

Social Network Analysis

a crash mini-course

Dino Pedreschi

KDD LAB, University of Pisa and ISTI-CNR

<http://kdd.isti.cnr.it>

MAINS – DM & CRM

Complex

[adj., v. kuh m-pleks, kom-pleks; n. kom-pleks]

–adjective

1.

composed of many interconnected parts; compound; composite: a complex highway system.

2.

characterized by a very complicated or involved arrangement of parts, units, etc.: complex machinery.

3.

so complicated or intricate as to be hard to understand or deal with: a complex problem.

Source: Dictionary.com

Complexity, a **scientific theory** which asserts that some systems display behavioral phenomena that are completely inexplicable by any conventional analysis of the systems' constituent parts. These phenomena, commonly referred to as emergent behaviour, seem to occur in many complex systems involving living organisms, such as a stock market or the human brain.

*Source: [John L. Casti](#), *Encyclopædia Britannica**

Complexity

Behind each complex system there is a **network**, that defines the interactions between the components

US DEMOCRACY

PROTONS

SEISMOLOGY

INTERNATIONAL REPTILES

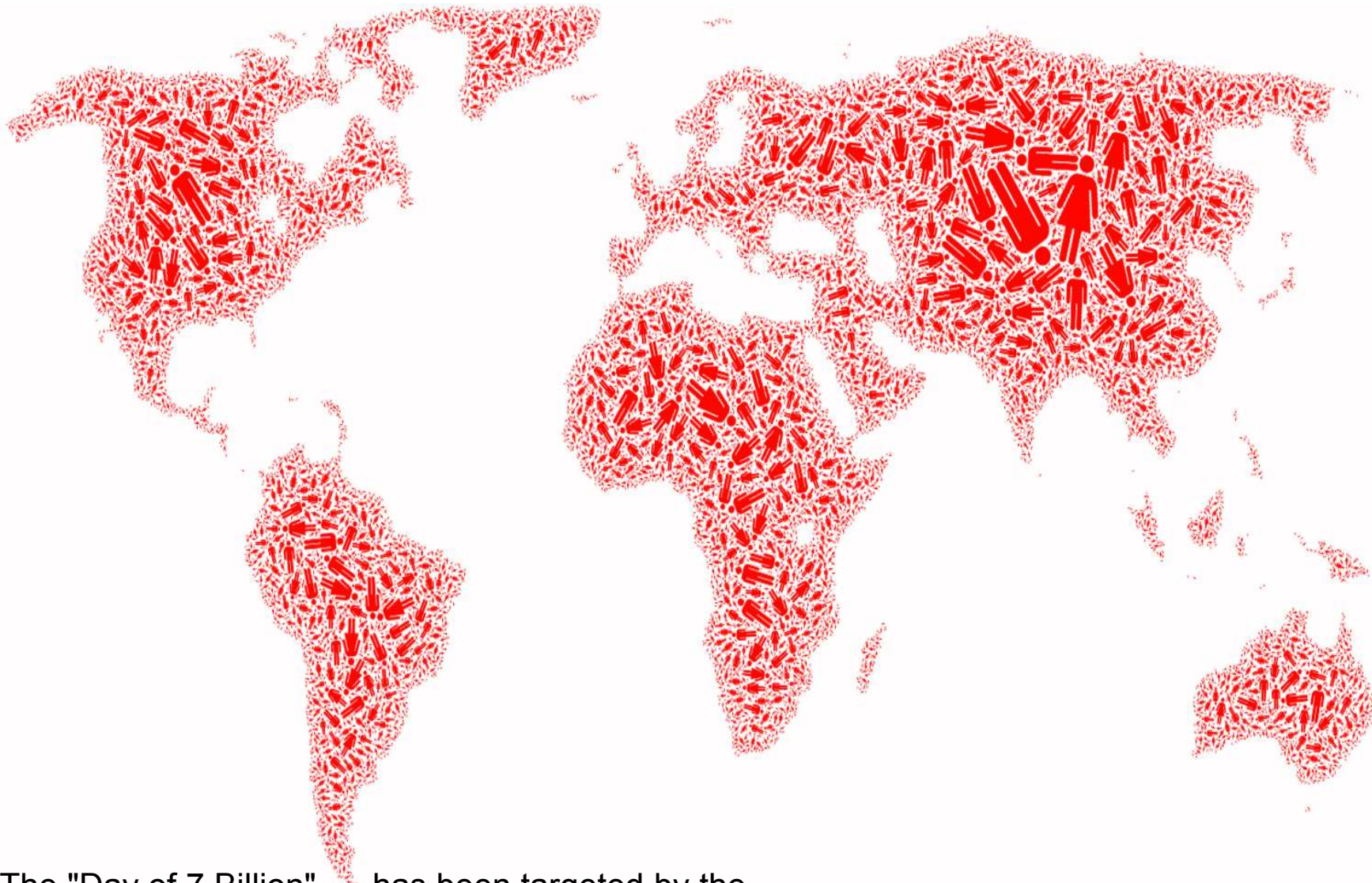
WISCONSIN

MISSISSIPPI

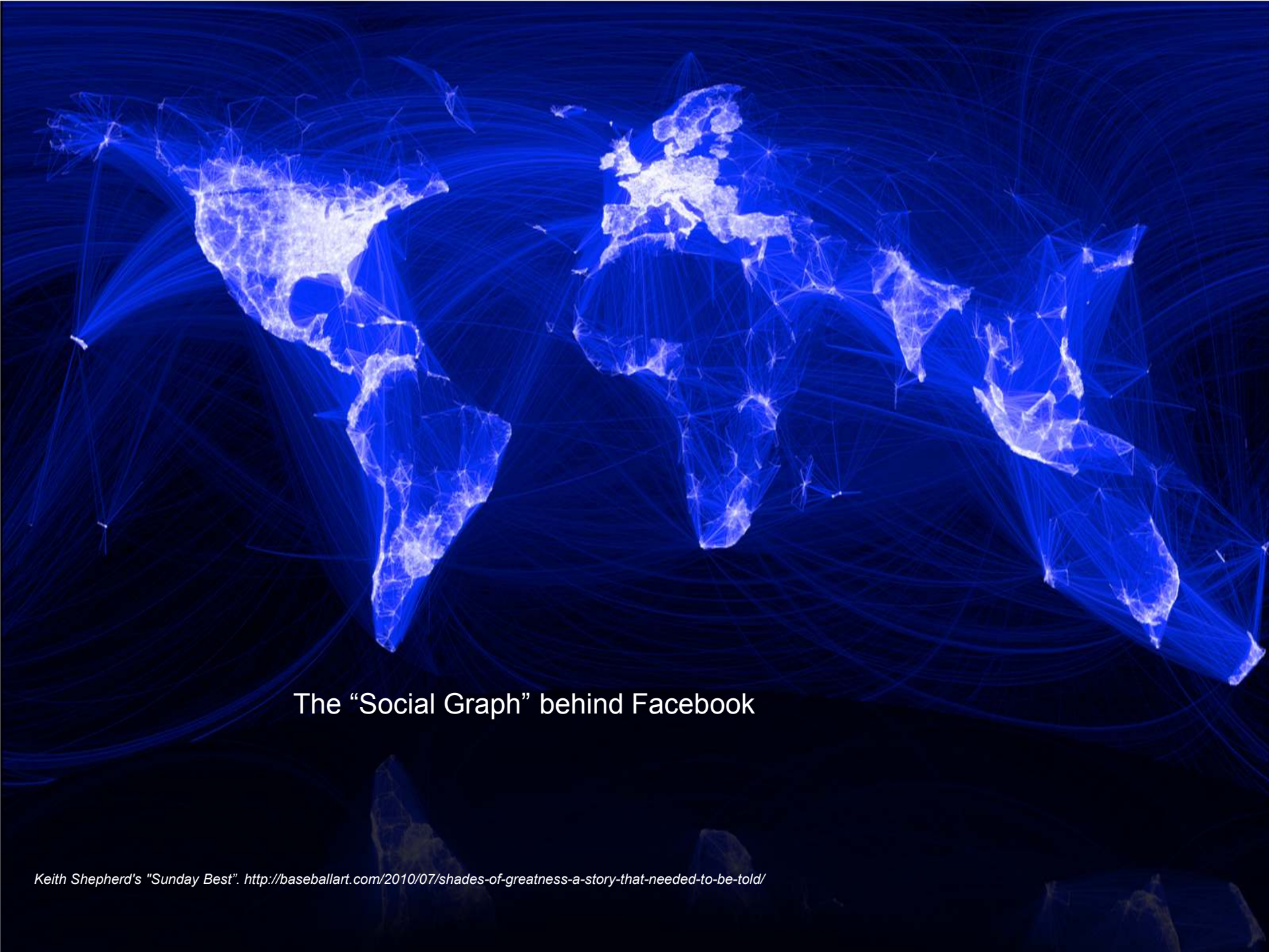
WOMEN OF LETTERS

WIKI

Social, informational,
technological, biological networks

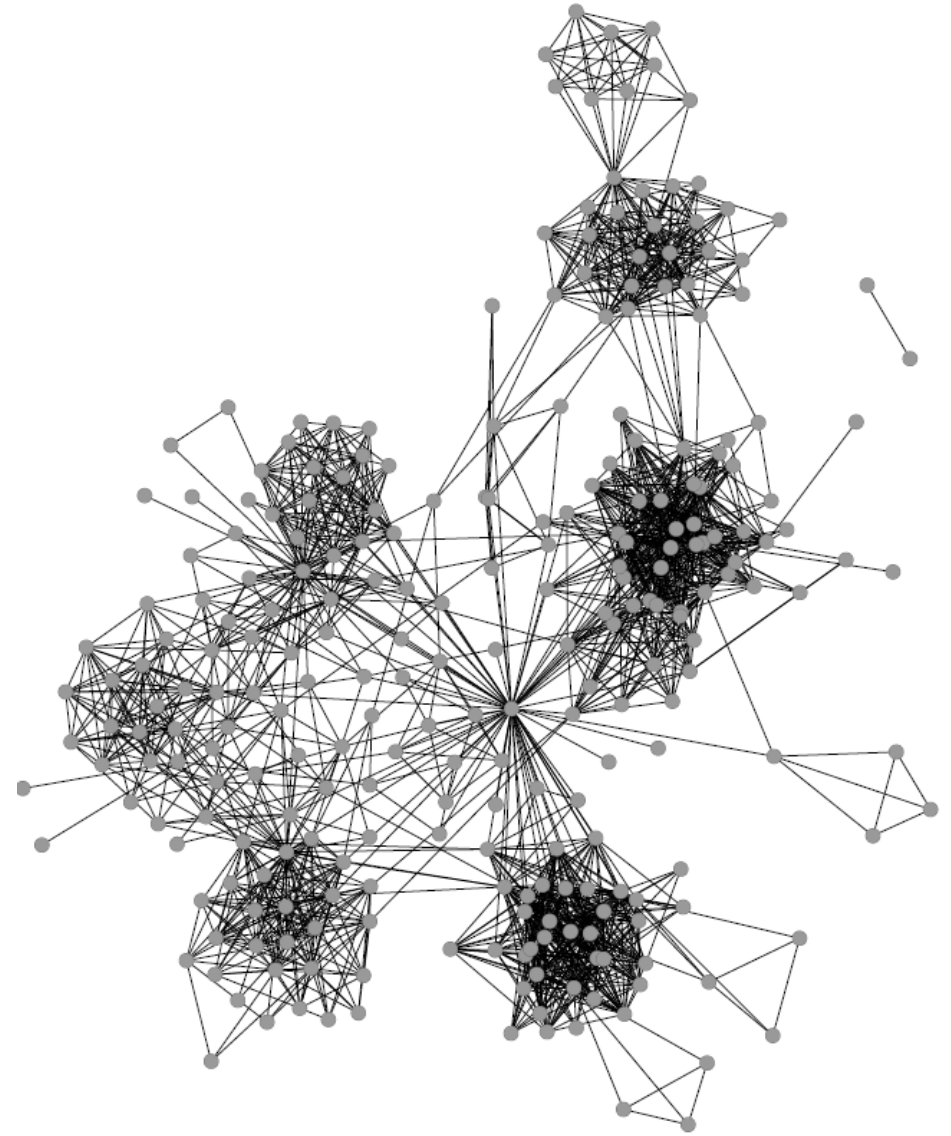
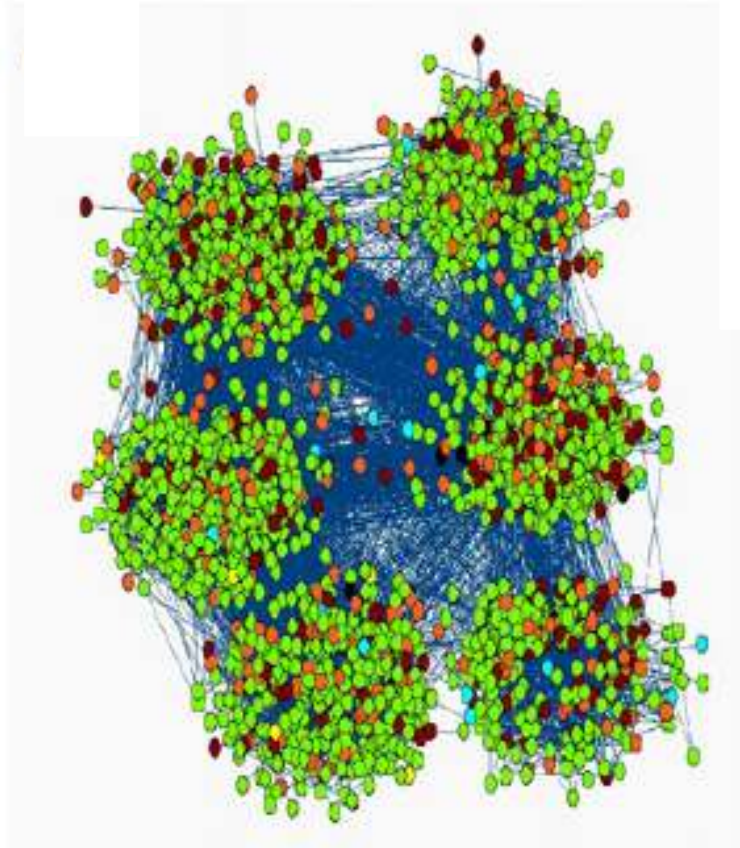


The "Day of 7 Billion" has been targeted by the United States Census Bureau to be in July 2012. Wrong! It was in October 2011



The “Social Graph” behind Facebook

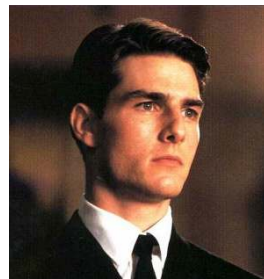
Keith Shepherd's "Sunday Best". <http://baseballart.com/2010/07/shades-of-greatness-a-story-that-needed-to-be-told/>



COLLABORATION NETWORKS: ACTOR NETWORK

Nodes: actors

Links: cast jointly



Days of Thunder (1990)
Far and Away (1992)
Eyes Wide Shut (1999)

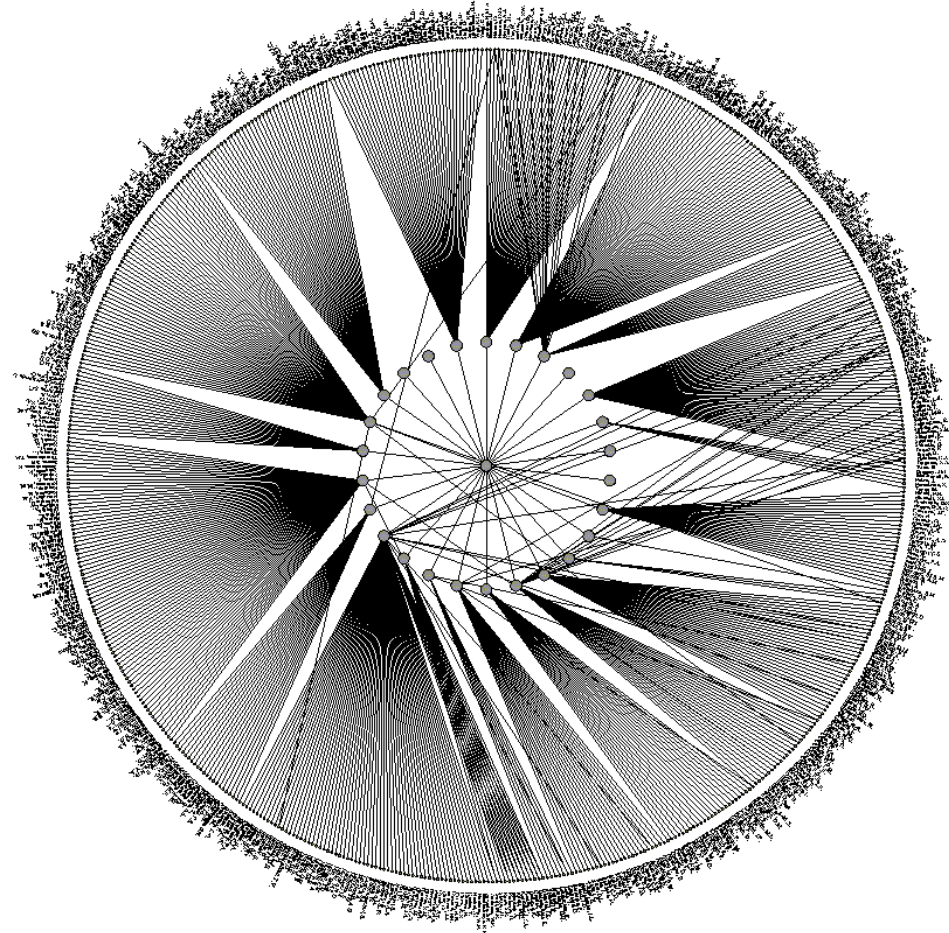


$N = 212,250$ actors $\langle k \rangle = 28.78$

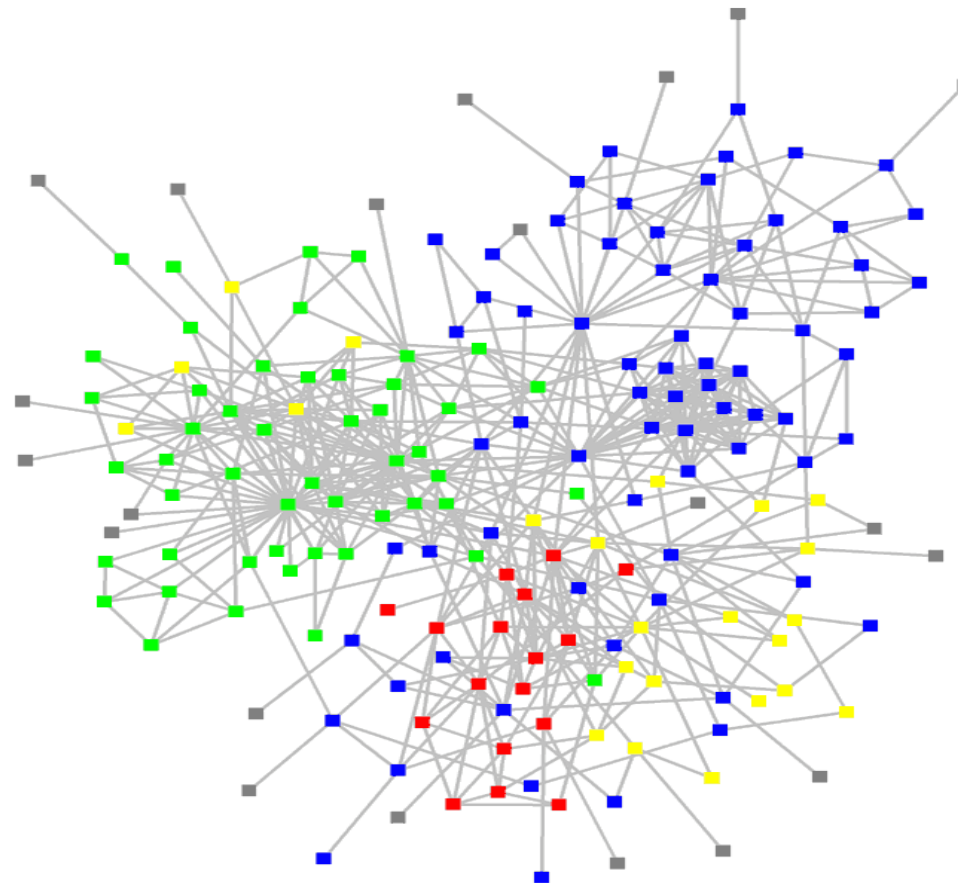
COLLABORATION NETWORKS: SCIENCE CO-AUTHORSHIP

Nodes: scientist (authors)

Links: write paper together



STRUCTURE OF AN ORGANIZATION



- ■ ■ : departments
- : consultants
- : external experts

www.orgnet.com

BUSINESS TIES IN US BIOTECH-INDUSTRY

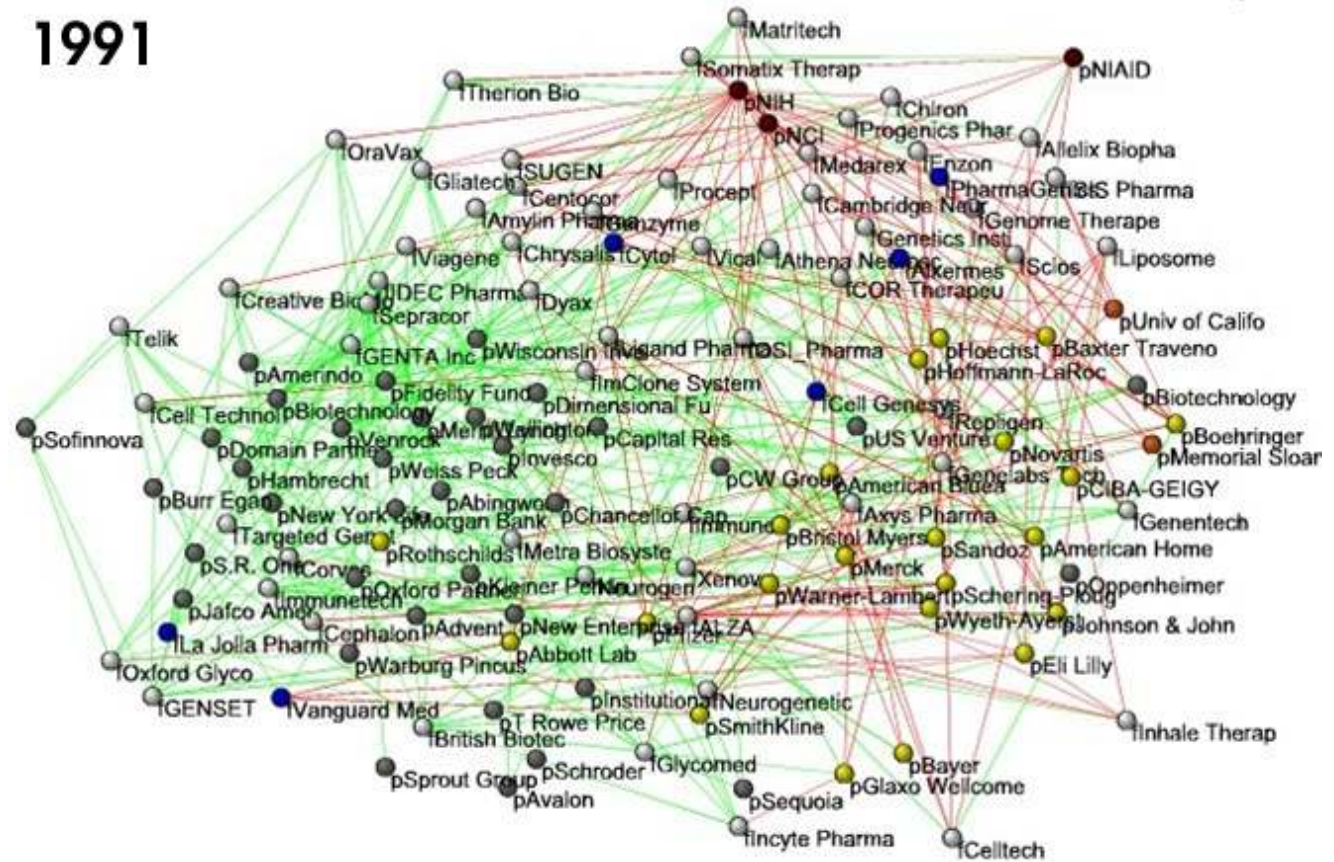
1991

Nodes:

