COMPLEX NETWORKS

WHO AM I

- Rémy Cazabet
- Associate Professor (Maître de conférences)
  - Université Lyon 1
  - LIRIS, DM2L Team (Data Mining & Machine Learning)
- Computer Scientist => Network Scientist
- Member of IXXI

CLASS OVERVIEW

- Previous Lecturer: Marton Karsai
- Lectures: 24h
- Tutorials (TD)
  - 3x2h
  - Lorenza Pacini
- Evaluation:
  - Lectures: Writing exam
  - Tutorials: projects during semester

COMPLEX NETWORKS

WHAT?
WHY?
WHY NOW?
WHAT FOR?

SCIENCE

- Science: understand how things work
  - The human body, the motion/characteristics of objects, societies, etc.
- 1)Experiment with the object (macro-level)
  - What if I throw a ball from that height ? From a moving platform ? If it's a dice ? In wood or in glass ?
  - What if I give this substance to eat/drink ? Is sickness related to cold ? Humidity ? etc.

SCIENCE

- 2)Great success of the 19/20 centuries: **Reductionism**
- To understand things, I need to understand what they are made of:
  - A human body: organs, vessels => cells => DNA, proteins & stuff => Nucleotides ….
  - Objects: Organic compounds => atoms => protons/electrons/neutrons => stuff
- => Now we know. And then what ?

SCIENCE

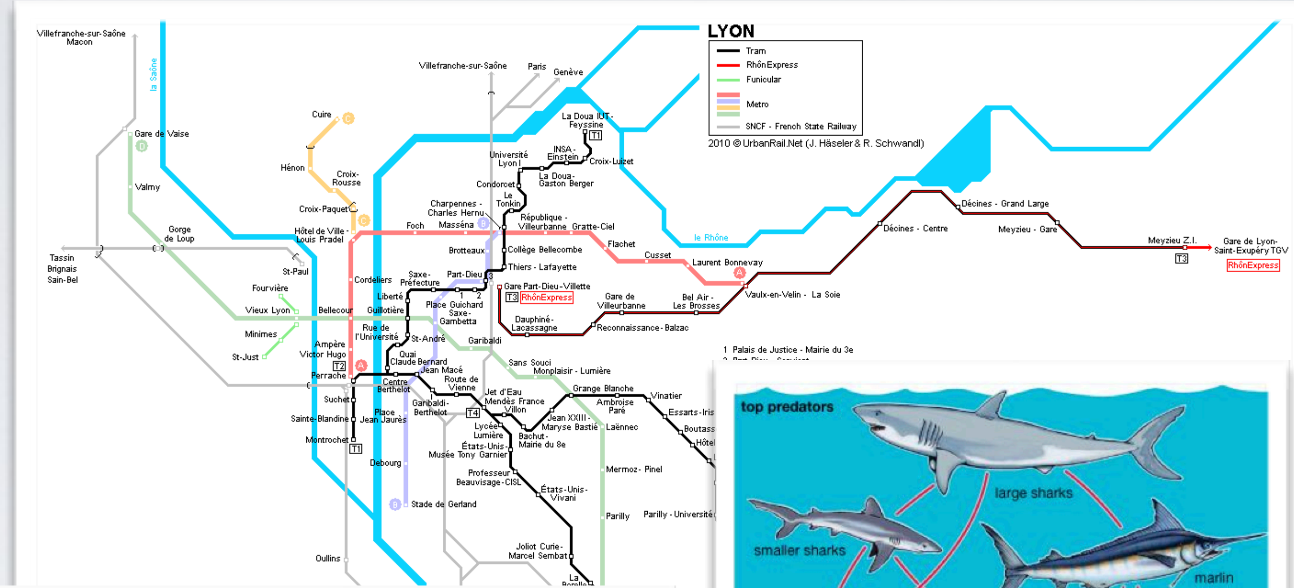
- 3) Two situations:
  - The system is **homogeneous** and/or has a **regular** structure
    - => You can explain it with a bunch of equations
  - The system is **heterogeneous** and/or **has a complex structure**
    - => Understanding each component is not enough to understand the system
    - Understanding each cell tells you little about how the brain works.
    - Understanding how each individual works/behave tells you little about societies
    - etc.

- => The structure/relations/interactions matters.
  - Networks represent structures

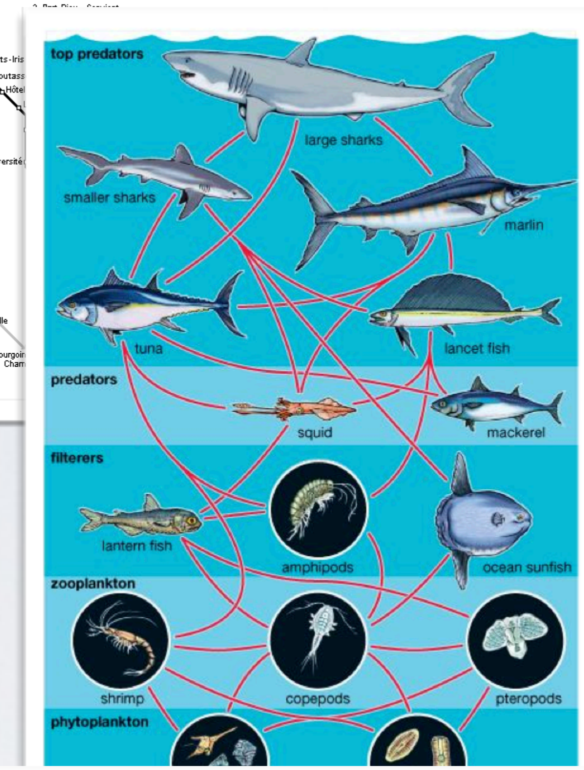
COMPLEX SYSTEMS

- **Complex systems**: Systems composed of multiple **parts** in **interactions**

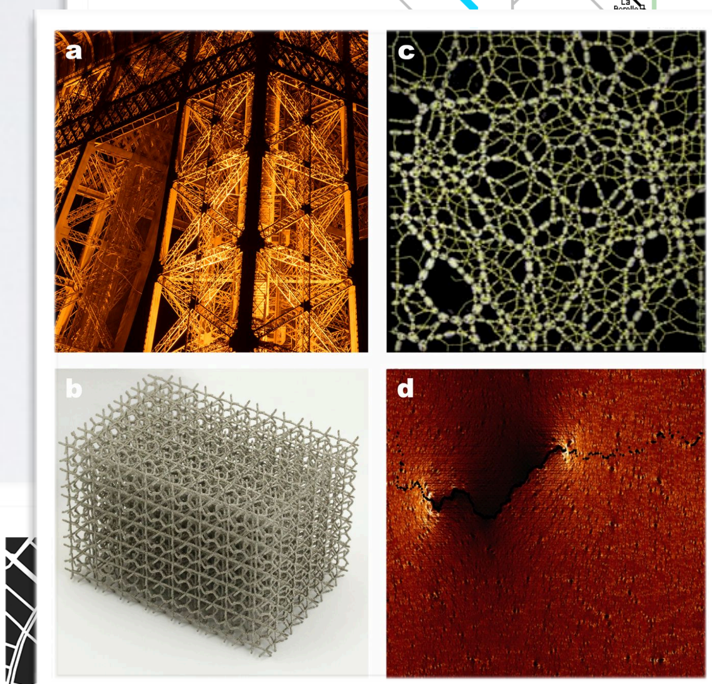
- Complex networks model the interactions between the parts
  - A common framework applicable to many systems
  - =>Many networks share similar characteristics
  - =>Similar processes shape the networks



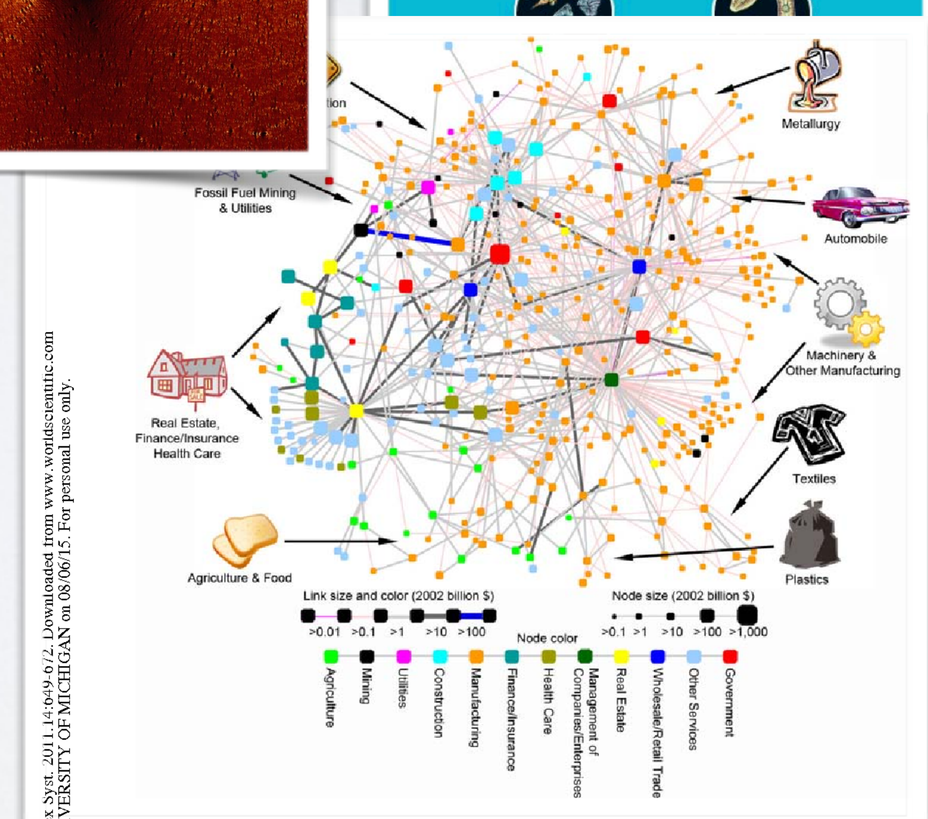
LYON
Tram
RhônExpress
Funicular
Metro
SNCF - French State Railway
2010 © UrbanRail.Net (J. Häseler & R. Schwandl)



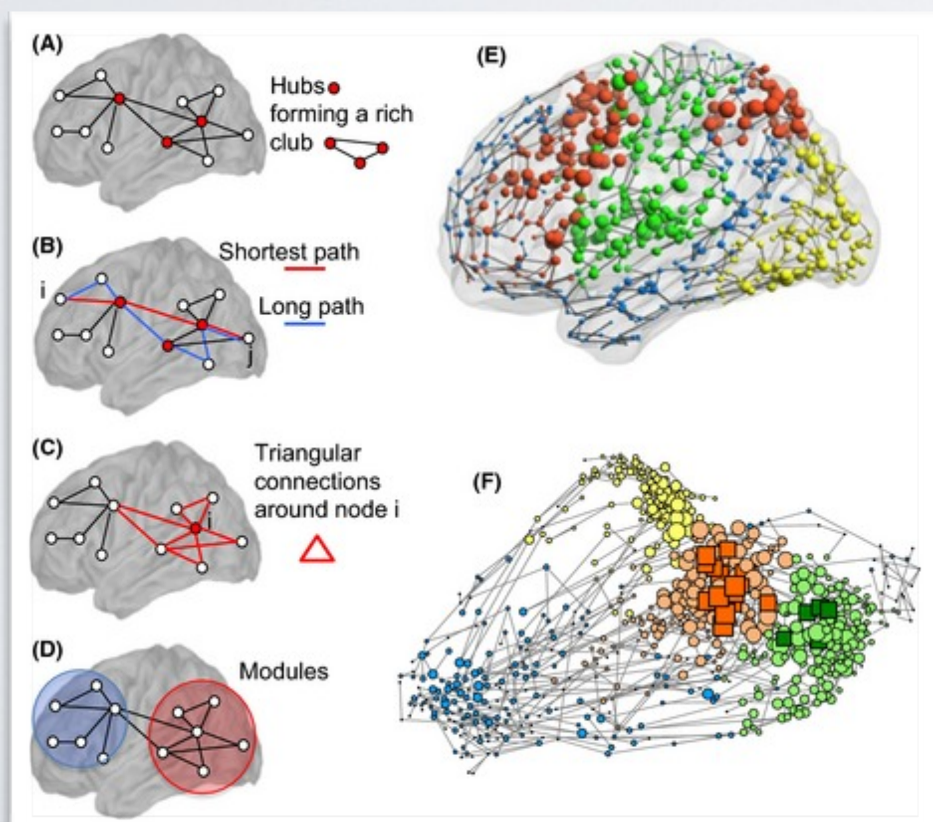
top predators
large sharks
smaller sharks
marlin
tuna
lancet fish
predators
squid
mackerel
filterers
lantern fish
amphipods
ocean sunfish
zooplankton
shrimp
copepods
pteropods
phytoplankton



