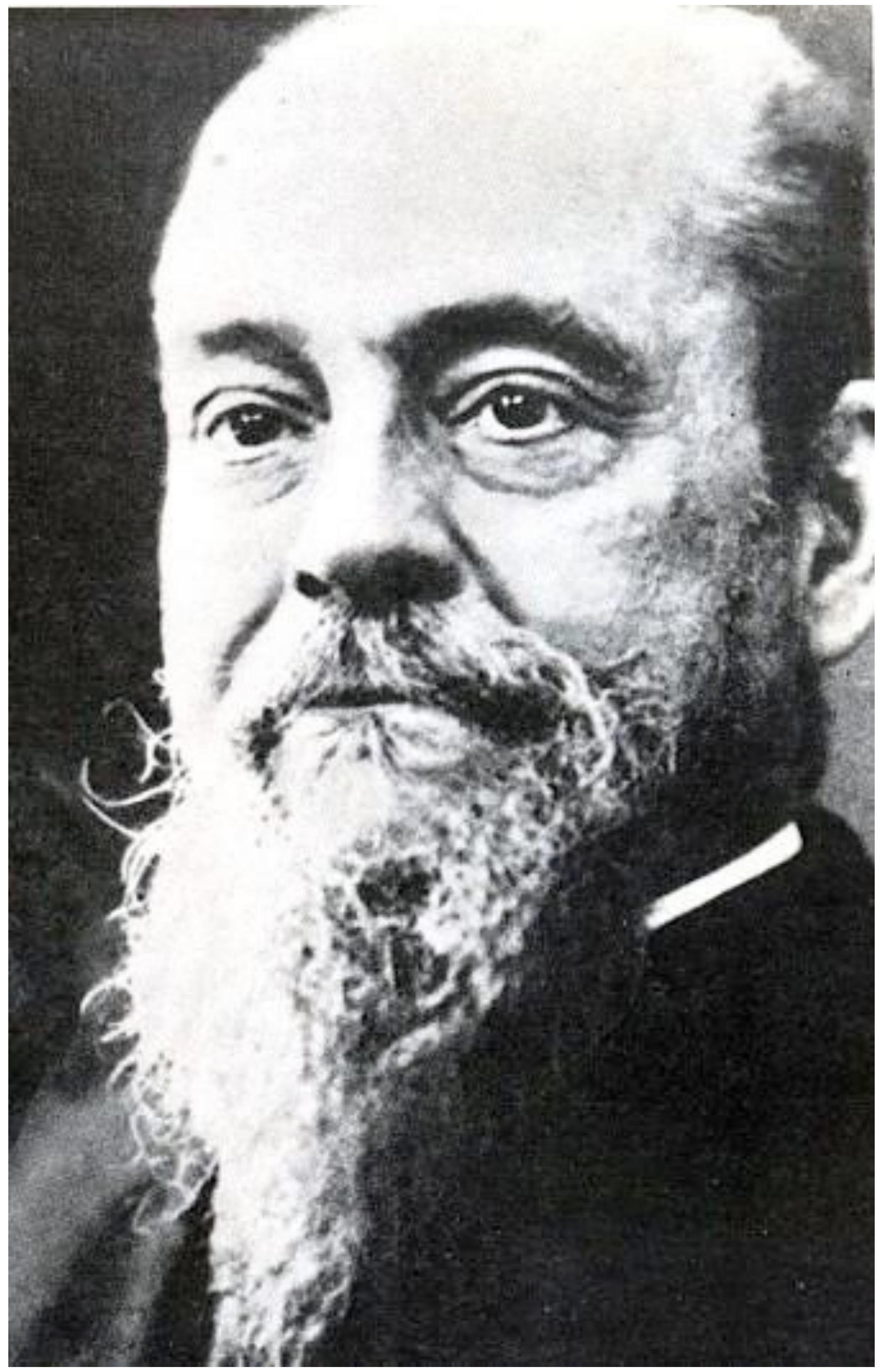


Inequality, heterogeneity, and the 1%



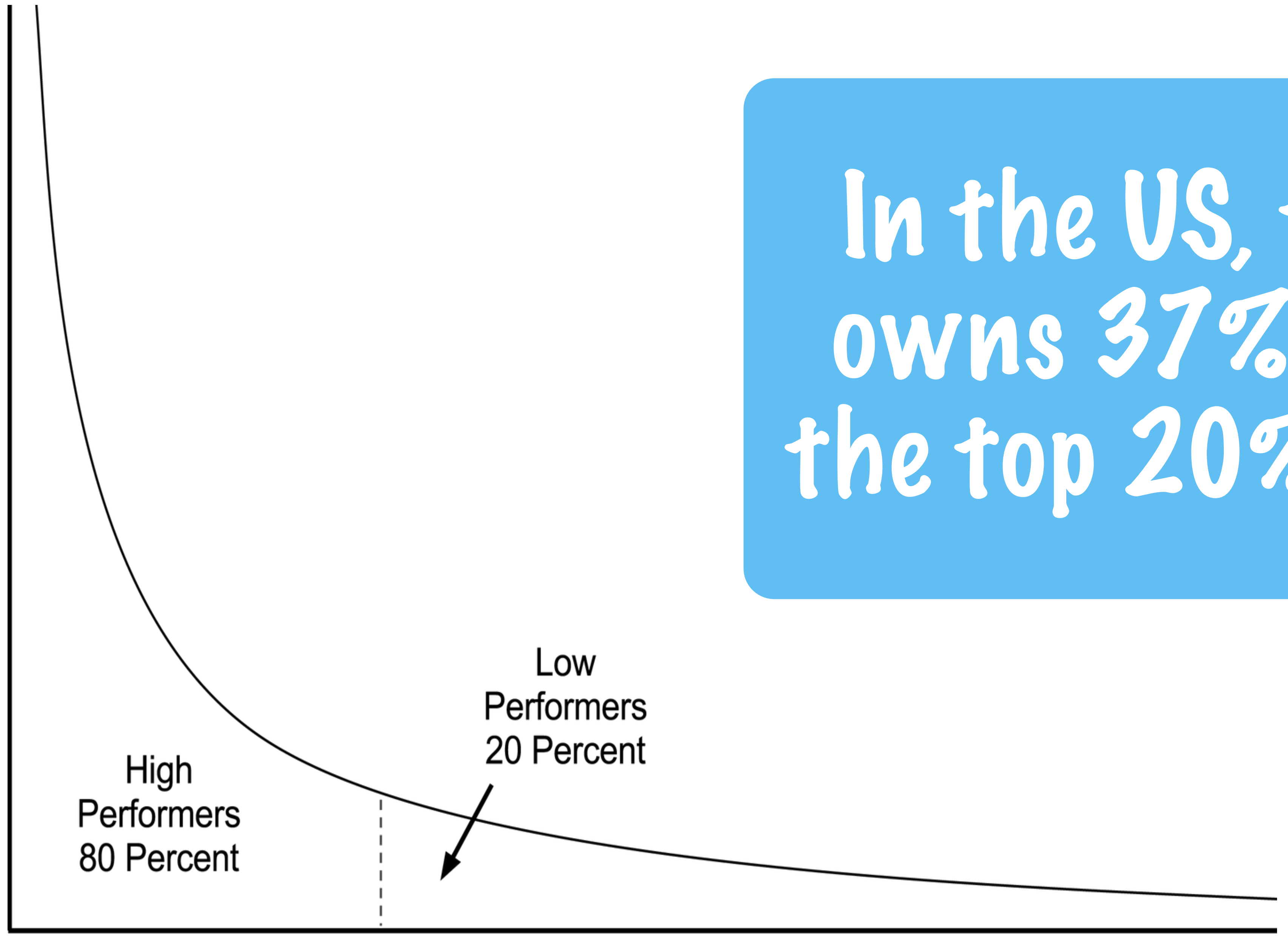
Vilfredo Federico Damaso Pareto (1848 – 1923), Italian economist, political scientist and philosopher, who had important contributions to our understanding of income distribution and to the analysis of individuals choices. A number of fundamental principles are named after him, like Pareto efficiency, Pareto distribution (another name for a power-law distribution), the Pareto principle (or 80/20 law).

80/20 RULE



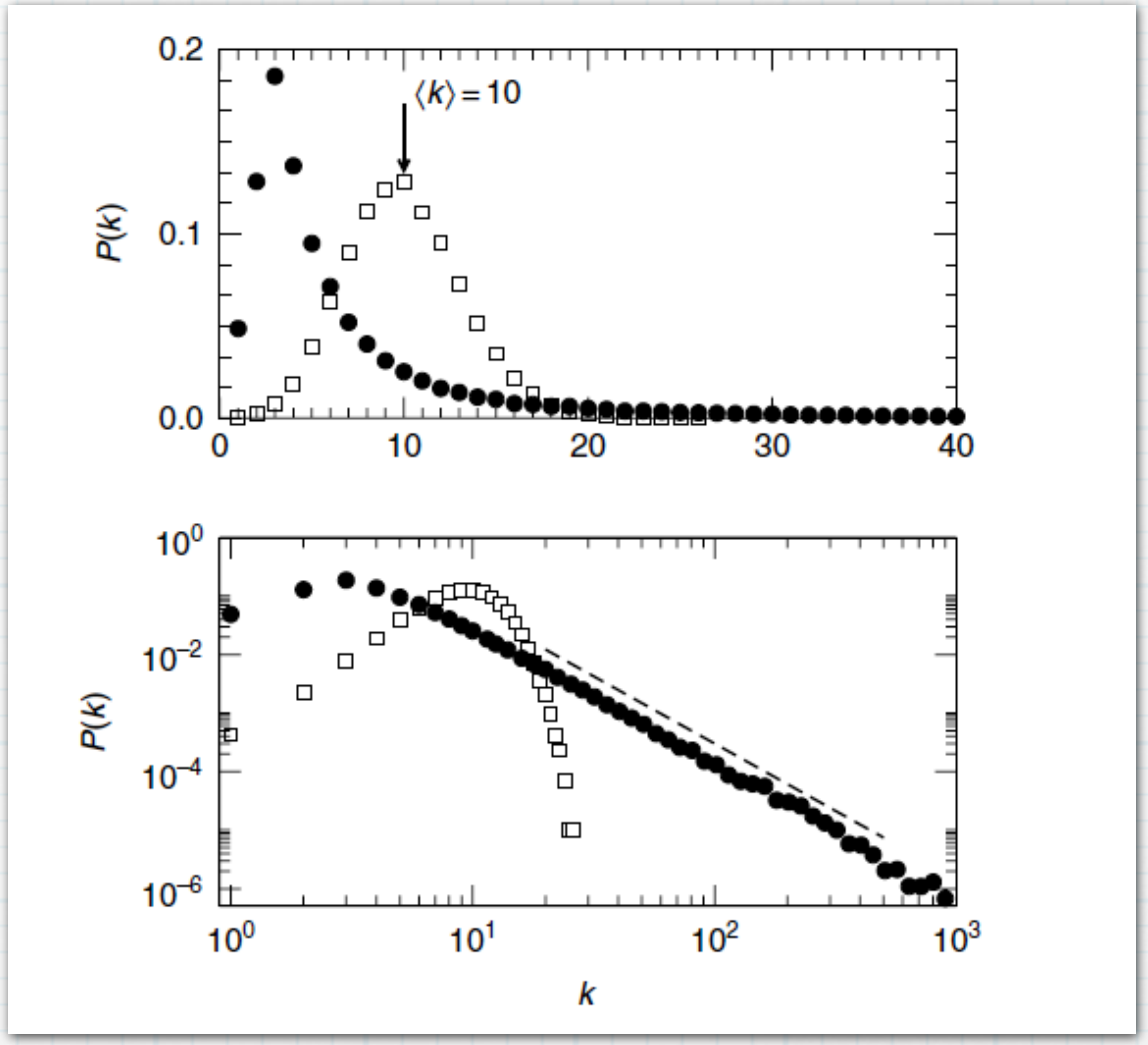
Inequality, heterogeneity, and the 1%

In the US, the top 1% owns 37% of wealth, the top 20% owns 88%



Vilfredo Federico Damaso Pareto (1848 – 1923), Italian economist, political scientist and philosopher, who had important contributions to our understanding of income distribution and to the analysis of individuals choices. A number of fundamental principles are named after him, like Pareto efficiency, Pareto distribution (another name for a power-law distribution), the Pareto principle (or 80/20 law).

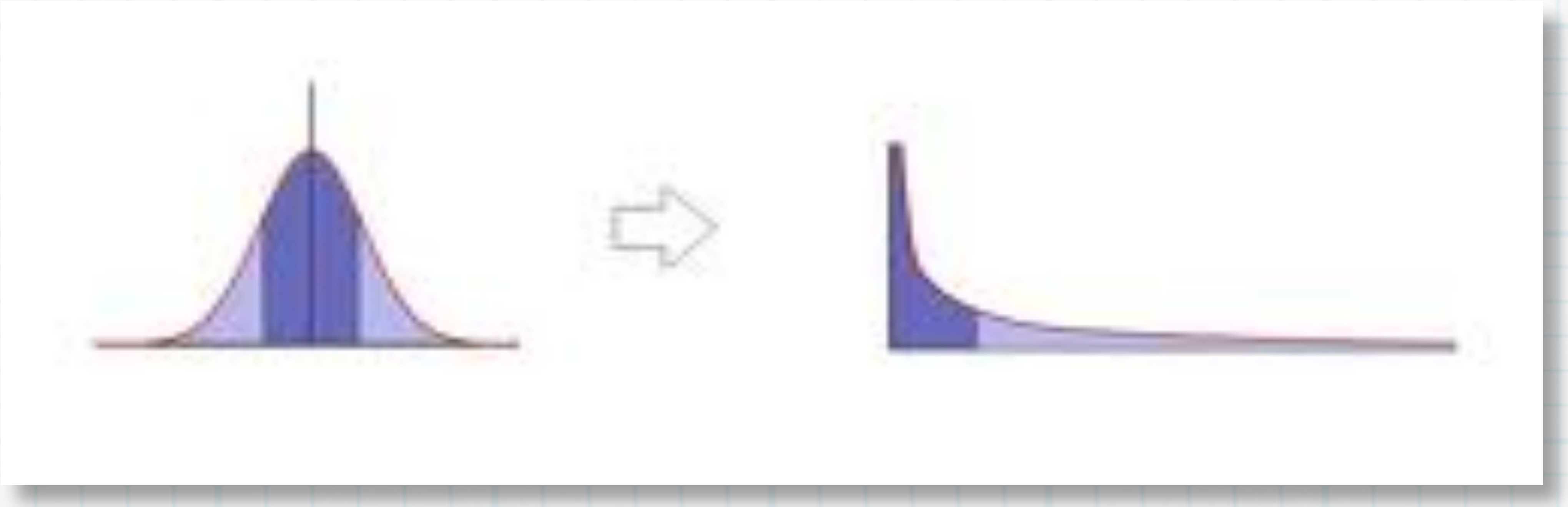
Heavy tails (a.k.a. long tails)



- For instance: POWER-LAW behavior

$$P(k) \sim k^{-\gamma}$$

Summary: noise and tails



Normal -vs- heavy-tail

* Heavy-tail distributions : skewed, appreciable probability for large events (hubs!)
* Standard deviation tends to be large, average is NOT a good descriptor (no "scale")

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Scale-free networks: Definition

Definition:

Networks with a power law tail in their degree distribution are called ‘scale-free networks’