- Companies
- Investment
- Pharma
- Research Labs
- Public
- Biotechnology

Links:

- Collaborations
- Financial
- R&D



<http://ecclectic.ss.uci.edu/~drwhite/Movie>

Information networks: the Web and Science Citation Indexes

1,000 Most Cited Physicists

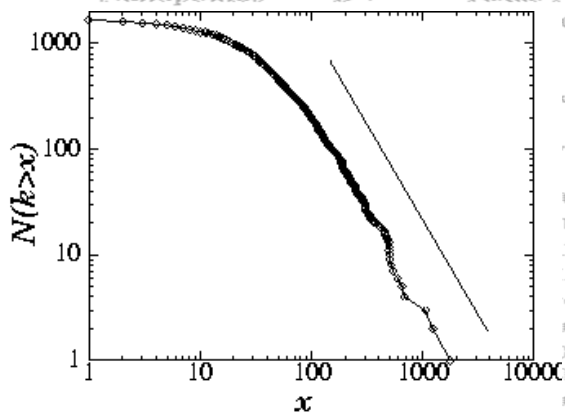
Out of over 500,000 E
(see <http://www.sst.tn>)

Author name	Institution	Country	Field
Witten	MIT (U)	USA, NJ	High
Gossard	UCSB (U)	USA, CA	Semi
Cava	MIT (U)	USA, NJ	Supe
Ballogg	MIT (U)	USA, NJ	Supe
Ploog	Max-Planck (NL)	Germany	Semi
	Nuclear Cent.	Switzerland	Astro
	State (U)	USA, FL	Solic
	Max-Planck (NL)	Germany	Semi
Nanopoulos	Texas A&M (U)	USA, TX	High
	(U)	USA, CA	Poly
	on (U)	USA, NJ	Solic
	estern (U)	USA, IL	S
	Univ. (U)	Switzerland	S
	bs (I)	USA, NJ	S
	I/NL)	USA, CA	C
	(U)	USA, IL	S
	d (U)	USA, CA	S
	n Univ. (U)	USA, TX	S
		Switzerland	S
	LBL (U/NL)	USA, CA	S
	n Univ. (U)	USA, TX	S
Waszczak	JV AT&T (I)	USA, NJ	S
Shirane	G Brookhaven (U)	USA, NY	S
Wiegmann	W Johns Hopkins (U)	USA, MD	S
Vandover	RB Bell Labs (I)	USA, NJ	M
Uchida*	S SRI (U)	USA, CA	S
Hor	PE SRI (U)	USA, TX	S
Murphy	DW SRI (U)	USA, CA	S
Birgeneau	RJ MIT (U)	USA, MA	S
Jorgensen	JD Argonne (NL)	USA, IL	S
Hinks	DG Argonne (NL)	USA, IL	S

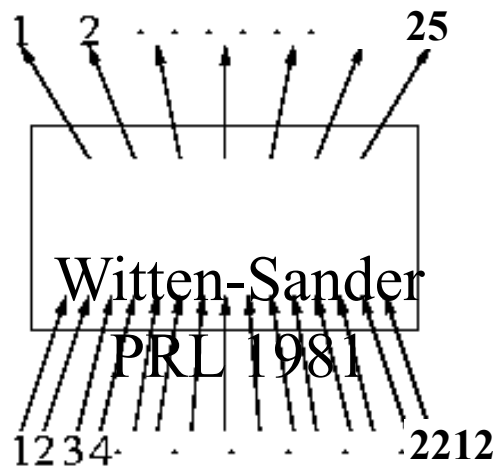
Nodes: papers

Links: citations

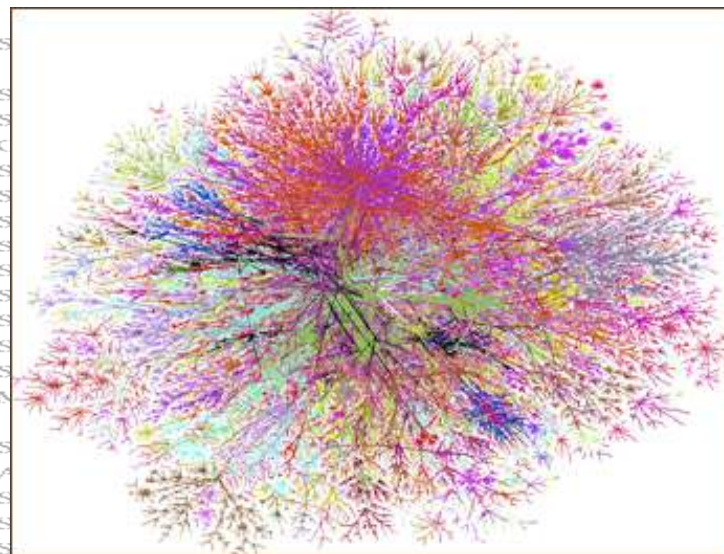
1736 PRL papers (1988)



Nodes: web pages
Links: ditto ;-)

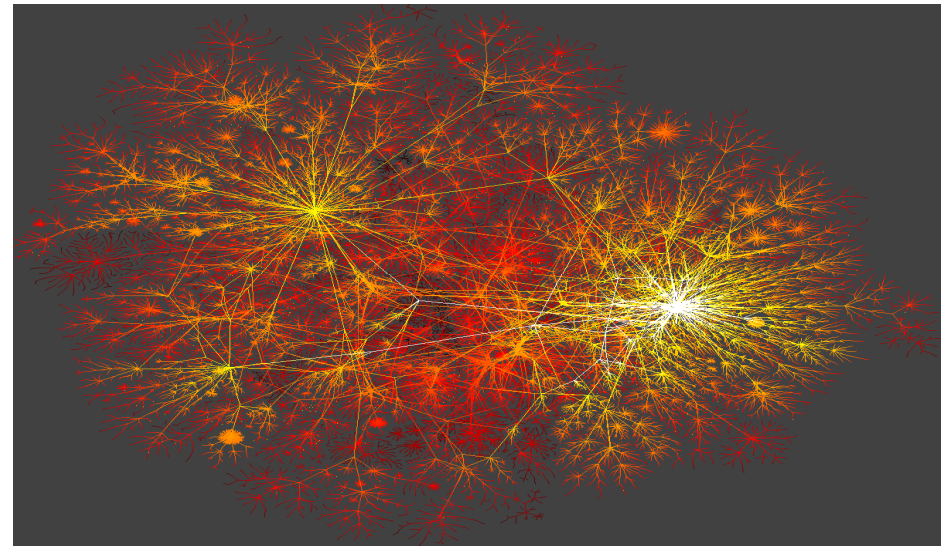
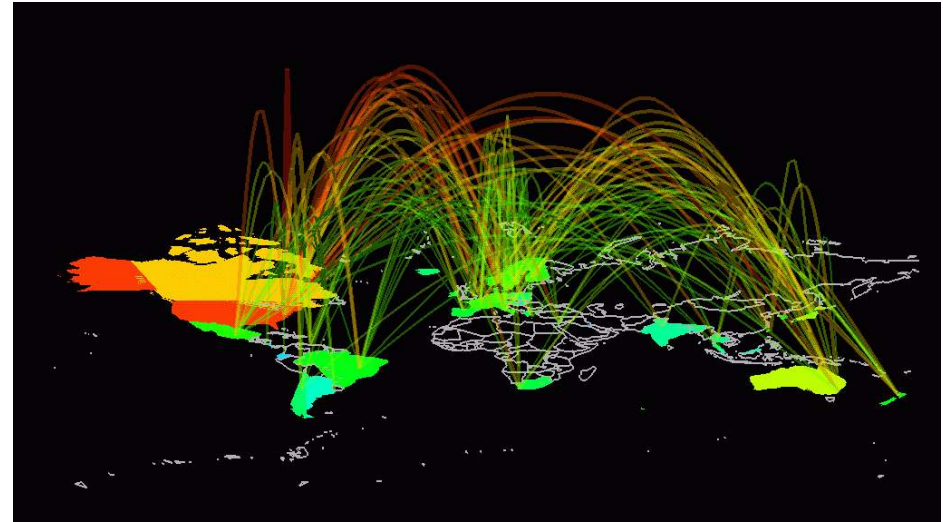
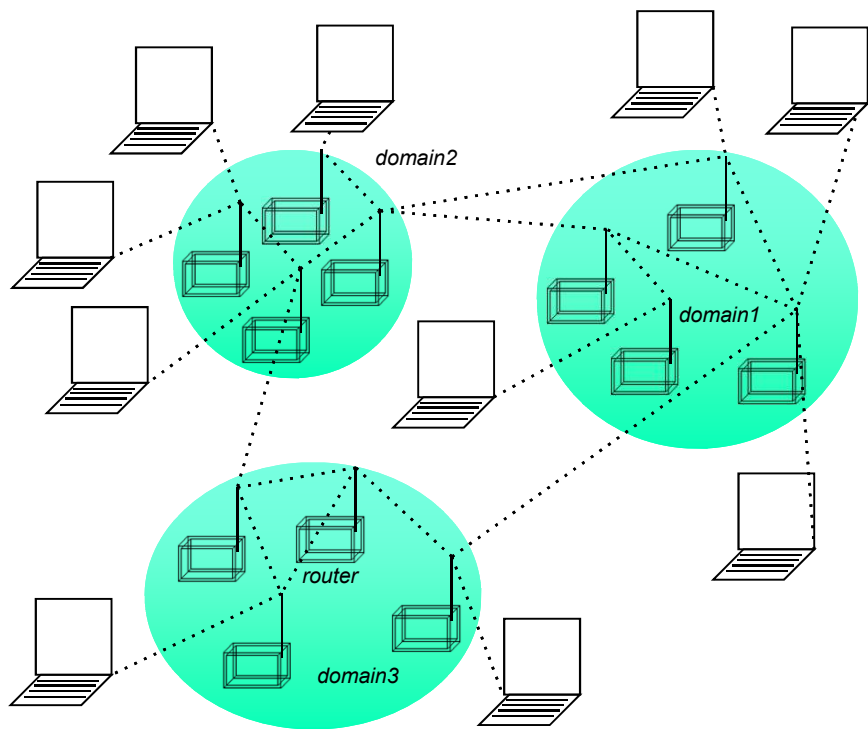


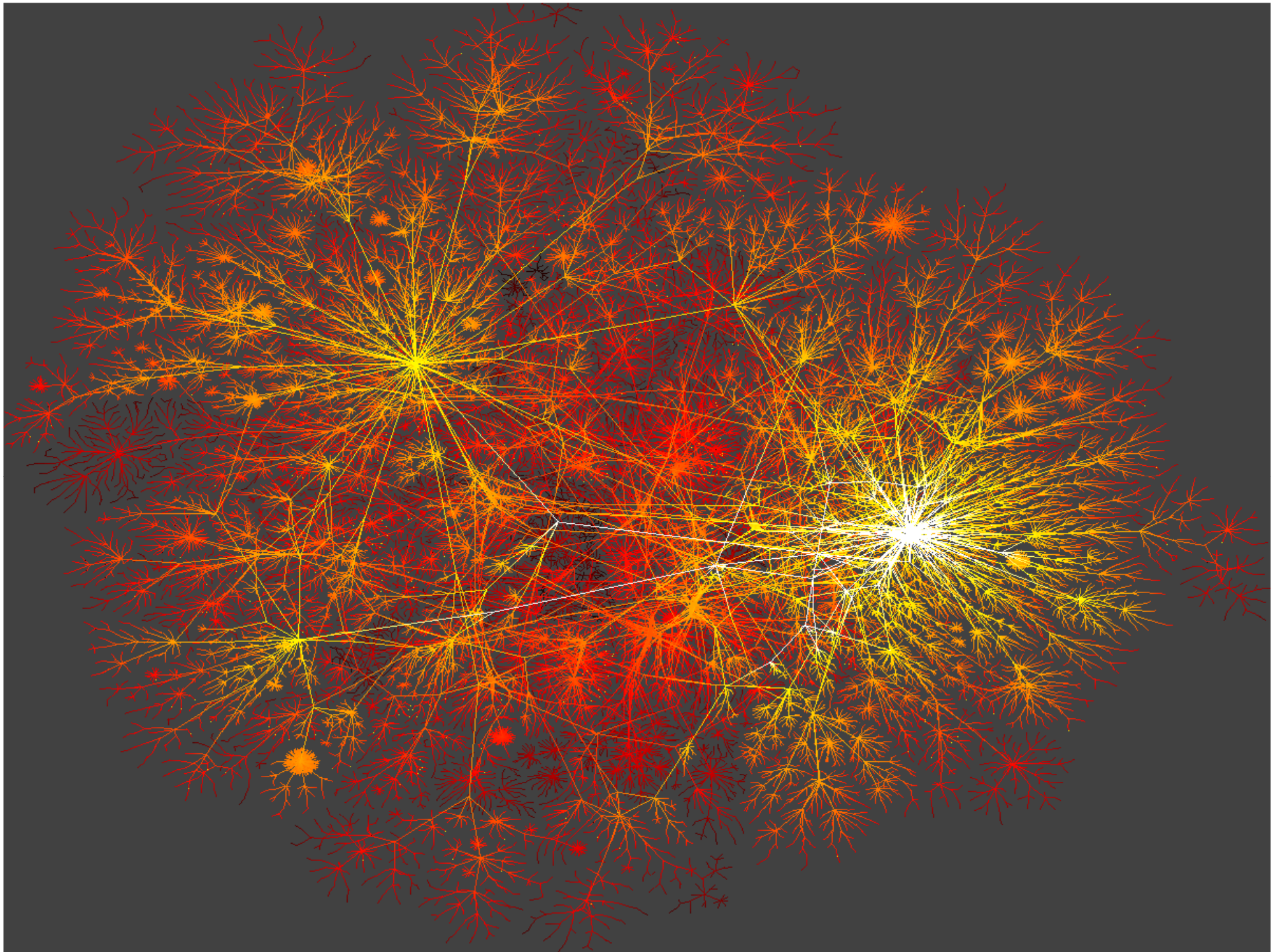
rank by total cit.
1
2
3
4
5
6
7
8
9
10
11
12
13



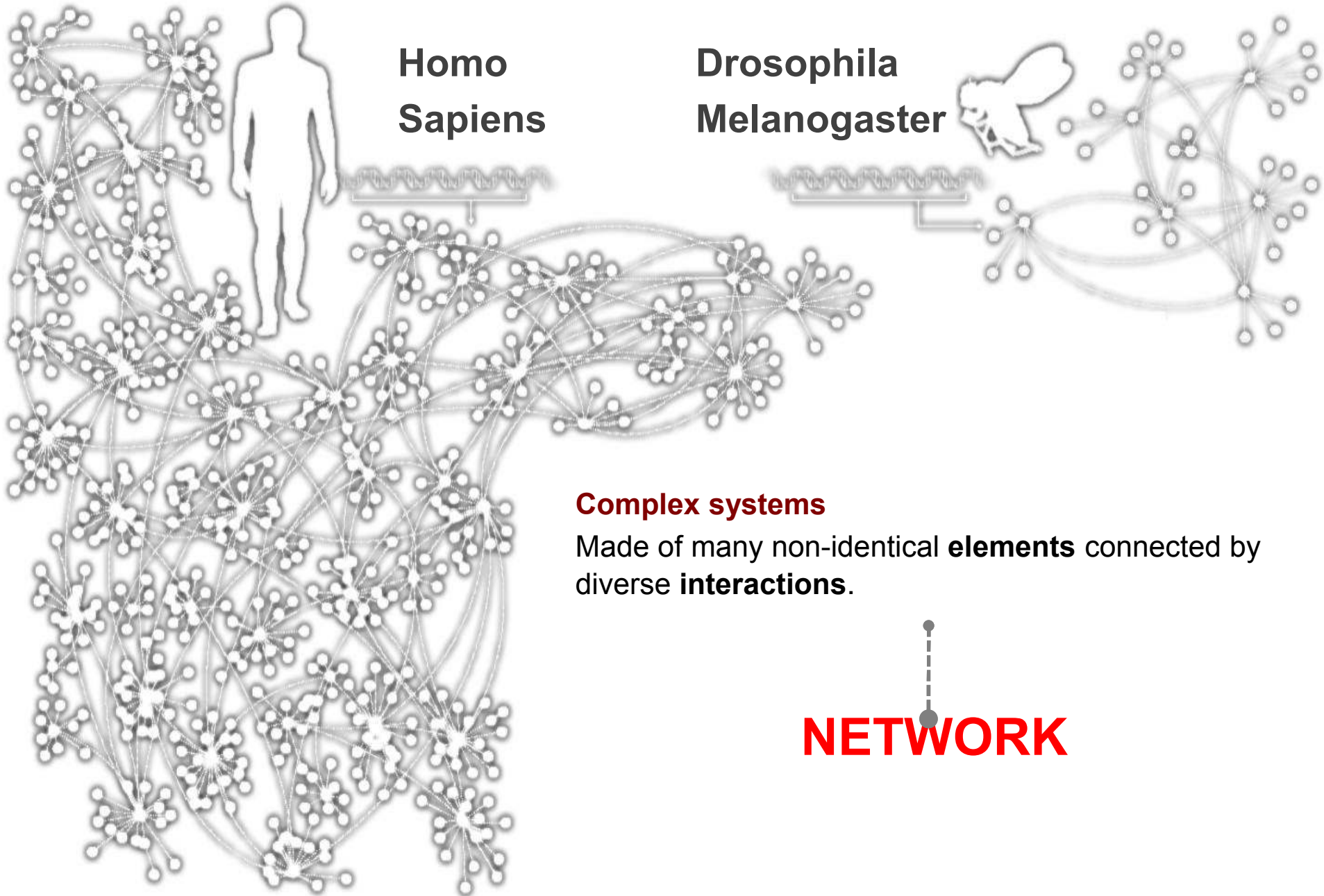
* citation total may be skewed because of multiple authors with the same name

INTERNET



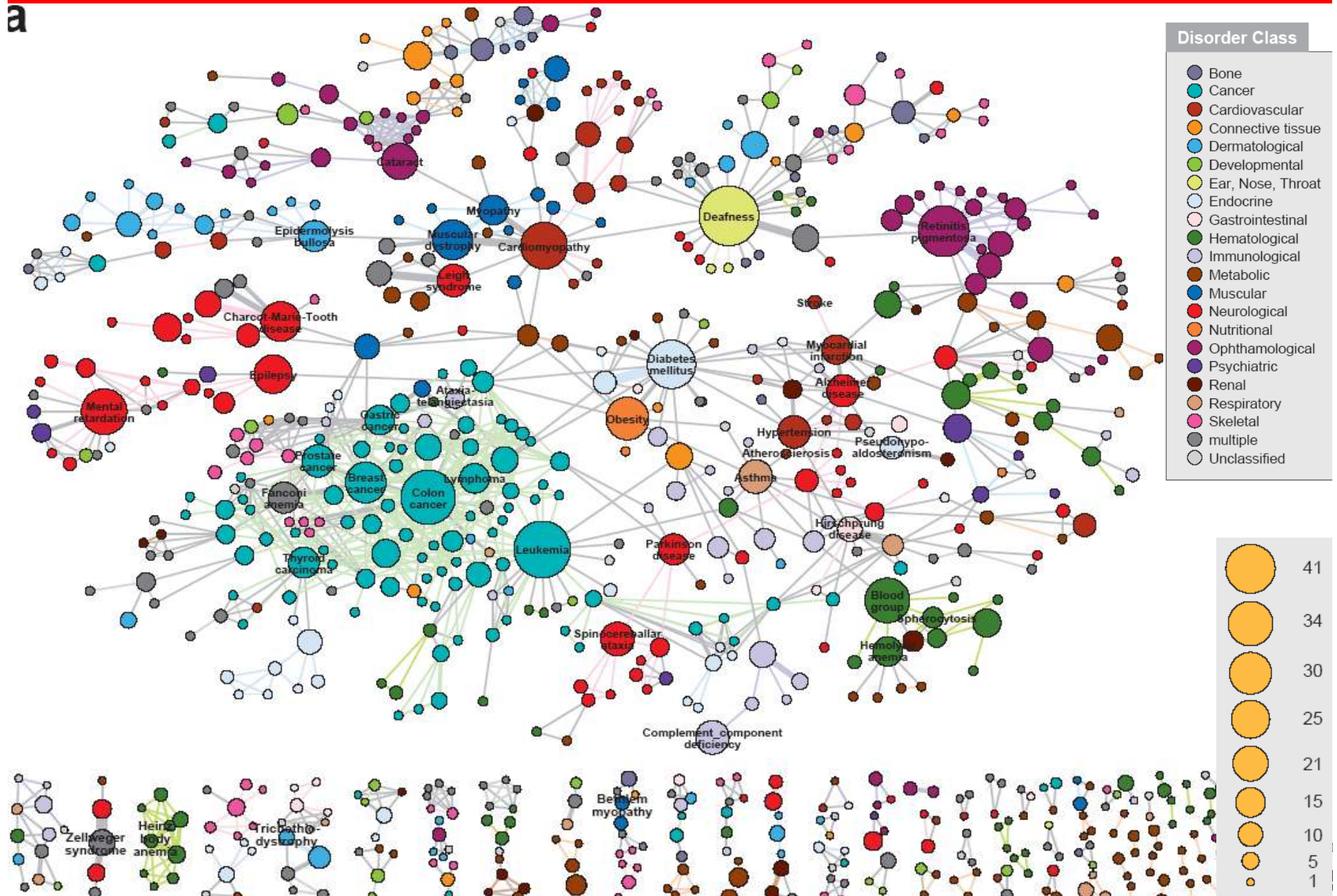


HUMANS GENES



HUMAN DISEASE NETWORK

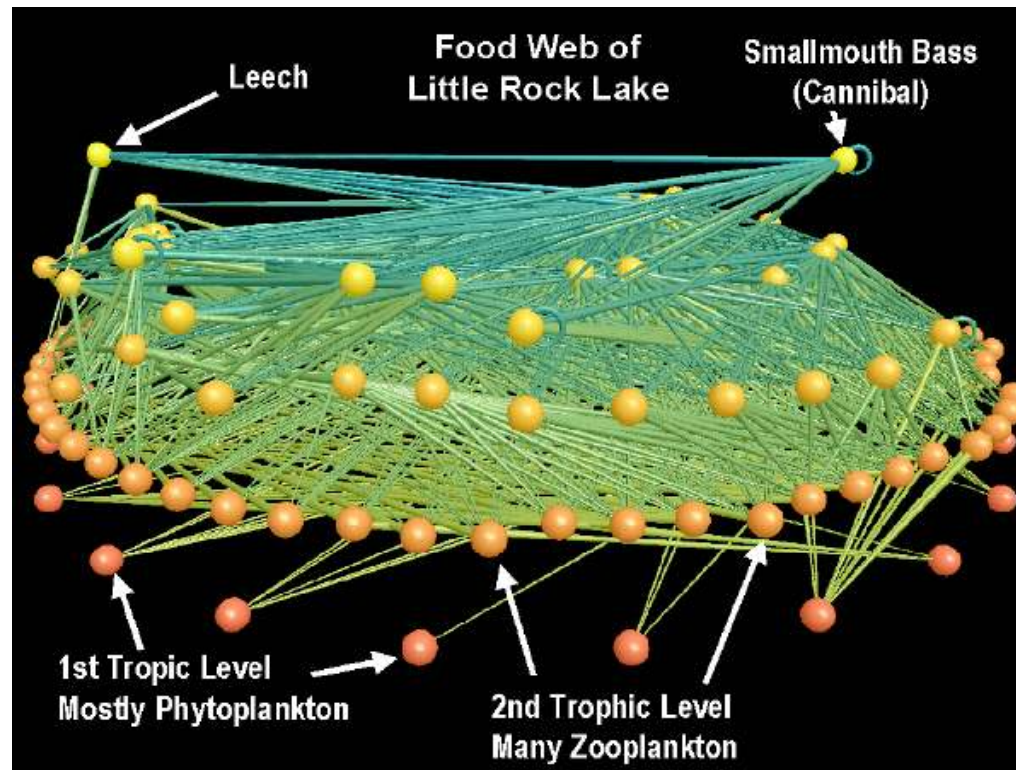
a



Biological networks: Food Web

Nodes: species

Links: trophic interactions



R. Sole (cond-mat/0011195)

R.J. Williams, N.D. Martinez *Nature* (2000)