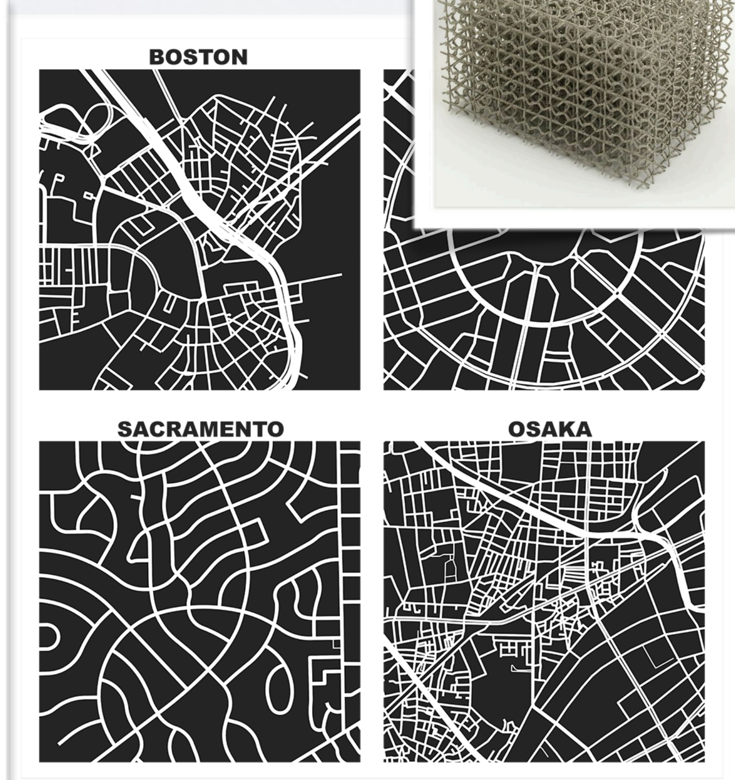
a
c
b
d



Metallurgy
Fossil Fuel Mining & Utilities
Automobile
Machinery & Other Manufacturing
Real Estate, Finance/Insurance Health Care
Textiles
Agriculture & Food
Plastics
Link size and color (2002 billion $)
>0.01 >0.1 >1 >10 >100
Node size (2002 billion $)
>0.1 >1 >10 >100 >1,000
Node color
Agriculture
Mining
Utilities
Construction
Manufacturing
Finance/Insurance
Health Care
Management of Companies/Enterprises
Real Estate
Wholesale/Retail Trade
Other Services
Government
x Syst. 2011.14:649-672. Downloaded from www.worldscientific.com
VERSITY OF MICHIGAN on 08/06/15. For personal use only.



(A)
Hubs forming a rich club
(B)
Shortest path
Long path
i
j
(C)
Triangular connections around node i
(D)
Modules
(E)
(F)



BOSTON
SACRAMENTO
OSAKA

WHO ?

- Network scientists:
  - Physicists
  - Computer scientists
  - Mathematicians
  - => Work on the same problems, with converging vocabularies and references
- Applied network scientists
  - Geographers, biologists, social scientists, etc.
  - =>Experts of i)their domain, and ii)complex networks analysis

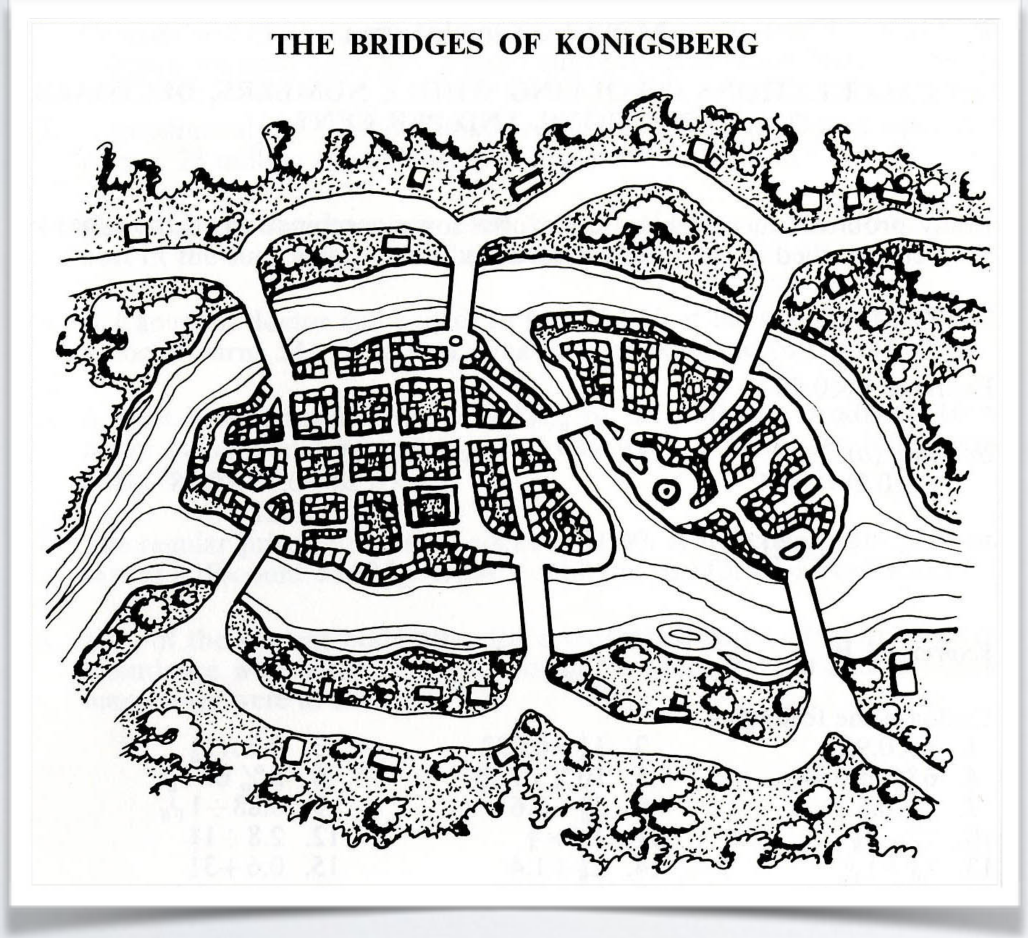
TO CONCLUDE

- Complex Network Analysis *is/should be/will become* (in my opinion) one of the basic tools of the modern scientist (and Data scientist), much as *statistics* or *linear algebra*.

A BRIEF HISTORY

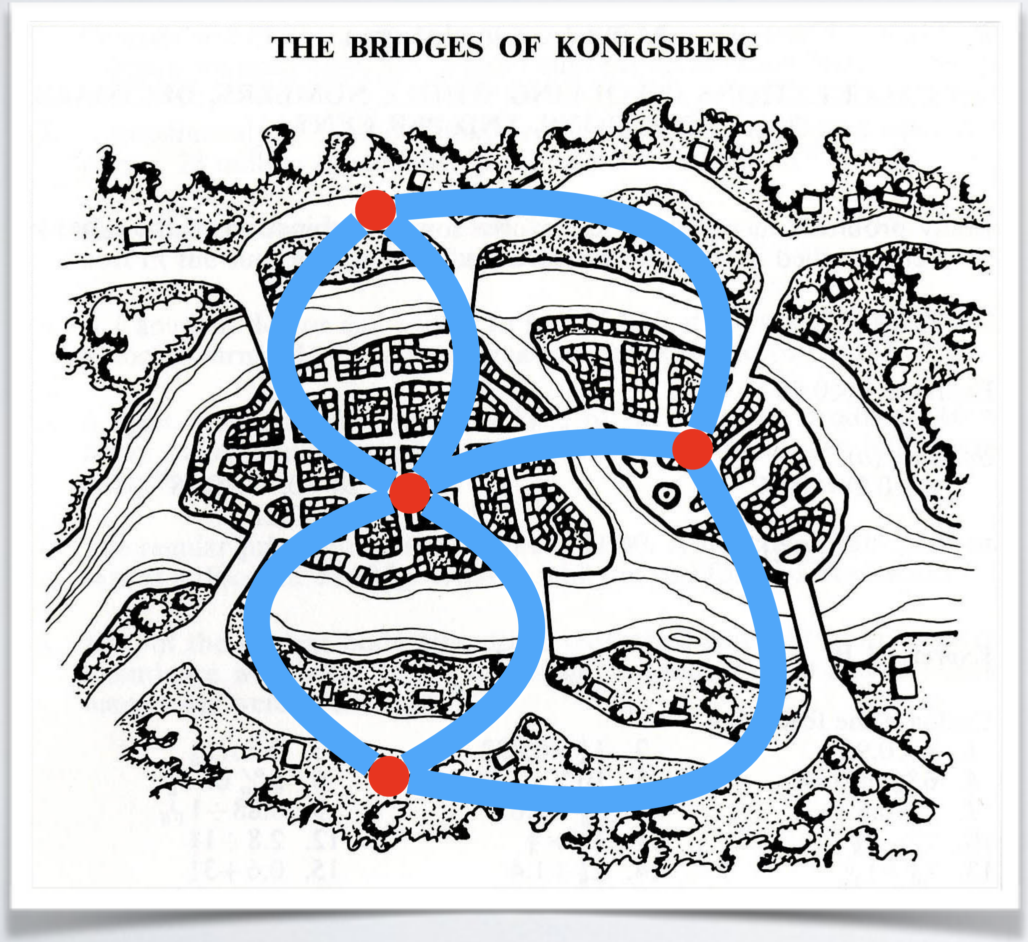
A BRIEF HISTORY

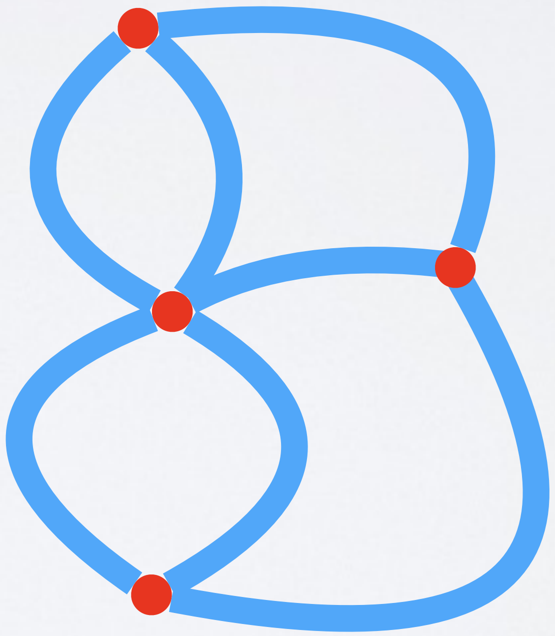
- Graph theory: 1736 - Euler and the bridges of konigsberg



Can one walk across the seven bridges and never cross the same bridge twice?