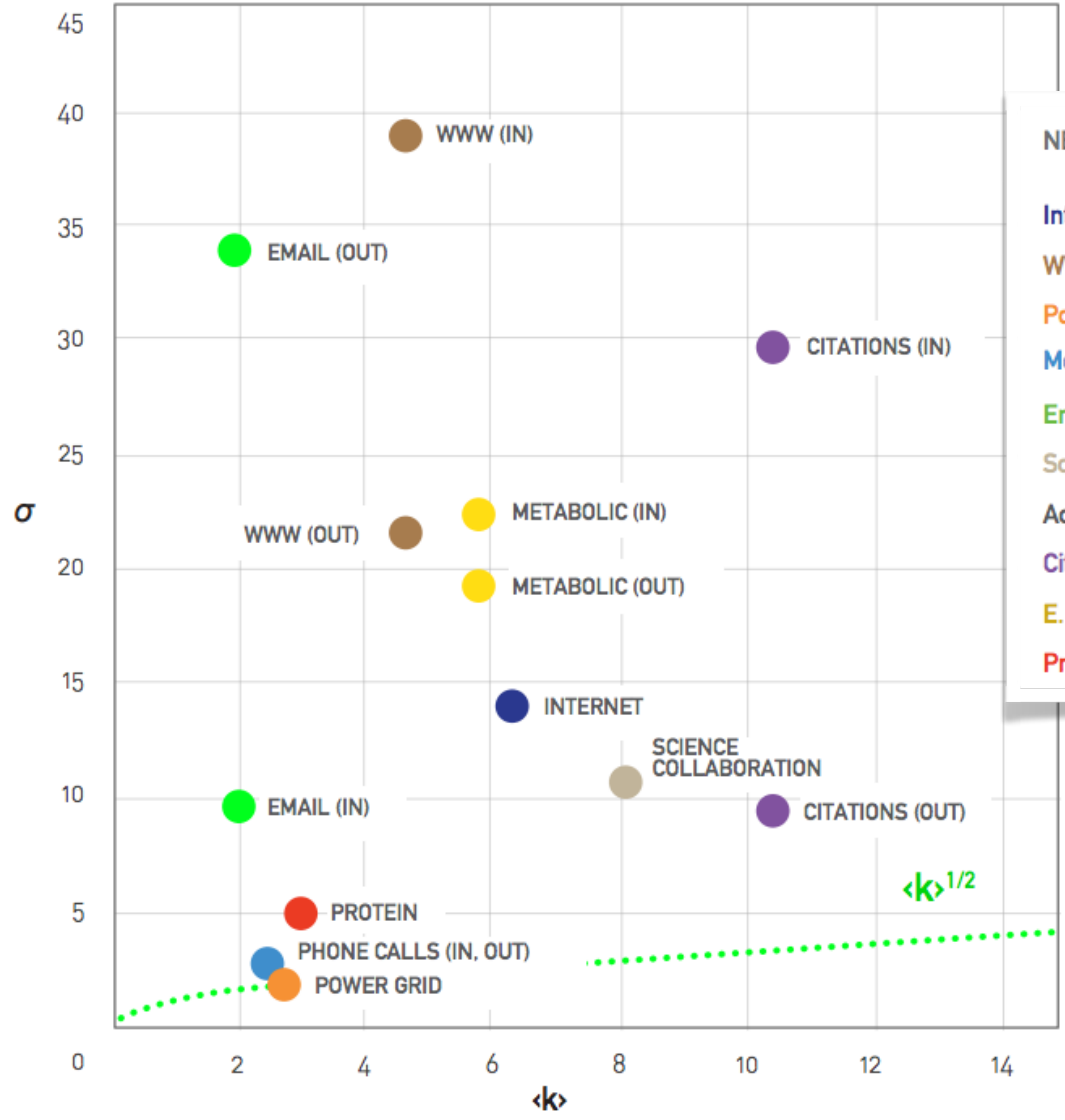
Where does the name come from?

Critical Phenomena and scale-invariance

- **Scale invariance**: there is no characteristic scale for the fluctuation **(scale-free behavior).**
- **Universality**: exponents are independent of the system’s details.

Slides after **Dante R. Chialvo**

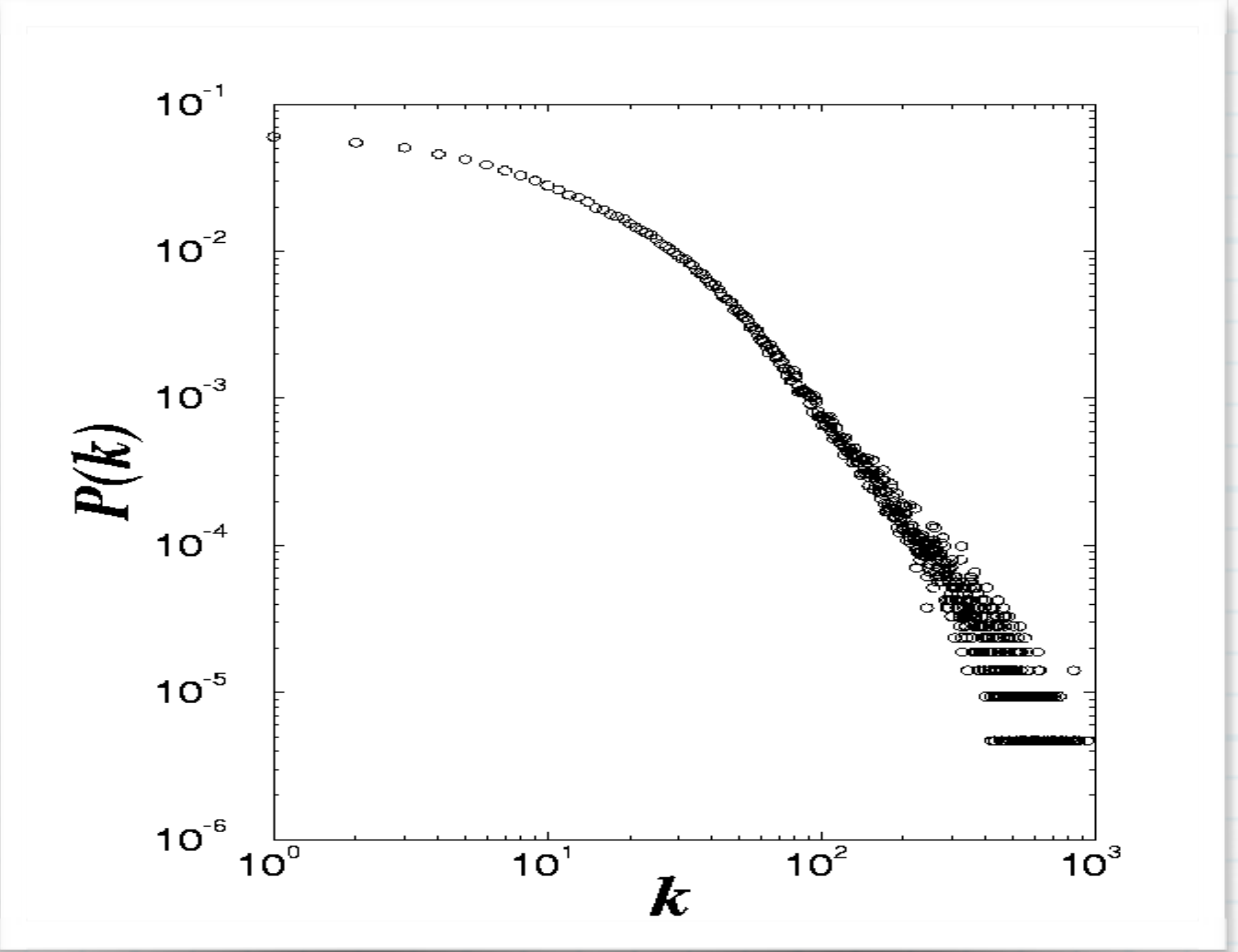
THE MEANING OF SCALE-FREE



NETWORK	N	L	$\langle k \rangle$	$\langle k_{in}^2 \rangle$	$\langle k_{out}^2 \rangle$	$\langle k^2 \rangle$
Internet	192,244	609,066	6.34	-	-	240.1
WWW	325,729	1,497,134	4.60	1546.0	482.4	-
Power Grid	4,941	6,594	2.67	-	-	10.3
Mobile Phone Calls	36,595	91,826	2.51	12.0	11.7	-
Email	57,194	103,731	1.81	94.7	1163.9	-
Science Collaboration	23,133	93,439	8.08	-	-	178.2
Actor Network	702,388	29,397,908	83.71	-	-	47,353.7
Citation Network	449,673	4,689,479	10.43	971.5	198.8	-
E. Coli Metabolism	1,039	5,802	5.58	535.7	396.7	-
Protein Interactions	2,018	2,930	2.90	-	-	32.3

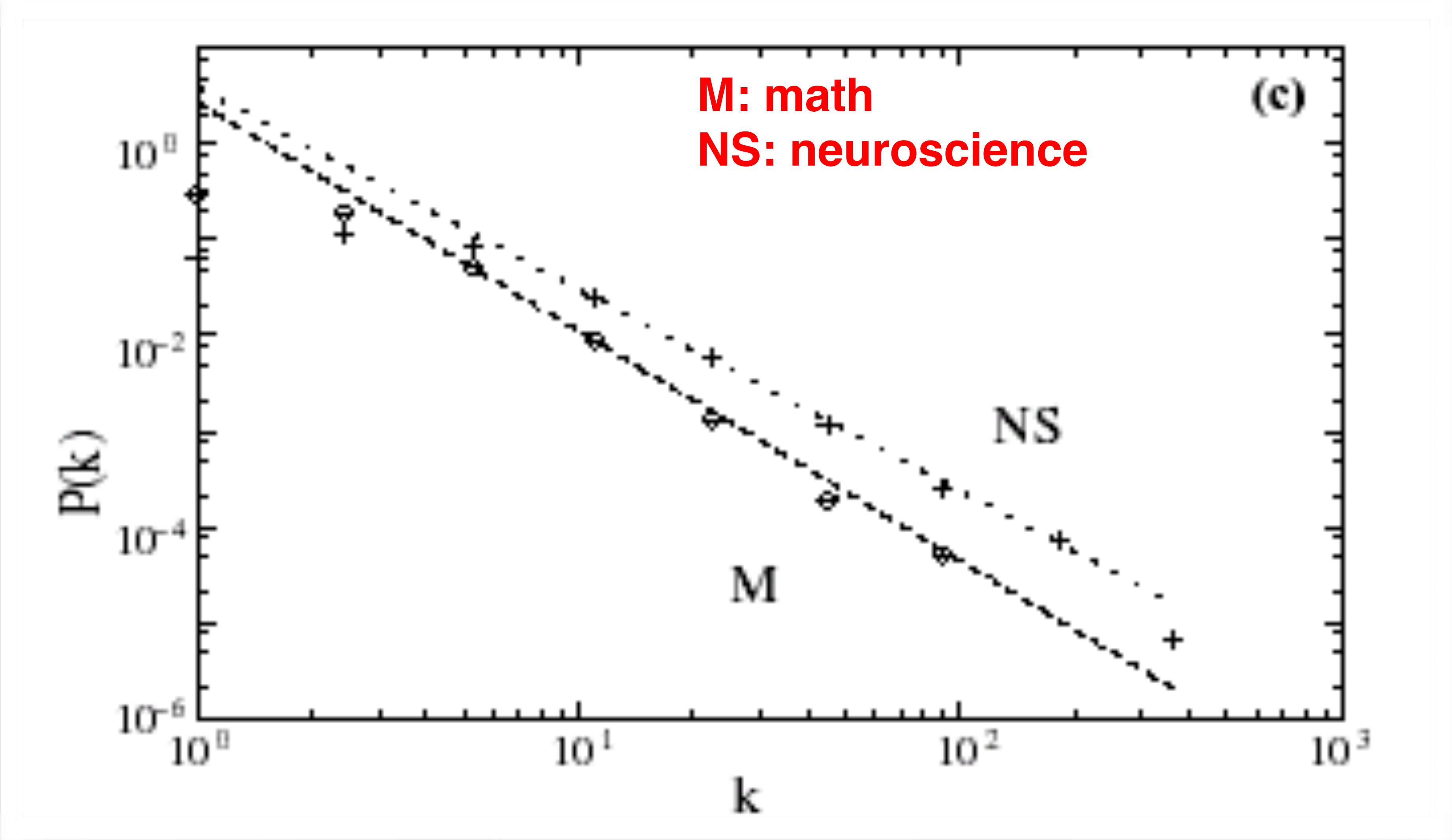
Universality

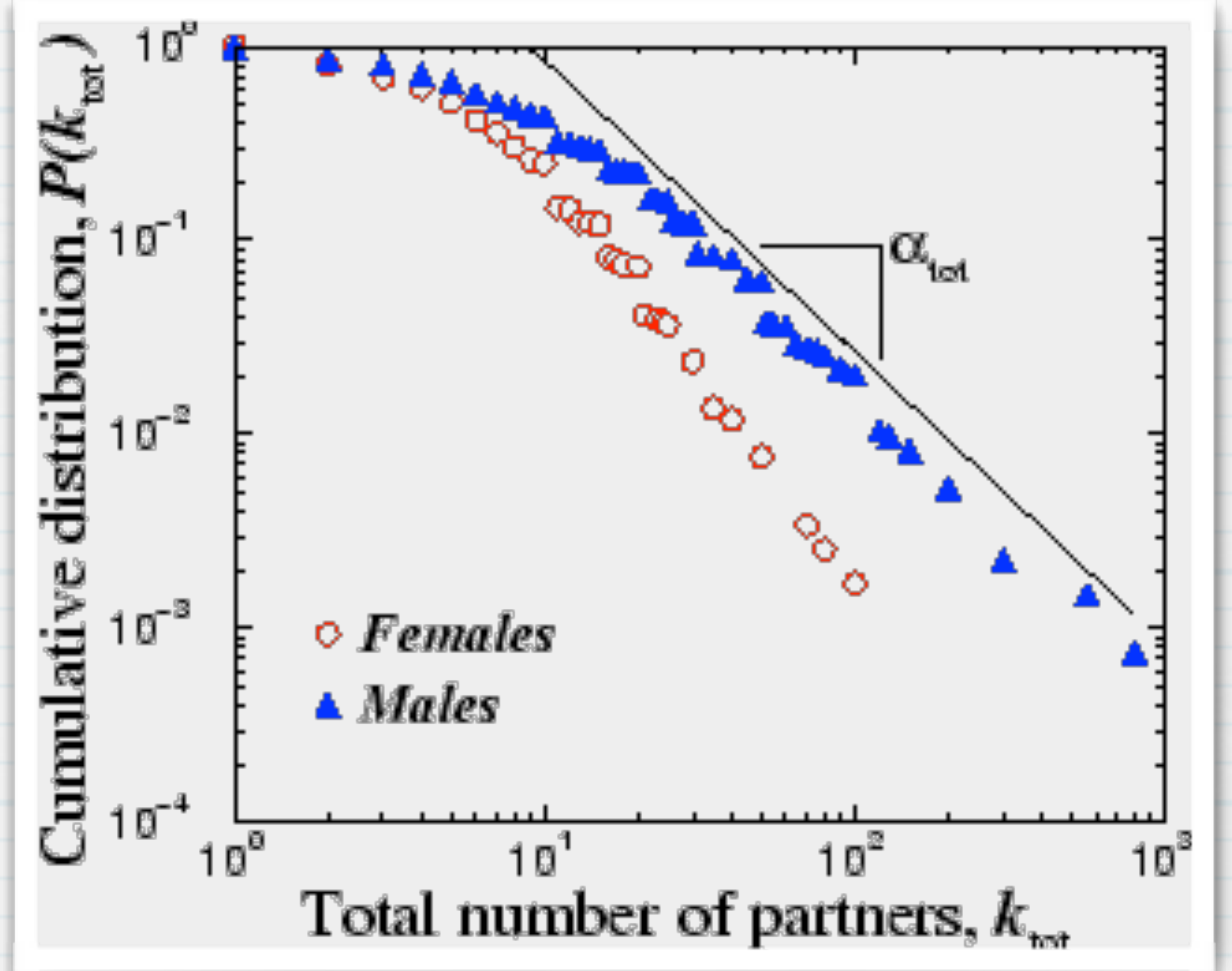
© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering



Actors

Coauthors



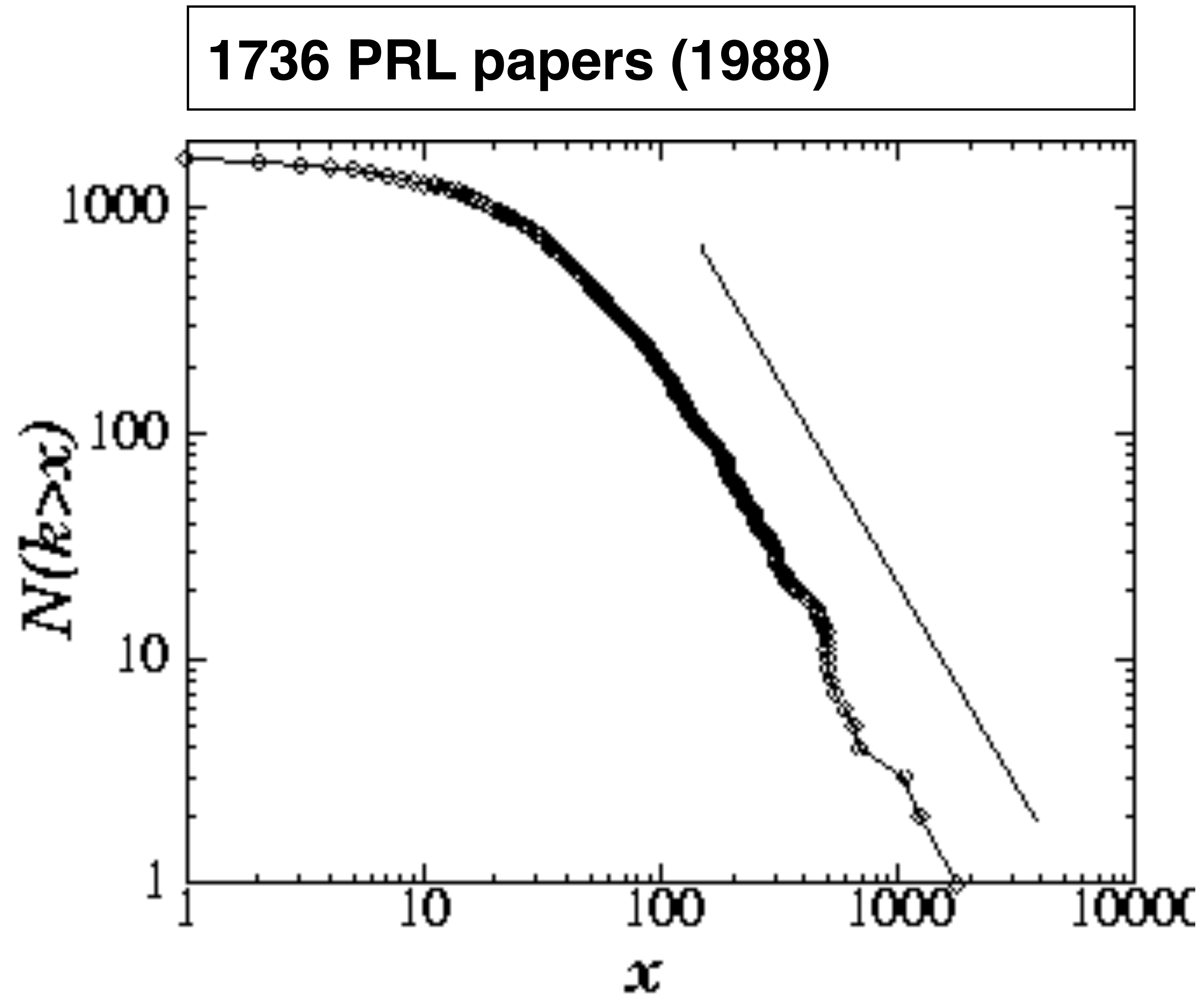


Sex

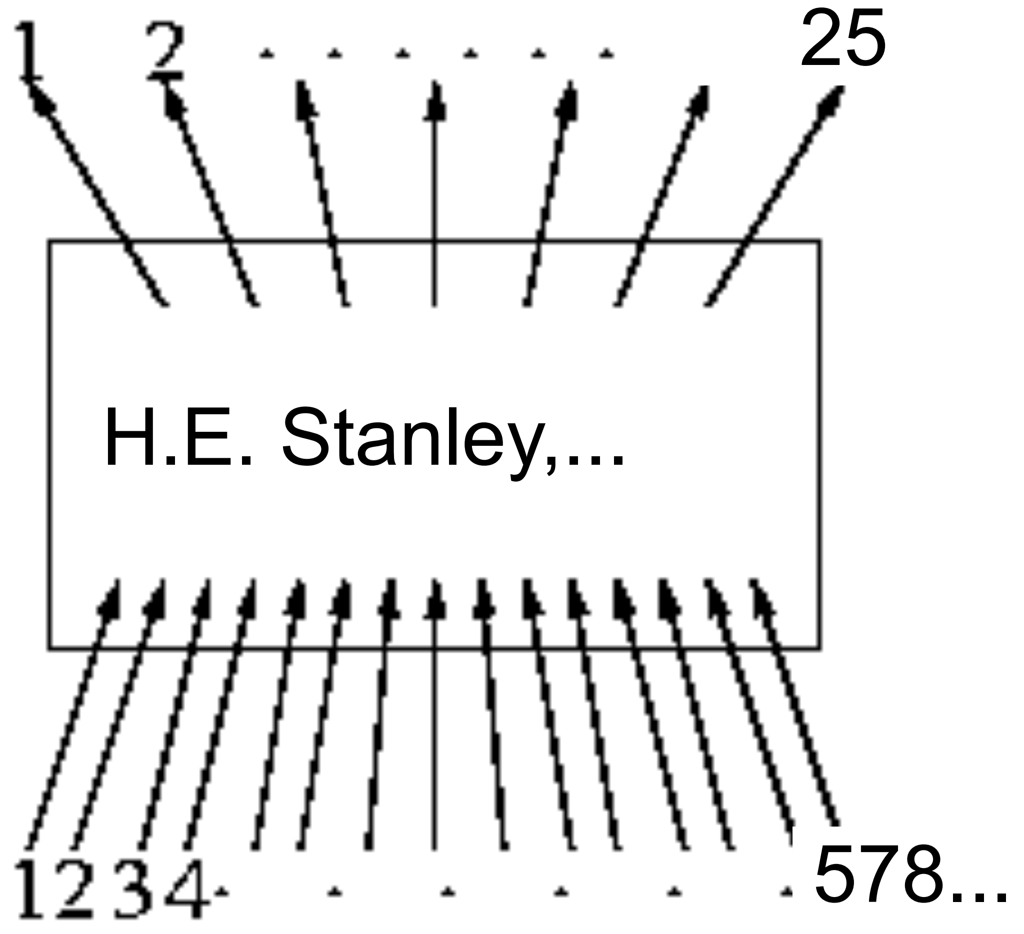
© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

SCIENCE CITATION NETWORKS

Nodes: papers
Links: citations



(S. Redner, 1998)



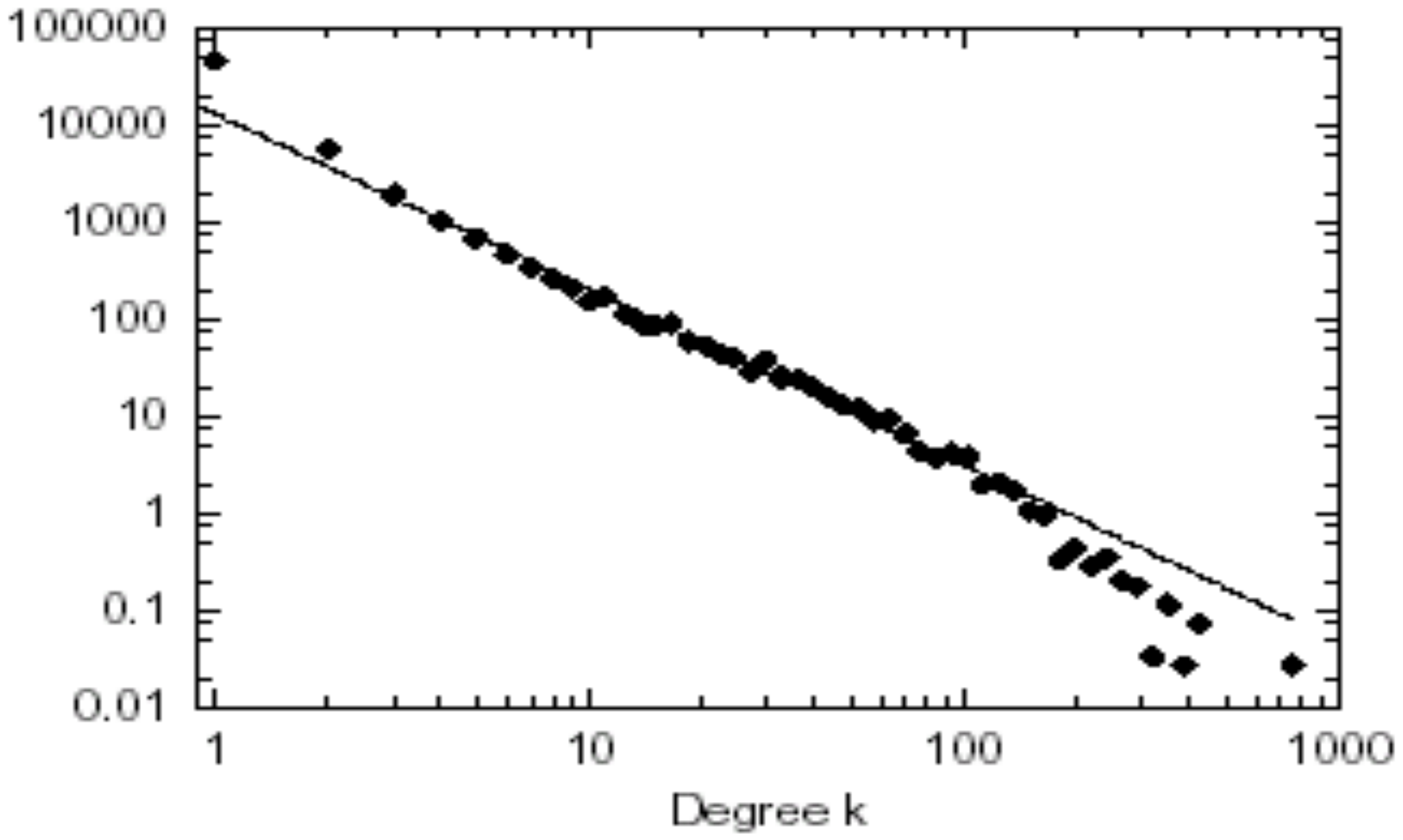
$P(k) \sim k^{-\gamma}$

($\gamma = 3$)

ONLINE COMMUNITIES

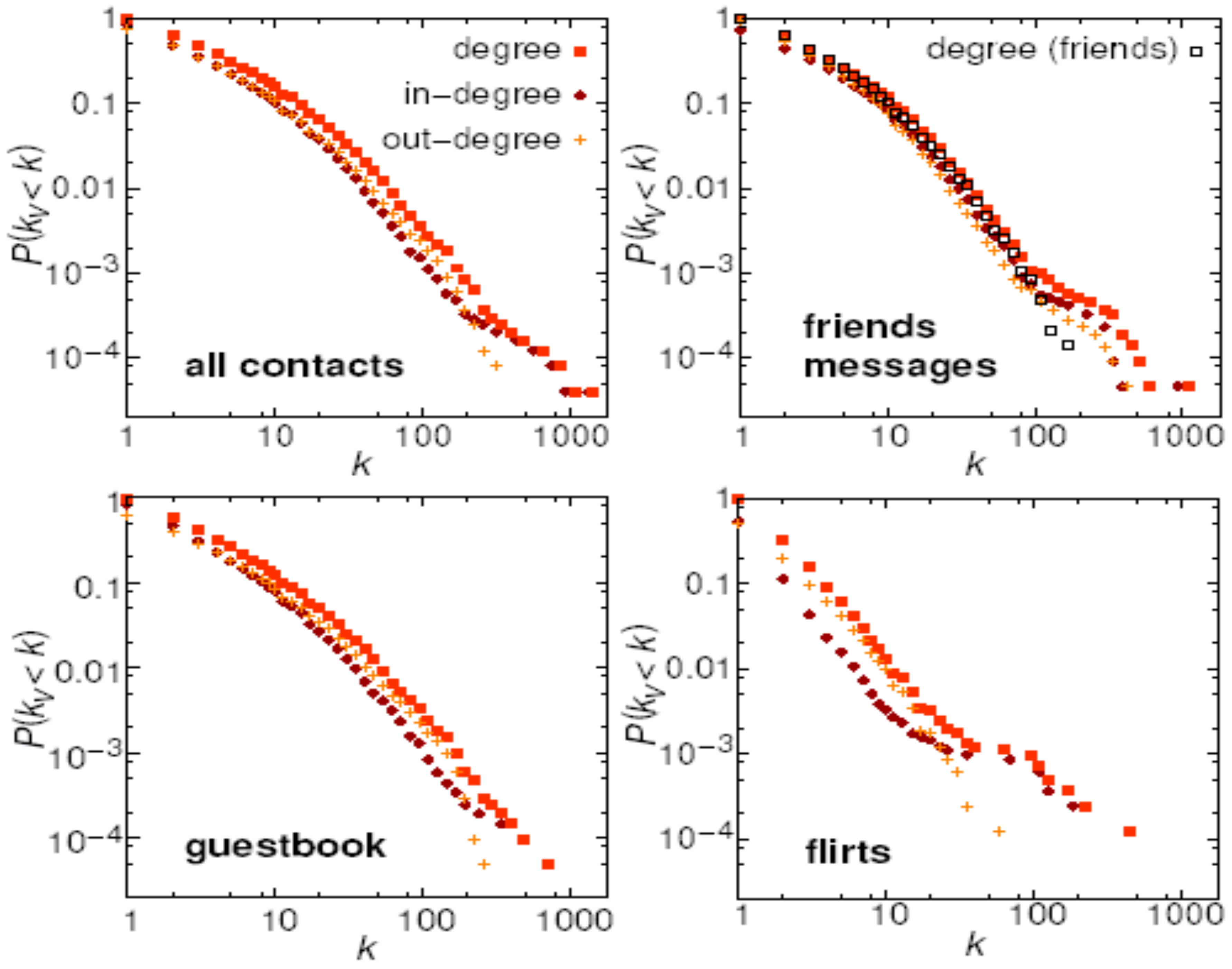
Nodes: online user
Links: email contact

Kiel University log files
112 days, N=59,912 nodes



Ebel, Mielsch, Bornholdtz, PRE 2002.

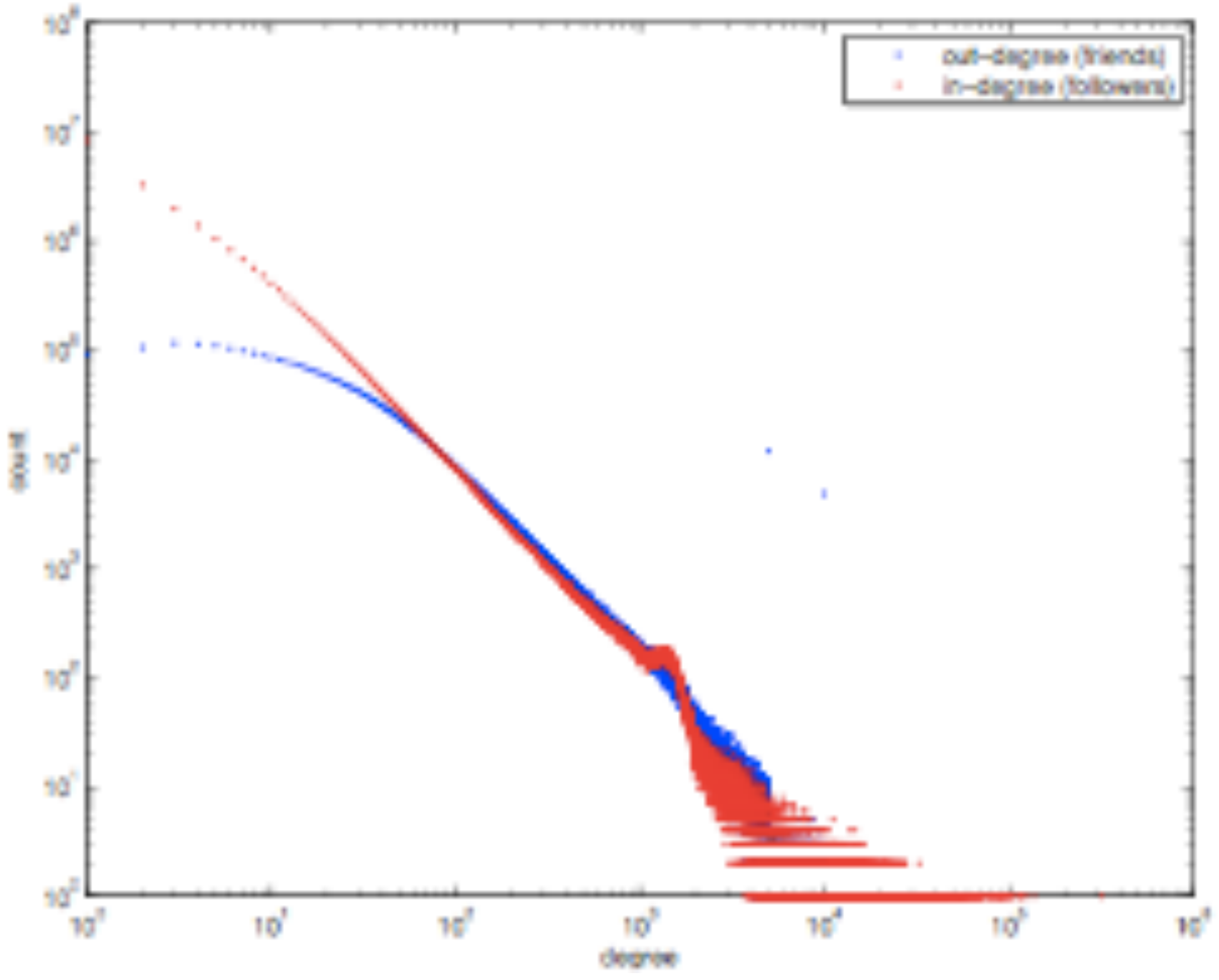
Pussokram.com online community;
512 days, 25,000 users.