Basic network measures

Degree of a node

Distance between two nodes

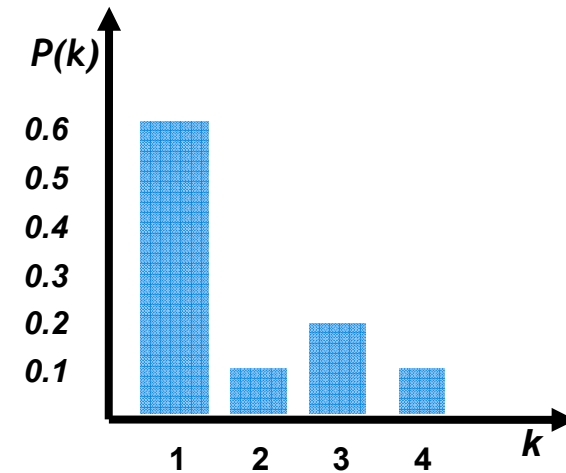
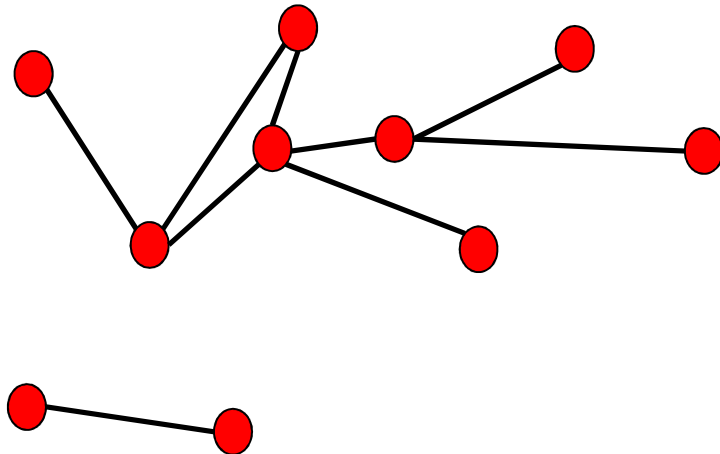
Clustering among three nodes

DEGREE DISTRIBUTION

Degree distribution $P(k)$: probability that a randomly chosen vertex has degree k

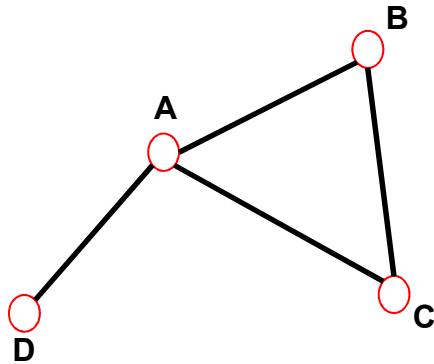
$N_k = \#$ nodes with degree k

$P(k) = N_k / N \rightarrow$ plot



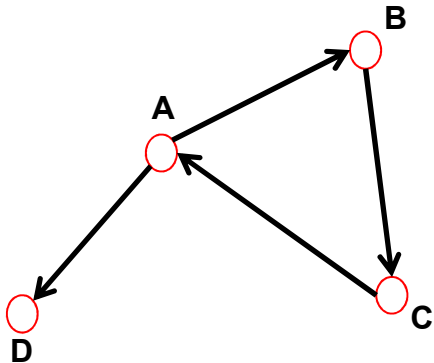
DISTANCE IN A GRAPH

Shortest Path, Geodesic Path



The *distance (shortest path, geodesic path)* between two nodes is defined as the **number of edges along the shortest path connecting them.**

*If the two nodes are disconnected, the distance is infinity.



In **directed graphs** each path needs to follow the direction of the arrows.

Thus in a digraph the distance from node A to B (on an AB path) is generally different from the distance from node B to A (on a BCA path).

NETWORK DIAMETER AND AVERAGE DISTANCE

Diameter: the maximum distance between any pair of nodes in the graph.

Average path length/distance for a **connected graph** (component) or a **strongly connected** (component of a) **digraph**.

where l_{ij} is the distance from node i to node j

$$\langle l \rangle \equiv \frac{1}{2L_{\max}} \sum_{i,j \neq i} l_{ij}$$

In an undirected (symmetrical) graph $l_{ij} = l_{ji}$, we only need to count them once

$$\langle l \rangle \equiv \frac{1}{L_{\max}} \sum_{i,j > i} l_{ij}$$

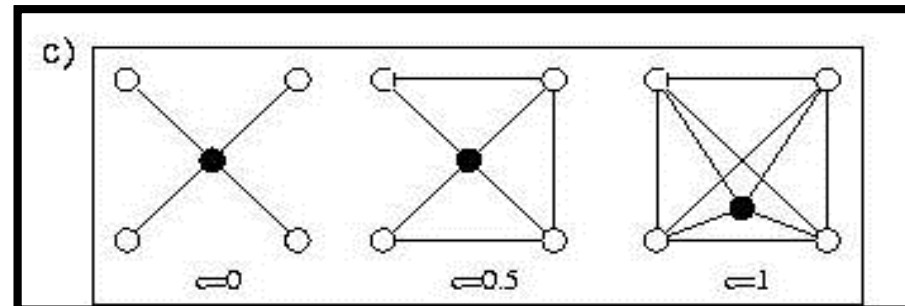
CLUSTERING COEFFICIENT

* Clustering coefficient:

what portion of your neighbors are connected?

- * Node i with degree k_i
- * C_i in $[0,1]$

$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$

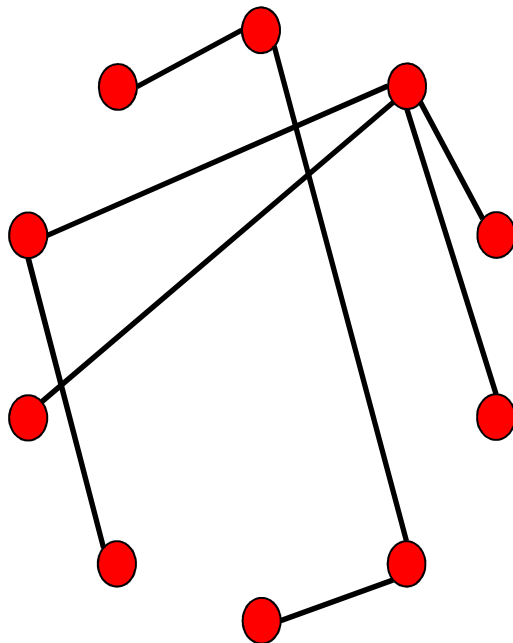
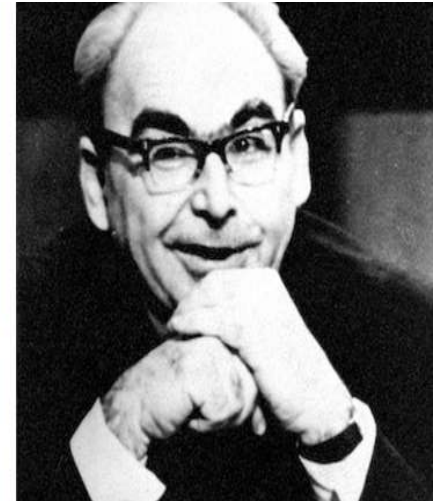


Random graphs

What are the expected basic measures emerging from random?

RANDOM NETWORK MODEL

Pául Erdős
(1913-1996)



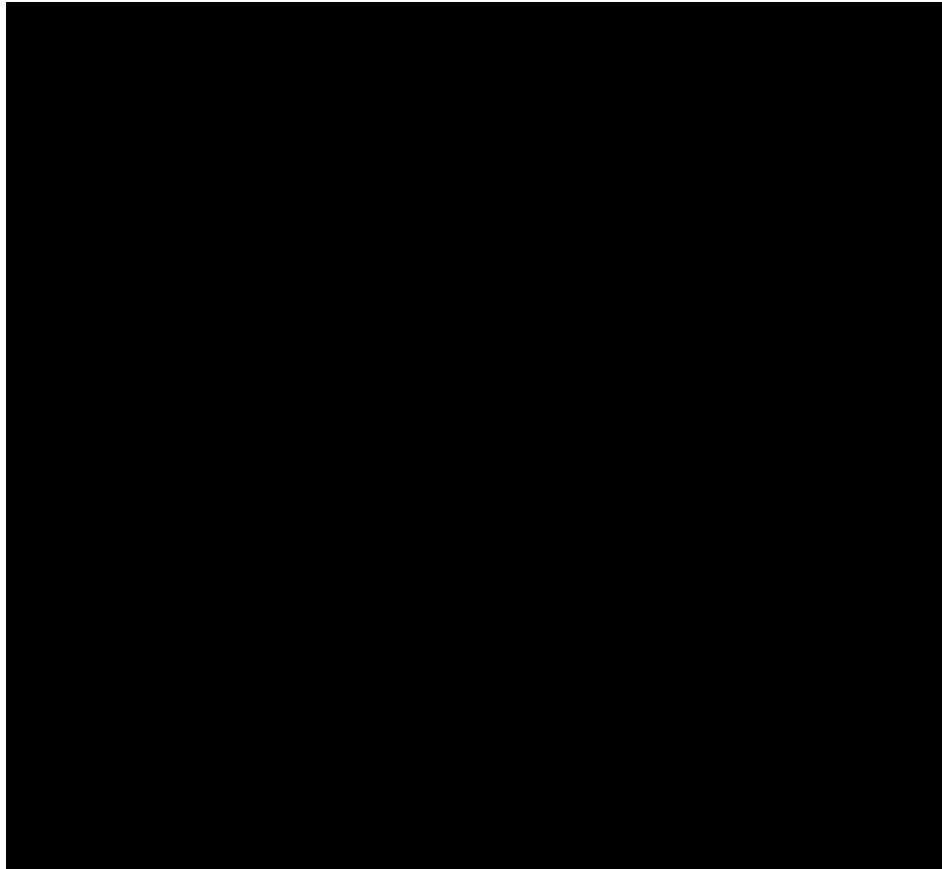
Erdős-Rényi model (1960)

Connect with probability p

$$p = 1/6 \quad N = 10$$

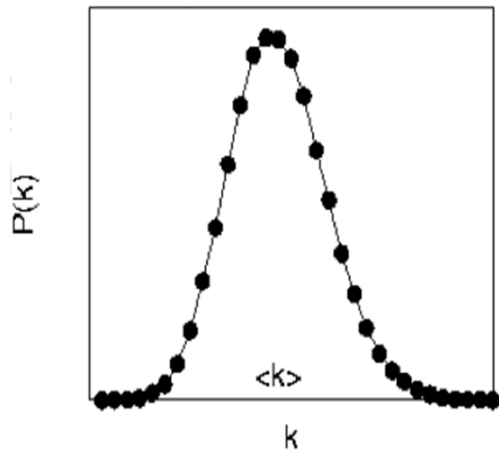
$$\langle k \rangle \sim 1.5$$

RANDOM NETWORK MODEL



Definition: A **random graph** is a graph of N labeled nodes where each pair of nodes is connected by a preset probability p .

DEGREE DISTRIBUTION OF A RANDOM GRAPH



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

Select k
nodes from $N-1$

probability of
having k edges

probability of
missing $N-1-k$
edges

$$\langle k \rangle = p(N-1)$$

$$\sigma_k^2 = p(1-p)(N-1)$$

$$\frac{\sigma_k}{\langle k \rangle} = \left[\frac{1-p}{p} \frac{1}{(N-1)} \right]^{1/2} \approx \frac{1}{(N-1)^{1/2}}$$

As the network size increases, the distribution becomes increasingly narrow—we are increasingly confident that the degree of a node is in the vicinity of $\langle k \rangle$.

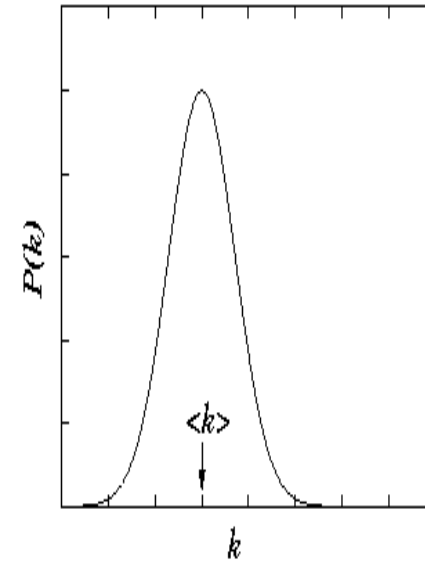
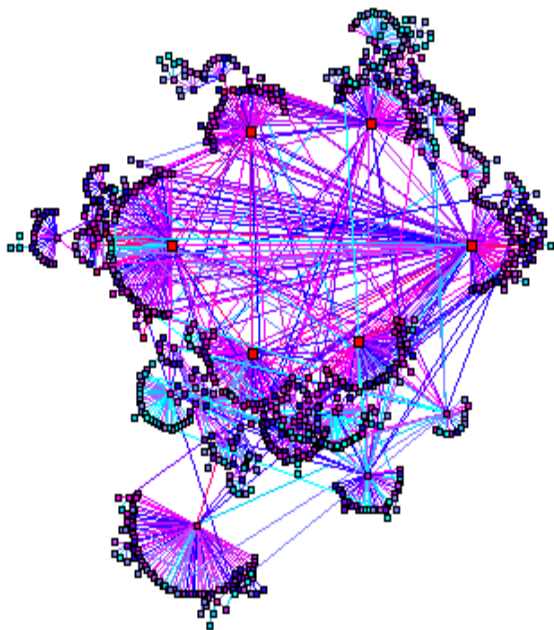
WORLD WIDE WEB

Nodes: **WWW documents**

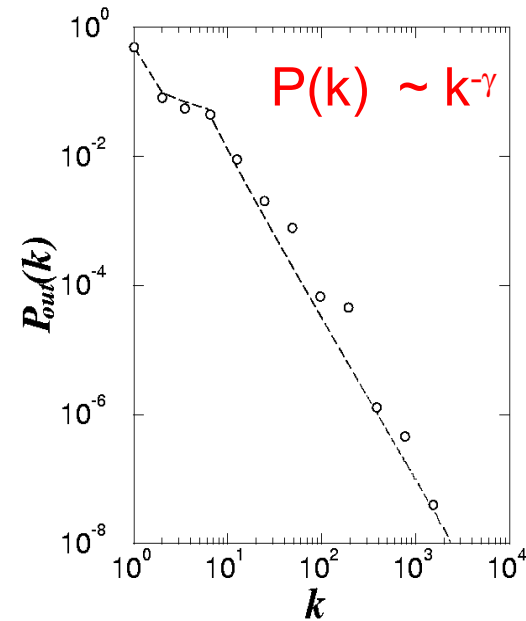
Links: **URL links**

Over 3 billion documents

ROBOT: collects all URL's found in a document and follows them recursively

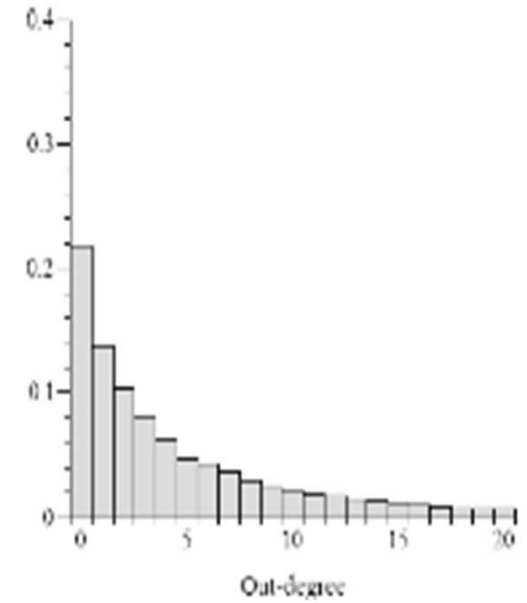
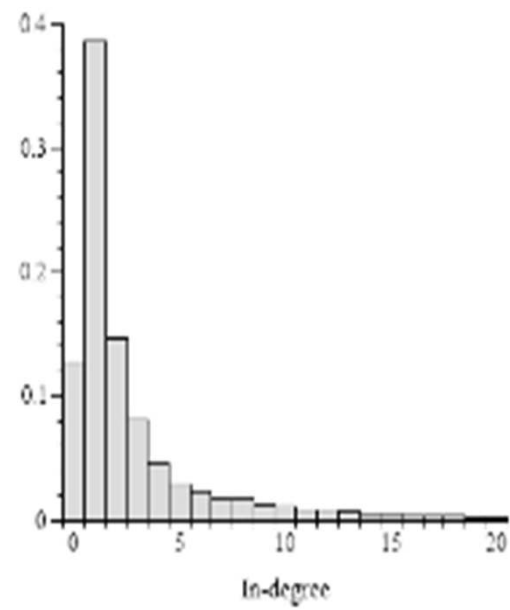
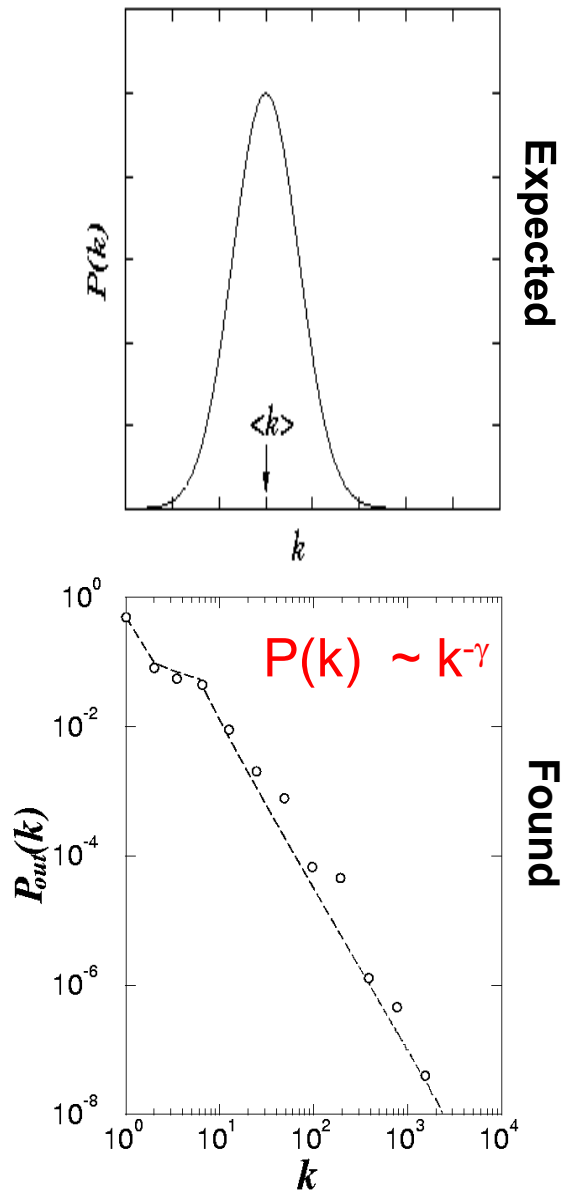


Expected



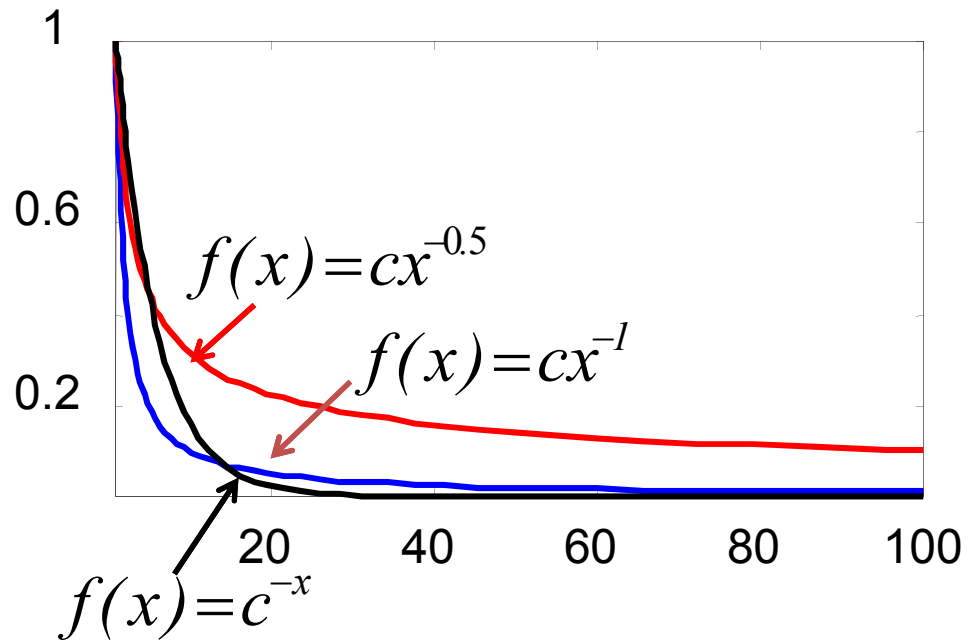
Found

Degree distribution of the WWW



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

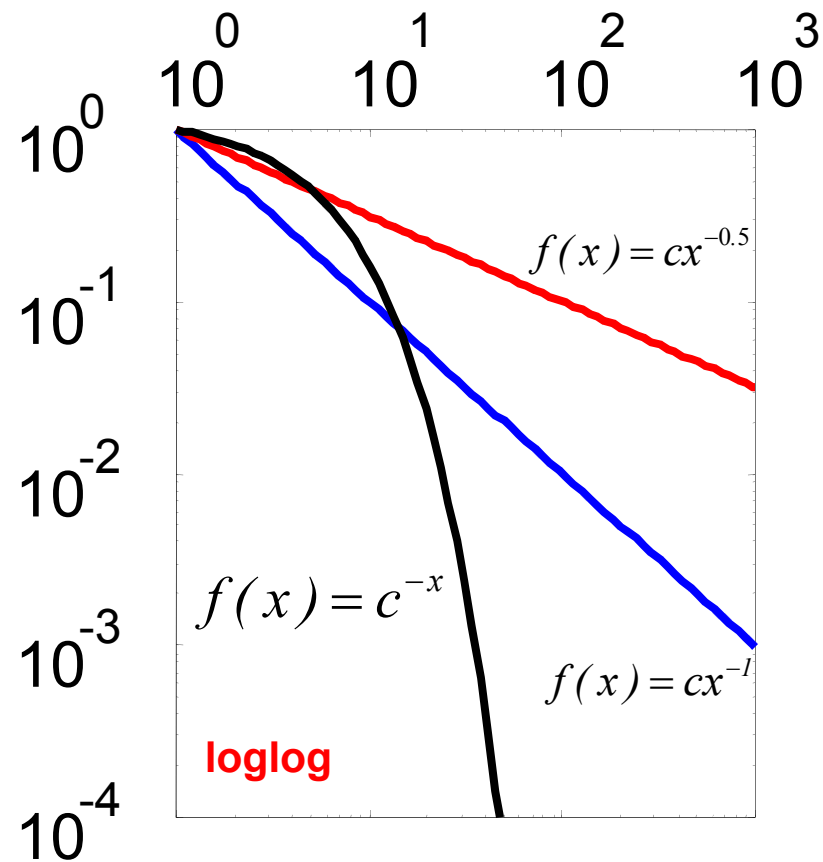
The difference between a power law and an exponential distribution



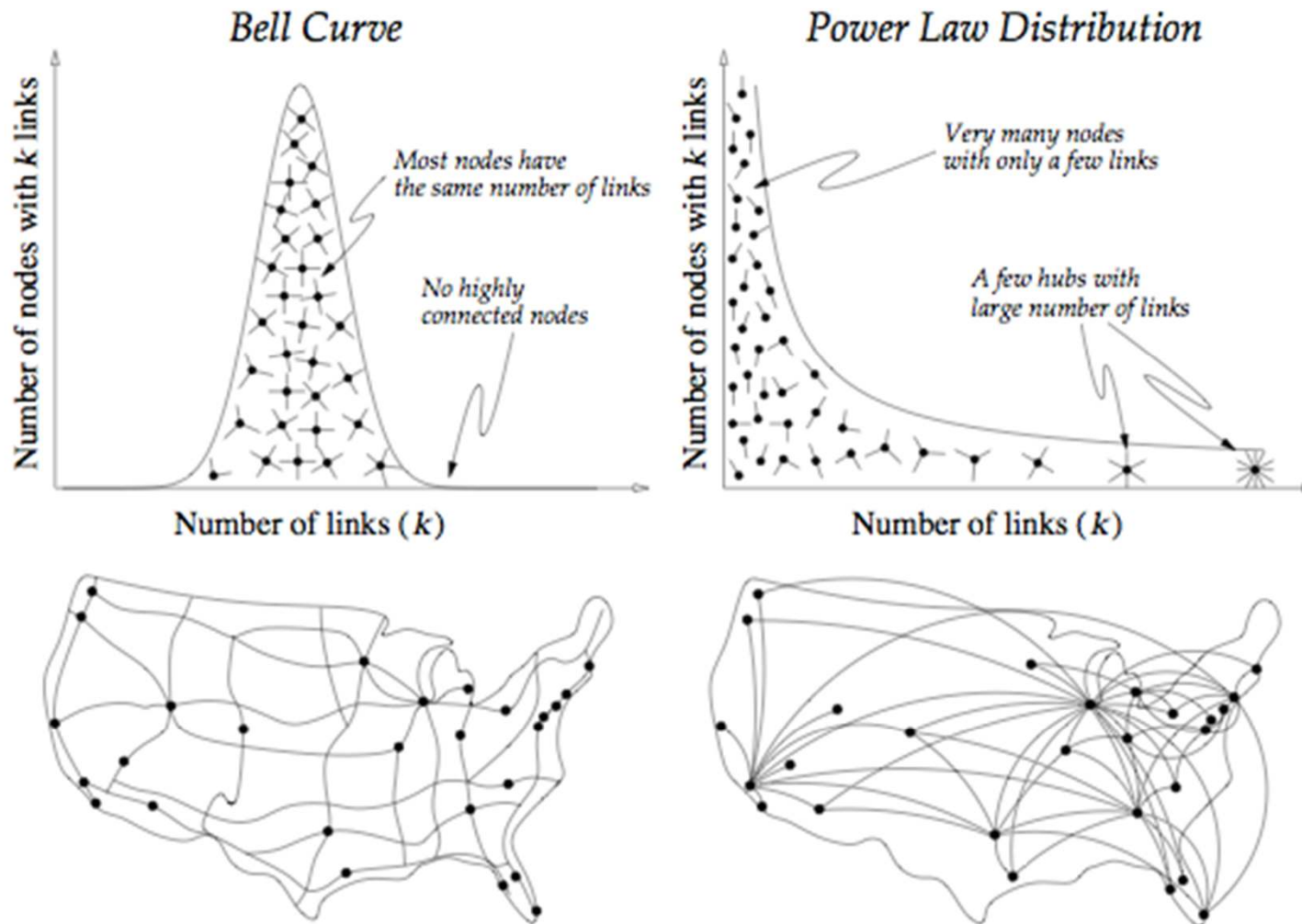
Above a certain x value, the power law is always higher than the exponential.