A BRIEF HISTORY

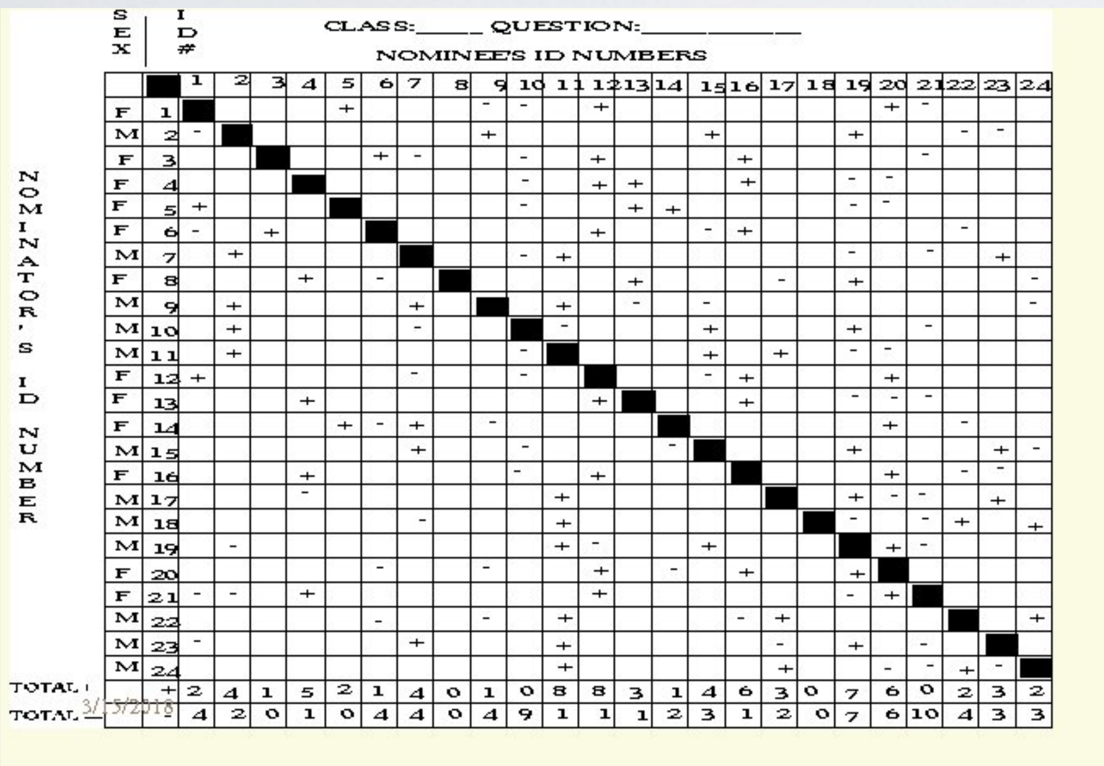




Answer: **No**

A BRIEF HISTORY

- Social networks: 1934 - Jacob Moreno



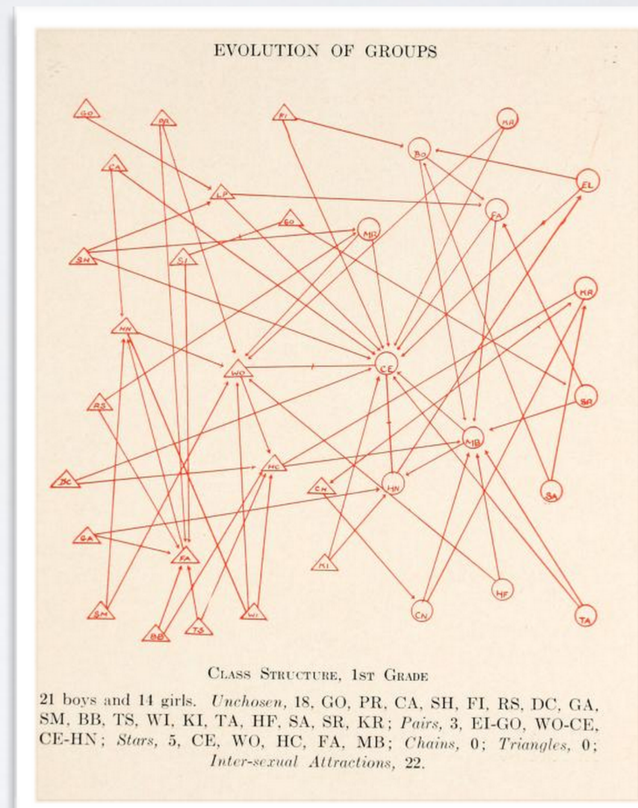
CLASS:____ QUESTION:__________

NOMINEE'S ID NUMBERS

NOMINATOR'S ID NUMBER

SEX	ID #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
F	1	■				+				-	-		+								+	-			
M	2	-	■							+						+				+			-	-	
F	3			■			+	-			-		+				+					-			
F	4				■						-		+	+			+			-	-				
F	5	+				■					-			+	+					-	-				
F	6	-		+			■						+			-	+						-		
M	7		+					■			-	+								-		-		+	
F	8				+		-		■					+				-		+					-
M	9		+					+		■		+		-		-									-
M	10		+					-			■	-				+				+		-			
M	11		+								-	■				+		+		-	-				
F	12	+						-			-		■			-	+				+				
F	13				+								+	■			+			-	-	-			
F	14					+	-	+		-					■						+		-		
M	15							+			-				-	■				+				+	-
F	16				+						-		+				■				+		-	-	
M	17				-							+						■		+	-	-		+	
M	18							-				+							■	-		-	+		+
M	19		-									+	-			+				■	+	-			
F	20						-			-			+		-		+			+	■				
F	21	-	-		+								+							-	+	■			
M	22						-			-		+					-	+					■		+
M	23	-						+				+						-		+		-		■	
M	24											+						+			-	-	+	-	■
TOTAL	+	2	4	1	5	2	1	4	0	1	0	8	8	3	1	4	6	3	0	7	6	0	2	3	2
TOTAL	-	4	2	0	1	0	4	4	0	4	9	1	1	1	2	3	1	2	0	7	6	10	4	3	3

Sociomatrix



CLASS STRUCTURE, 1ST GRADE

21 boys and 14 girls. *Unchosen*, 18, GO, PR, CA, SH, FI, RS, DC, GA, SM, BB, TS, WI, KI, TA, HF, SA, SR, KR; *Pairs*, 3, EI-GO, WO-CE, CE-HN; *Stars*, 5, CE, WO, HC, FA, MB; *Chains*, 0; *Triangles*, 0; *Inter-sexual Attractions*, 22.

Sociograms

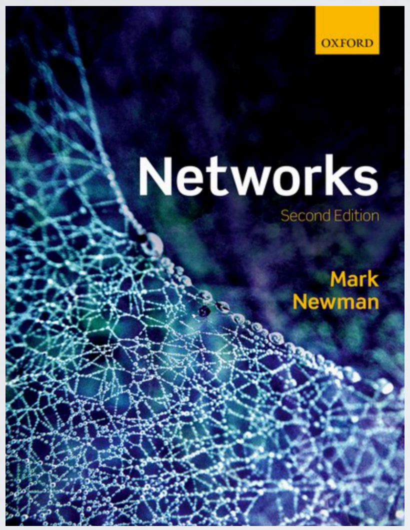
KEY PUBLICATIONS

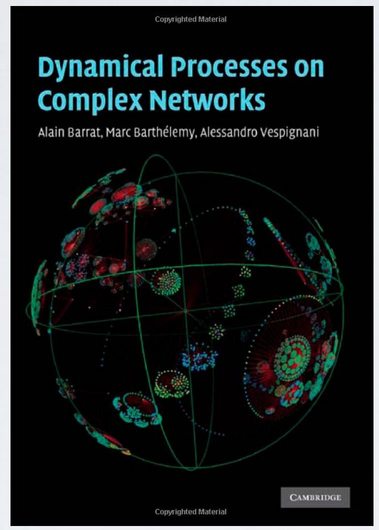
- 1998: Watts & Strogatz - Small-World:
  - 2nd Most cited paper of the year in Nature
- 1999: Barabasi & Albert - scale-free networks:
  - Most cited paper of the year in Science
- 2002: Girvan & Newman - Community detection:
  - Most cited paper of the year in PNAS
- 2004: Barabasi & Oltvai - Network Biology:
  - Most cited paper (ever) in Nature genetics
- 2010: Kwak et al. - What is Twitter, a Social Network or a News Media?
  - Most cited paper (ever) of the WWW conference
- …

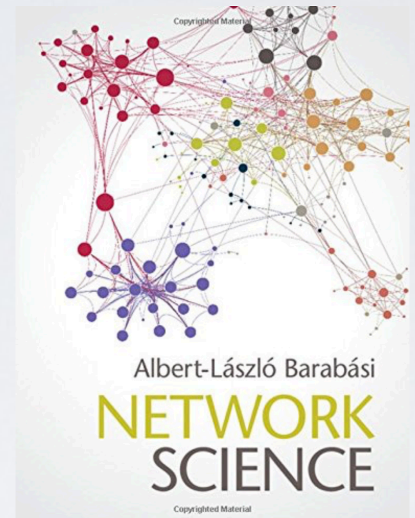
(As of 2019)

Materials

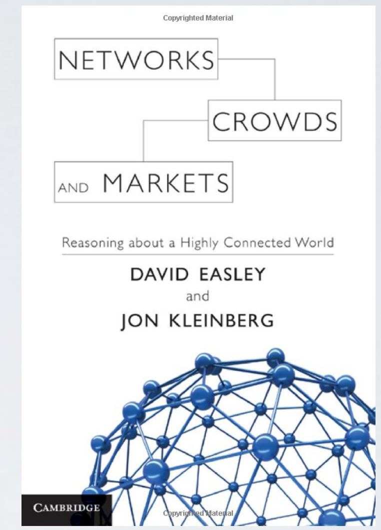
Lecture books







available free online



available free online

Reviews

SIAM REVIEW
Vol. 45, No. 2, pp. 167–256
© 2003 Society for Industrial and Applied Mathematics

The Structure and Function of Complex Networks*

M. E. J. Newman†

REVIEWS OF MODERN PHYSICS, VOLUME 74, JANUARY 2002

Statistical mechanics of complex networks

Réka Albert* and Albert-László Barabási
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556

Characterization and Modeling of weighted networks

Marc Barthélemy[1], Alain Barrat[2], Romualdo Pastor-Satorras[3], and Alessandro Vespignani[2]

Physics Reports 486 (2010) 75–174

Contents lists available at ScienceDirect
ELSEVIER
Physics Reports
journal homepage: www.elsevier.com/locate/physrep

Community detection in graphs

Santo Fortunato *
Complex Networks and Systems Lagrange Laboratory, ISI Foundation, Viale S. Severo 65, 10133, Torino, I, Italy

Physics Reports 519 (2012) 97–125

Contents lists available at SciVerse ScienceDirect
ELSEVIER
Physics Reports
journal homepage: www.elsevier.com/locate/physrep

Temporal networks

Petter Holme [a,b,c,*], Jari Saramäki [d]
[a] *IceLab, Department of Physics, Umeå University, 901 87 Umeå, Sweden*
[b] *Department of Energy Science, Sungkyunkwan University, Suwon 440-746, Republic of Korea*
[c] *Department of Sociology, Stockholm University, 106 91 Stockholm, Sweden*
[d] *Department of Biomedical Engineering and Computational Science, School of Science, Aalto University, 00076 Aalto, Espoo, Finland*

Contents lists available at ScienceDirect
ELSEVIER
Physics Reports
journal homepage: www.elsevier.com/locate/physrep

Spatial networks

Marc Barthélemy *

Contents lists available at ScienceDirect
ELSEVIER
Physics Reports
journal homepage: www.elsevier.com/locate/physrep

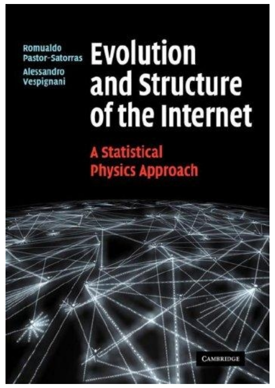
The structure and dynamics of multilayer networks

S. Boccaletti [a,b,*], G. Bianconi [c], R. Criado [d,e], C.I. del Genio [f,g,h], J. Gómez-Gardeñes [i], M. Romance [d,e], I. Sendiña-Nadal [j,e], Z. Wang [k,l], M. Zanin [m,n]

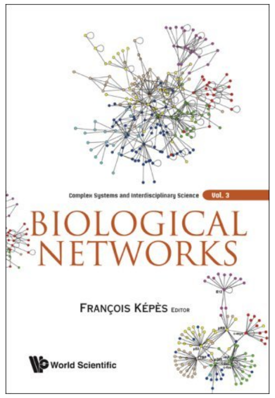
…and many more…all of them on arXiv.org!

Materials

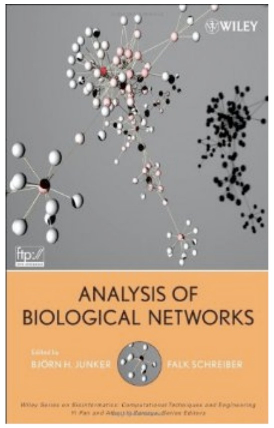
Related books



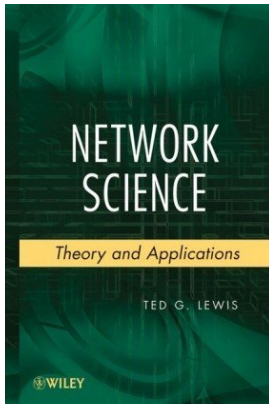
R. Pastor-Satorras, A. Vespignani, Evolution and Structure of the Internet: A Statistical Physics Approach (Cambridge University Press, 2007), rst edn.



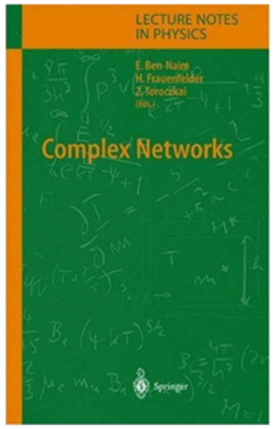
F. Kopos, Biological Networks (Complex Systems and Interdisciplinary Science) (World Scientific Publishing Company, 2007), rst edn.



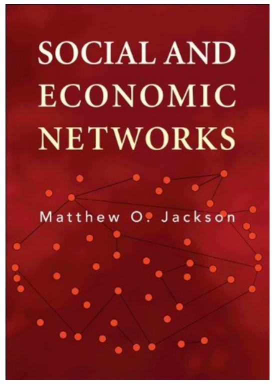
B. H. Junker, F. Schreiber, Analysis of Biological Networks (Wiley Series in Bioinformatics) (Wiley-Interscience, 2008).



T. G. Lewis, Network Science: Theory and Applications (Wiley, 2009).



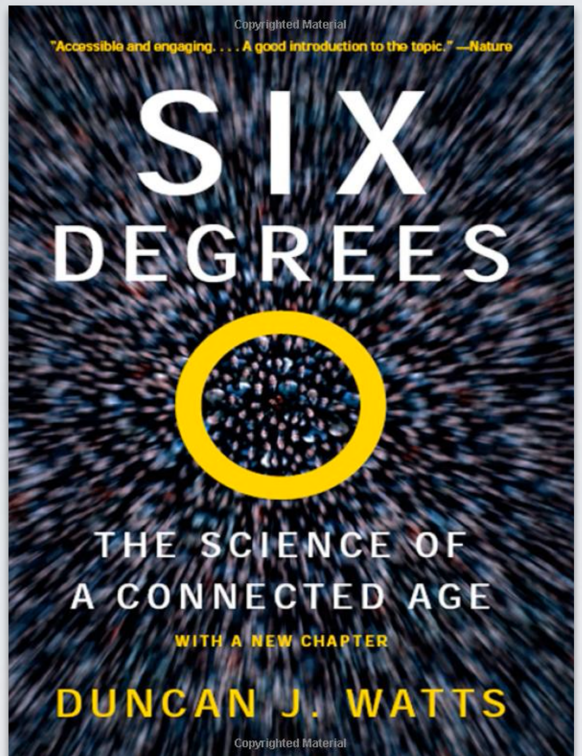
E. Ben Naim, H. Frauenfelder, Z.Torotzai, Complex Networks (Lecture Notes in Physics) (Springer, 2010), rst edn.