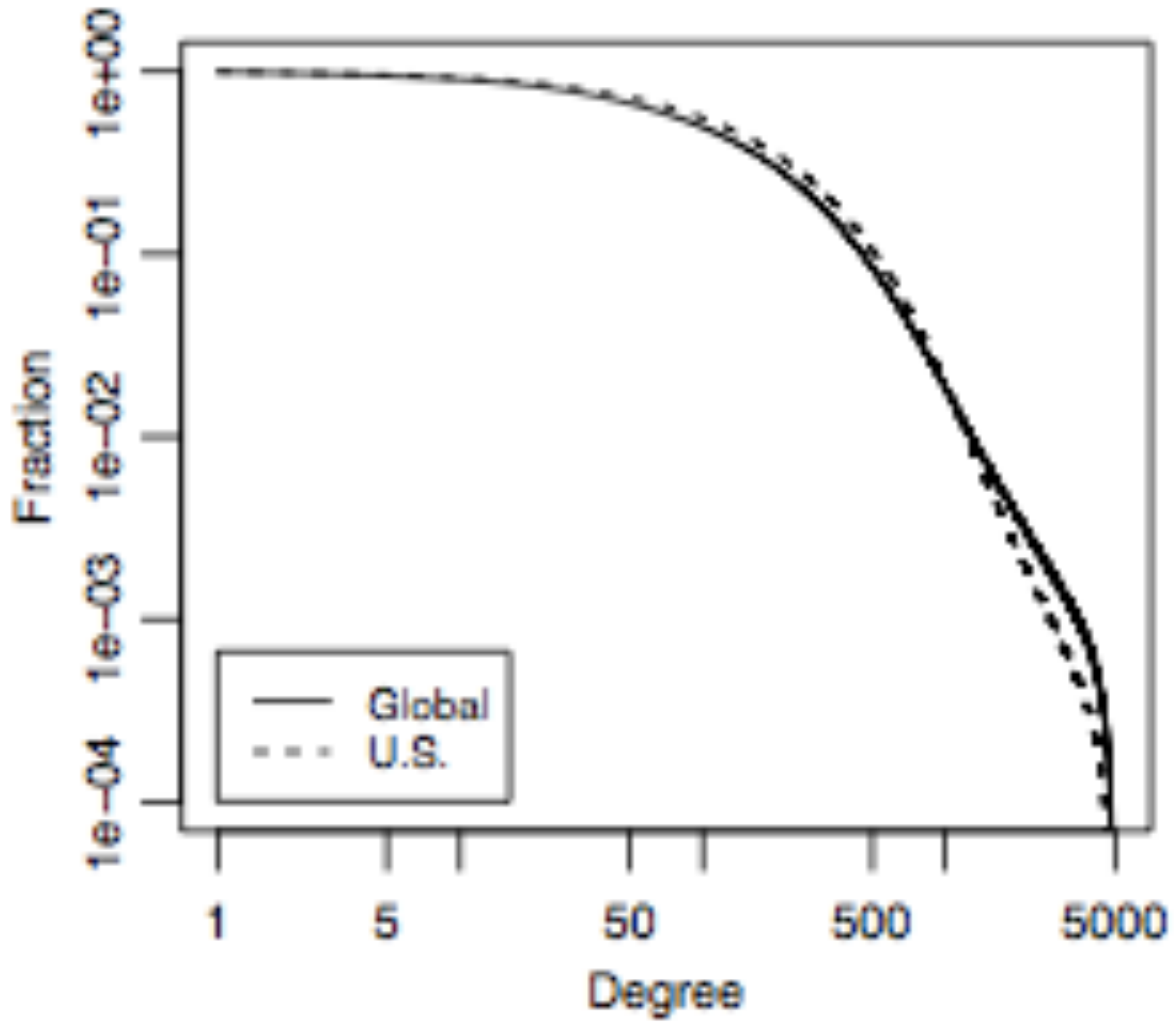
Holme, Edling, Liljeros, 2002.

ONLINE COMMUNITIES

Twitter:

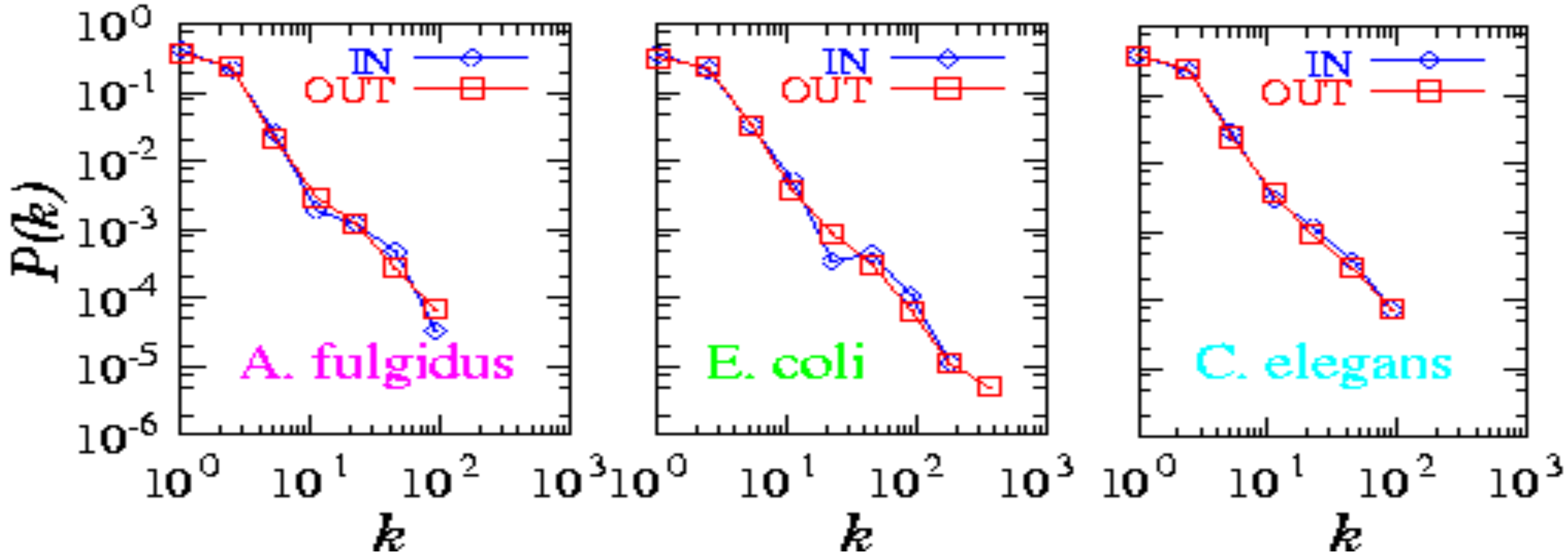


Facebook



Brian Karrer, Lars Backstrom, Cameron Marlowm 2011

METABOLIC NETWORKS



Archaea

Bacteria

Eukaryotes

Organisms from all three domains of life are scale-free!

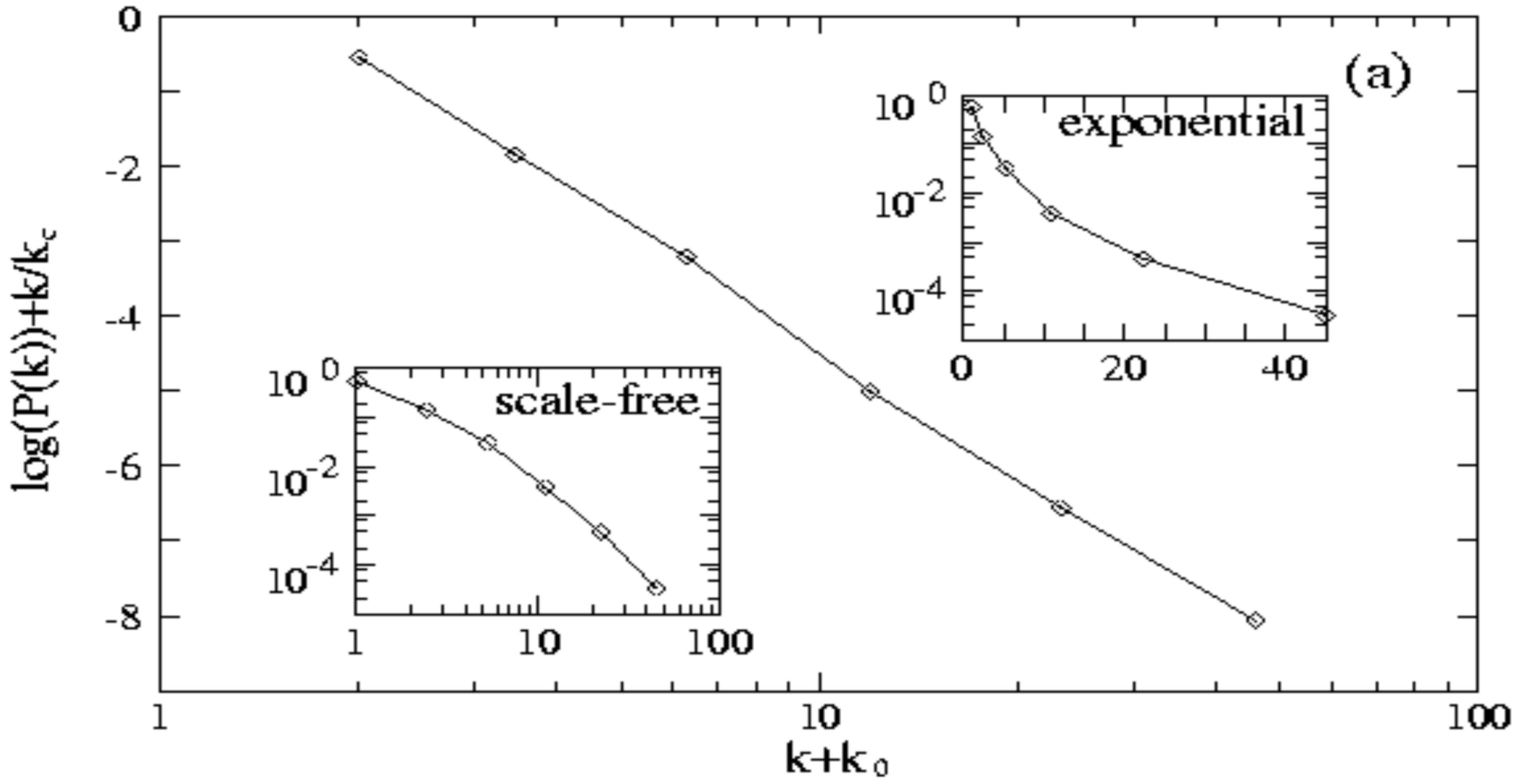
$$P_{in}(k) \approx k^{-2.2}$$

$$P_{out}(k) \approx k^{-2.2}$$

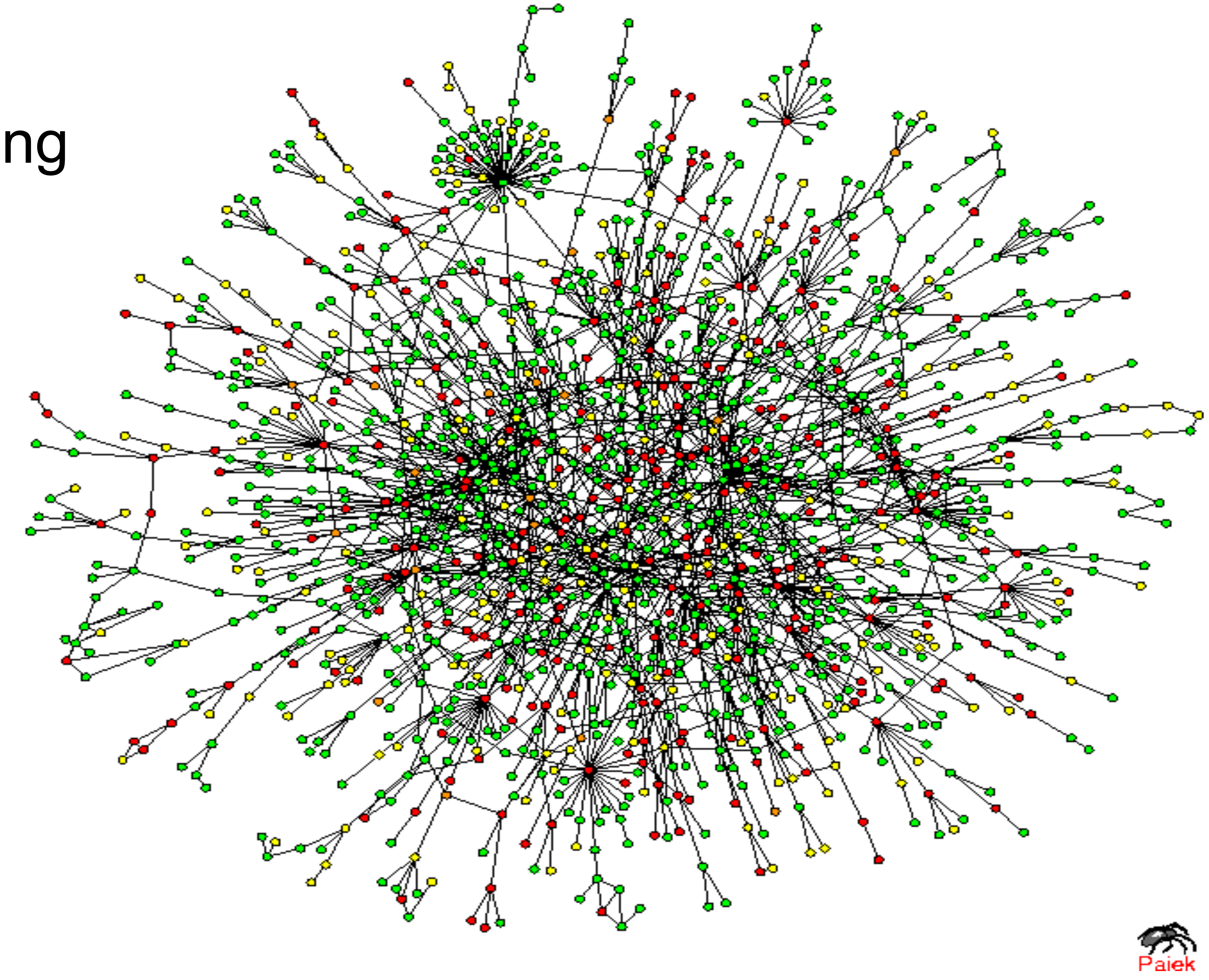
H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi, *Nature*, 407 651 (2000)