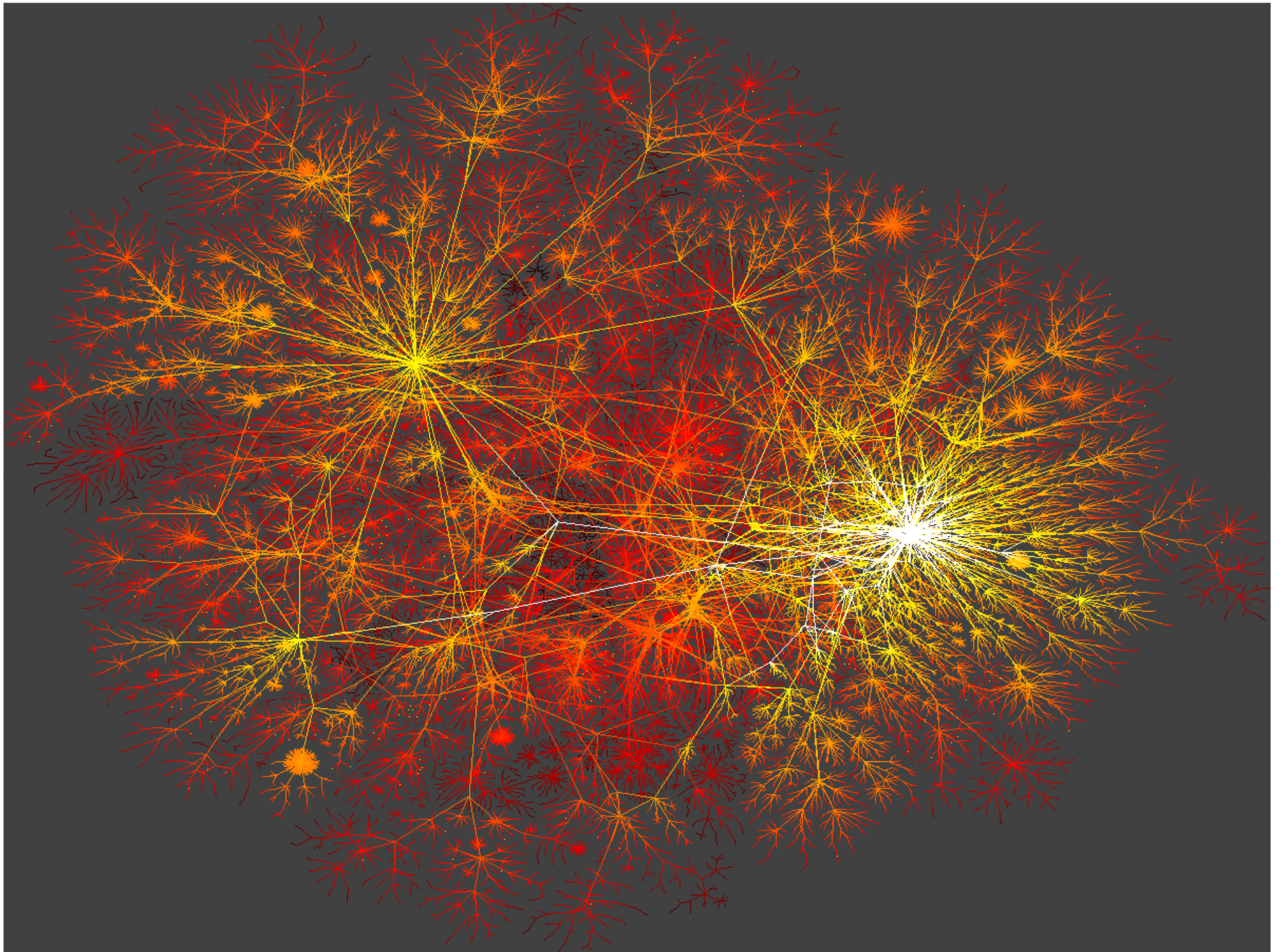
The difference between a power law and an exponential distribution

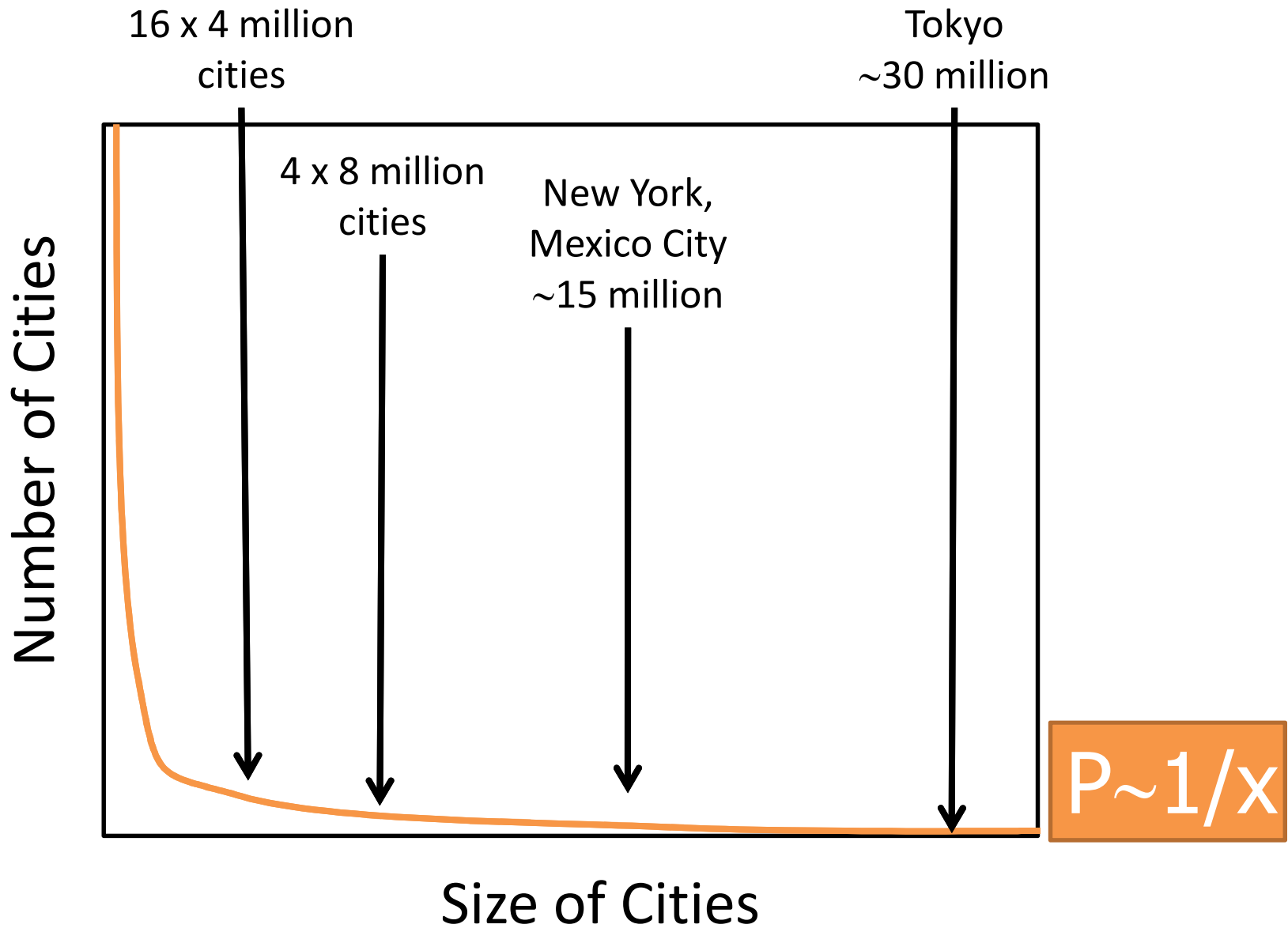
This difference is particularly obvious if we plot them on a log vertical scale: for large x there are orders of magnitude differences between the two functions.



Exponential vs Power law distributions







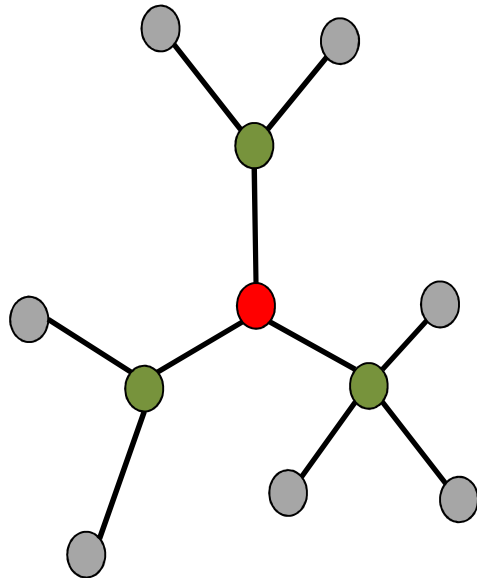
After Bill enters the arena the average income of the public ~ USD \$1,000,000

~ \$50 billion



DISTANCES IN RANDOM GRAPHS

Random graphs tend to have a tree-like topology with almost constant node degrees.



- nr. of first neighbors:

$$N_1 \cong \langle k \rangle$$

- nr. of second neighbors:

$$N_2 \cong \langle k \rangle^2$$

- nr. of neighbours at distance d:

$$N_d \cong \langle k \rangle^d$$

- estimate maximum distance:

$$1 + \sum_{l=1}^{l_{max}} \langle k \rangle^l = N \quad \Rightarrow \quad l_{max} = \frac{\log N}{\log \langle k \rangle}$$

DISTANCES IN RANDOM GRAPHS

compare with real data

$$l_{max} = \frac{\log N}{\log \langle k \rangle}$$

Network	Size	(k)	l	l _{rand}	C	C _{rand}	Reference	Nr
www, site level, undir	153127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015-6209	3.52-4.11	3.7-3.76	6.36-6.18	0.18-0.3	0.001	Yook e al., 2001a, Pastor-Satorras et al., 2001	2
Movie actors	225226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52909	9.7	5.9	4.79	0.43	1.8 x 10 ⁻⁴	Newman, 2001a, 2001b, 2001c	4
MEDLINE eo-authorship	1520251	18.1	4.6	4.91	0.066	1.1 x 10 ⁻⁵	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11994	3.59	9.7	7.34	0.496	3 x 10 ⁻⁴	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70975	3.9	9.5	8.2	0.59	5.4 x 10 ⁻⁵	Barabasi et al, 2001	8
Neurosci. co-authorship	209293	11.5	6	5.01	0.76	5.5 x 10 ⁻⁵	Barabasi et al, 2001	9
E. coli, sustrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Sole, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Sole, 2000	13
Words, co-occurrence	460902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Sole, 2001	14
Words, synonyms	22311	13.48	4.5	3.84	0.7	0.0006	Yook et al. 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C.Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

Given the huge differences in scope, size, and average degree, the agreement is excellent.

CLUSTERING COEFFICIENT

$$C_i \equiv \frac{2n_i}{k_i(k_i - 1)}$$

Since edges are independent and have the same probability p ,

$$n_i \cong p \frac{k_i(k_i - 1)}{2} \quad \Rightarrow \quad C \cong p = \frac{\langle k \rangle}{N}$$

The clustering coefficient of random graphs is small.

For fixed degree C decreases with the system size N .

CLUSTERING IN RANDOM GRAPHS

compare with real data

Network	Size	(k)	l	l_{rand}	C	C_{rand}	Reference	Nr
www, site level, undir	153127	35.21	3.1	3.35	0.1078	0.00023	Adamic, 1999	1
Internet, domain level	3015-6209	3.52-4.11	3.7-3.76	6.36-6.18	0.18-0.3	0.001	Yook et al., 2001a, Pastor-Satorras et al., 2001	2
Movie actors	225226	61	3.65	2.99	0.79	0.00027	Watts and Strogatz, 1998	3
LANL co-authorship	52909	9.7	5.9	4.79	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c	4
MEDLINE eo-authorship	1520251	18.1	4.6	4.91	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c	5
SPIRES co-authorship	56627	173	4.0	2.12	0.726	0.003	Newman, 2001a, 2001b, 2001c	6
NCSTRL co-authorship	11994	3.59	9.7	7.34	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c	7
Math. co-authorship	70975	3.9	9.5	8.2	0.59	5.4×10^{-5}	Barabasi et al, 2001	8
Neurosci. co-authorship	209293	11.5	6	5.01	0.76	5.5×10^{-5}	Barabasi et al, 2001	9
E. coli, sustrate graph	282	7.35	2.9	3.04	0.32	0.026	Wagner and Fell, 2000	10
E. coli, reaction graph	315	28.3	2.62	1.98	0.59	0.09	Wagner and Fell, 2000	11
Ythan estuary food web	134	8.7	2.43	2.26	0.22	0.06	Montoya and Sole, 2000	12
Silwood Park food web	154	4.75	3.40	3.23	0.15	0.03	Montoya and Sole, 2000	13
Words, co-occurrence	460902	70.13	2.67	3.03	0.437	0.0001	Ferrer i Cancho and Sole, 2001	14
Words, synonyms	22311	13.48	4.5	3.84	0.7	0.0006	Yook et al. 2001b	15
Power grid	4941	2.67	18.7	12.4	0.08	0.005	Watts and Strogatz, 1998	16
C.Elegans	282	14	2.65	2.25	0.28	0.05	Watts and Strogatz, 1998	17

**Are real networks like
random graphs?
NO!**

ARE REAL NETWORKS LIKE RANDOM GRAPHS? NO!

As quantitative data about real networks became available, we can compare their topology with the predictions of random graph theory.

Note that once we have N and $\langle k \rangle$ for a random network, from it we can derive every measurable property. Indeed, we have:

Average path length:

$$\langle l_{rand} \rangle \approx \frac{\log N}{\log \langle k \rangle}$$



Clustering Coefficient:

$$C_{rand} = p = \frac{\langle k \rangle}{N}$$

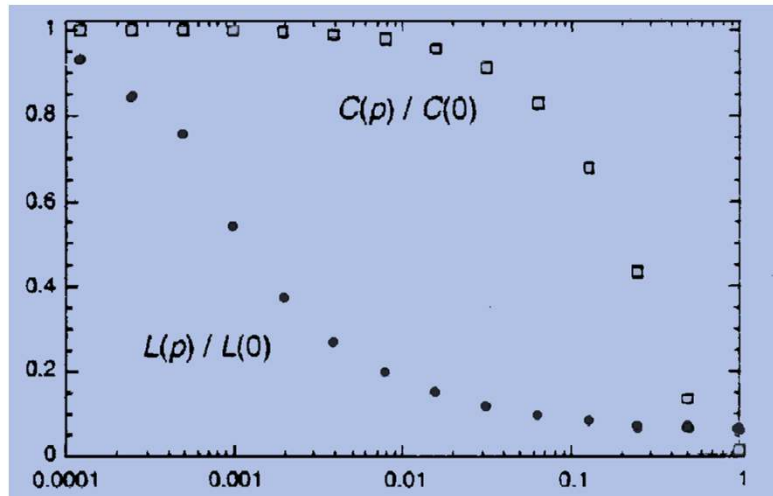
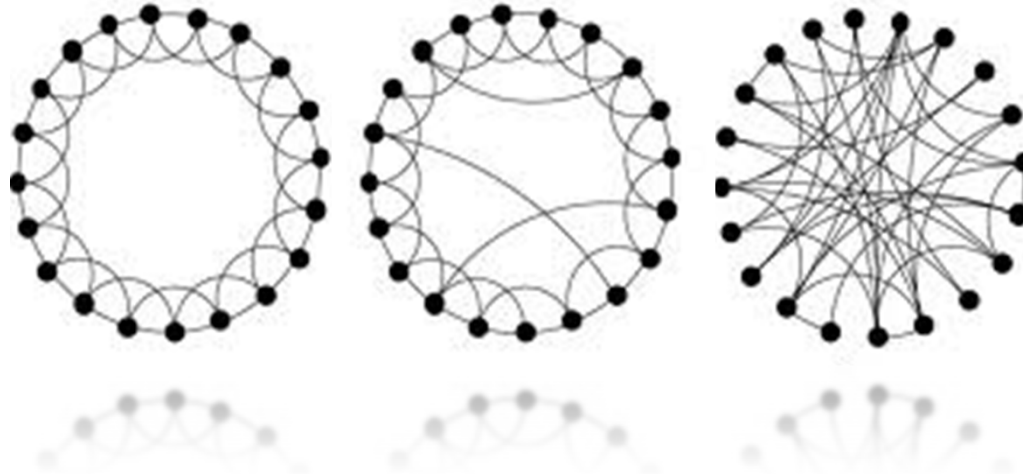


Degree Distribution:

$$P_{rand}(k) \cong C_{N-1}^k p^k (1-p)^{N-1-k}$$



Models for «real» networks: small world



The Watts Strogatz Model:

It takes a lot of randomness to ruin the clustering, but a very small amount to overcome locality

Models for real networks: Preferential Attachment

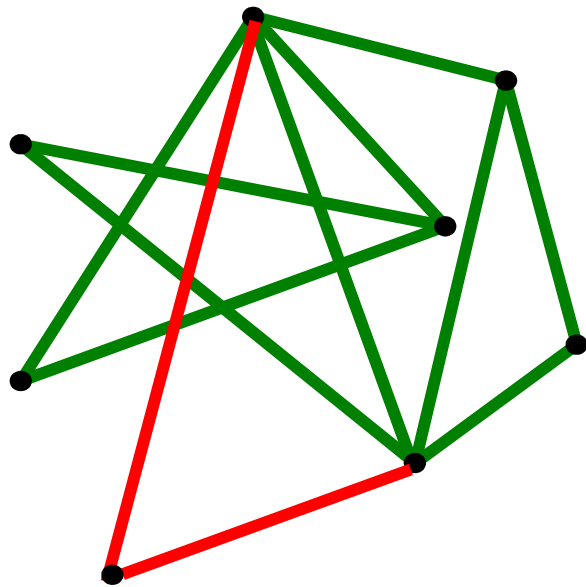
Where will the new node link to?

ER, WS models: choose randomly.

New nodes prefer to link to highly connected nodes (www, citations, IMDB).

PREFERENTIAL ATTACHMENT:

the probability that a node connects to a node with k links is proportional to k .