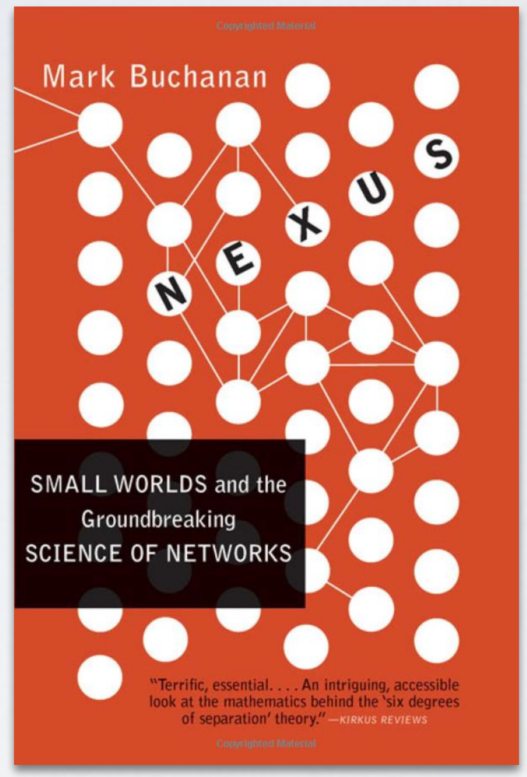


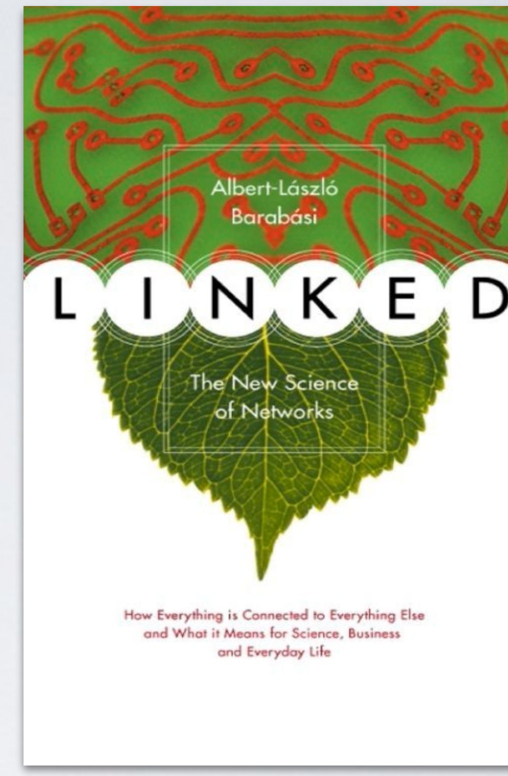
M. O. Jackson, Social and Economic Networks (Princeton University Press, 2010).

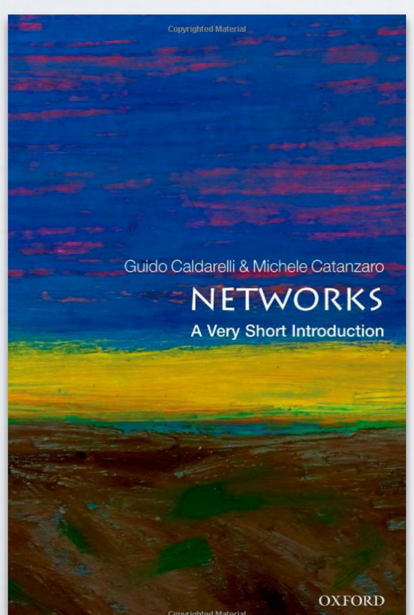
Materials

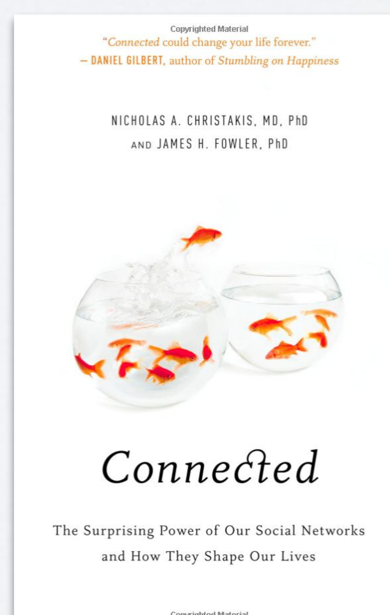
Pop-science books

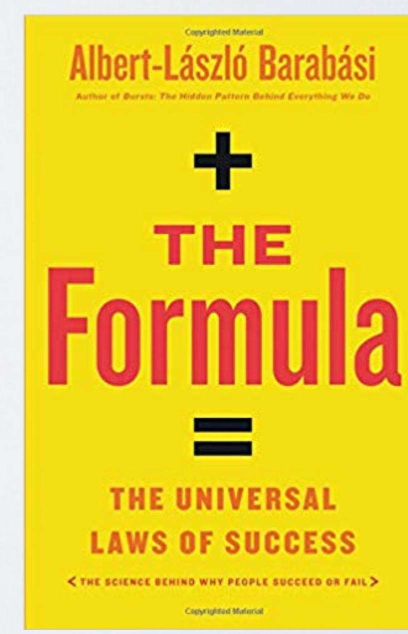






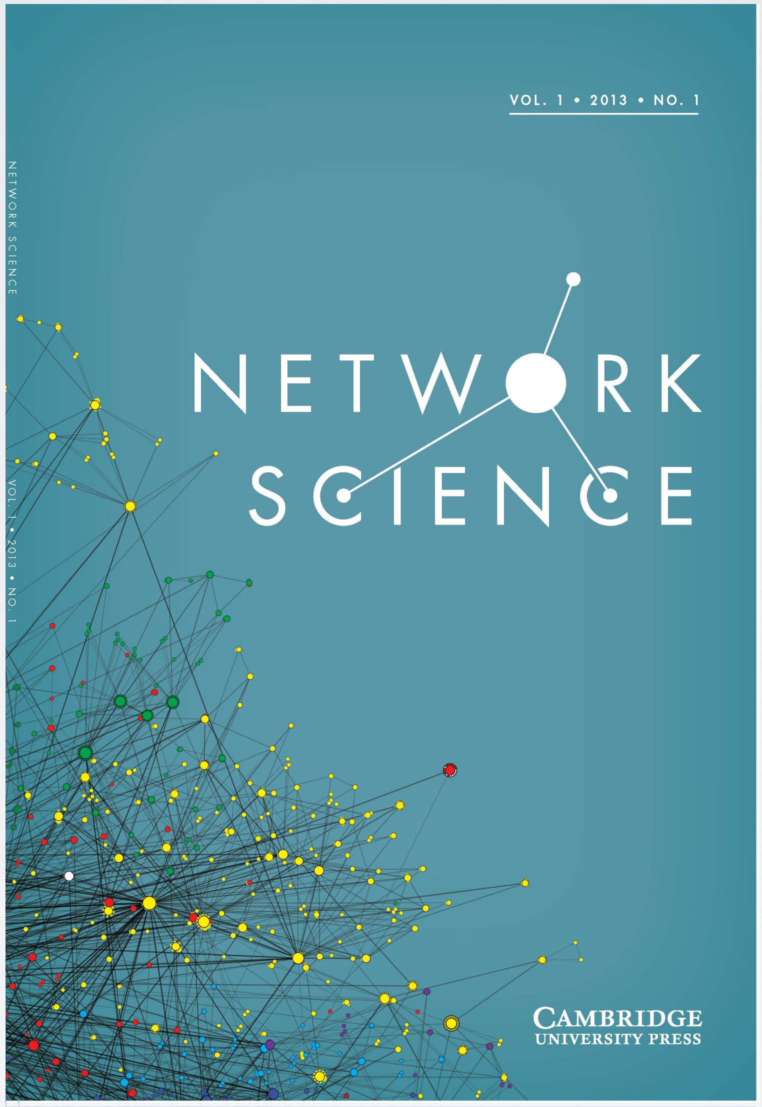


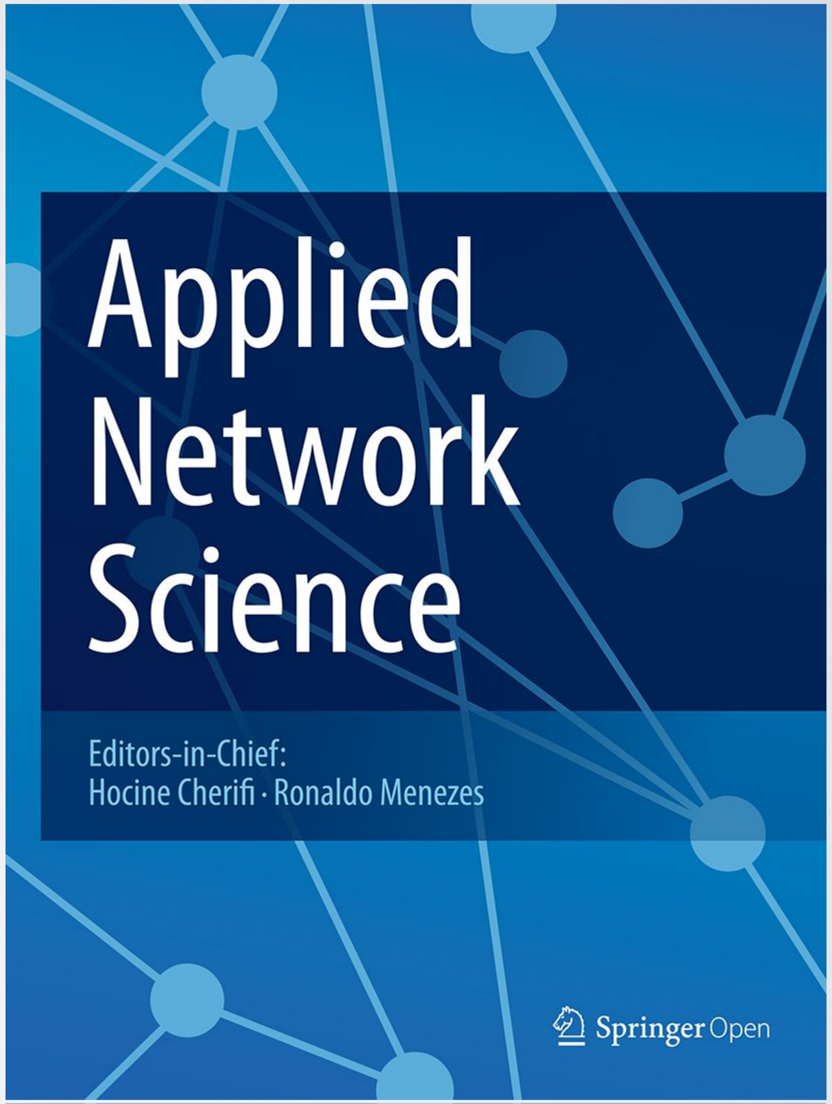




Materials

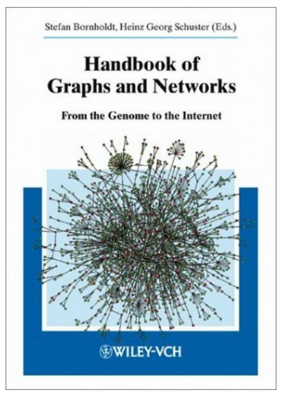
Journals



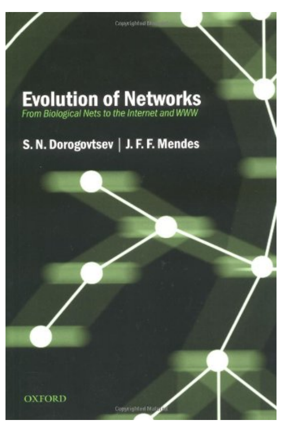


Materials

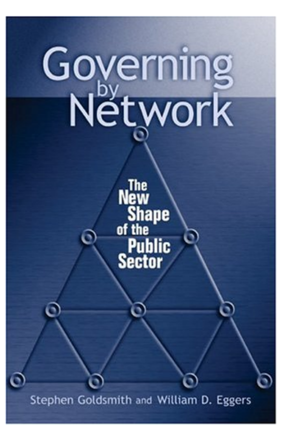
Related books



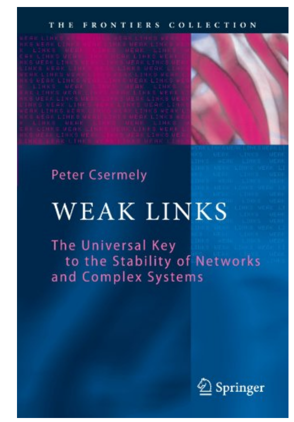
Handbook of Graphs and Networks: From the Genome to the Internet (Wiley-VCH, 2003).



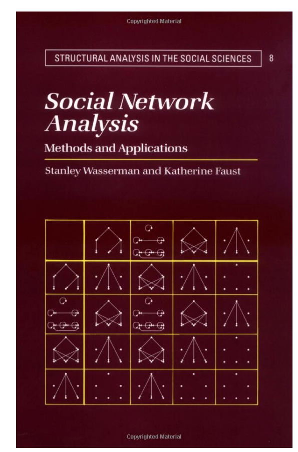
S. N. Dorogovtsev and J. F. F. Mendes, Evolution of Networks: From Biological Nets to the Internet and WWW (Oxford University Press, 2003).