TOPOLOGY OF THE PROTEIN NETWORK

Nodes: proteins
Links: physical interactions-binding



$$P(k) \sim (k + k_0)^{-\gamma} \exp(-\frac{k + k_0}{k_\tau})$$



H. Jeong, S.P. Mason, A.-L. Barabasi, Z.N. Oltvai, Nature 411, 41-42 (2001)

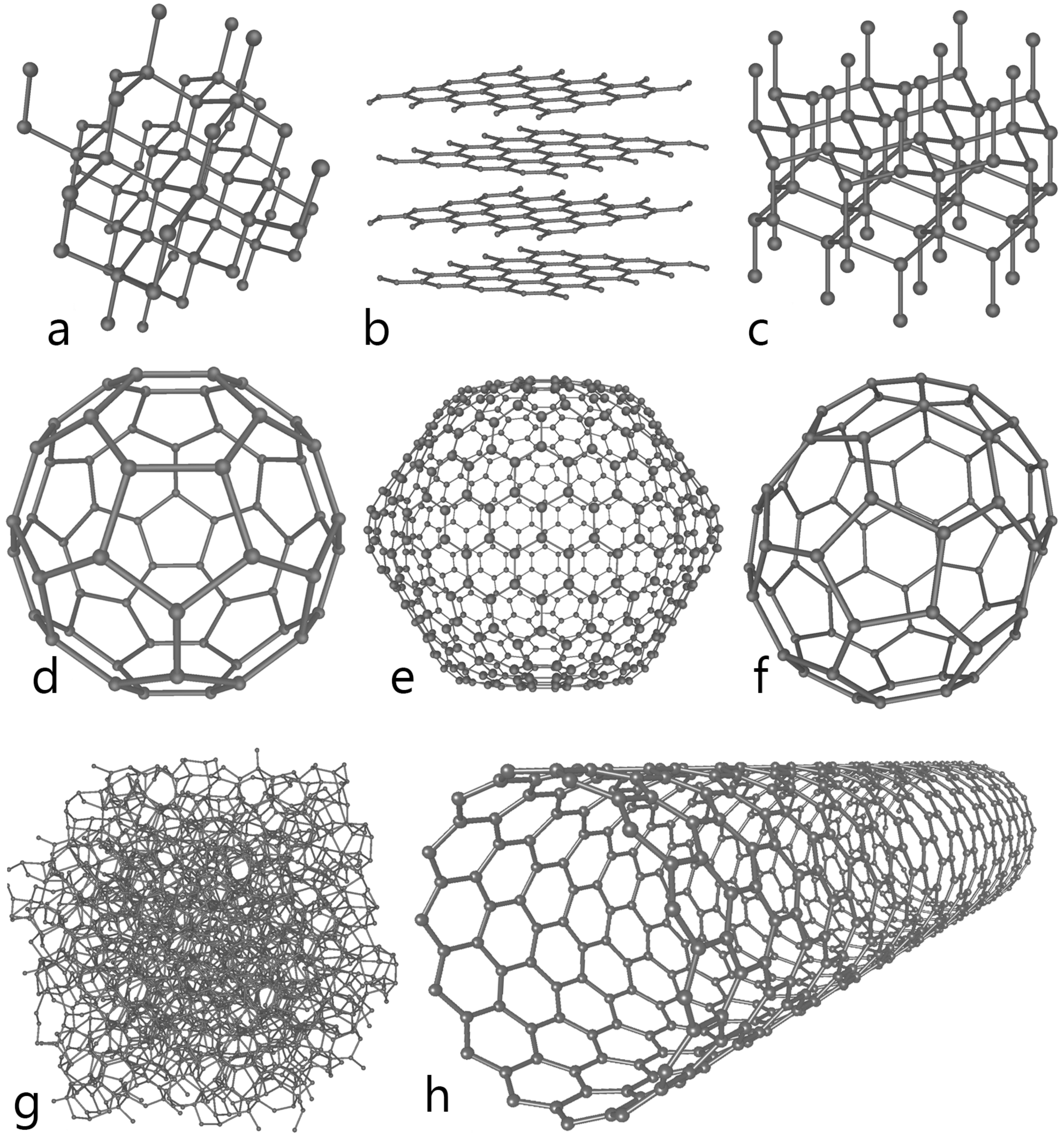
Not all networks are scale-free

•*Networks appearing in material science, like the network describing the bonds between the atoms in crystalline or amorphous materials, where each node has exactly the same degree.*

•*The neural network of the C.elegans worm.*

•*The power grid, consisting of generators and switches connected by transmission lines.*

•*Roads!*

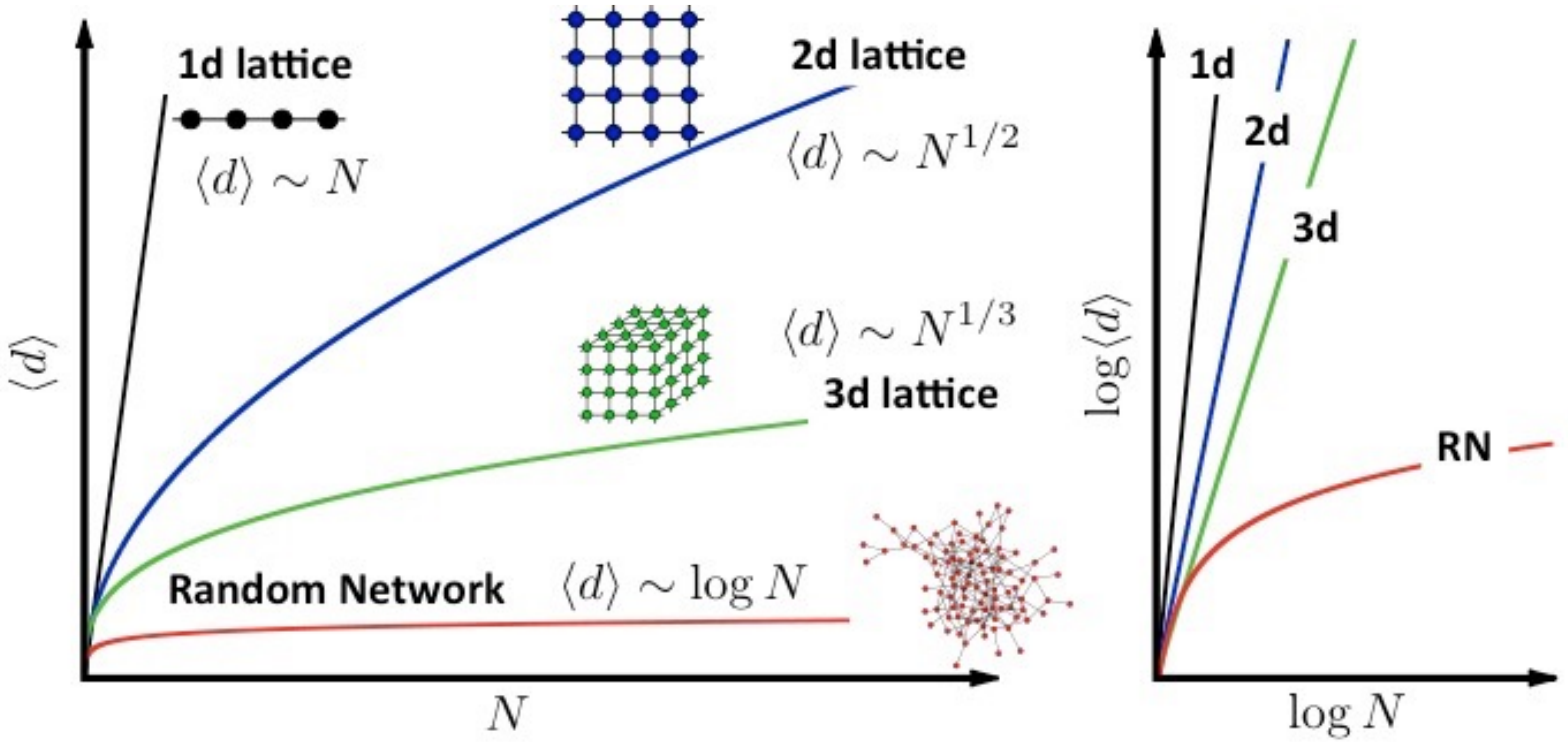


- Review
  - Scale-free property
  - Universality
- Quiz
  - Robustness

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Short paths

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering


1d lattice
$\langle d \rangle \sim N$
2d lattice
$\langle d \rangle \sim N^{1/2}$
$\langle d \rangle \sim N^{1/3}$
3d lattice
Random Network
$\langle d \rangle \sim \log N$
$\langle d \rangle$
N
1d
2d
3d
RN
$\log \langle d \rangle$
$\log N$

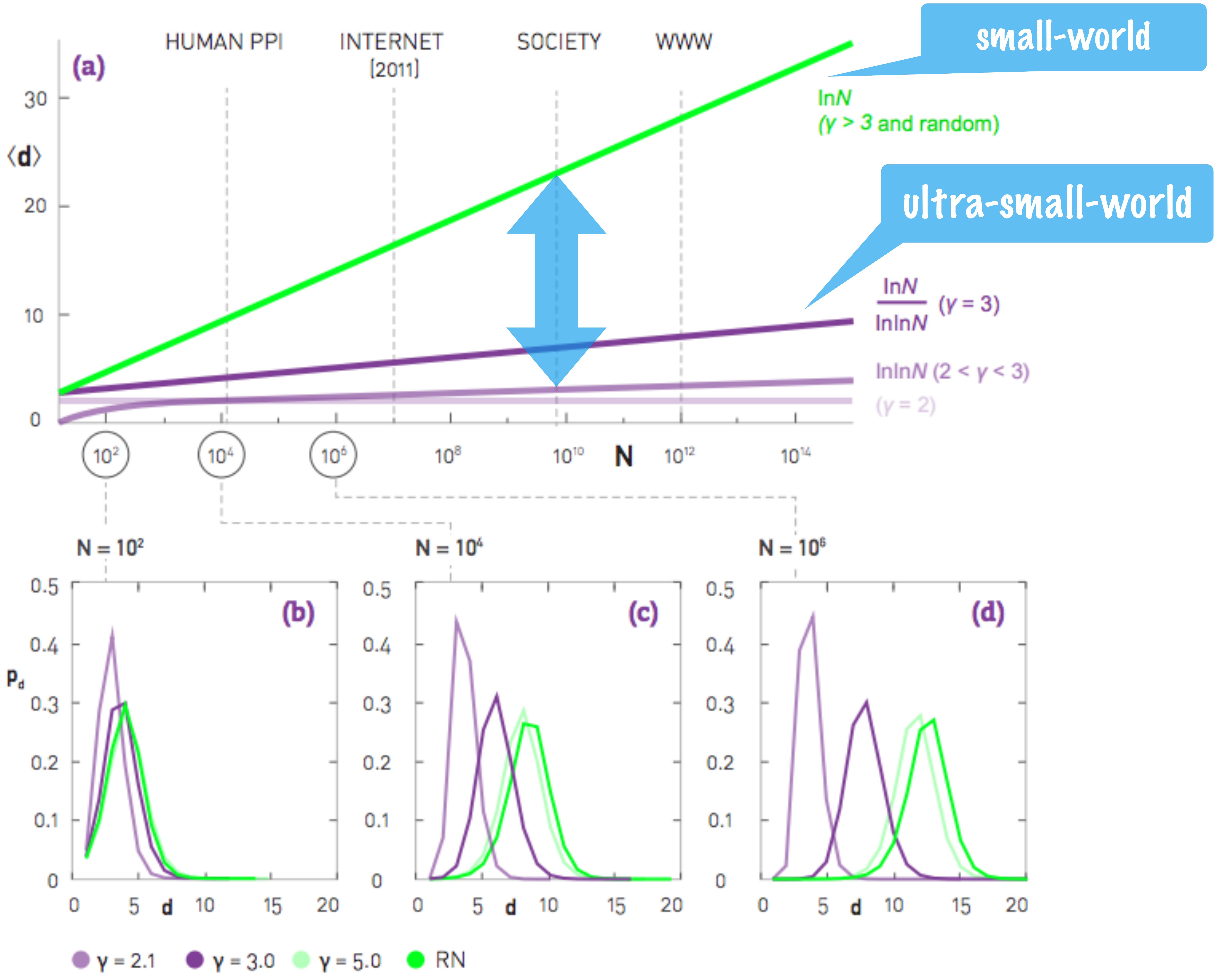
SMALL WORLD BEHAVIOR IN SCALE-FREE NETWORKS

$$
< l > \sim \begin{cases} const. & \gamma = 2 \\ \dfrac{\ln \ln N}{\ln(\gamma - 1)} & 2 < \gamma < 3 \\ \dfrac{\ln N}{\ln \ln N} & \gamma = 3 \\ \ln N & \gamma > 3 \end{cases}
$$

ultra-small-world

small-world

Cohen, Havlin Phys. Rev. Lett. 90, 58701(2003); Cohen, Havlin and ben-Avraham, in Handbook of Graphs and Networks, Eds. Bornholdt and Shuster (Willy-VCH, NY, 2002) Chap. 4; Confirmed also by: Dorogovtsev et al (2002), Chung and Lu (2002); (Bollobas, Riordan, 2002; Bollobas, 1985; Newman, 2001



small-world
ultra-small-world
(a)
HUMAN PPI
INTERNET [2011]
SOCIETY
WWW
⟨d⟩
lnN (γ > 3 and random)
lnN / lnlnN (γ = 3)
lnlnN (2 < γ < 3)
(γ = 2)
N
N = 10²
N = 10⁴
N = 10⁶
(b)
(c)
(d)
p_d
d
γ = 2.1
γ = 3.0
γ = 5.0
RN

Summary

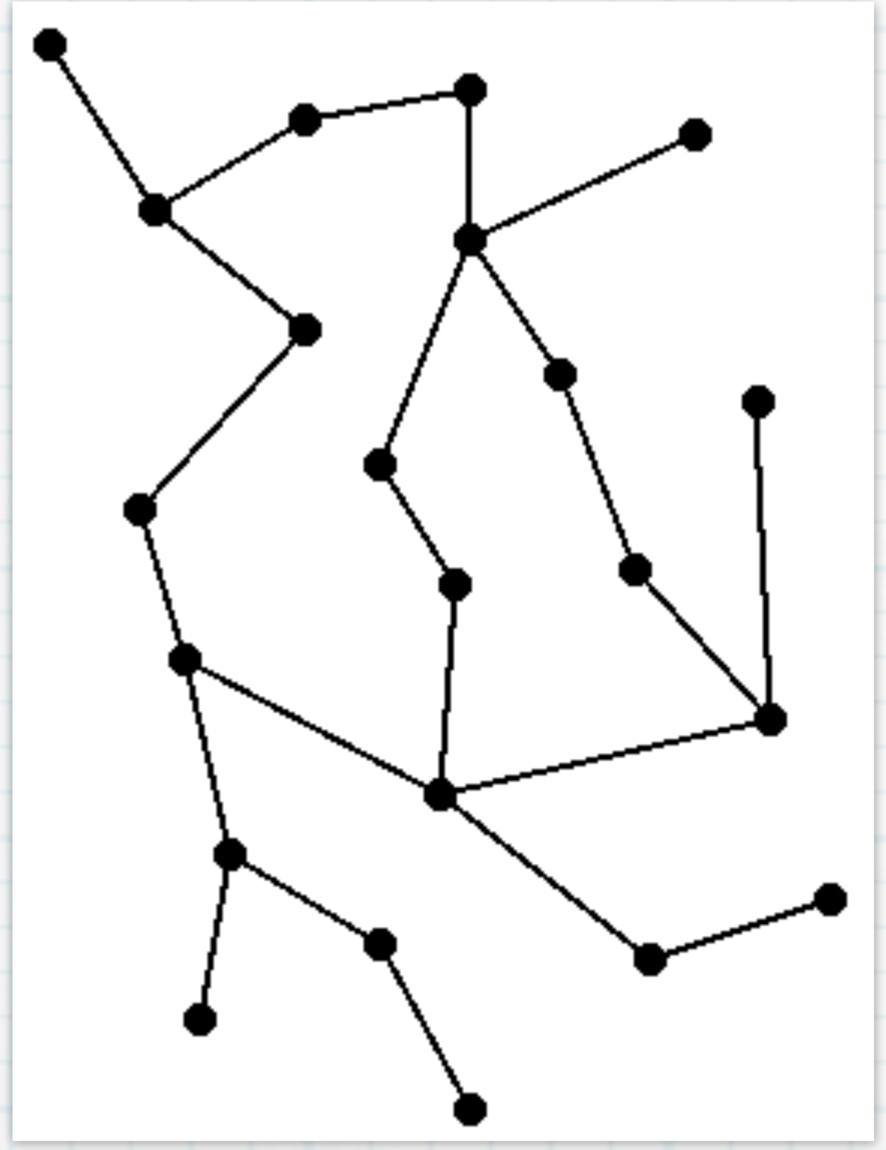
* Social networks have short paths, high clustering, and the scale free property
* The scale free property implies (ultra) short paths
* The scale free property does not imply high clustering
* The scale free property is universal: many (but not all) real networks are scale free!