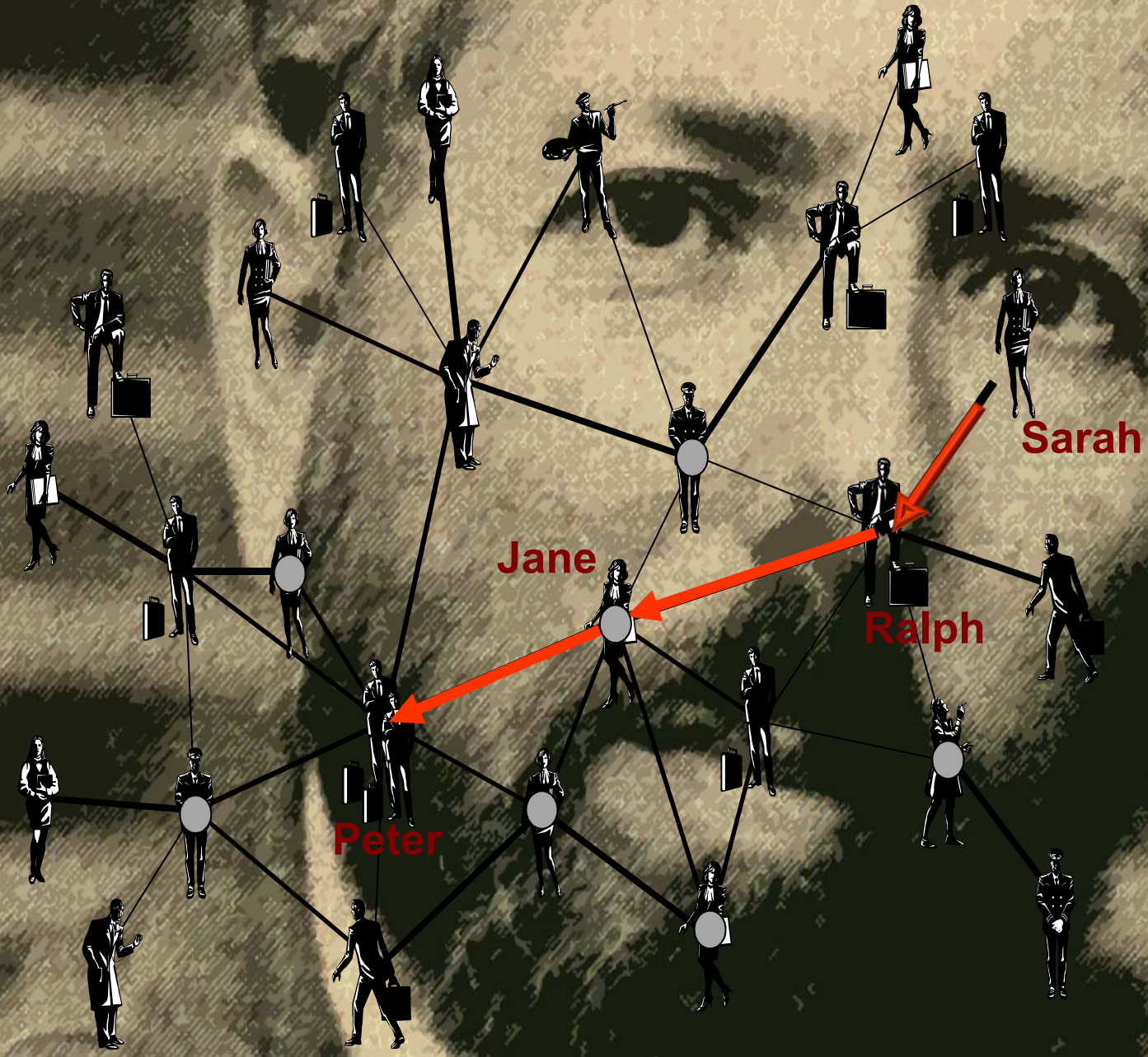


$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

Empirical validation of social theories on big data

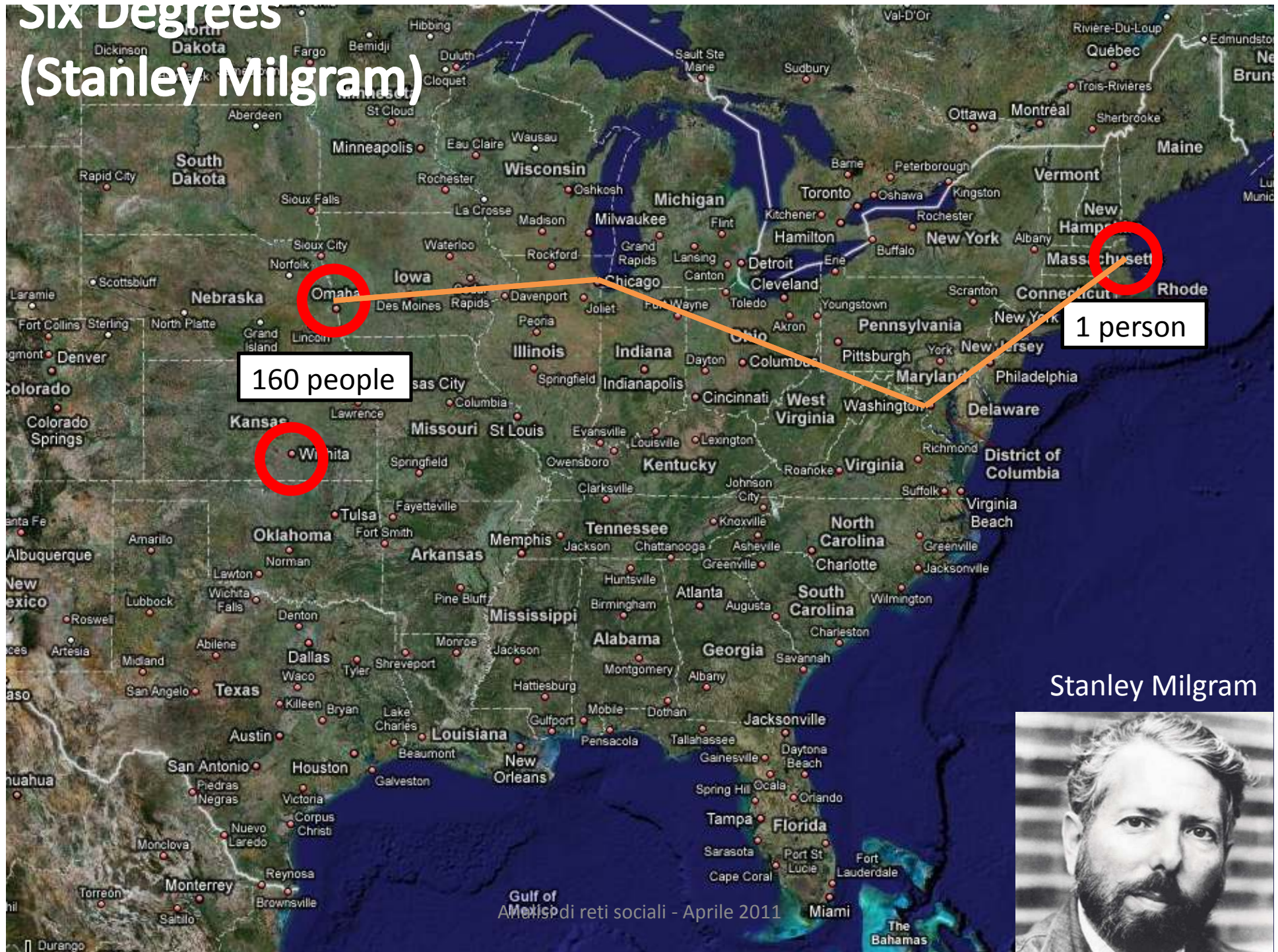
SIX DEGREES

small worlds



*Frigyes Karinthy, 1929
Stanley Milgram, 1967*

Six Degrees (Stanley Milgram)



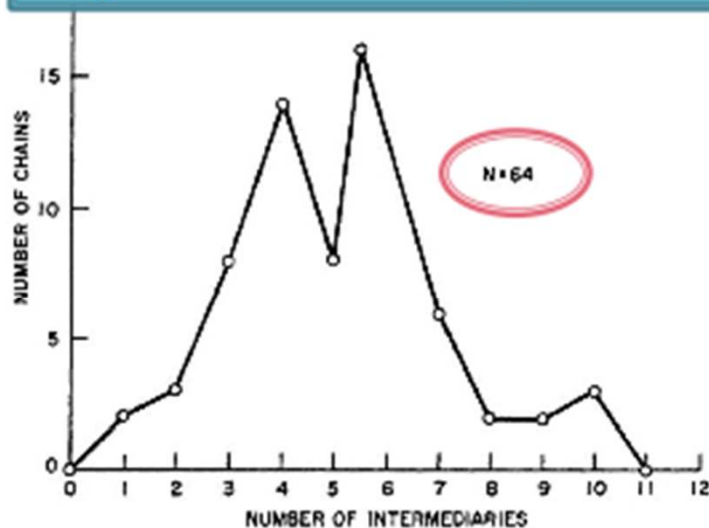
Stanley Milgram



The Small-world experiment

- 64 chains completed:
 - 6.2 on the average, thus “6 degrees of separation”
- Further observations:
 - People who owned stock had shortest paths to the stockbroker than random people: 5.4 vs. 5.7
 - People from the Boston area have even closer paths: 4.4

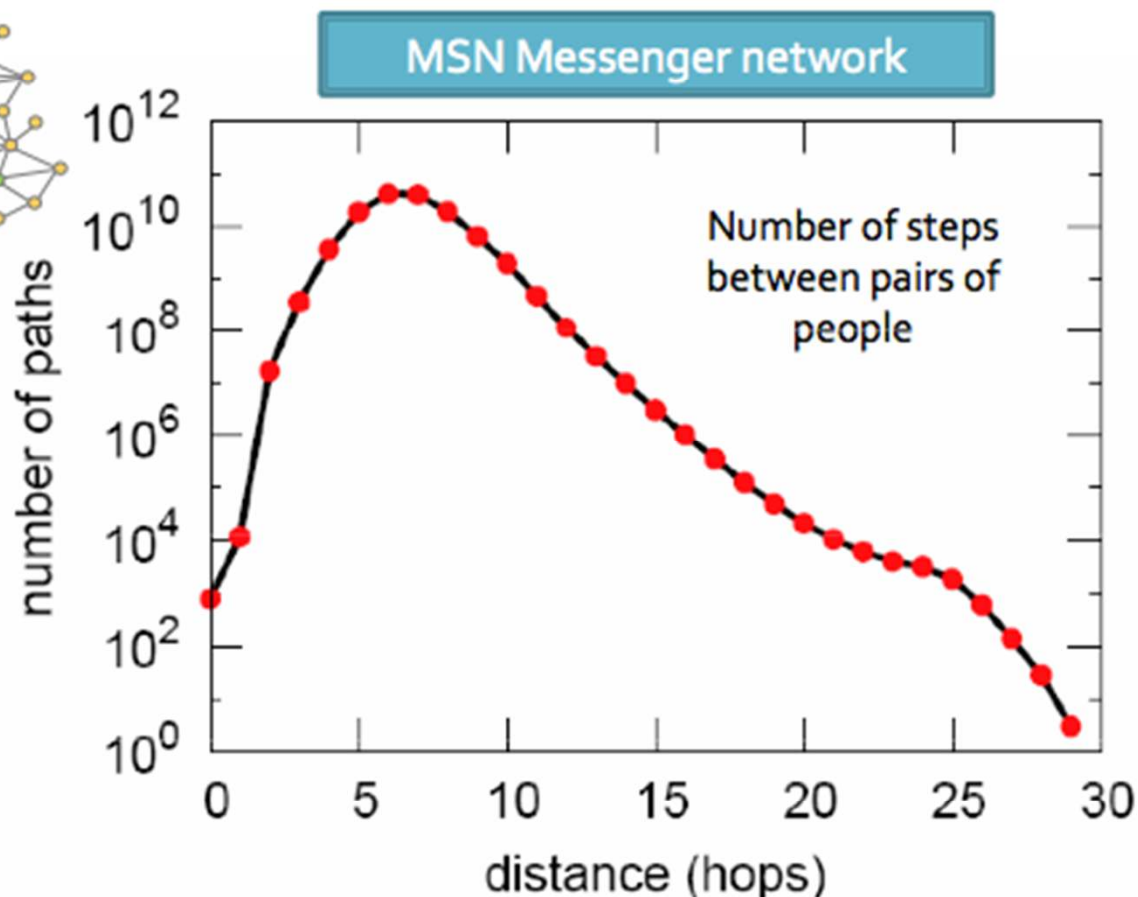
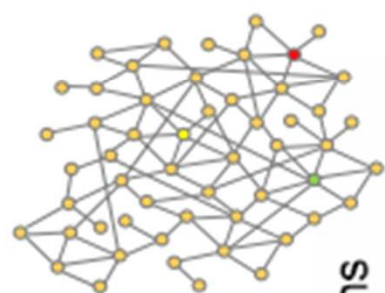
Milgram's small world experiment



IM communication network

- **Buddy graph**
 - 240 million people (people that login in June '06)
 - 9.1 billion buddy edges (friendship links)
- **Communication graph** (take only 2-user conversations)
 - Edge if the users exchanged at least 1 message
 - 180 million people
 - 1.3 billion edges
 - 30 billion conversations

MSN Network: Small world



Avg. path length 6.6
90% of the people can be reached in < 8 hops

Hops	Nodes
0	1
1	10
2	78
3	3,96
4	8,648
5	3,299,252
6	28,395,849
7	79,059,497
8	52,995,778
9	10,321,008
10	1,955,007
11	518,410
12	149,945
13	44,616
14	13,740
15	4,476
16	1,542
17	536
18	167
19	71
20	29
21	16
22	10
23	3
24	2
25	3

The strength of weak ties

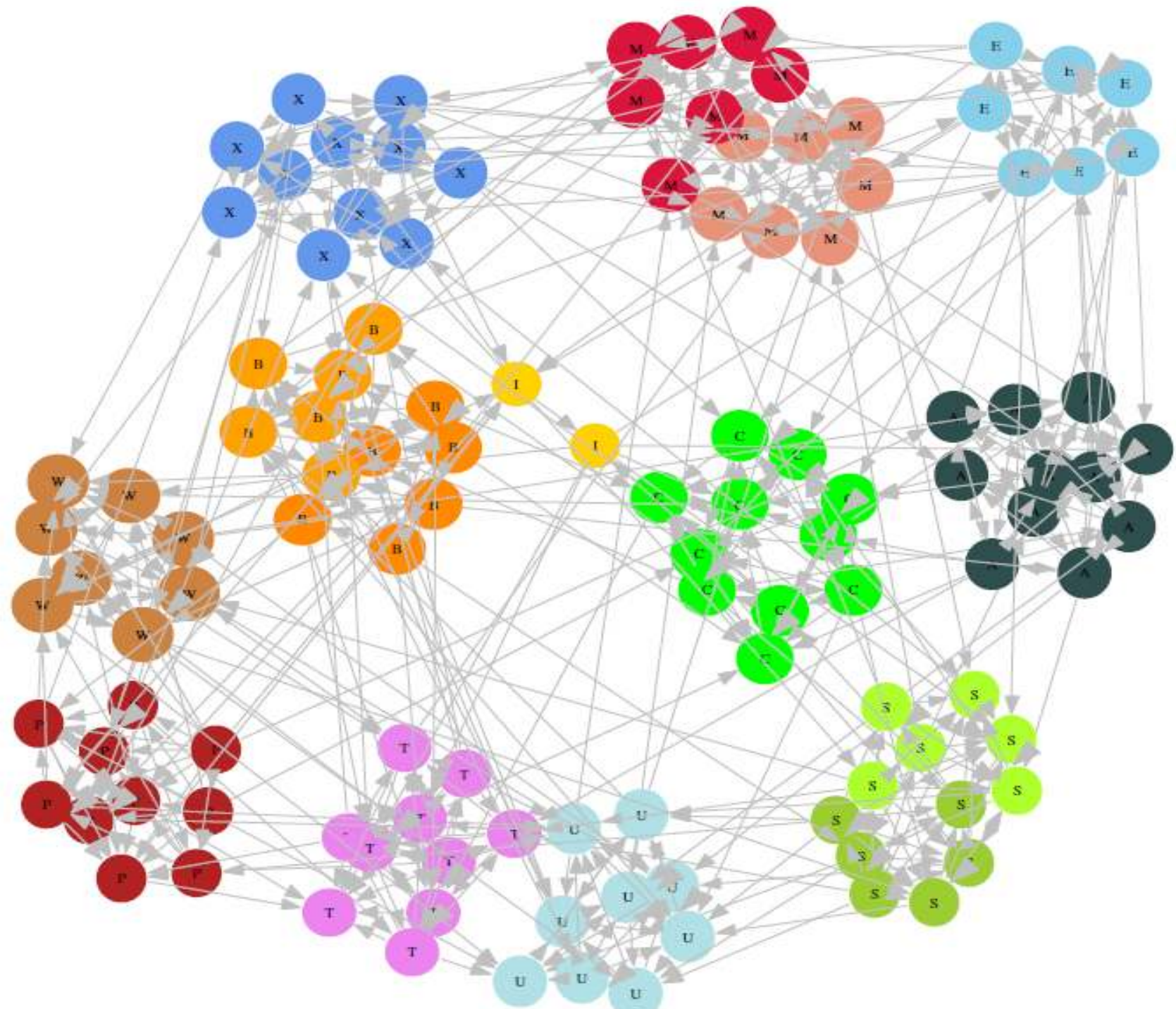
- Mark S. **Granovetter**, 1973
- His PhD thesis: how people get to know about new jobs?
- Through personal contacts
- Surprise: often acquaintances, **not** close friends

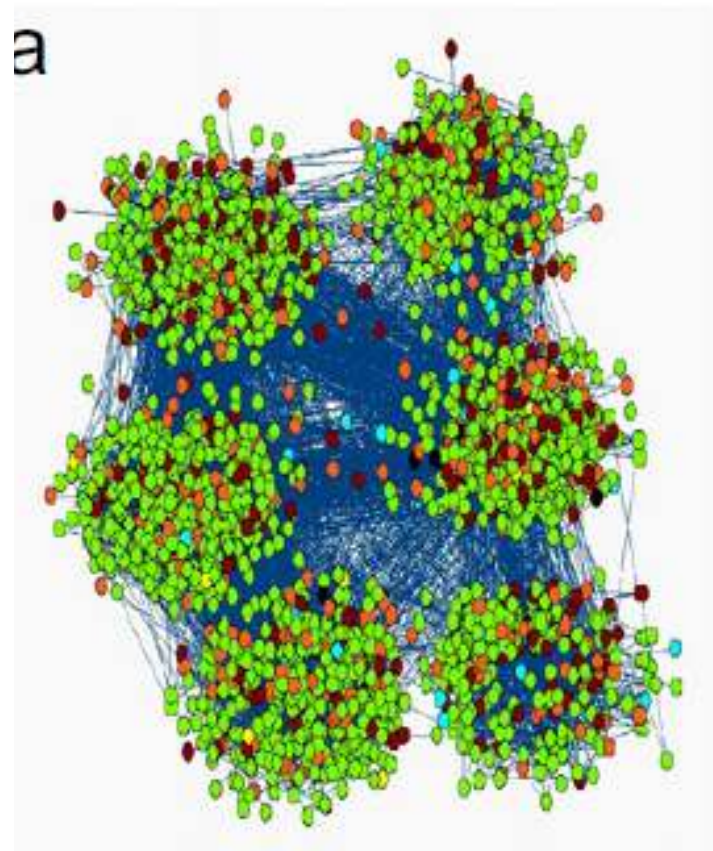
The Strength of Weak Ties

Mark S. Granovetter

American Journal of Sociology, Volume 78, Issue 6 (May, 1973), 1360-1380.







b

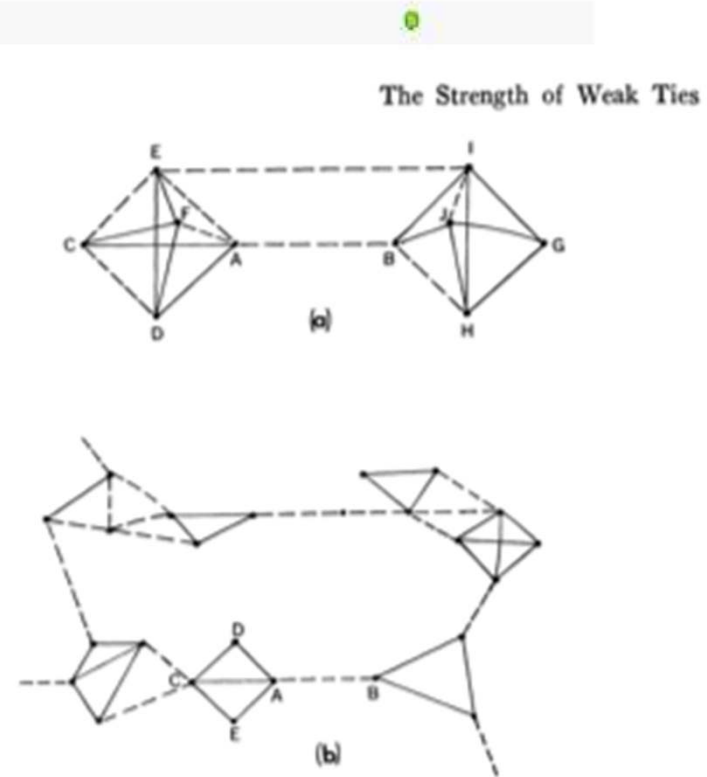
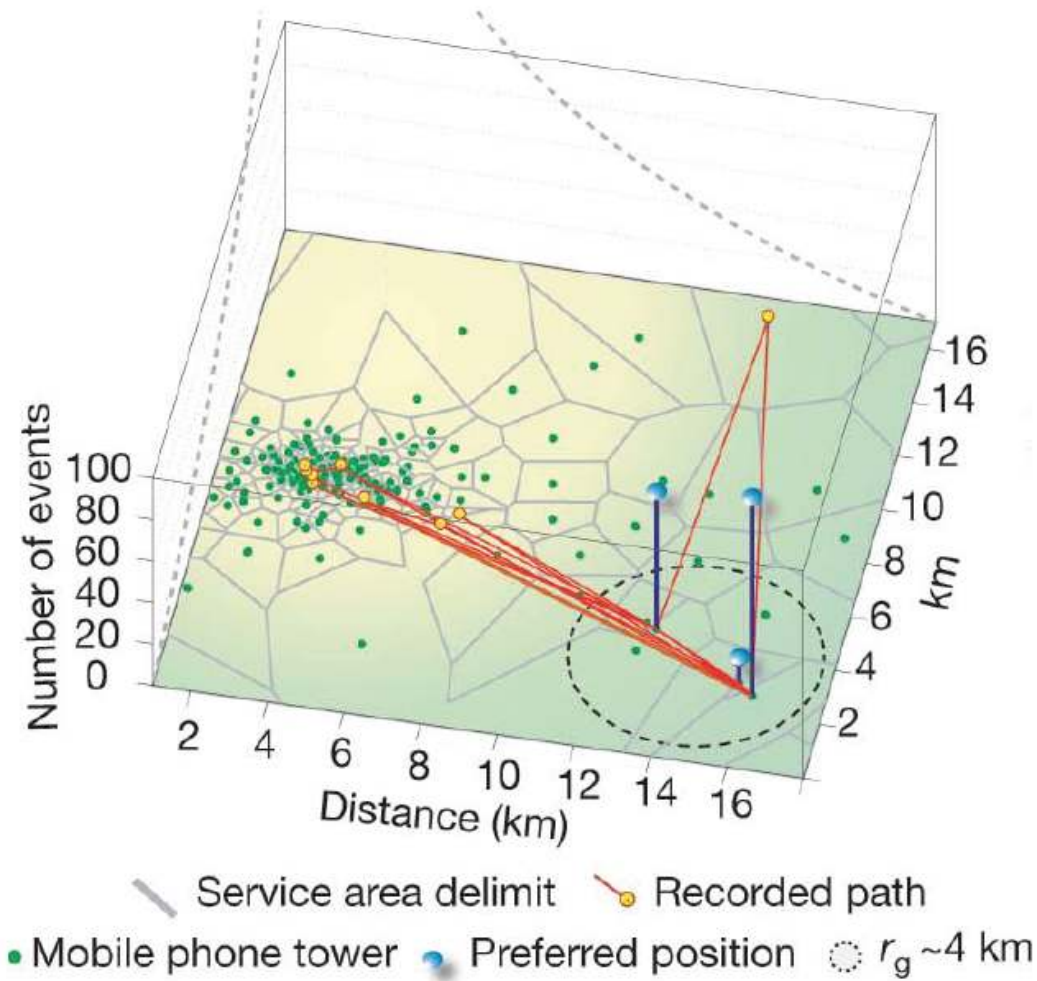


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. — = strong tie; - - - = weak tie.