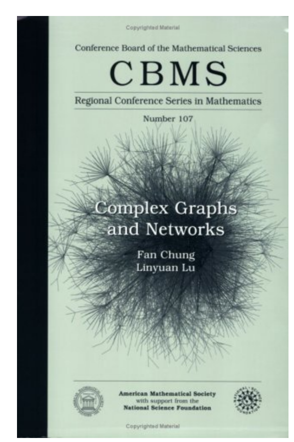
S. Goldsmith, W. D. Eggers, Governing by Network: The New Shape of the Public Sector (Brookings Institution Press, 2004).



P. Csermely, Weak Links: The Universal Key to the Stability of Networks and Complex Systems (The Frontiers Collection) (Springer, 2006), rst edn.



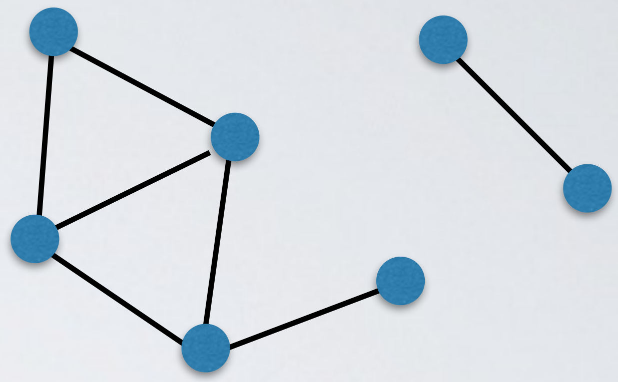
S. Wasserman and K. Faust
Social Network Analysis (Methods and Applications)
Cambridge University Press (1994)



L. L. F. Chung, Complex Graphs and Networks (CBMS Regional Conference Series in Mathematics) (American Mathematical Society, 2006).

GRAPHS & NETWORKS

GRAPHS & NETWORKS



Networks often refers to real systems

- www,
- social network
- metabolic network.
- <u>Language: (Network, node, link)</u>

Graph is the mathematical representation of a network

- <u>Language: (Graph, vertex, edge)</u>

Vertex	Edge
person	friendship
neuron	synapse
Website	hyperlink
company	ownership
gene	regulation

In most cases we will use the two terms interchangeably.

NETWORK REPRESENTATIONS

NETWORK REPRESENTATIONS

- $G = (V, E)$
  - edge: $(u, v) \in E$
  - Often encoded as **edge list** or **adjacency list**
  - Software: custom data structure and manipulation
    - add_nodes([i,j]), add_edge(i,j), …

```
1 2
2 3
2 4
3 4
4 5
4 7
5 6
5 8
9 10
```

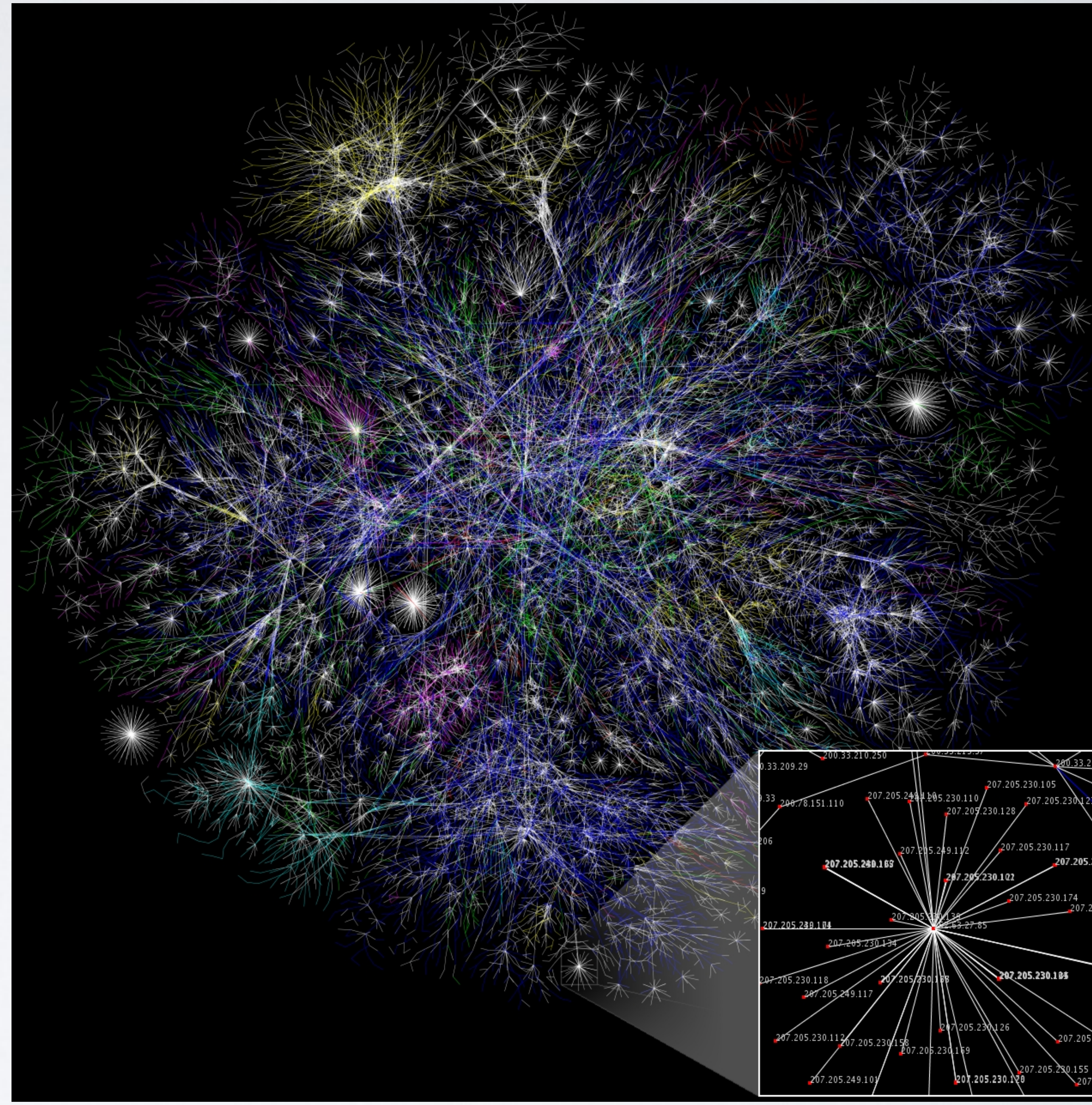
```
1    2
2    1 3 4
3    2 4
4    2 3 5 7
5    4 6 8
6    5
7    4
8    5
9    10
10   9
```

- Adjacency Matrix A
  - Edge: A_{ij}
  - Graph Laplacian $L = D - A$ with D the degree matrix
    - Powerful graph spectral properties, more later

Types of Networks

Undirected networks

Opte project

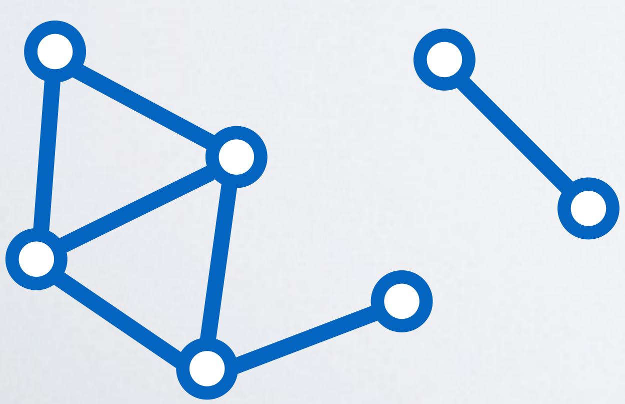


The Internet: Nodes - routers, Links - physical wires

$G=(V, E)$

$(u,v) \in E \equiv (v,u) \in E$

- The directions of edges do not matter
- Interactions are possible between connected entities in both directions

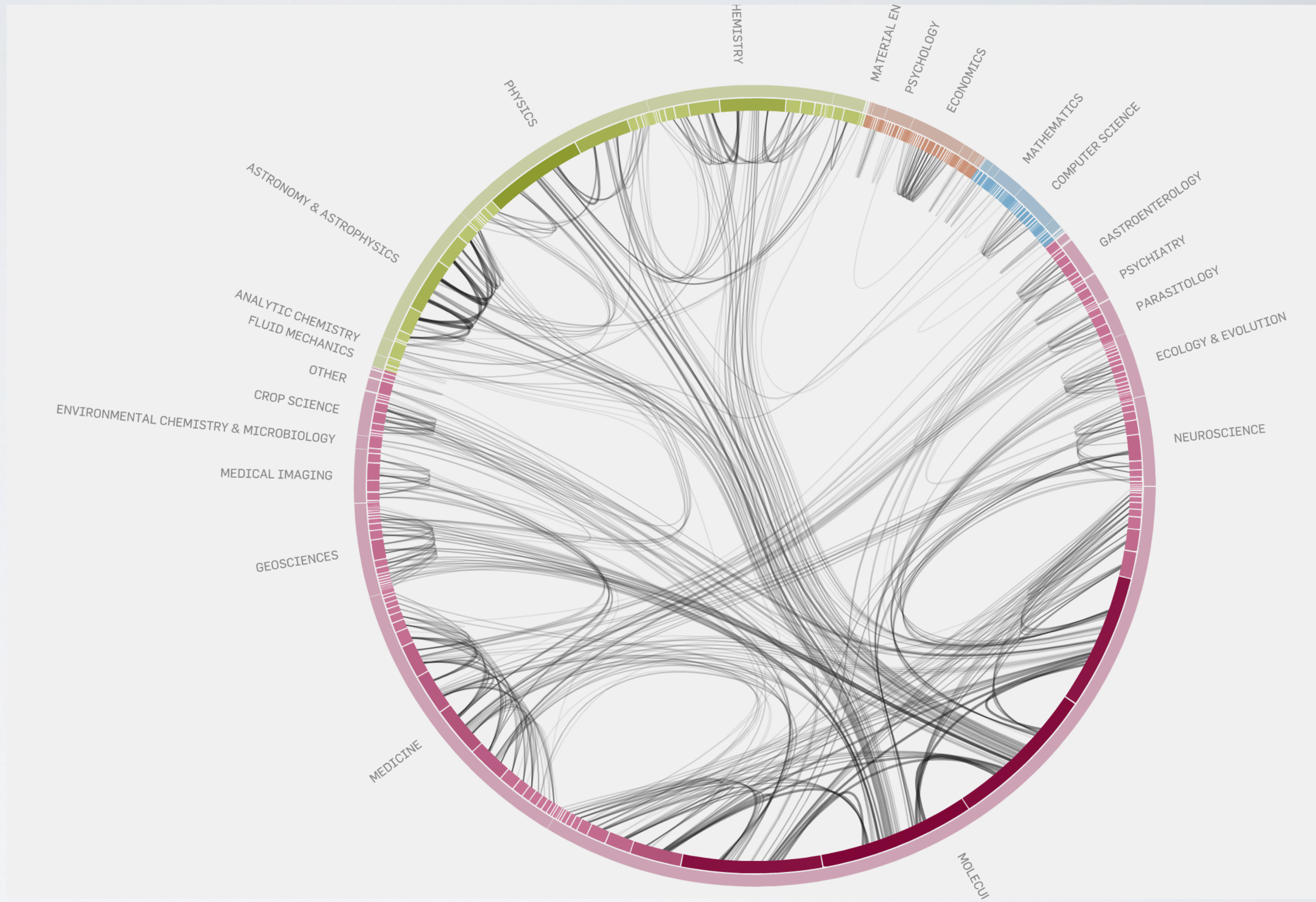


Directed networks

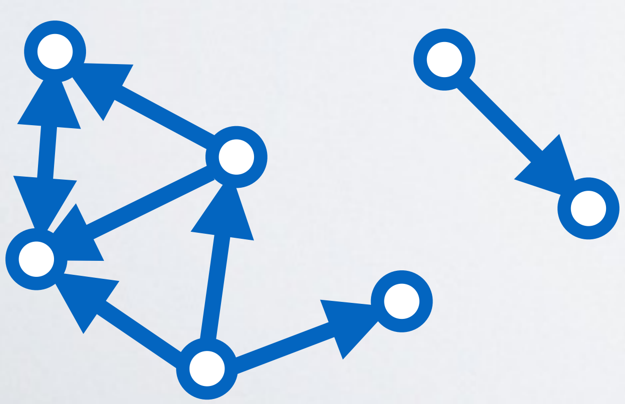
Moritz Stefaner, eigenfactor.com

$G=(V, E)$

$(u,v) \in E \not\equiv (v,u) \in E$



- The directions of edges matter
- Interactions are possible between connected entities only in specified directions