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Robustness

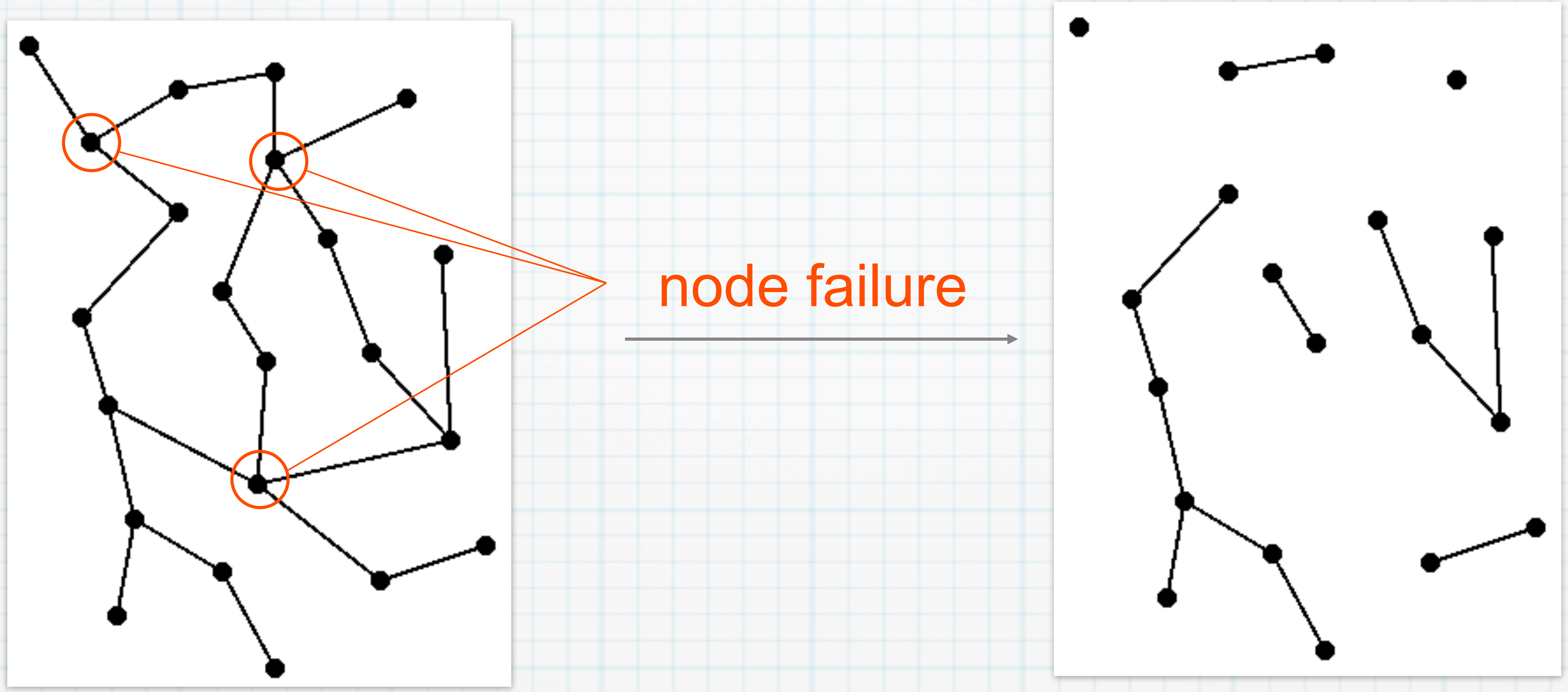
© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Physical properties: Robustness



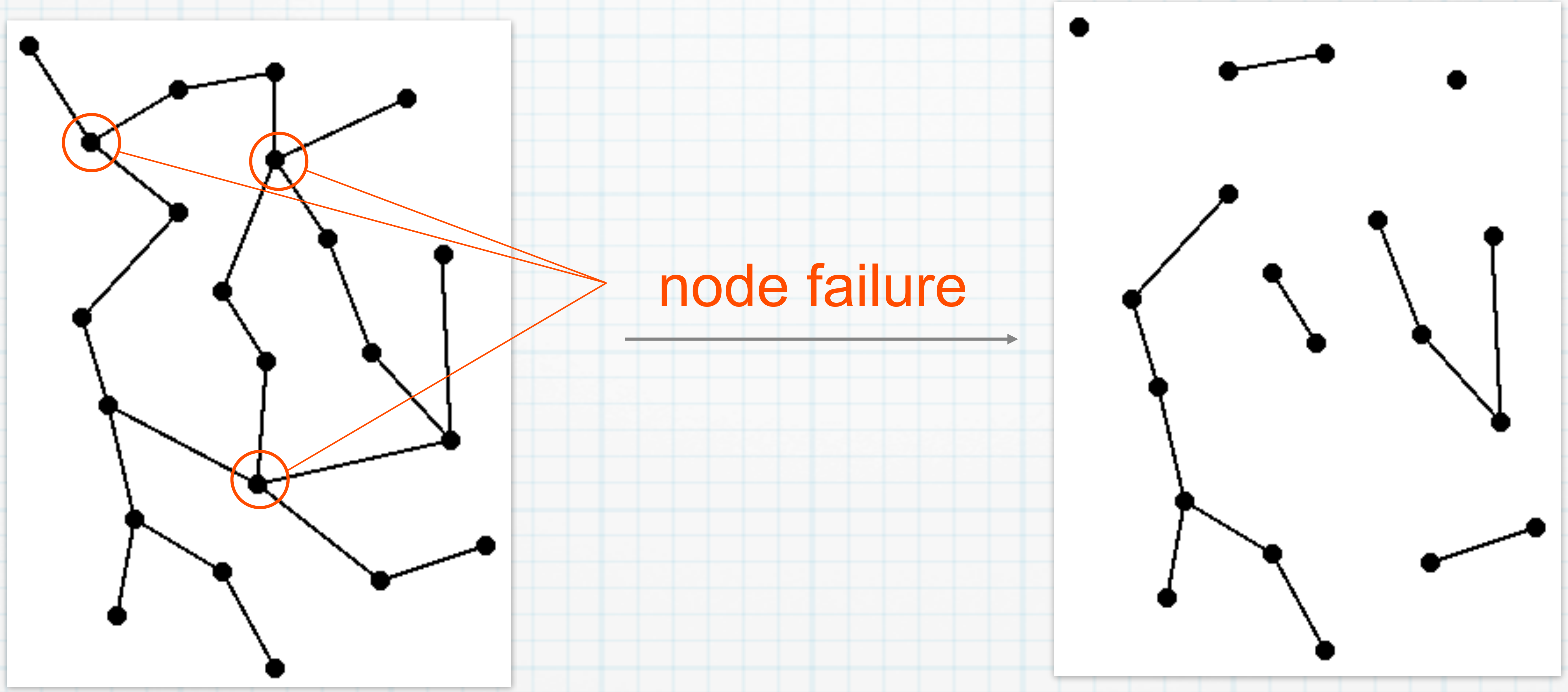
© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Physical properties: Robustness



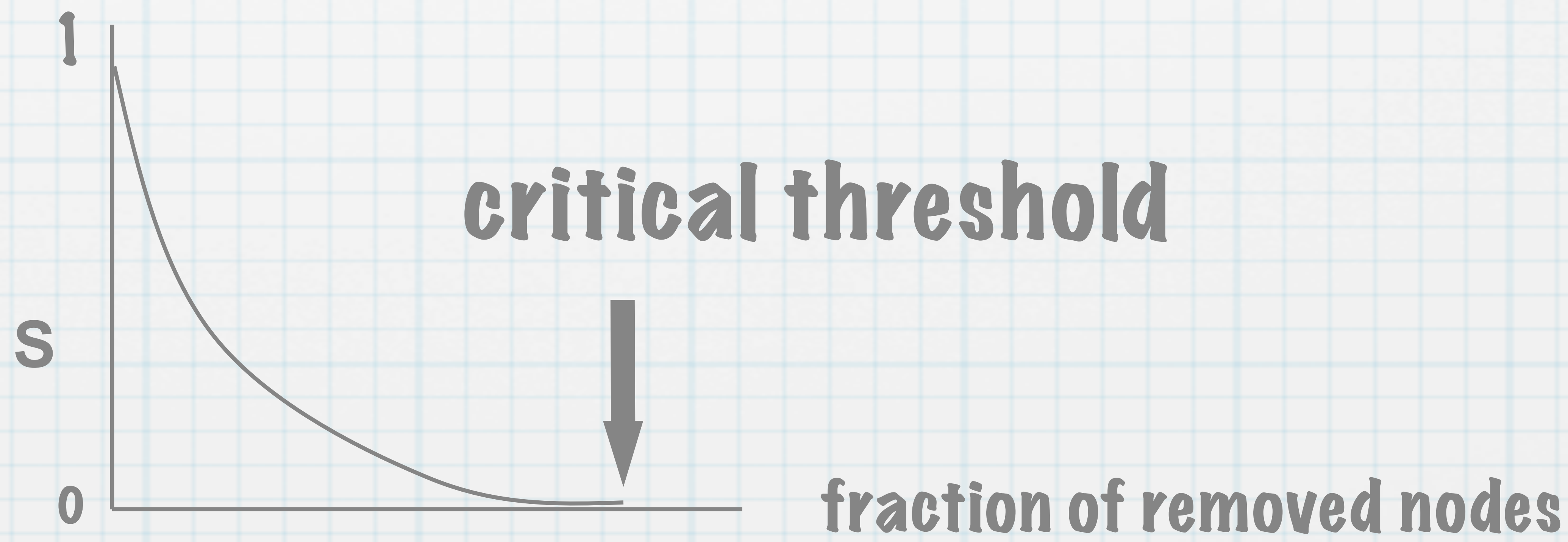
© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Physical properties: Robustness



Similar to percolation problem

S=size of giant component

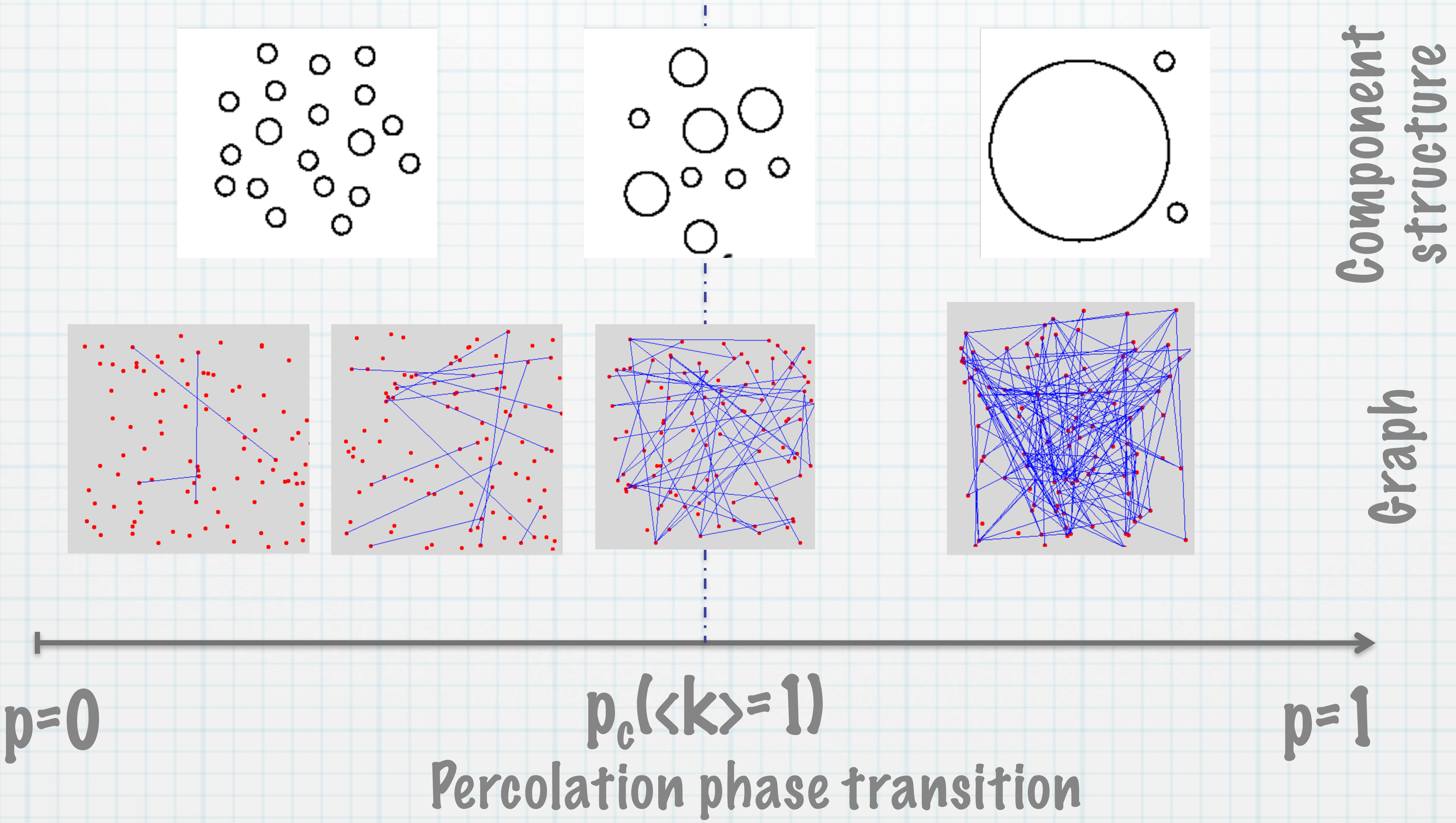


© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

INDIANA UNIVERSITY BLOOMINGTON

Remember the Erdös-Rényi model

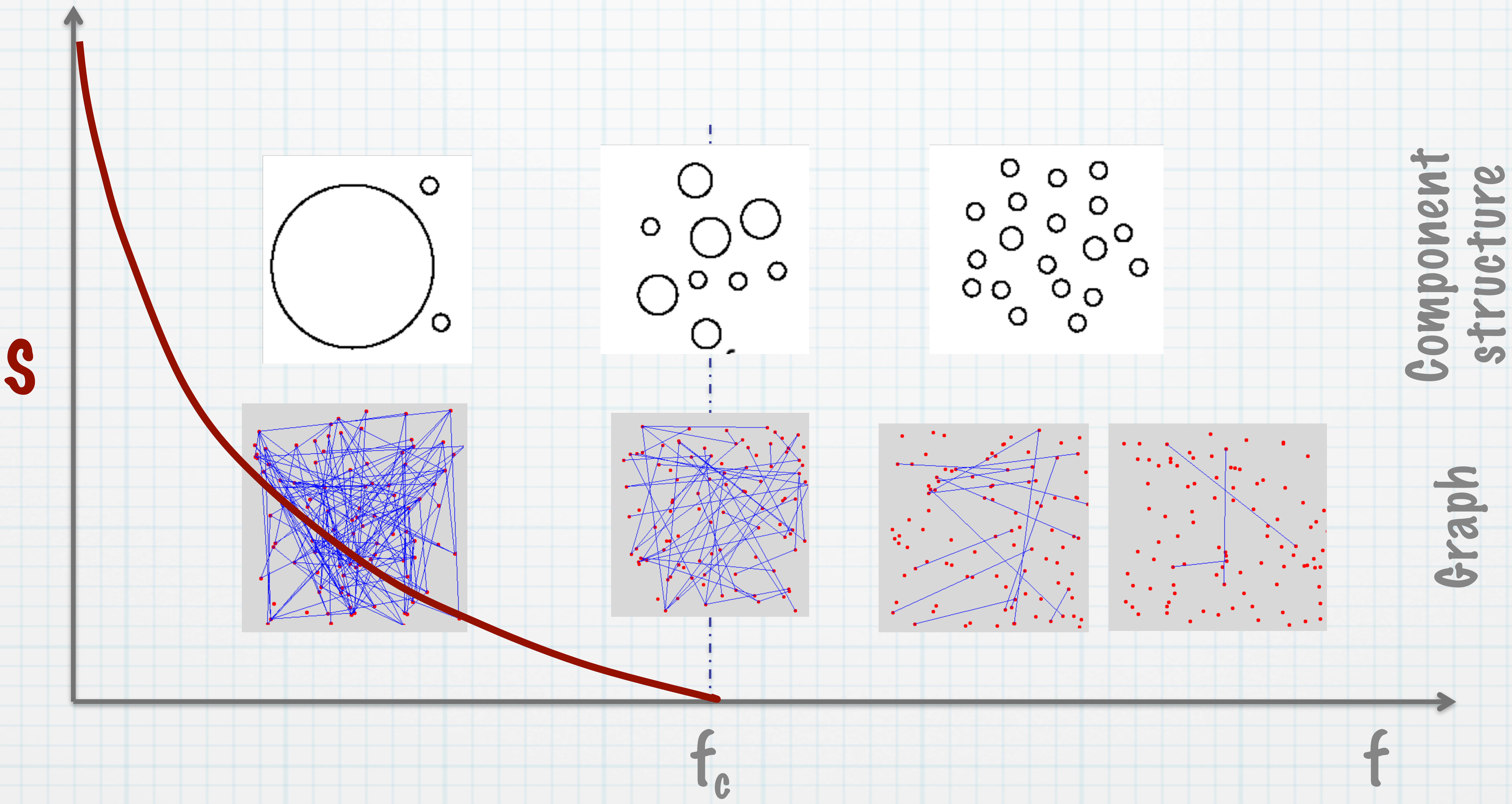
Graph structure changes as a function of p



© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Damage is modeled as an inverse percolation process

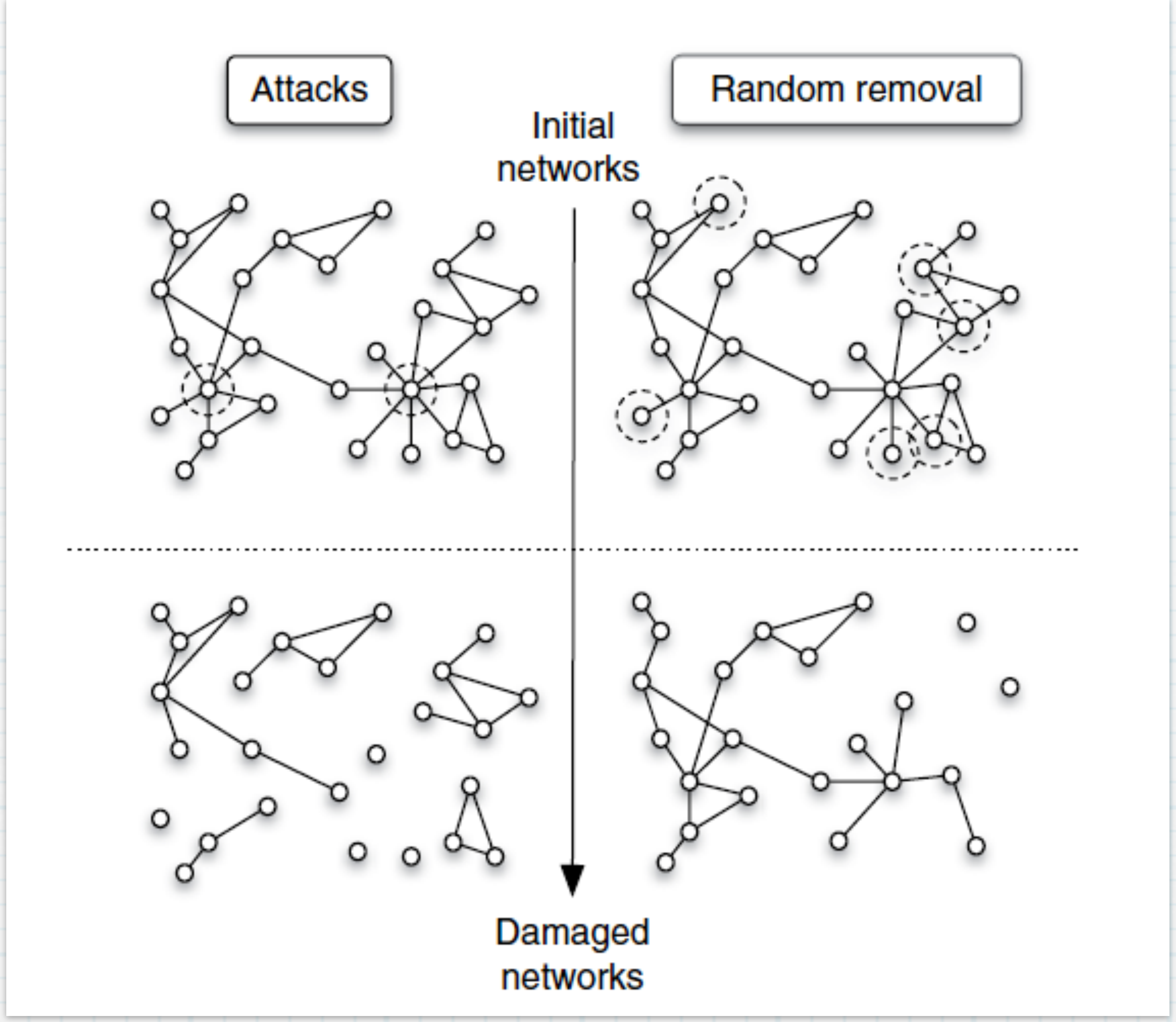
f = fraction of removed nodes



Inverse percolation phase transition

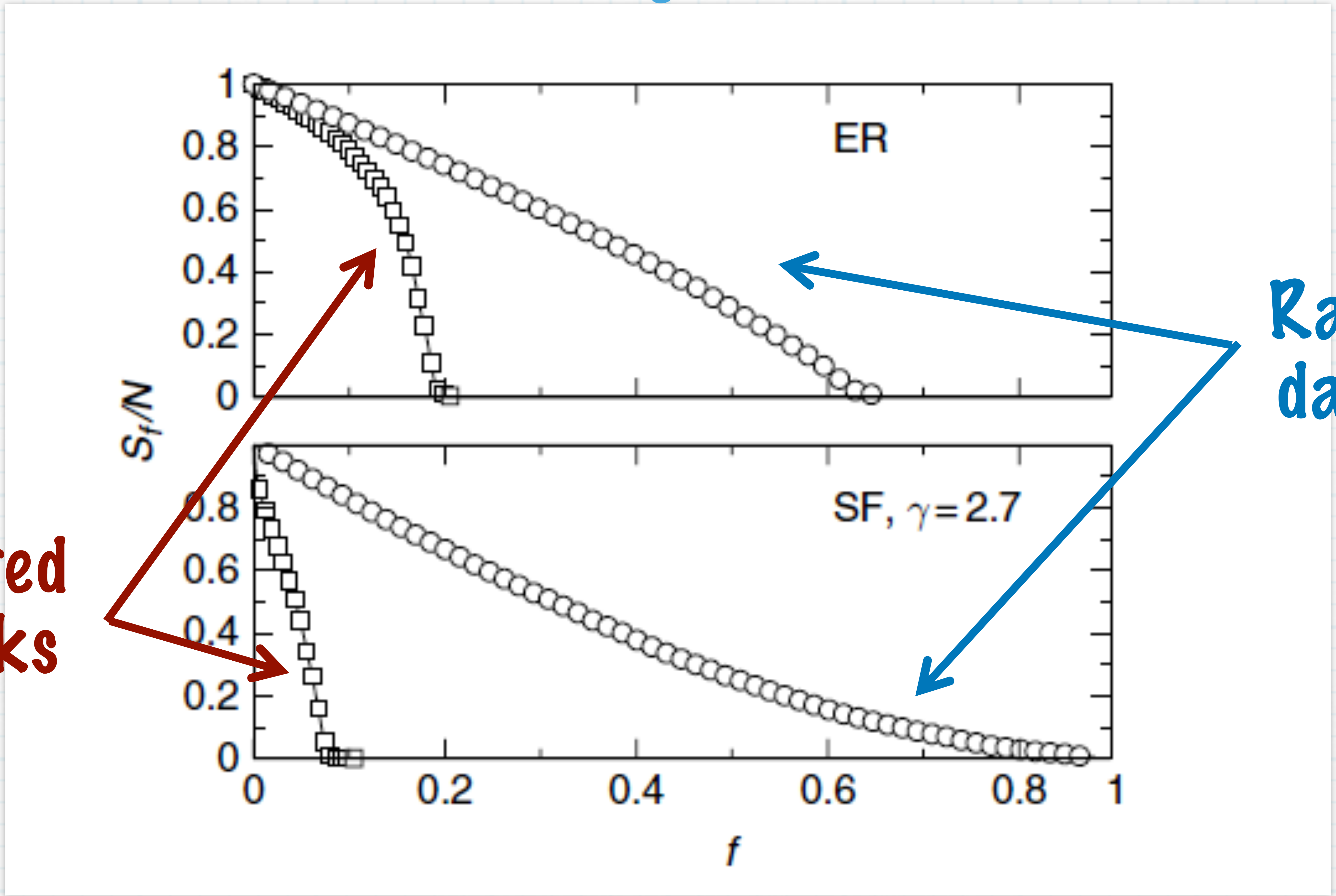
© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Different types of attack



© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Different types of attack

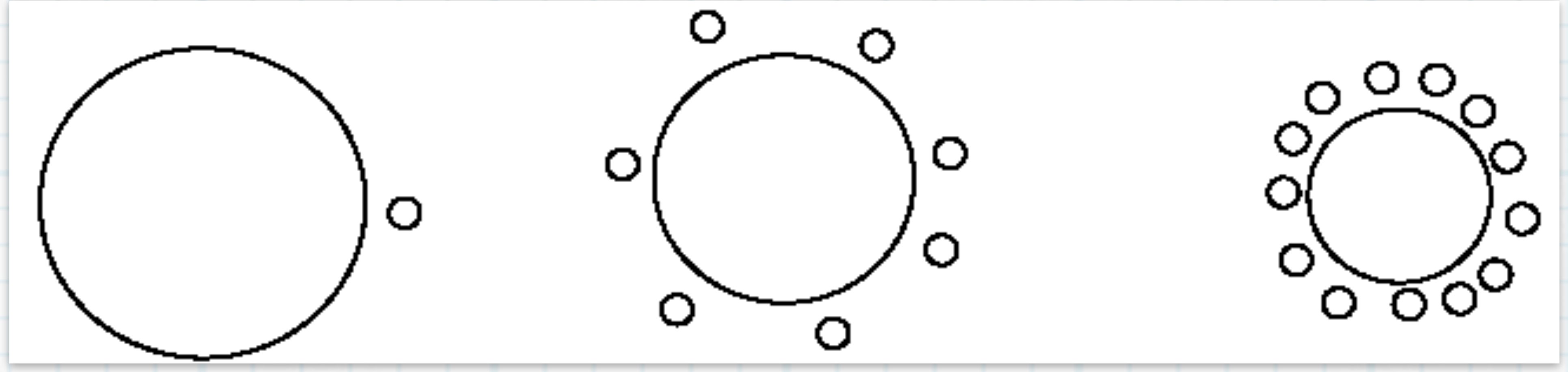


Targeted attacks

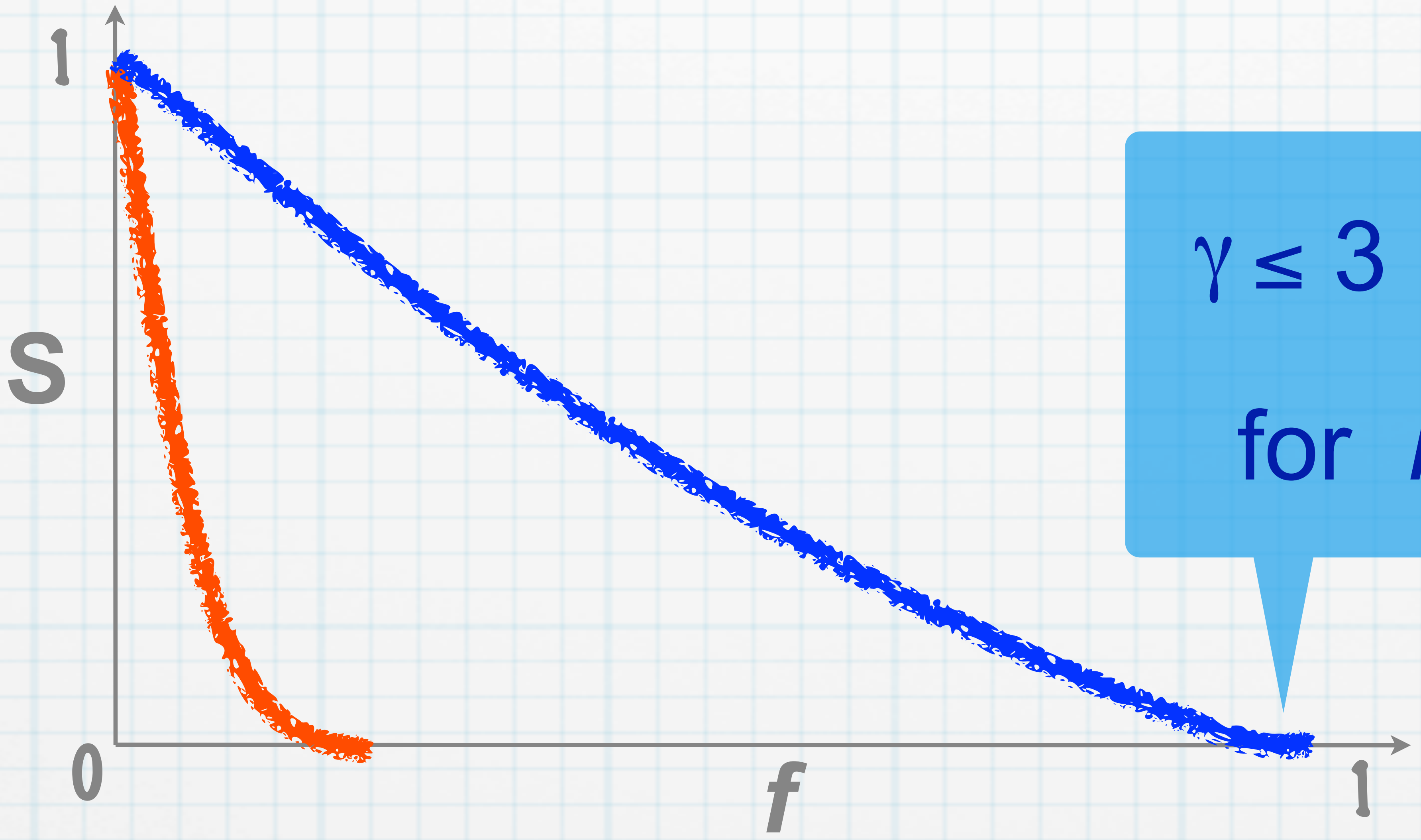
Random damage

INDIANA UNIVERSITY BLOOMINGTON

Resilience in heterogeneous networks



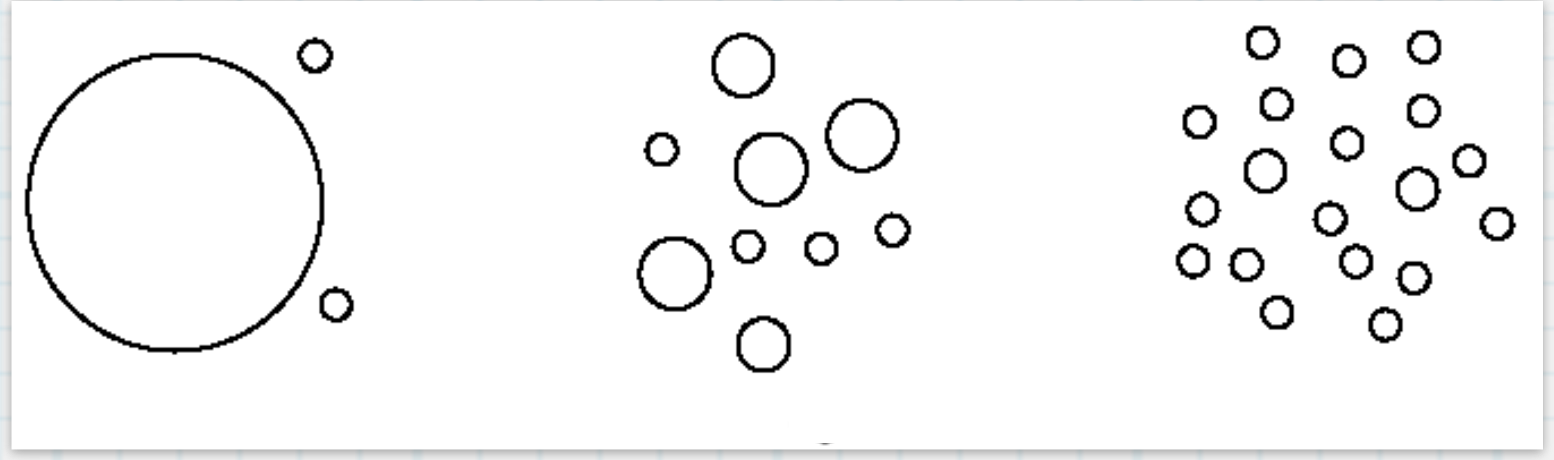
Failures



$\gamma \leq 3 : f_c \longrightarrow 1$

for $N \longrightarrow \infty$

Attacks



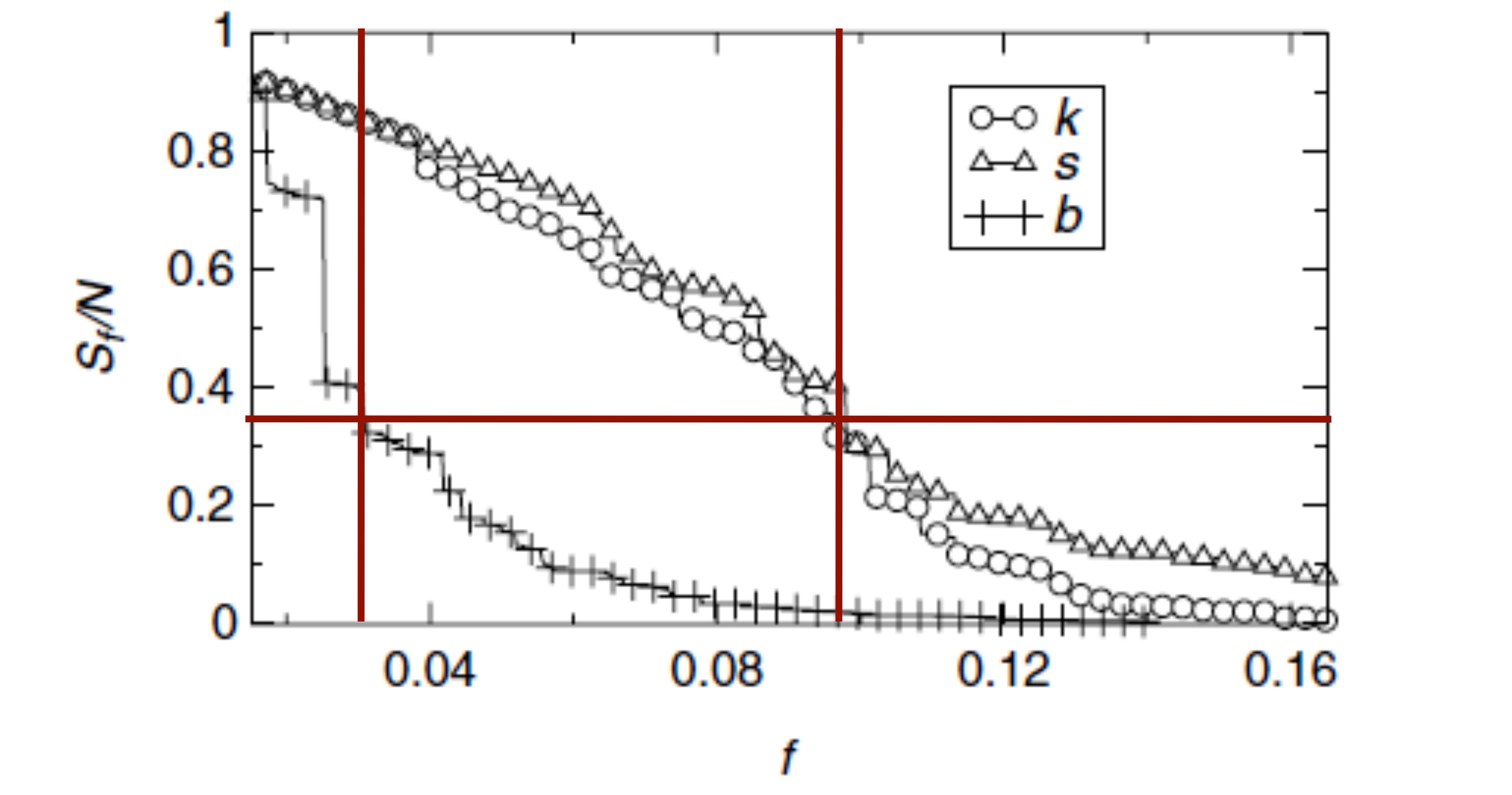
© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Review

* Scale free networks are **more or less** resilient than random networks with respect to random failures?
* Scale free networks are **more or less** resilient than random networks with respect to targeted attacks?

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Damage of a real network



* Worldwide airport network
* Targeting 10% of the nodes disrupts the network entirely
* Targeted betweenness is the most effective strategy

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Preferential Attachment

© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Preferential Attachment model

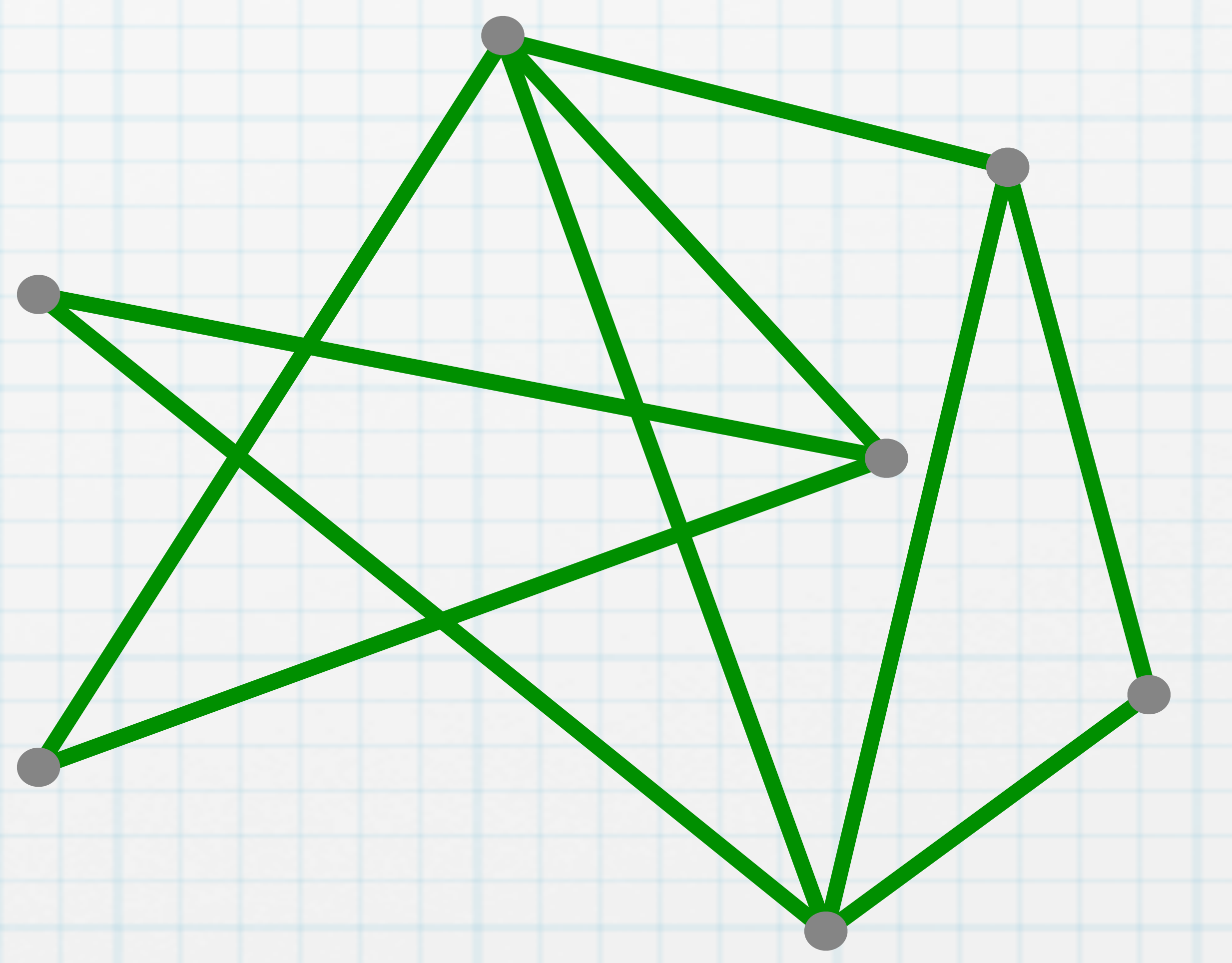
1. GROWTH

At every time step we add a new node with m edges (connected to the nodes already present in the system)

2. ATTACHMENT

The probability Π that a new node will be connected to node i depends on the degree k_i of that node

$$\Pi(k_i) = \frac{k_i}{\Sigma_j k_j}$$



a.k.a. "Barabasi-Albert model"

```
import networkx as nx
G = barabasi_albert_graph(n,m)
```


© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering

Preferential Attachment model

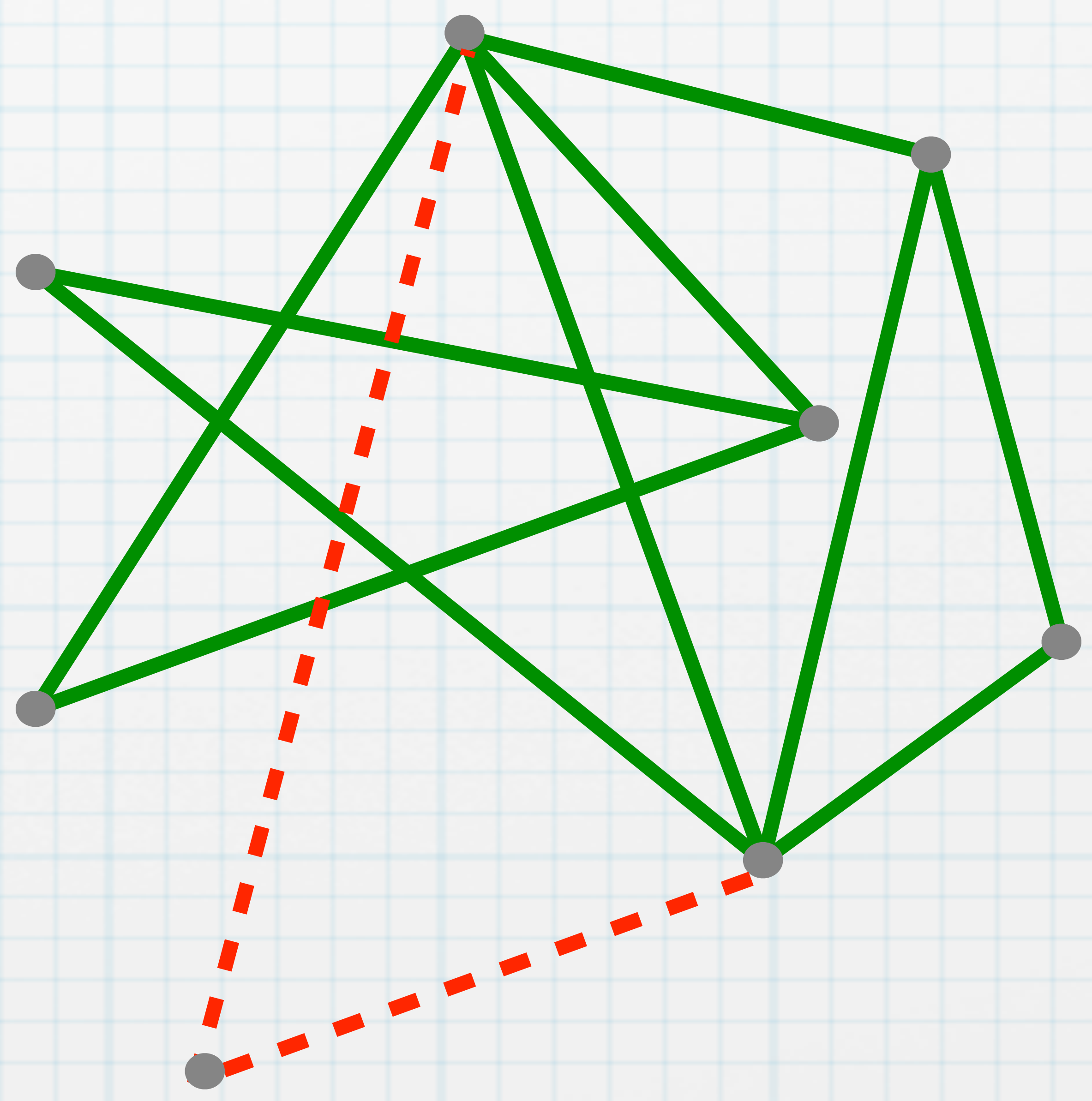
1. GROWTH

At every time step we add a new node with m edges (connected to the nodes already present in the system)

2. ATTACHMENT

The probability Π that a new node will be connected to node i depends on the degree k_i of that node

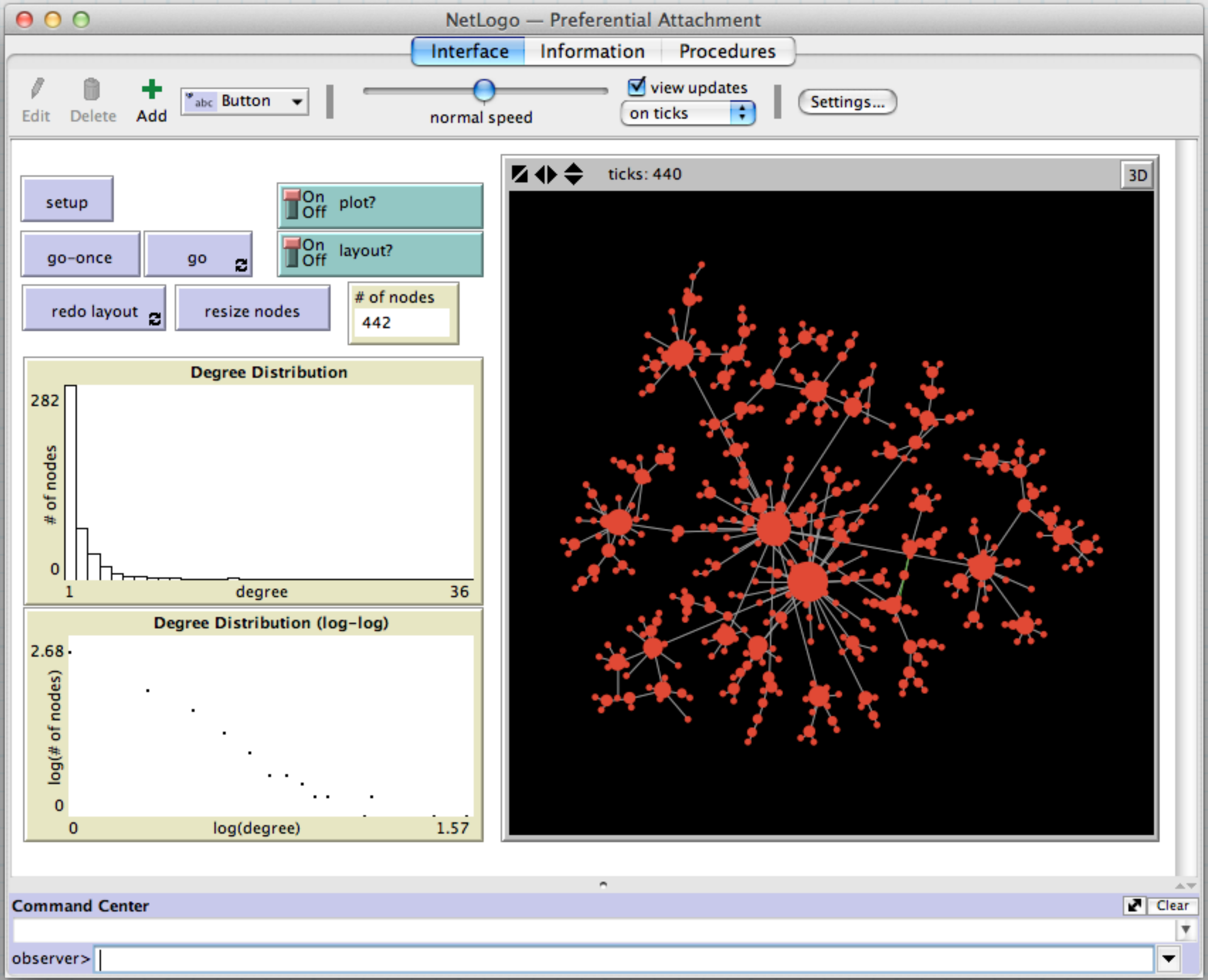
$$\Pi(k_i) = \frac{k_i}{\Sigma_j k_j}$$



a.k.a. "Barabasi-Albert model"

```
import networkx as nx
G = barabasi_albert_graph(n,m)
```


© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering



© 2010-2017 F. Menczer, School of Informatics, Computing, and Engineering