Tie strength in real data

- For many years the Granovetter's theory was not tested
- But, today we have large who-talks-to-whom graphs:
 - Email, Messenger, Cell phones, Facebook
- Onnela et al. 2007:
 - Cell-phone network of 20% of country's population

Country-wide mobile phone data



when
you
call



where
you
call



who
you
call

Social proximity and tie strength

- How connected are u and v in the social network.
 - Various well-established **measures of network proximity**, based on the common neighbors (Jaccard, Adamic-Adar) or the structure of the paths (Katz) connecting u and v in the who-calls-whom network.
- How intense is the interaction between u and v .
 - Number of calls as **strength of tie**

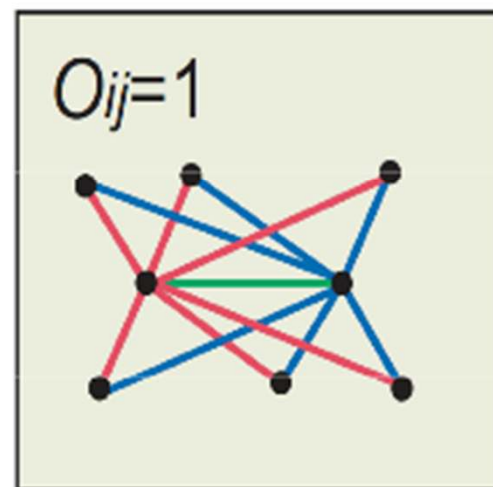
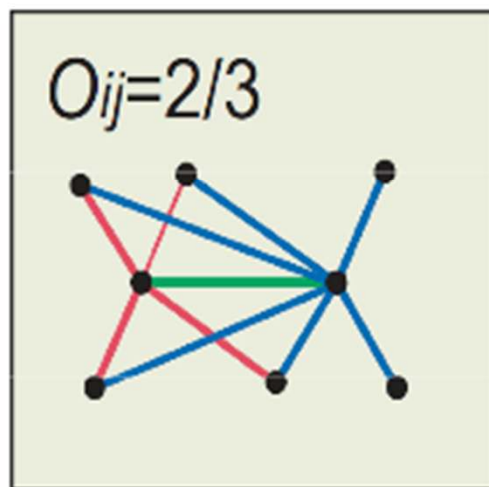
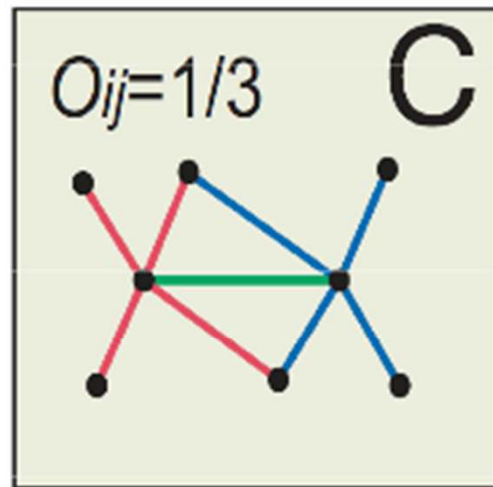
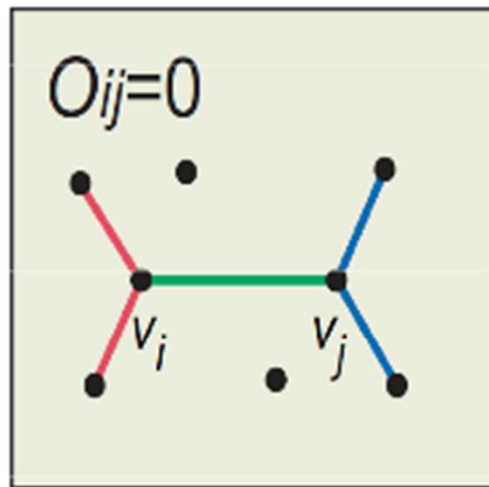
Neighborhood Overlap

- **Overlap:**

$$O_{ij} = \frac{n(i) \cap n(j)}{n(i) \cup n(j)}$$

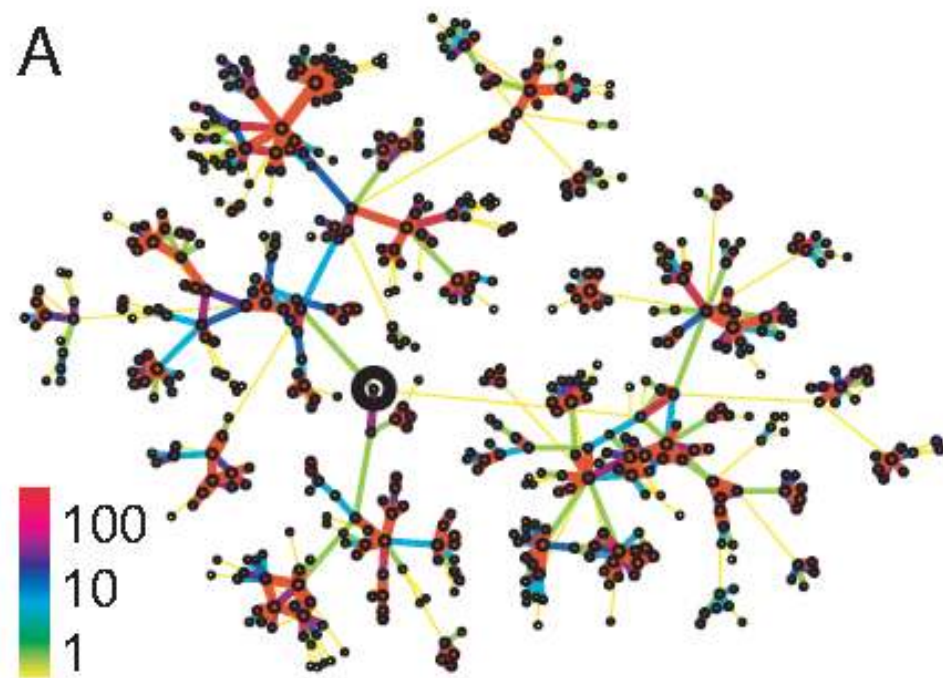
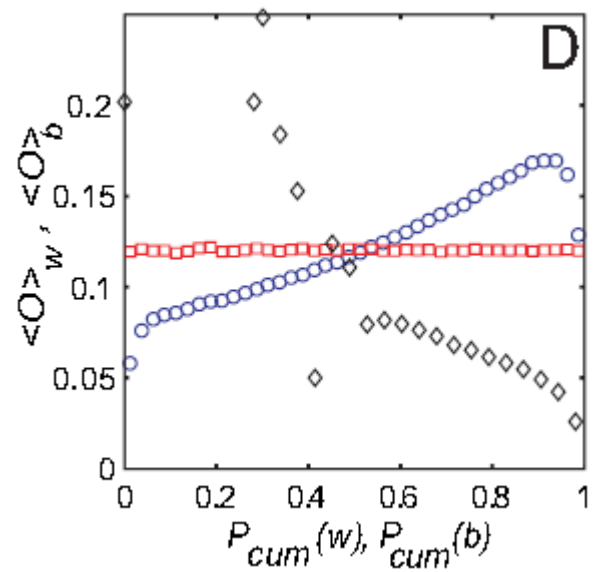
- $n(i)$... set of neighbors of A

- **Overlap = 0**
when an edge is
a **local bridge**



Strength of weak ties

- Large scale empirical validation of Granovetter's theory
 - Social proximity increases with tie strength
 - Weak ties span across different communities
- J.-P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, A.-L. Barabási. **Structure and tie strengths in mobile communication networks**. PNAS 104 (18), 7332-7336 (2007).



Human mobility, social ties and link prediction

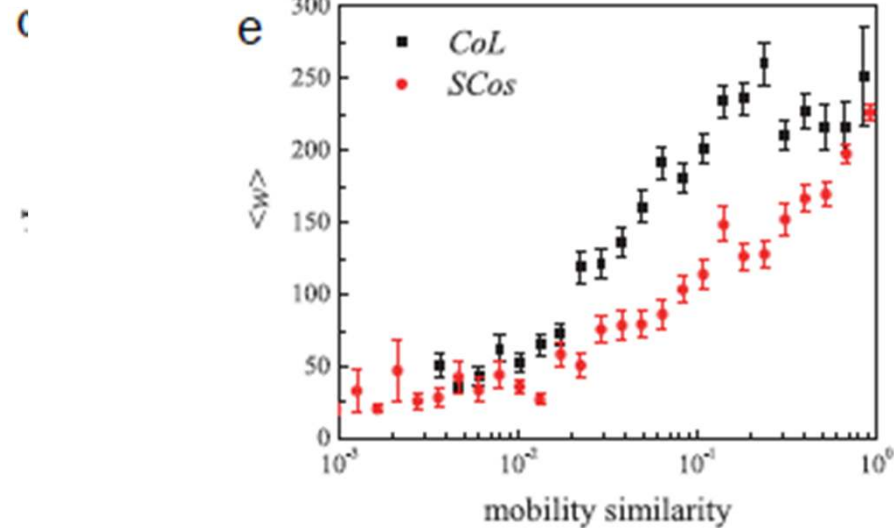
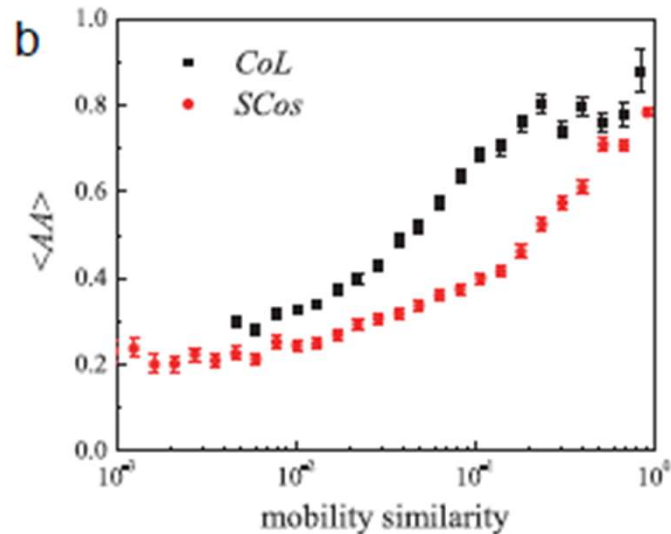
**Dashun Wang, Dino Pedreschi, Chaoming Song,
Fosca Giannotti, Albert-Lászlo Barabási**

**SIGKDD Int. Conf. on Knowledge Discovery and
Data Mining – KDD 2011**

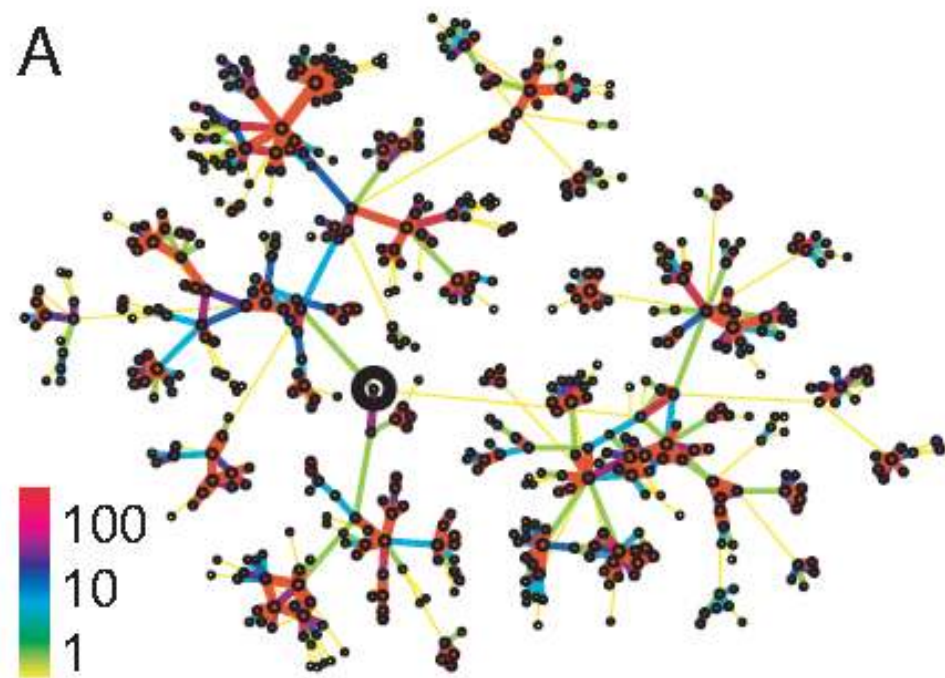
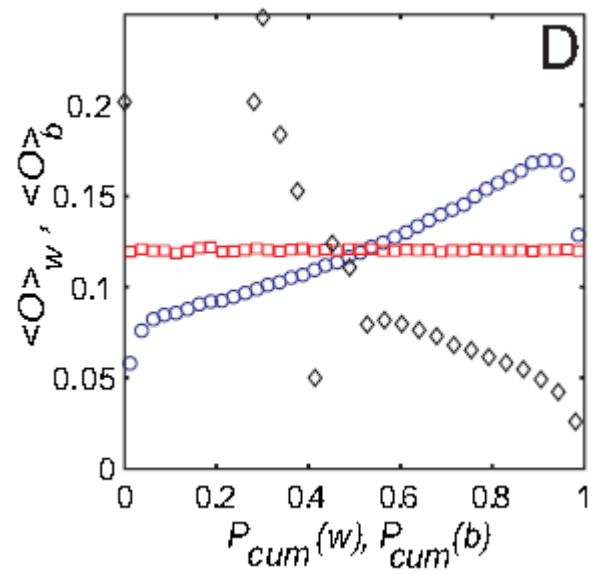
Colocation, social proximity, tie strength

- How similar is the movement of users u and v
 - Various **co-location measures**, quantifying the similarity between the movement routines of u and v (mobile homophily)
- How connected are u and v in the social network.
 - Various well-established **measures of network proximity**, based on the common neighbors (Jaccard, Adamic-Adar) or the structure of the paths (Katz) connecting u and v in the who-calls-whom network.
- How intense is the interaction between u and v .
 - Number of calls as **strength of tie**

mobility dimension of the “strength of weak ties”

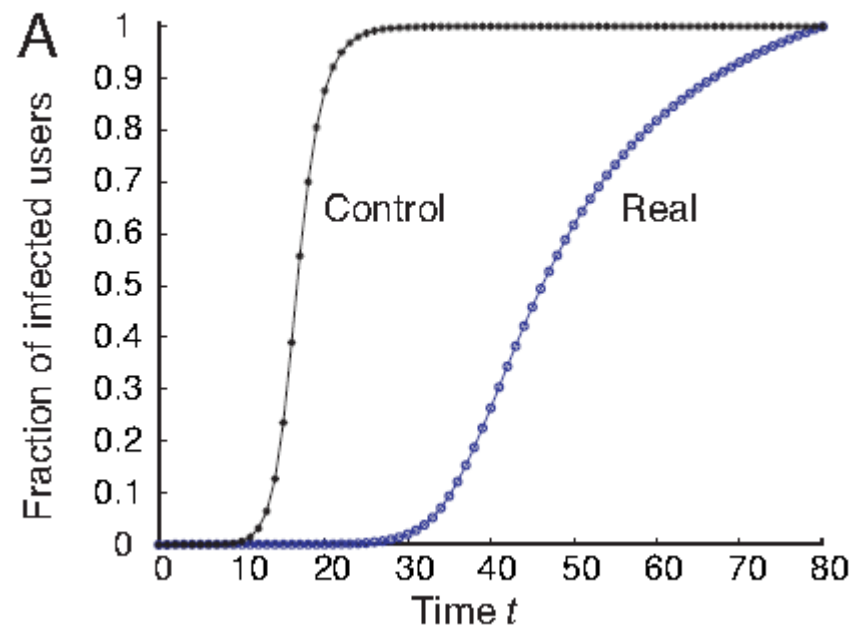


- co-location, network proximity and tie strength strongly correlate with each other
- measured on 3 months of calls, 6 Million users, nation-wide (large European country)



The strength of weak ties ...

- For information **diffusion** (spreading of news and rumors on a social network)



The weakness of weak ties

- Diffusion of **innovation / adoption**

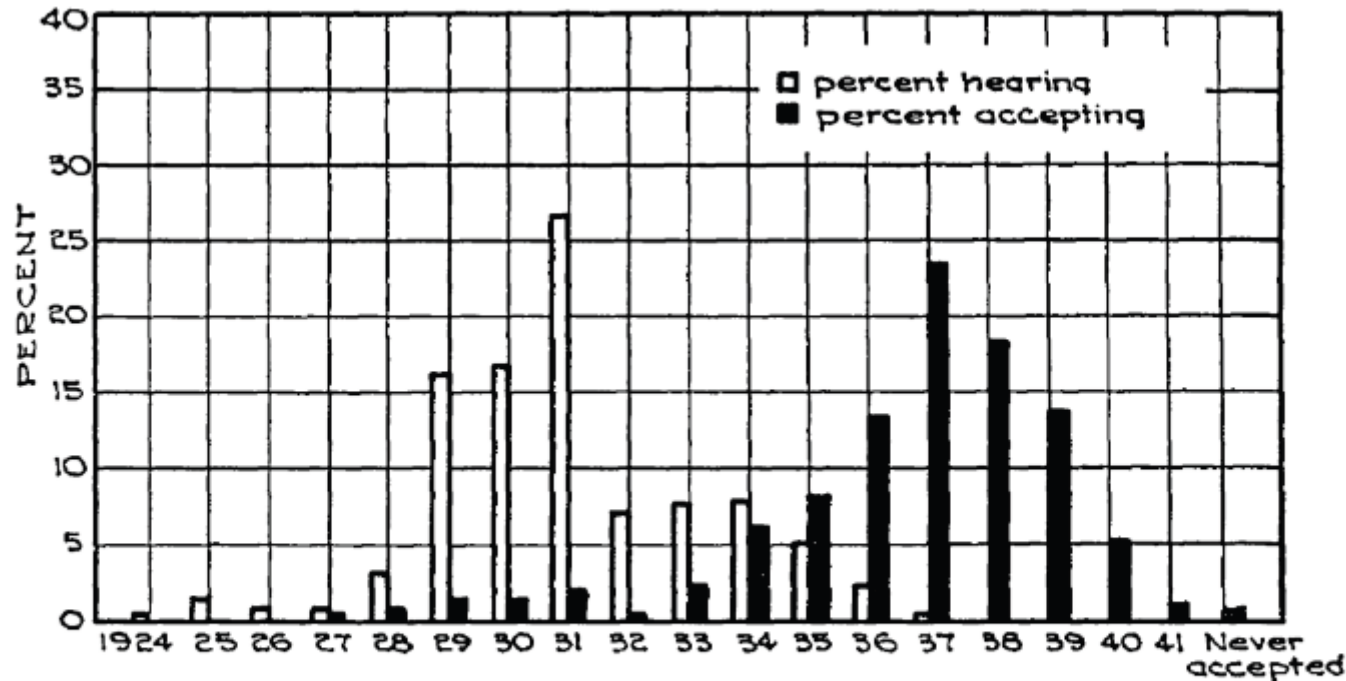
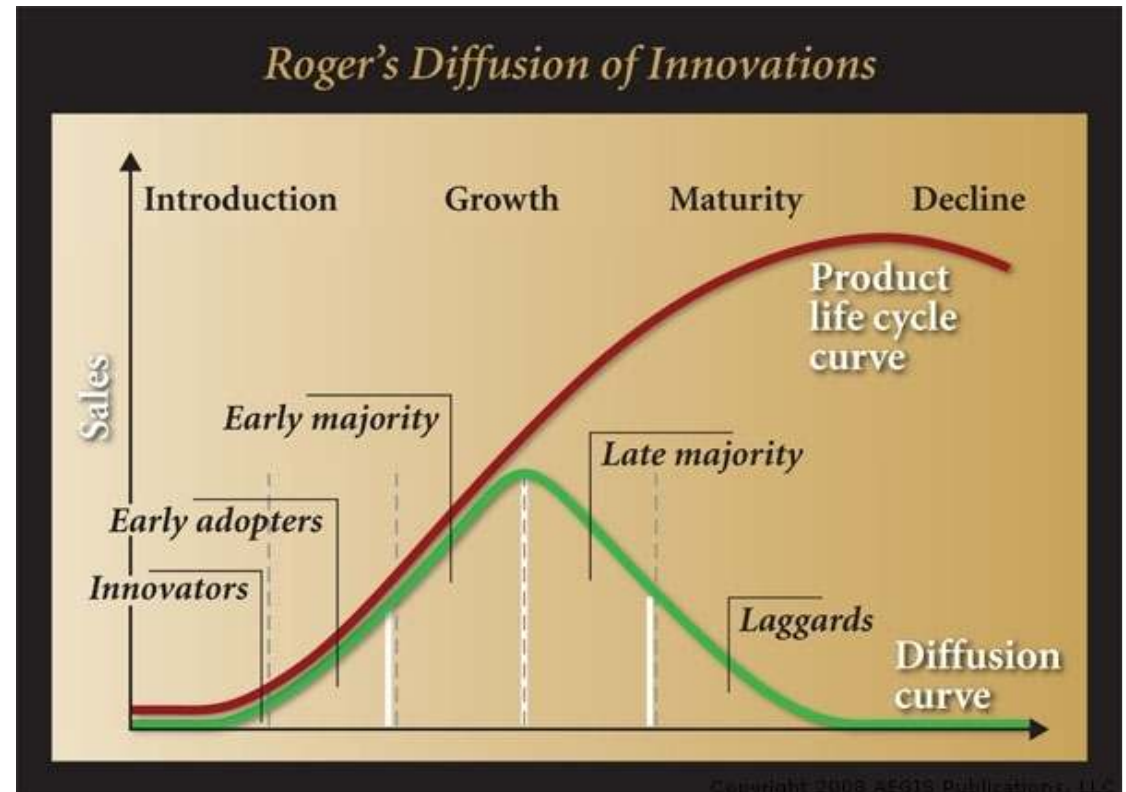
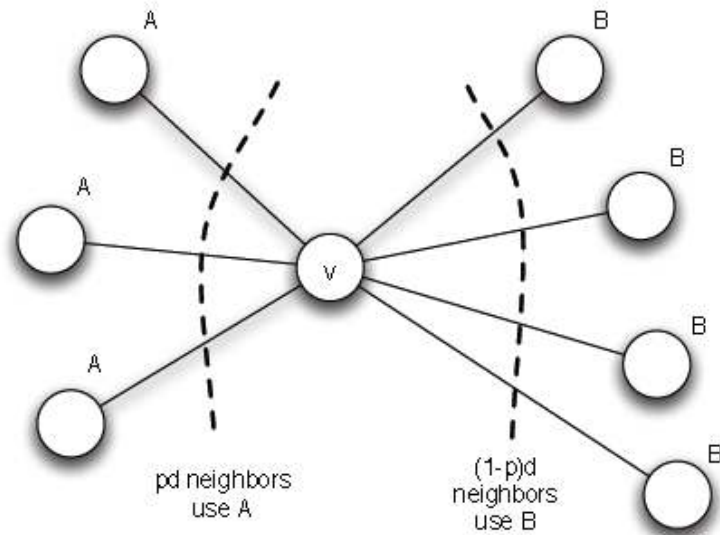
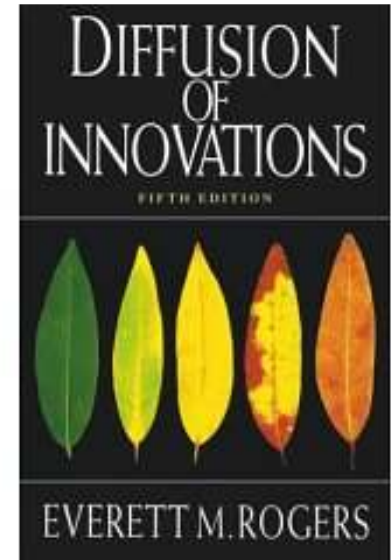


Figure 19.10: The years of first awareness and first adoption for hybrid seed corn in the Ryan-Gross study. (Image from [358].)