Citation network: Nodes - publications, Links - references

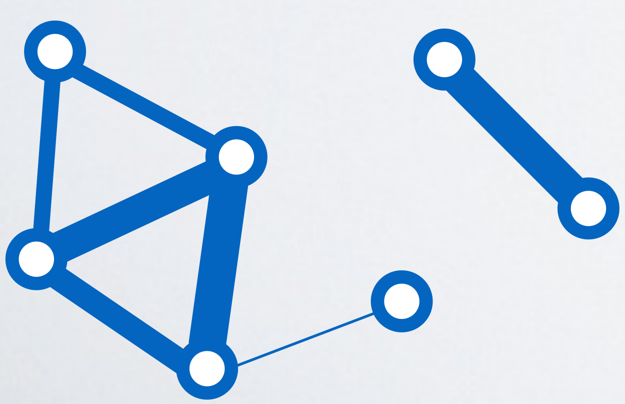
Weighted networks

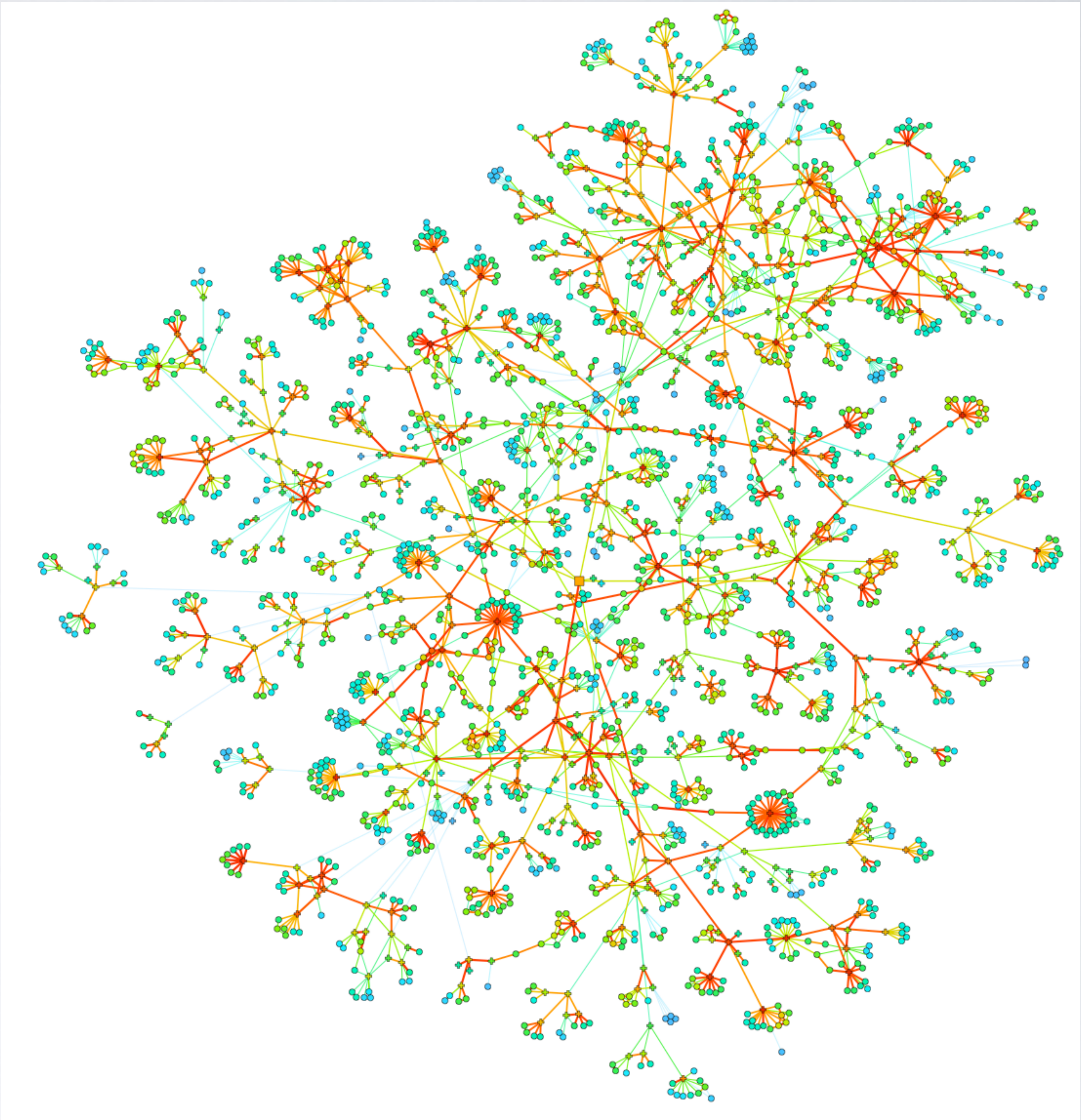
Onnela et.al. New Journal of Physics 9, 179 (2007).

$G=(V, E, w)$

$w: (u,v) \in E \Rightarrow R$

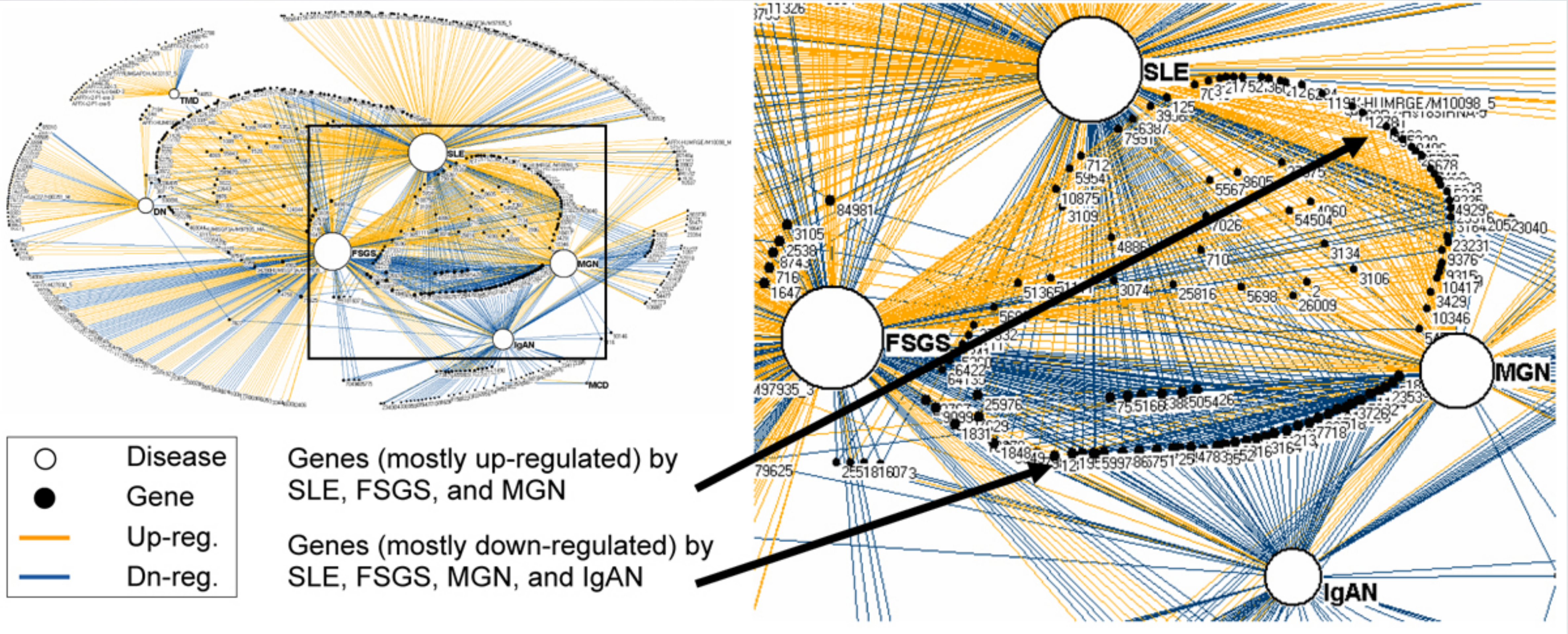
- Strength of interactions are assigned by the weight of links



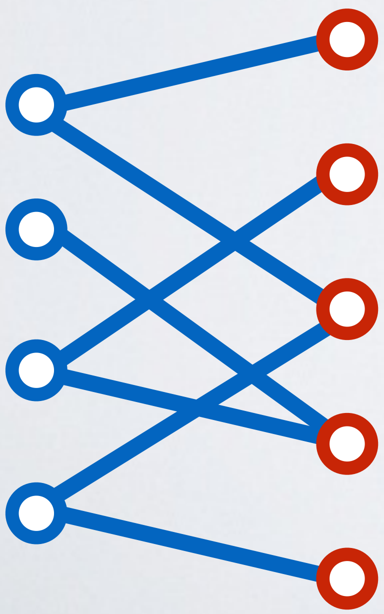


Social interaction network: Nodes - individuals
Links - social interactions

Bipartite network



Bhavnani et.al. BMC Bioinformatics 2009, **10**(Suppl 9):S3



Gene-desease network:

Nodes - Desease (7)&Genes (747)

Links - gene-desease relationship

$$G=(U, V, E)$$

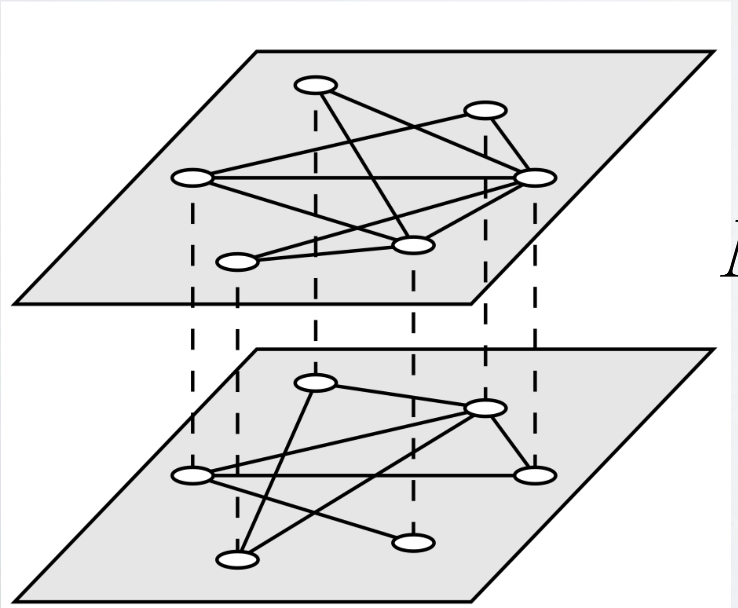
$$U \cap V = \emptyset$$

$$\forall (u,v) \in E,\ u \in U \ and \ v \in V$$

Multiplex and multilayer networks

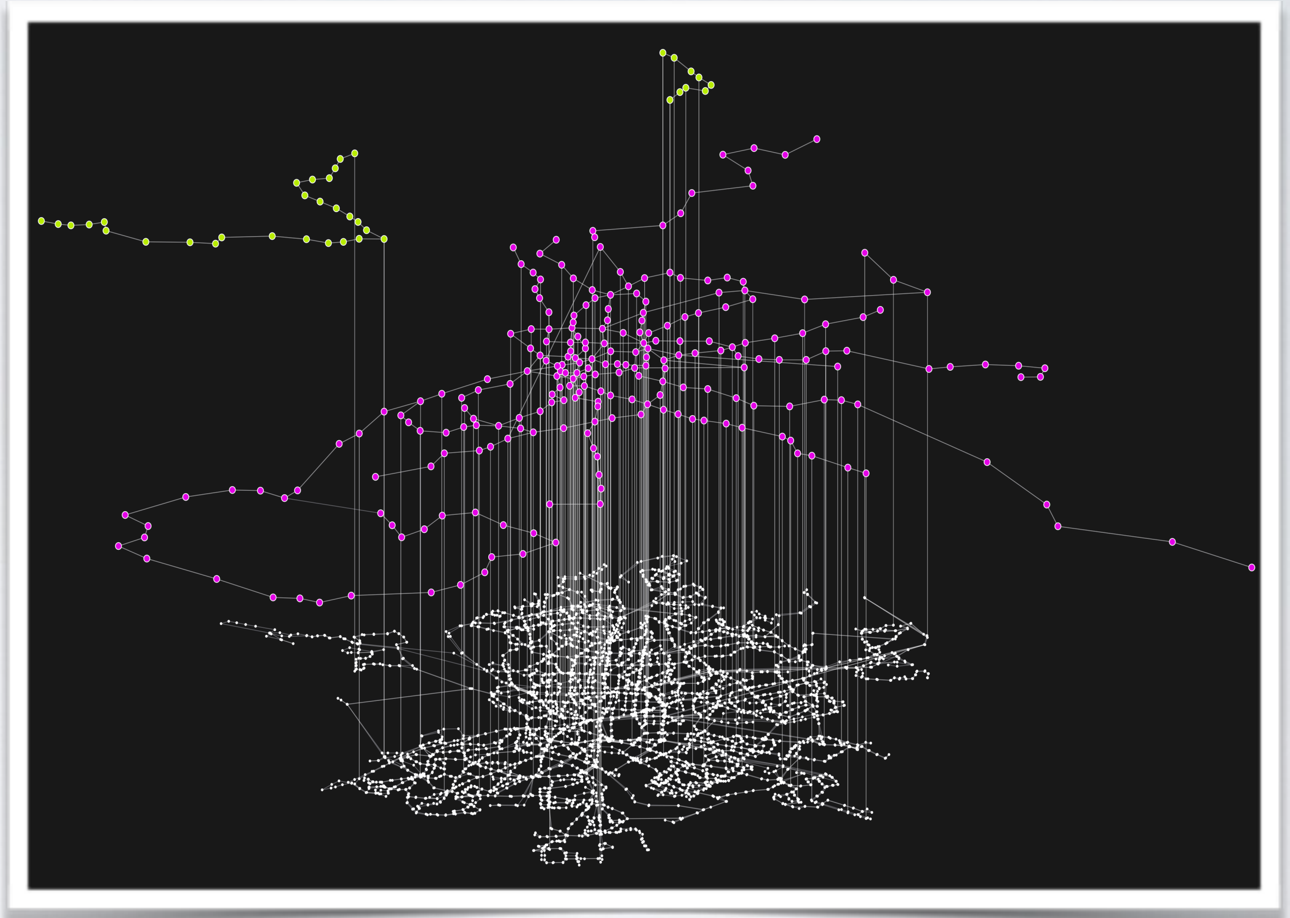
$G=(V, E_i),\ i=1\ldots M$

- Nodes can be present in multiple networks simultaneously
- These networks are connected (can influence each other) via the common nodes



$M=2$

Gomes et.al. Phys. Rev. Lett. 110, 028701 (2013)



[Mendez-Bermudez et al. 2017]

Temporal and evolving networks

$G=(V, E_t), (u,v,t,d) \in E_t$

t - time of interaction (u,v)
d - duration of interaction (u,v,t)

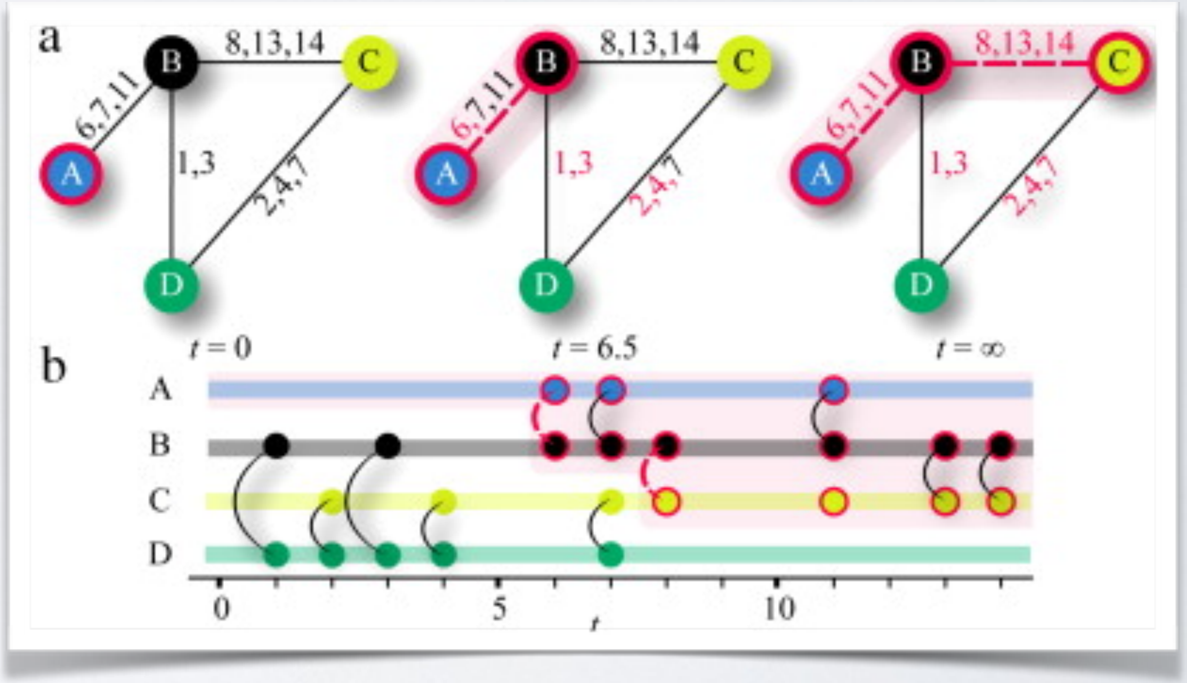
- Temporal links encode time varying interactions

$G=(V_{t'}, E_{t'})$

$v(t) \in V_{t'}$

$(u,v,t) \in E_{t'}$

- Dynamical nodes and links encode the evolution of the network



Mobile communication network
Nodes - individuals
Links - calls and SMS

DESCRIPTION OF GRAPHS

DESCRIPTION OF GRAPHS

- When confronted with a graph, how to describe it?
- How to compare graphs?
- What can we say about a graph?

SIZE

- A network is composed of nodes and edges.
- Size: How many nodes and edges ? (n & m)

	#nodes (n)	#edges (m)
Wikipedia HL	2M	30M
Twitter 2015	288M	60B
Facebook 2015	1.4B	400B
Brain c. Elegans	280	6393
Roads US	2M	2.7M
Airport traffic	3k	31k

DENSITY

Defined as:

Directed $$D = \frac{|E|}{|V|\,(|V|-1)}$$

Undirected $$D = \frac{2|E|}{|V|\,(|V|-1)}$$

Often more relevant: average degree (2|E| / |V|)

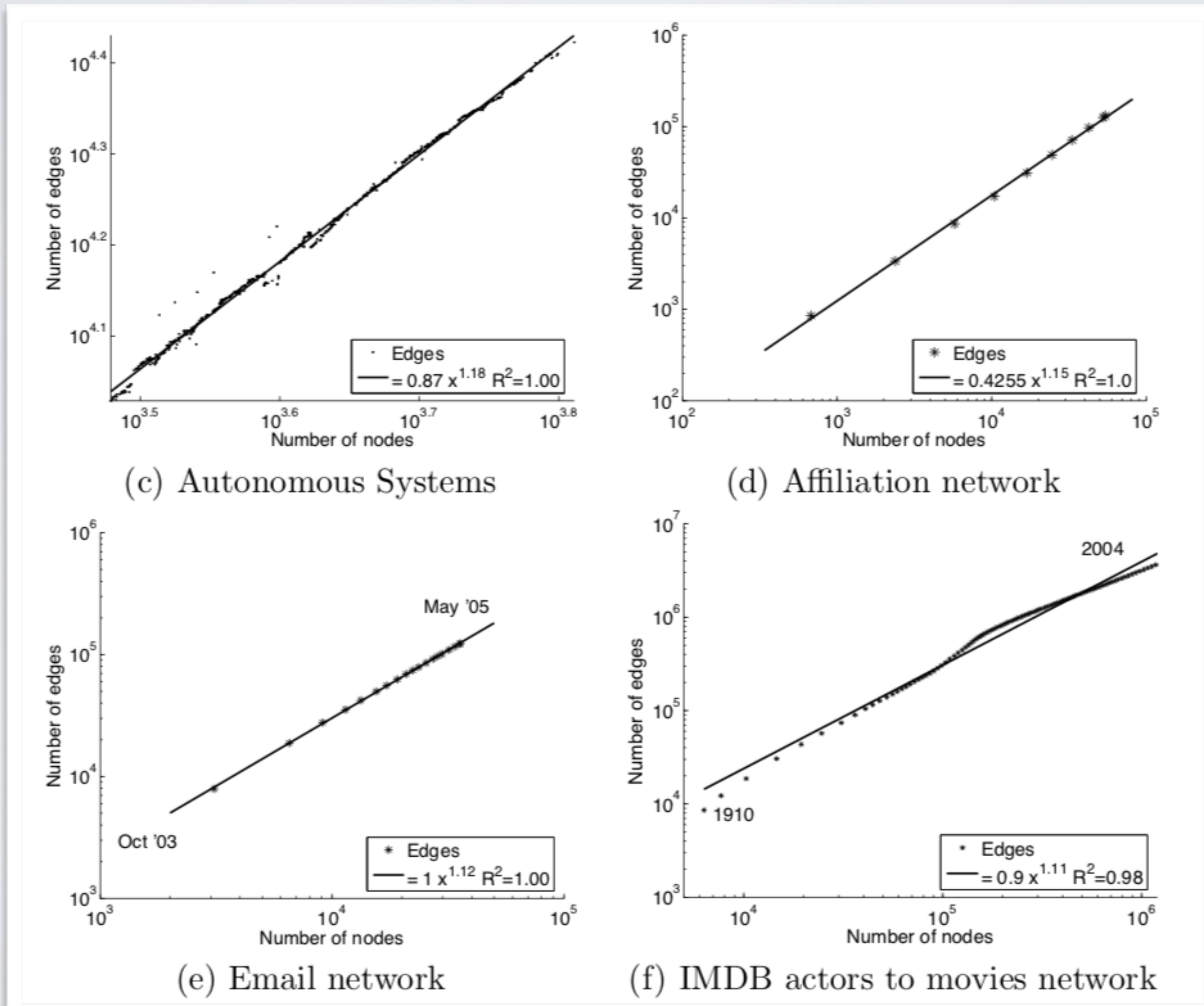
	#nodes	#edges	Density	avg. deg
Wikipedia	2M	30M	1.5×10^{-5}	30
Twitter 2015	288M	60B	1.4×10^{-6}	416
Facebook	1.4B	400B	4×10^{-9}	570
Brain c.	280	6393	0.16	46
Roads Calif.	2M	2.7M	6×10^{-7}	2.7
Airport	3k	31k	0.007	21

DENSITY

- It has been observed that: [Leskovec. 2006]
  - When graphs increase in size, the average degree increases
  - This increase is very slow

- Think of friends in a social network

DENSITY



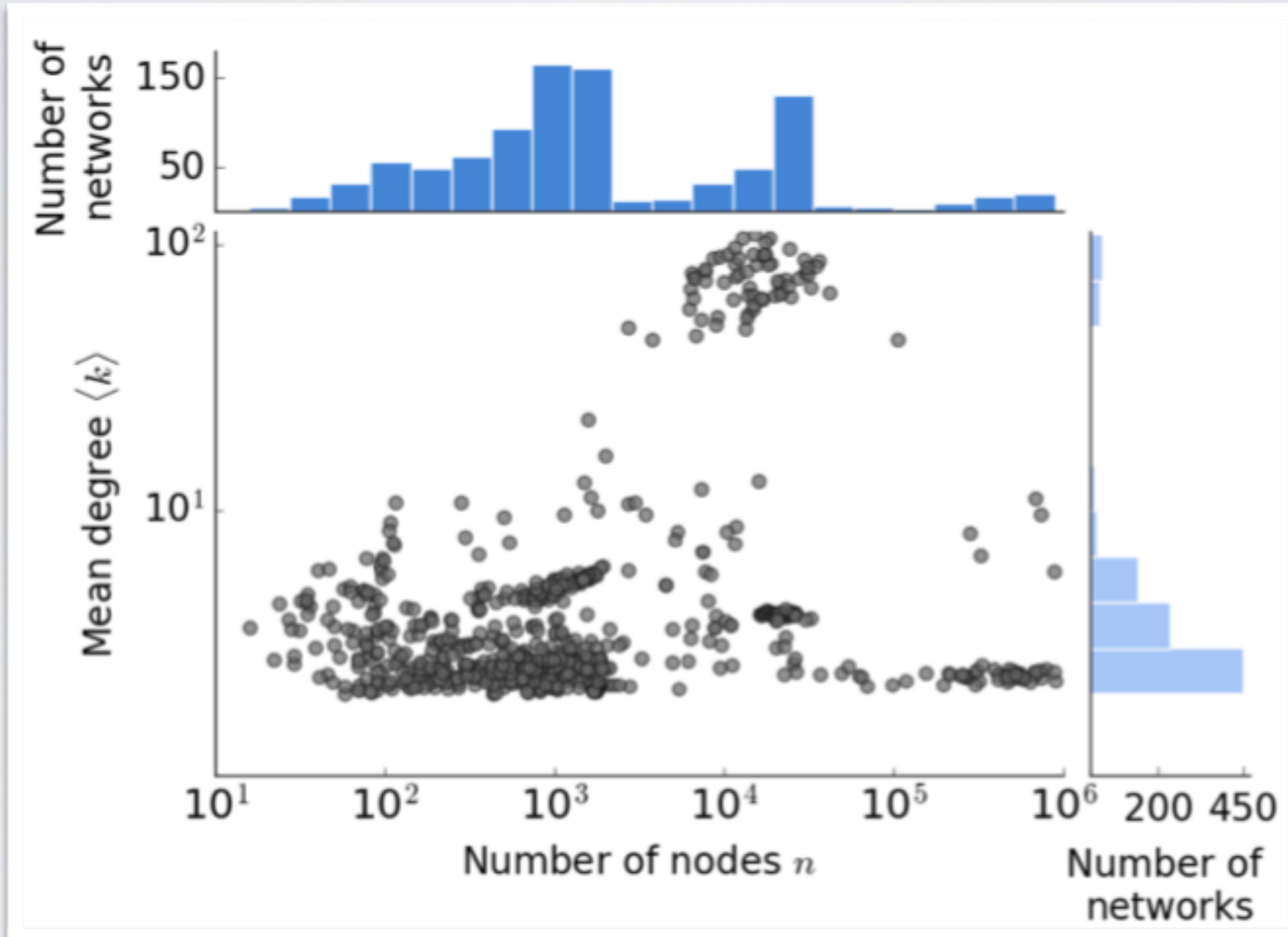
(c) Autonomous Systems

(d) Affiliation network

(e) Email network

(f) IMDB actors to movies network

DENSITY

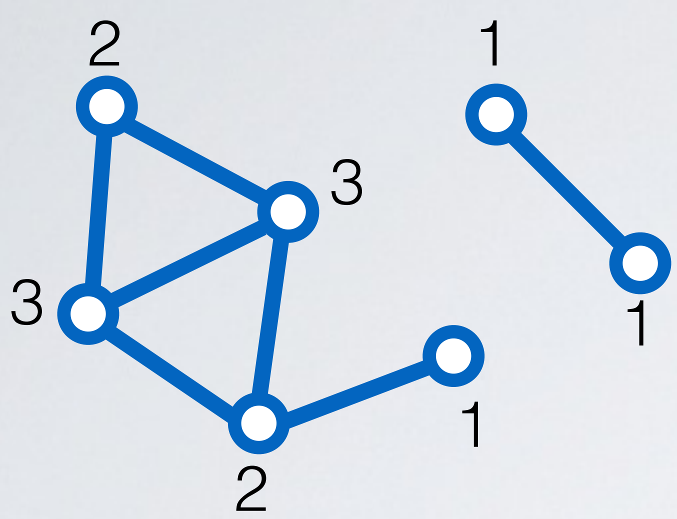


[Broido, Clauset 2018]