The strength of the strong ties for the



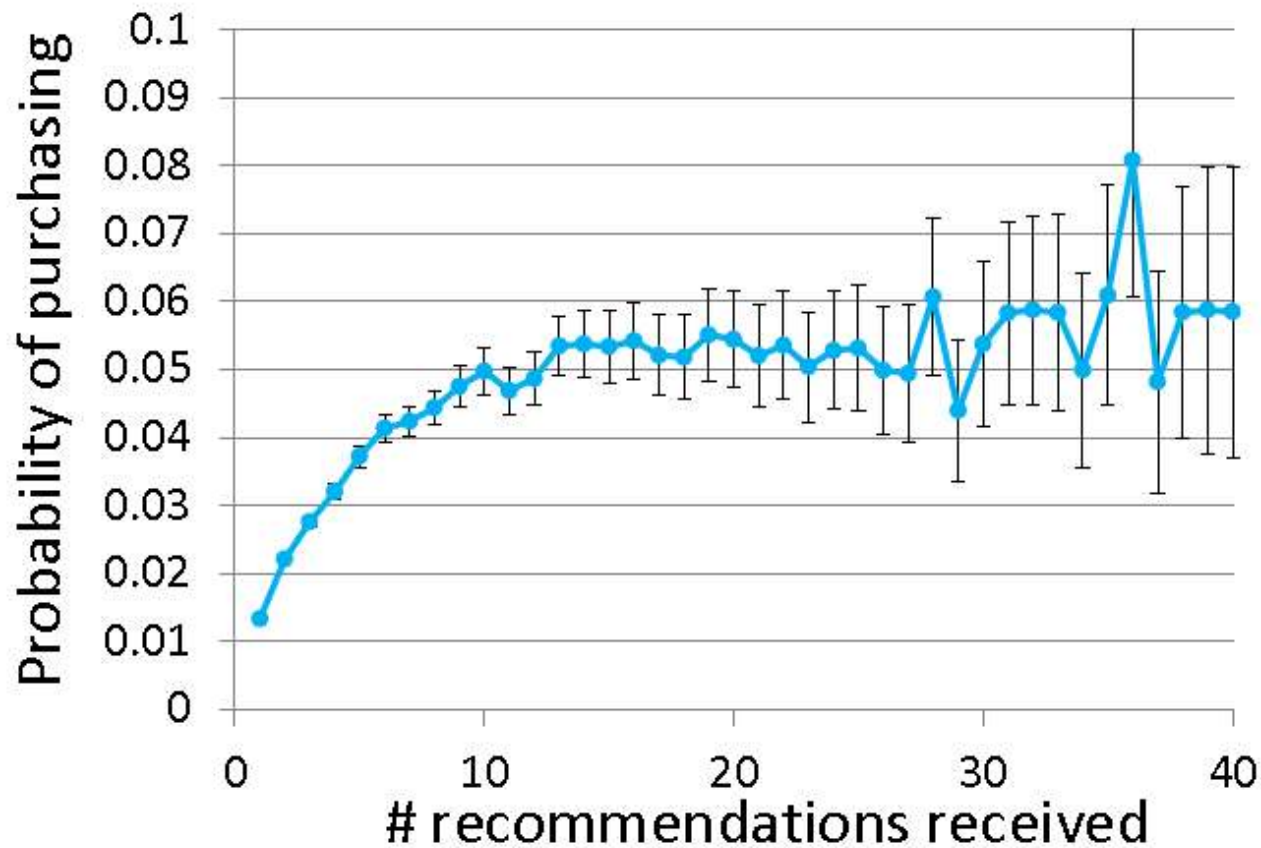
Diffusion in Viral Marketing

- Senders and followers of recommendations receive discounts on products



- **Data: Incentivized Viral Marketing program**
 - 16 million recommendations
 - 4 million people, 500k products

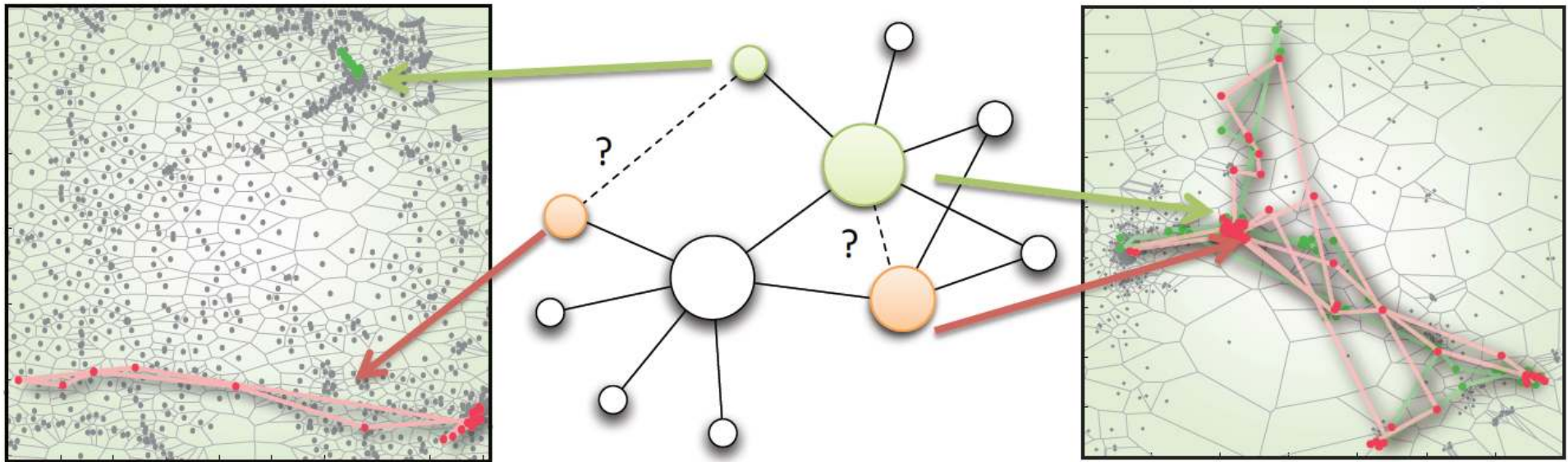
Adoption Curve: Validation



Social network mining 1: link prediction

Which new links will appear in the
social network?

Link prediction in social networks



Potential links with common neighbors

Unsupervised precision

Katz	9.1%
Adamic-Adar	7.8%
SCos	5.6%
Weighted SCos	5.6%
Extra-role CoL	5.1%
Weighted CoL	5.1%
CN	5.1%
CoL	5.0%
Jaccard	3.0%

Classification

	Pred. class=0	Pred. class=1
actual class=0	6,627	82
actual class=1	117	228

decision-tree: $AA > 0.5$ and $S\text{CoL} > 0.7$
73.5% precision and 66.1% recall

Combining topology and mobility measures is the key to achieving high precision and recall.

People is predictable!

- Probability of a new link between two (disconnected) random users:

10^{-6}

- Best prediction accuracy using only social features:

10%

- Best prediction accuracy using **social + mobility** features:

75%

A small detour on human
predicatability



To what degree is
human motion
predictable?



Predictability



0%

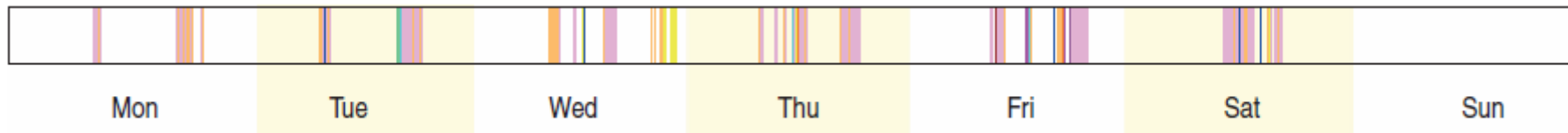
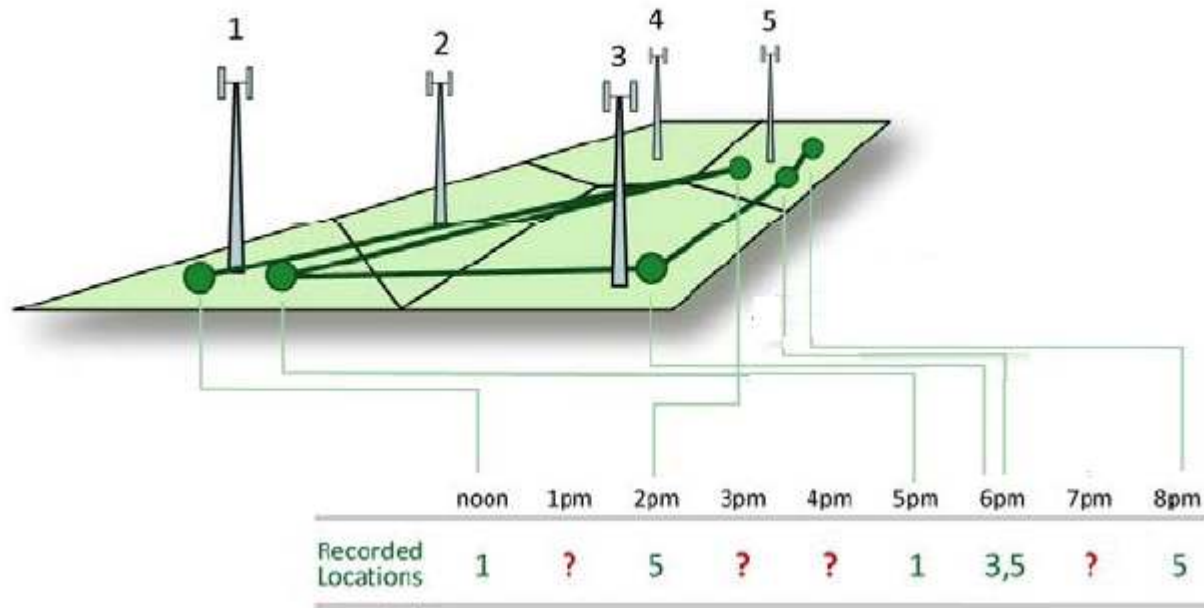
Random walk models (RW, LF, CTRW)

100%

Periodic motion

Entropy of human trajectories

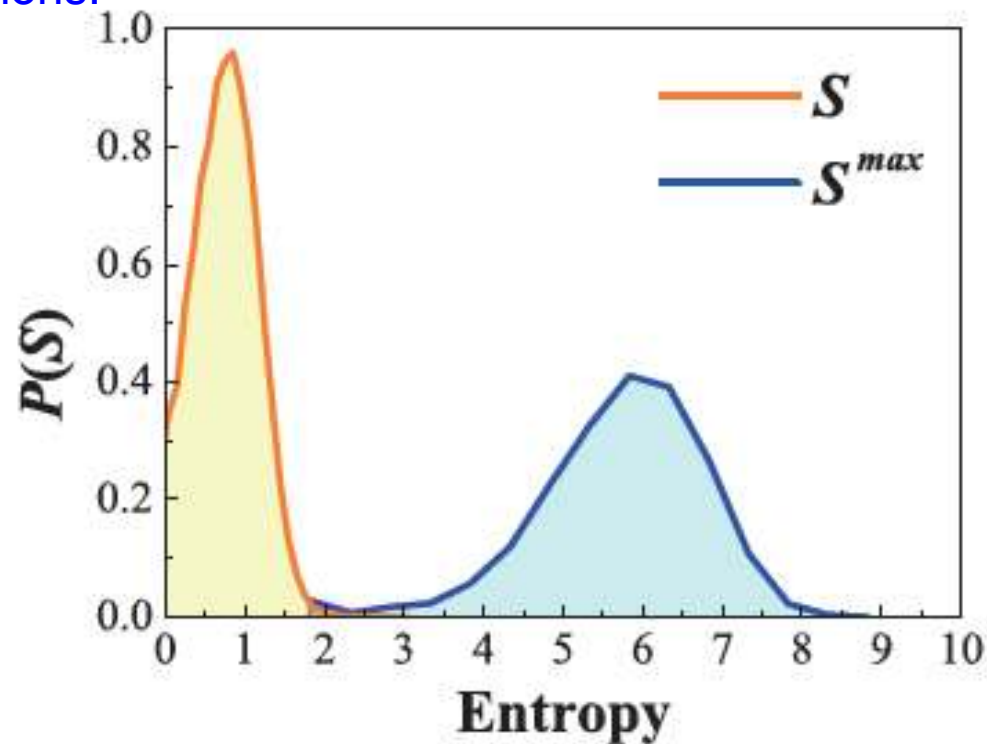
Recorded Trajectory



$$-\sum_{T'_i \subset T_i} P(T'_i) \log_2 [P(T'_i)]$$

Entropy Distribution Across the Population

$S^{\max} \sim 6 \rightarrow$ a random user could be found in any of $2^{S^{\max}} \sim 64$ locations.



$S = 0.8 \rightarrow$ the real uncertainty in the user's whereabouts is $2^{0.8} = 1.74$.

$$S \in [0, S^{\max}], S^{\max} = \log_2 N$$

Daily routines are highly predictable

- ❑ A potential **93%** average predictability in user mobility.
- ❑ Lack of variability in predictability across the population.
- ❑ Tiny dependence on demographic and external parameters

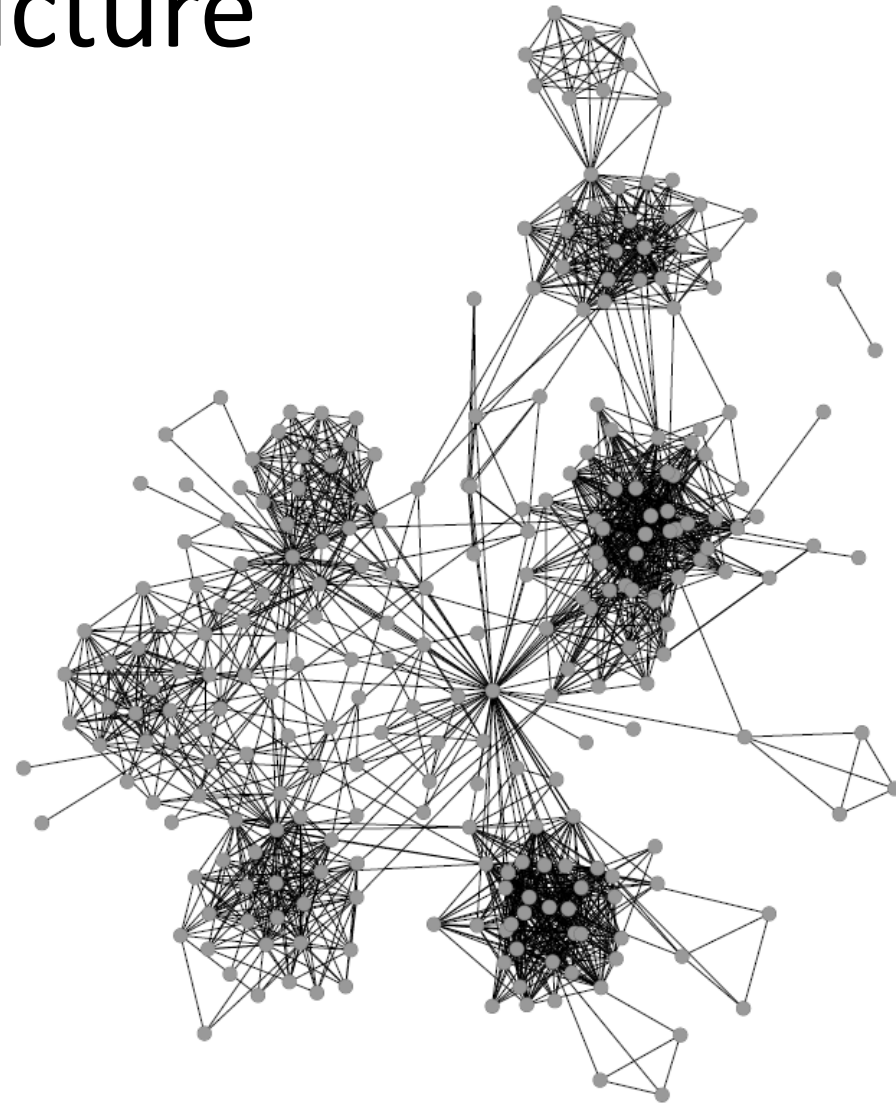
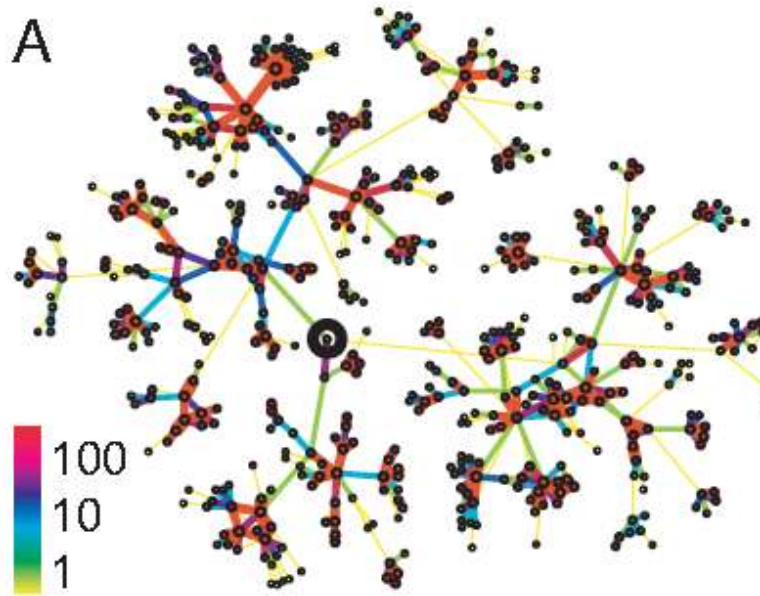
Song, Qu, Blumm, Barabasi, Science 327,108(2010)

Social network mining 2: community discovery

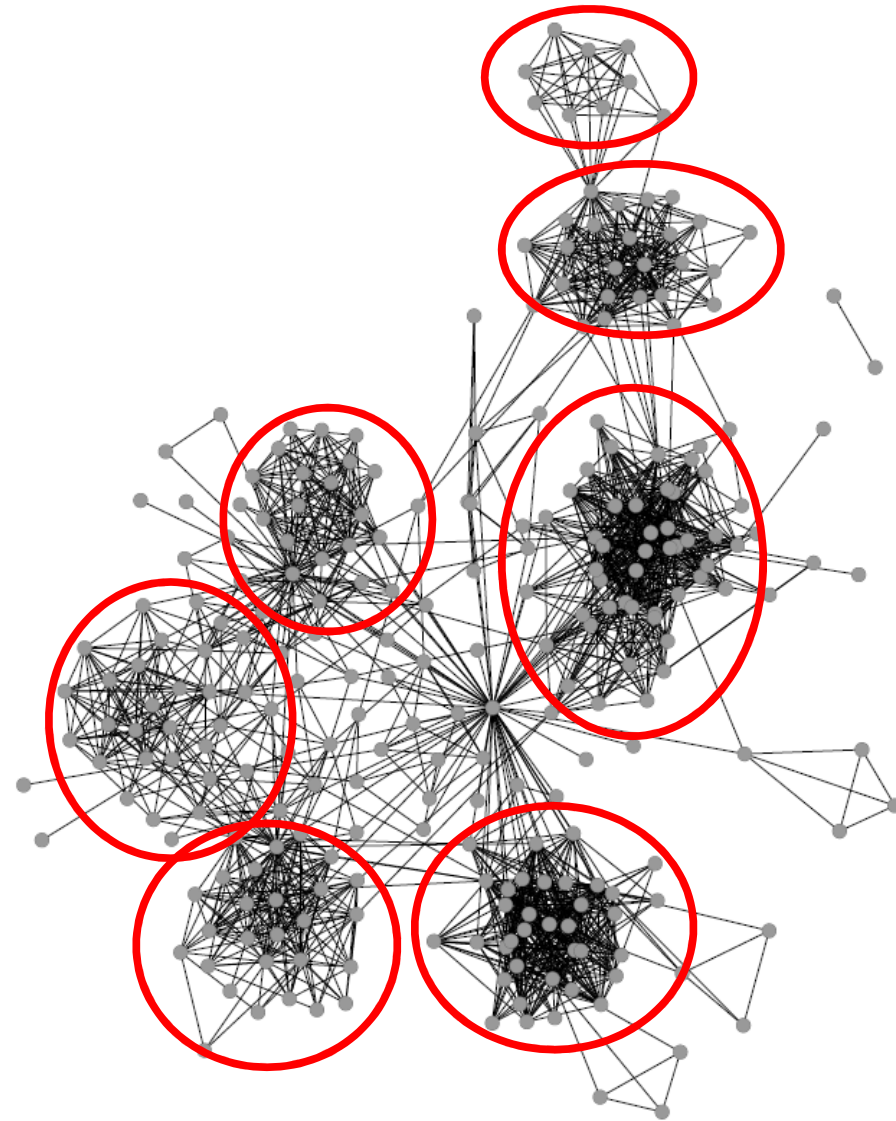
How to highlight the modular
structure of a network?

Community structure

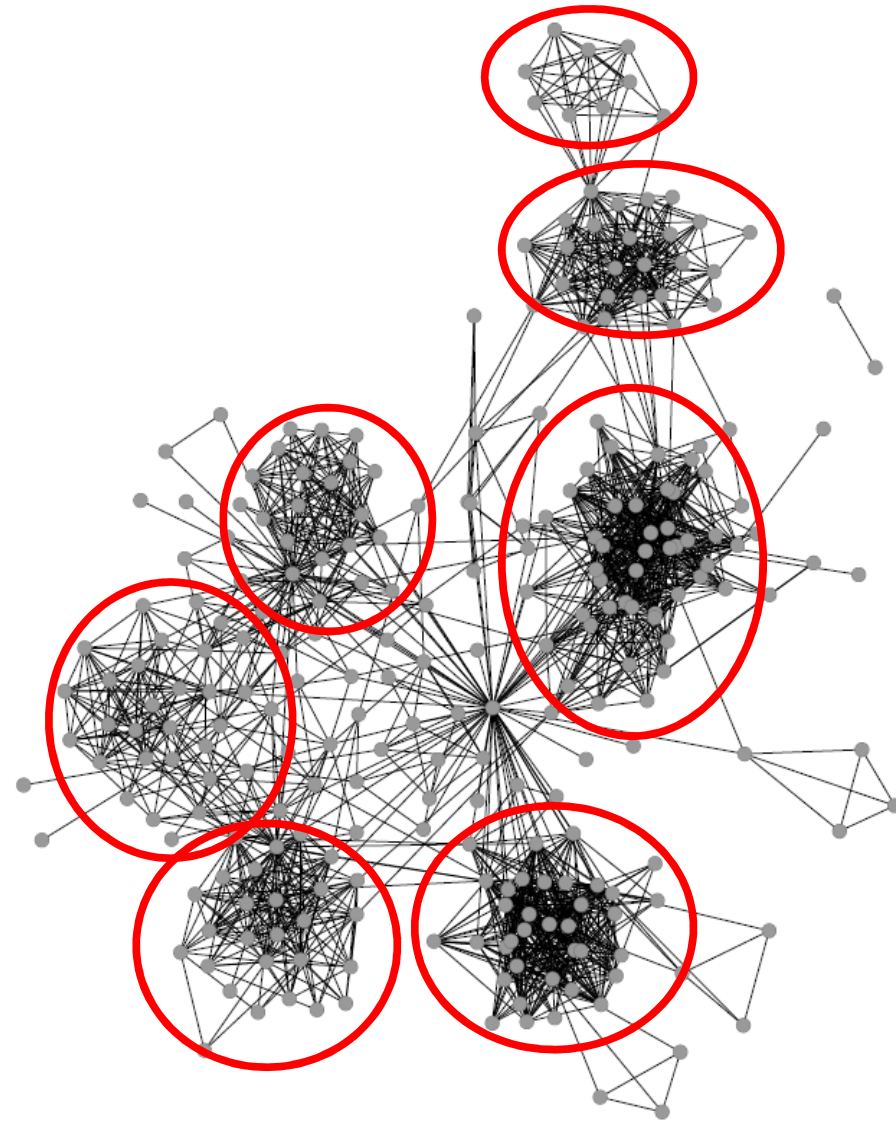
A

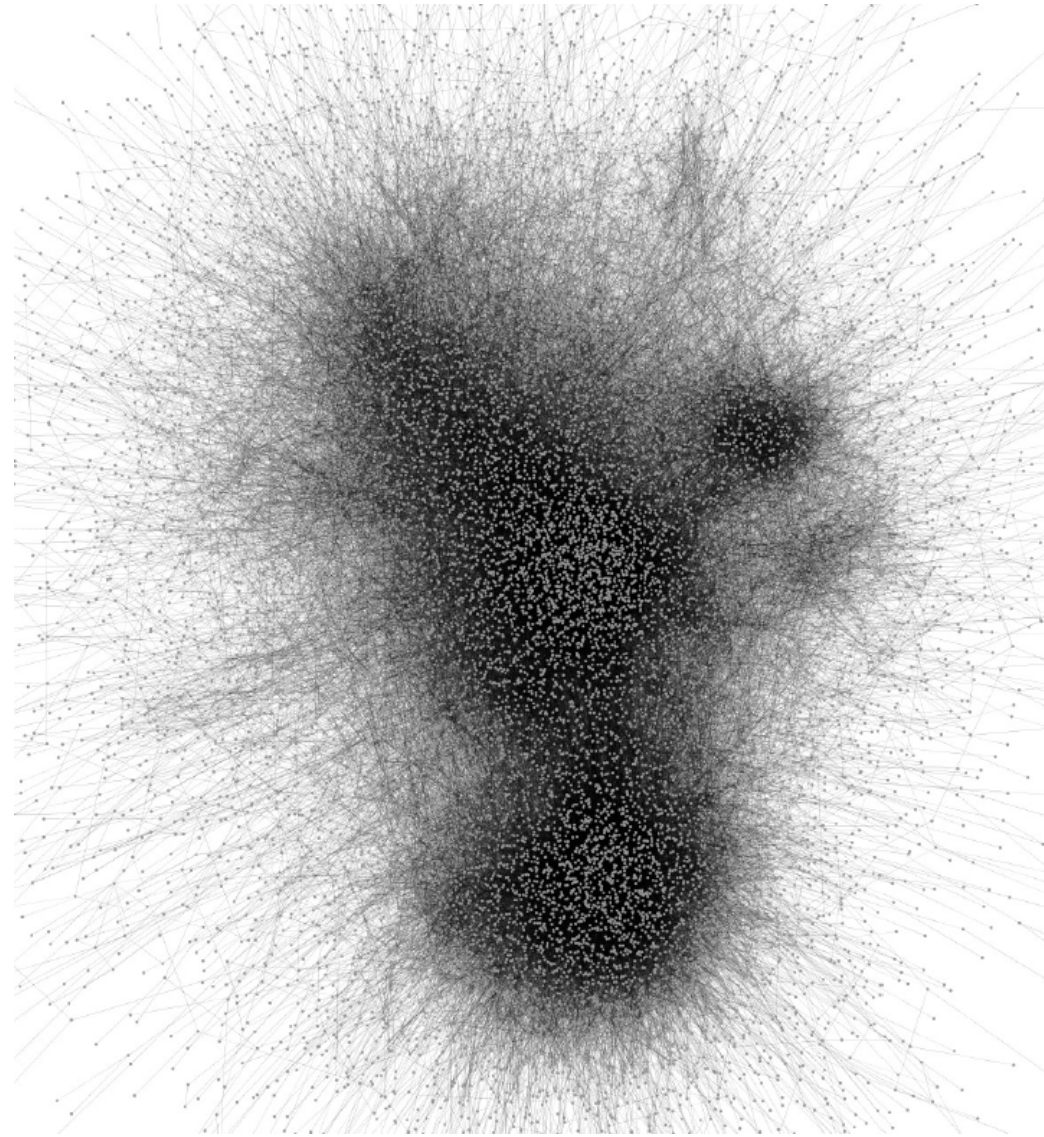


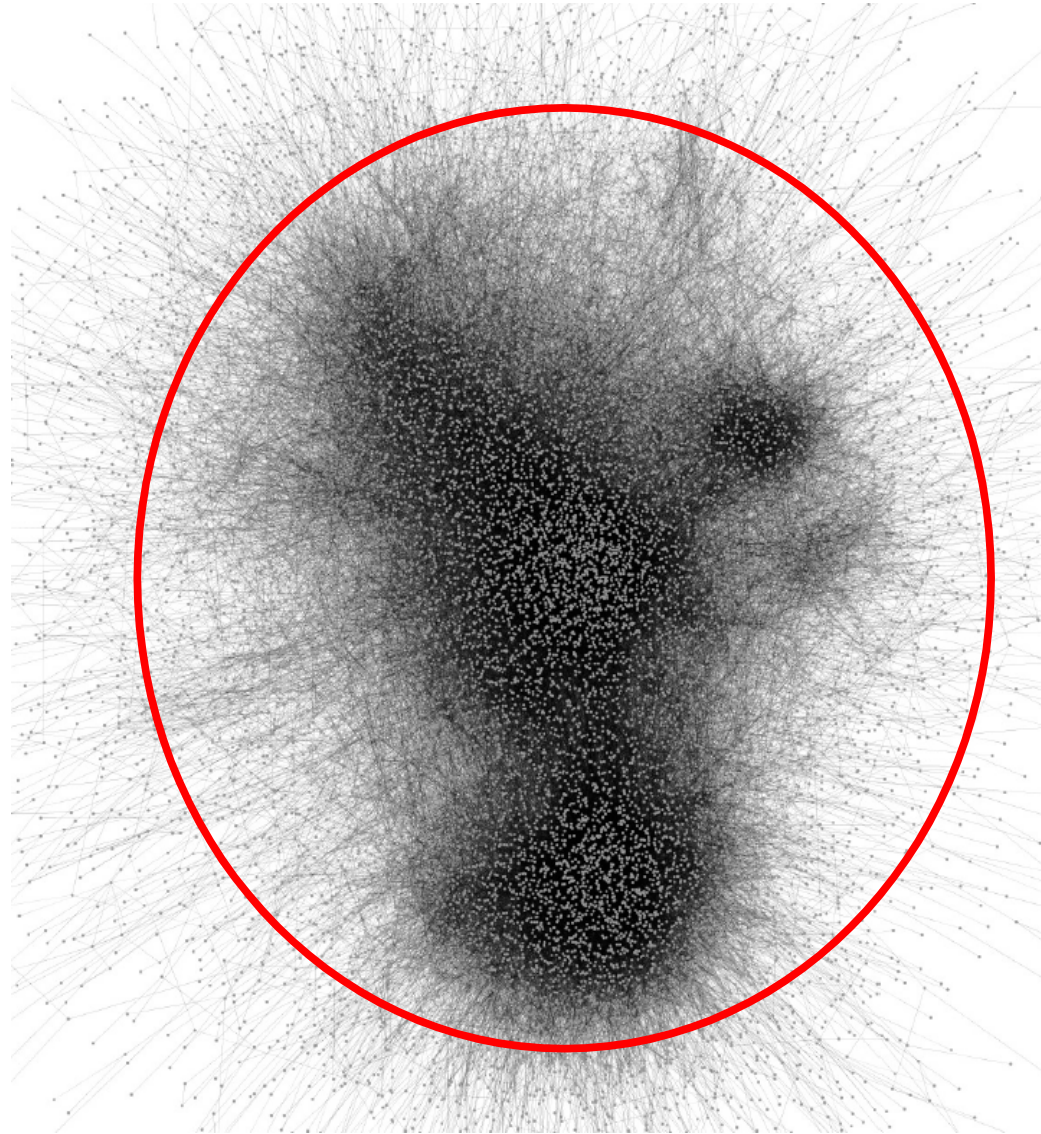
Communities



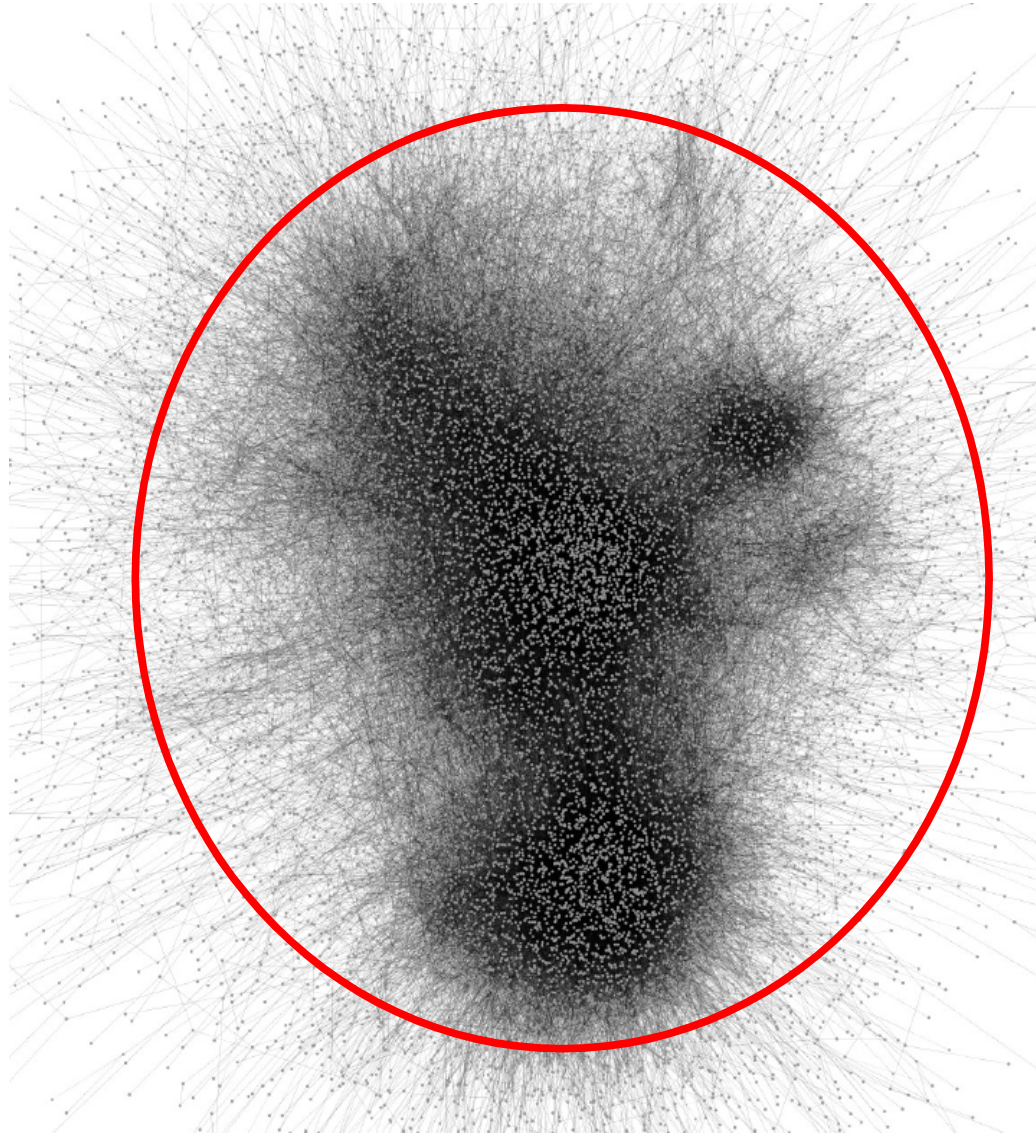
Communities



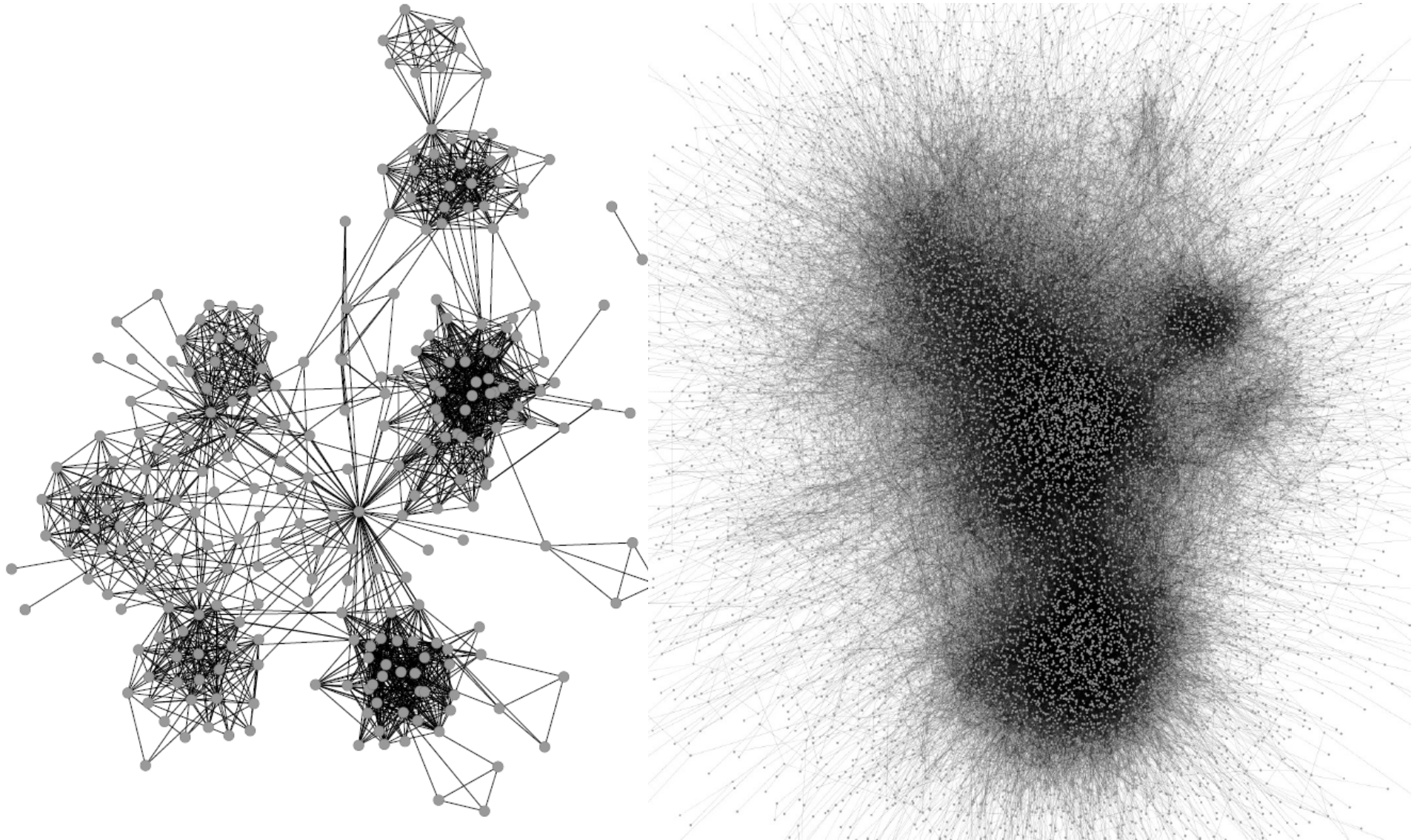




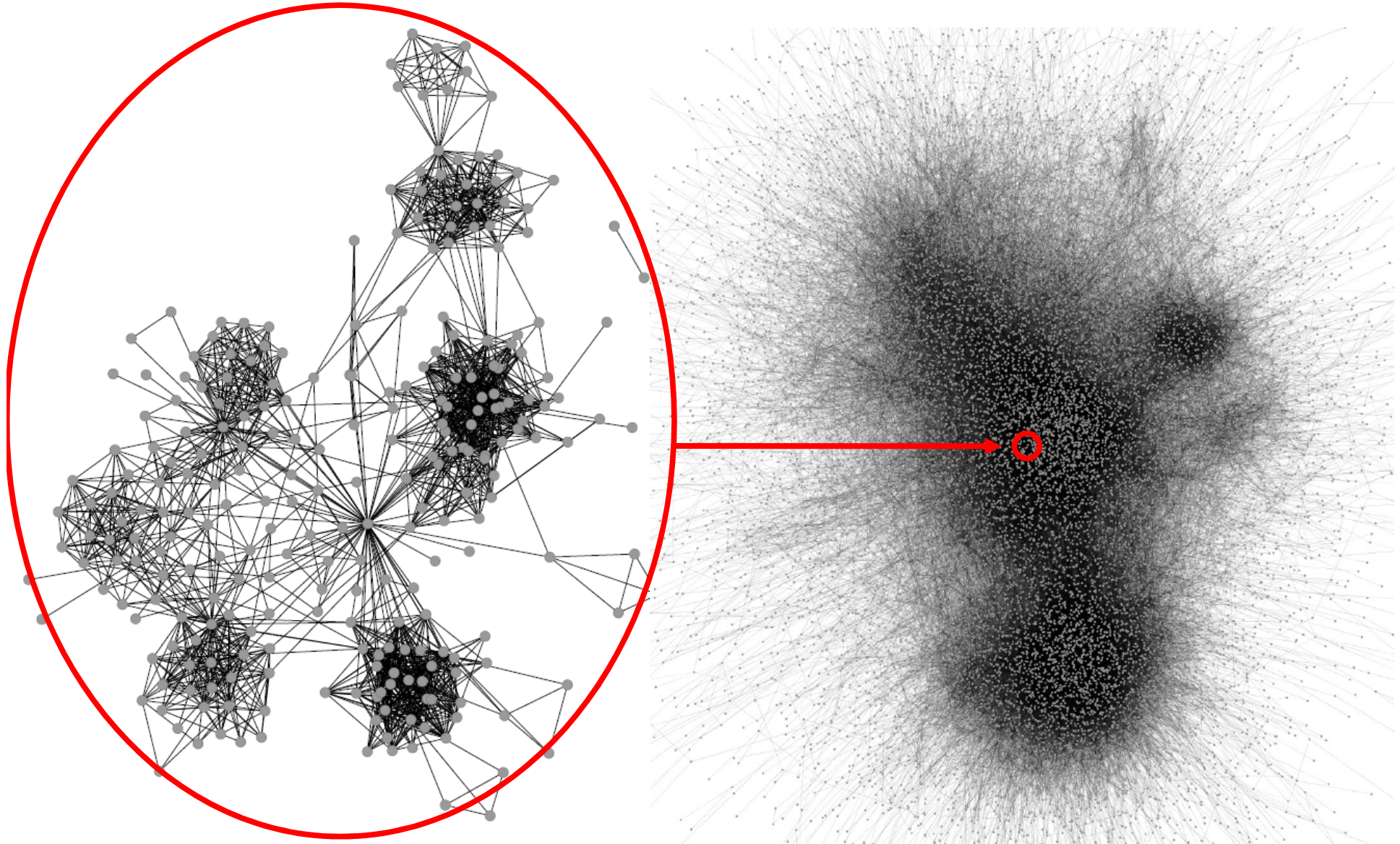
?



Are these two different networks?



No!



DEMON Algorithm

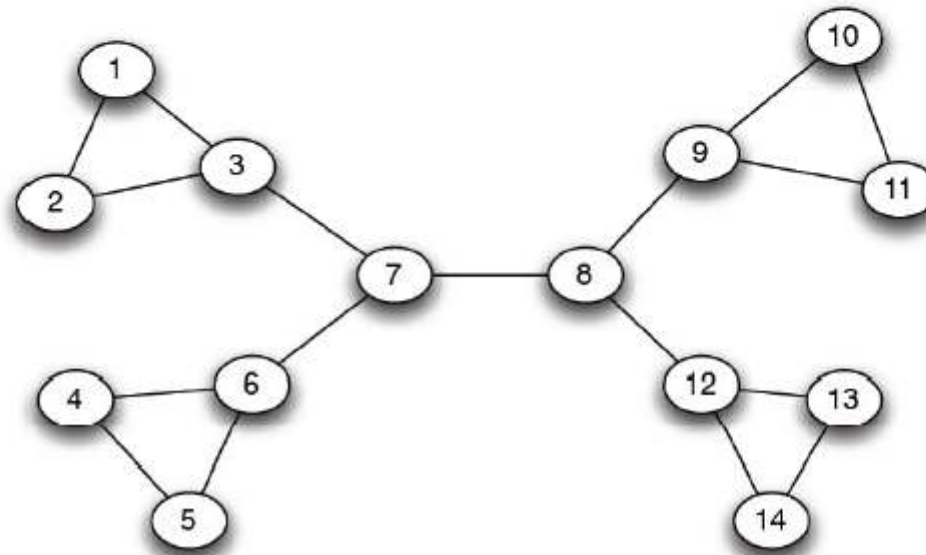
- For each node X:
 - Extract the Ego Network of X
 - Remove X from the Ego Network
 - Discover communities in the Ego Network (easy)
 - Put back X into each discovered community C
- Then, merge the discovered communities bottom-up
 - Coscia, Giannotti, Pedreschi, Rossetti. KDD 2012

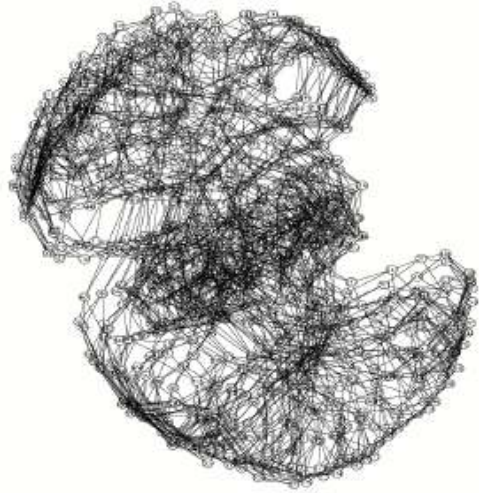
Community discovery

- Challenging task
- Many competing approaches
- Huge literature
- A recent survey:
 - Michele Coscia, Fosca Giannotti, Dino Pedreschi: A classification for community discovery methods in complex networks. *Statistical Analysis and Data Mining* 4(5): 512-546 (2011)

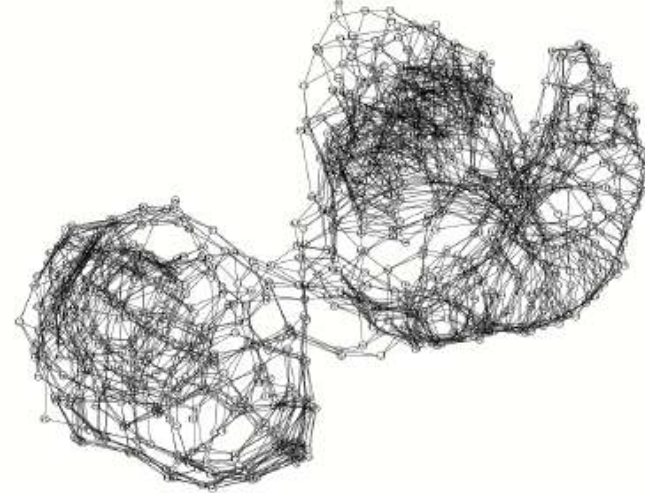
Method 1: Girvan-Newman

- Divisive hierarchical clustering based on the notion of edge **betweenness**:
 - Number of shortest paths passing through the edge
- Remove edges in decreasing betweenness
- Example:

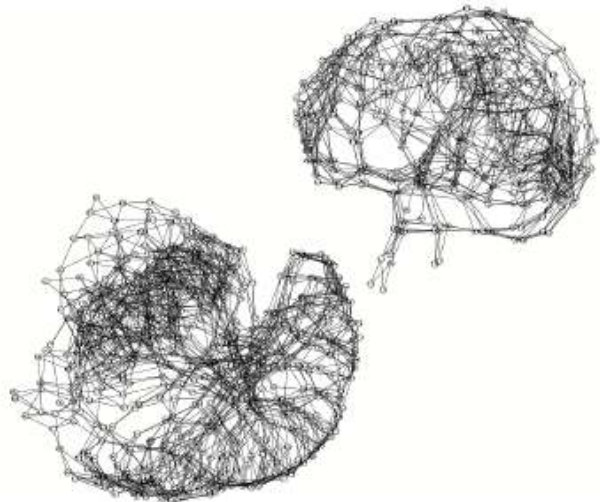




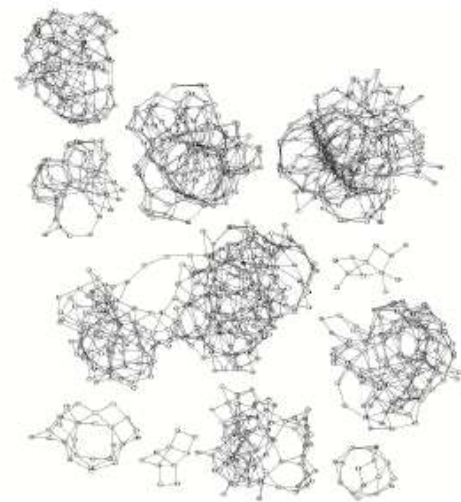
0 cuts



100 cuts



120 cuts



500 cuts

Textbooks and course-ware

Books

- David Easley, Jon Kleinberg: *Networks, Crowds, and Markets*.

<http://www.cs.cornell.edu/home/kleinber/networks-book/>

- M. E. J. Newman: The structure and function of complex networks, *SIAM Review*, Vol. 45, p. 167-256, 2003.

http://didawiki.cli.di.unipi.it/lib/exe/fetch.php/wma/newman_2003.pdf

- A.-L. Barabasi. *Linked*. Plume, 2002

Courses

- Pedreschi + Giannotti @ University of Pisa
 - <http://didawiki.cli.di.unipi.it/doku.php/wma/start>
- Barabasi @ Northeastern University
 - <http://barabasilab.neu.edu/courses/phys5116/>
- Leskovec @ Stanford University
 - <http://www.stanford.edu/class/cs224w/handouts.html>
- Slides from this course are freely adapted from those of Laszlo Barabasi, Jure Leskovec, Fosca Giannotti, besides my own. Thanks!

Knowledge Discovery and Data Mining Laboratory

Web Site: <http://kdd.isti.cnr.it>

Personnel

Lab Head



Cappelli Amedeo



Giannotti Fosca



Pedreschi Dino



Turini Franco

Post Doc



Berlingerio Michele



Pinelli Fabio



Trasarti Roberto

Research Staff



Nanni Mirco



Renso Chiara



Rinzivillo Salvatore



Ruggieri Salvatore

PhD Student



Coscia Michele



Monreale Anna



Ong Rebecca



Pennacchioli Diego



Caterina D'angelo



Schifani Claudio



Falchi Chiara



Qu Zehui



Furletti Barbara, Romei Andrea, Barsocchi Sergio