Node degree

Number of connections of a node

- Undirected network



$$k_i = A_{i1} + A_{i2} + \ldots + A_{iN} = \sum_{j}^{N} A_{ij}$$

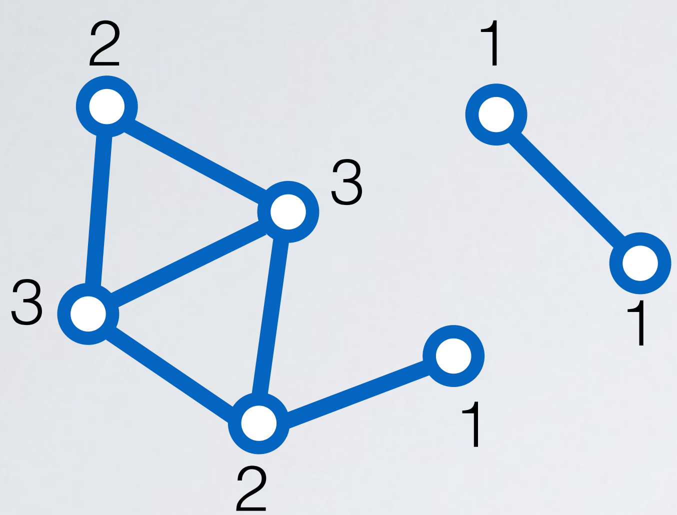
$$m = \frac{\sum_i k_i}{2}$$ where $$m = |E|$$

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0	0
4	0	1	1	0	1	0	1	0	0	0
5	0	0	0	1	0	1	0	1	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0

Node degree

Number of connections of a node

- Undirected network



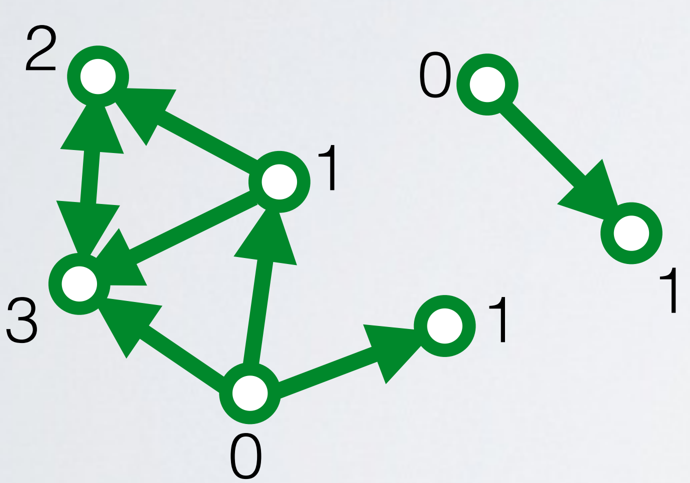
$$k_i = A_{i1} + A_{i2} + \dots + A_{iN} = \sum_{j}^{N} A_{ij}$$

$$m = \frac{\sum_i k_i}{2} \quad \text{where} \quad m = |E|$$

mean degree $$\langle k \rangle = \frac{1}{N} \sum_{i}^{N} k_i$$

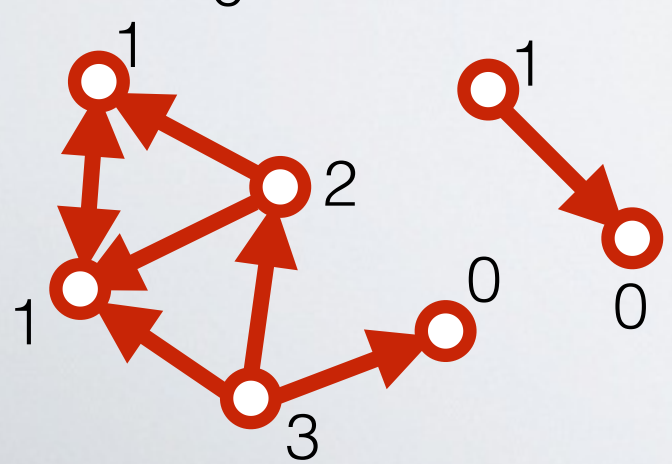
	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0	0
4	0	1	1	0	1	0	1	0	0	0
5	0	0	0	1	0	1	0	1	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0

- Directed network



In degree $$k_i^{in} = \sum_{j}^{N} A_{ij}$$

	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0	0
4	0	1	1	0	1	0	1	0	0	0
5	0	0	0	1	0	1	0	1	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0



Out degree $$k_j^{out} = \sum_{i}^{N} A_{ij}$$

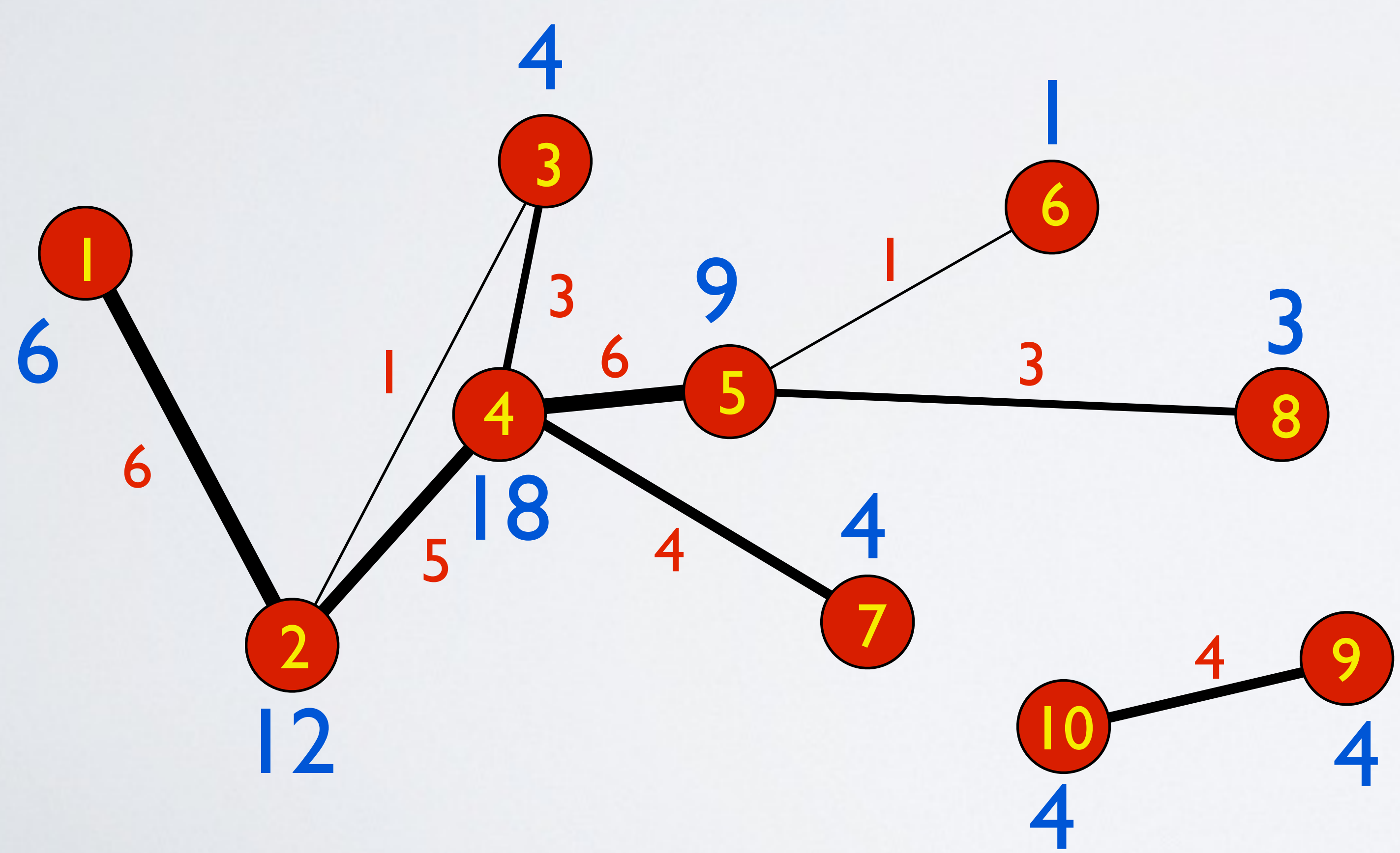
	1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	0	0	0	0	0	0
3	0	1	0	1	0	0	0	0	0	0
4	0	1	1	0	1	0	1	0	0	0
5	0	0	0	1	0	1	0	1	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	1	0

Weighted degree: strength

- Weighted networks

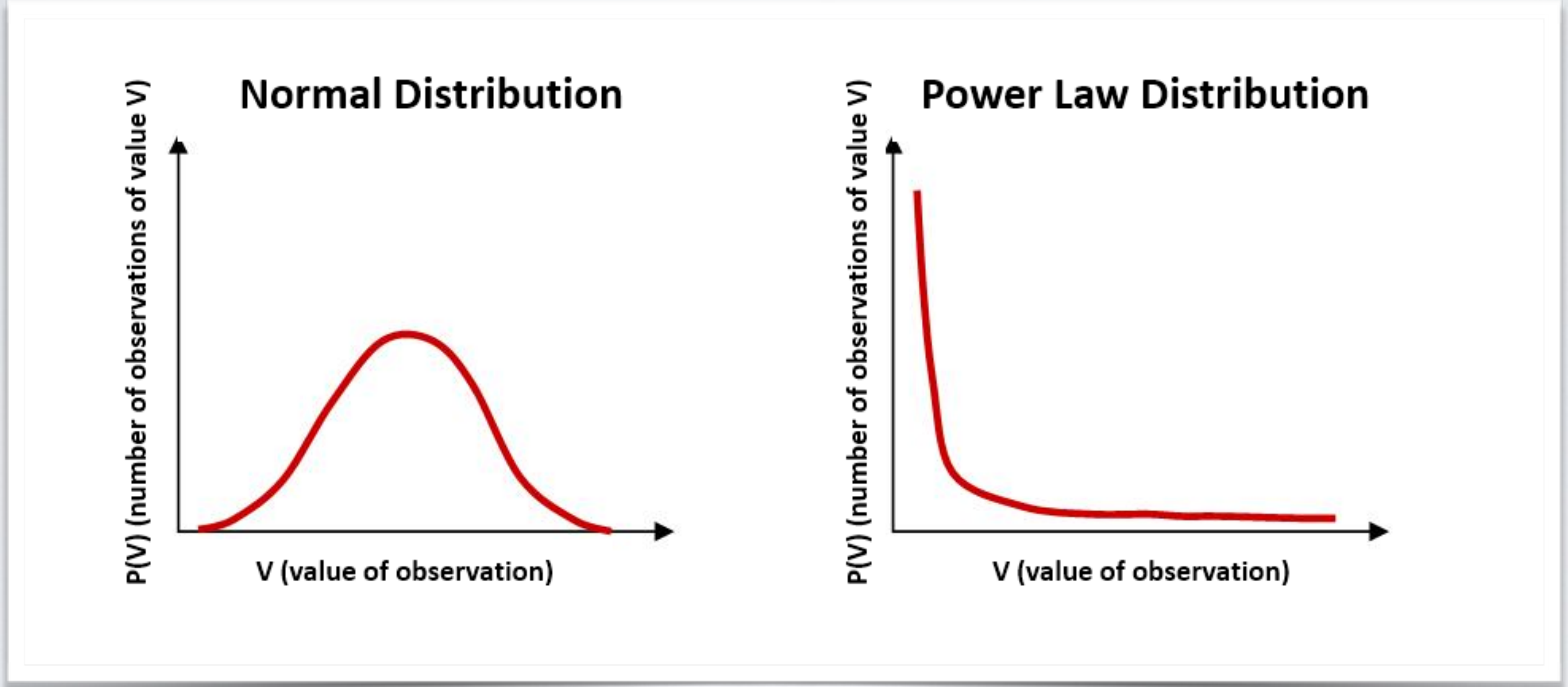
The sum of the weights of links connected to node i

$$s_i = w_{i1} + w_{i2} + \ldots + w_{iN} = \Sigma_j \, w_{ij}$$



	1	2	3	4	5	6	7	8	9	10
1	0	6	0	0	0	0	0	0	0	0
2	6	0	1	3	0	0	0	0	0	0
3	0	1	0	3	0	0	0	0	0	0
4	0	3	3	0	6	0	4	0	0	0
5	0	0	0	6	0	1	0	3	0	0
6	0	0	0	0	1	0	0	0	0	0
7	0	0	0	4	0	0	0	0	0	0
8	0	0	0	0	3	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	4
10	0	0	0	0	0	0	0	0	4	0

DEGREE DISTRIBUTION



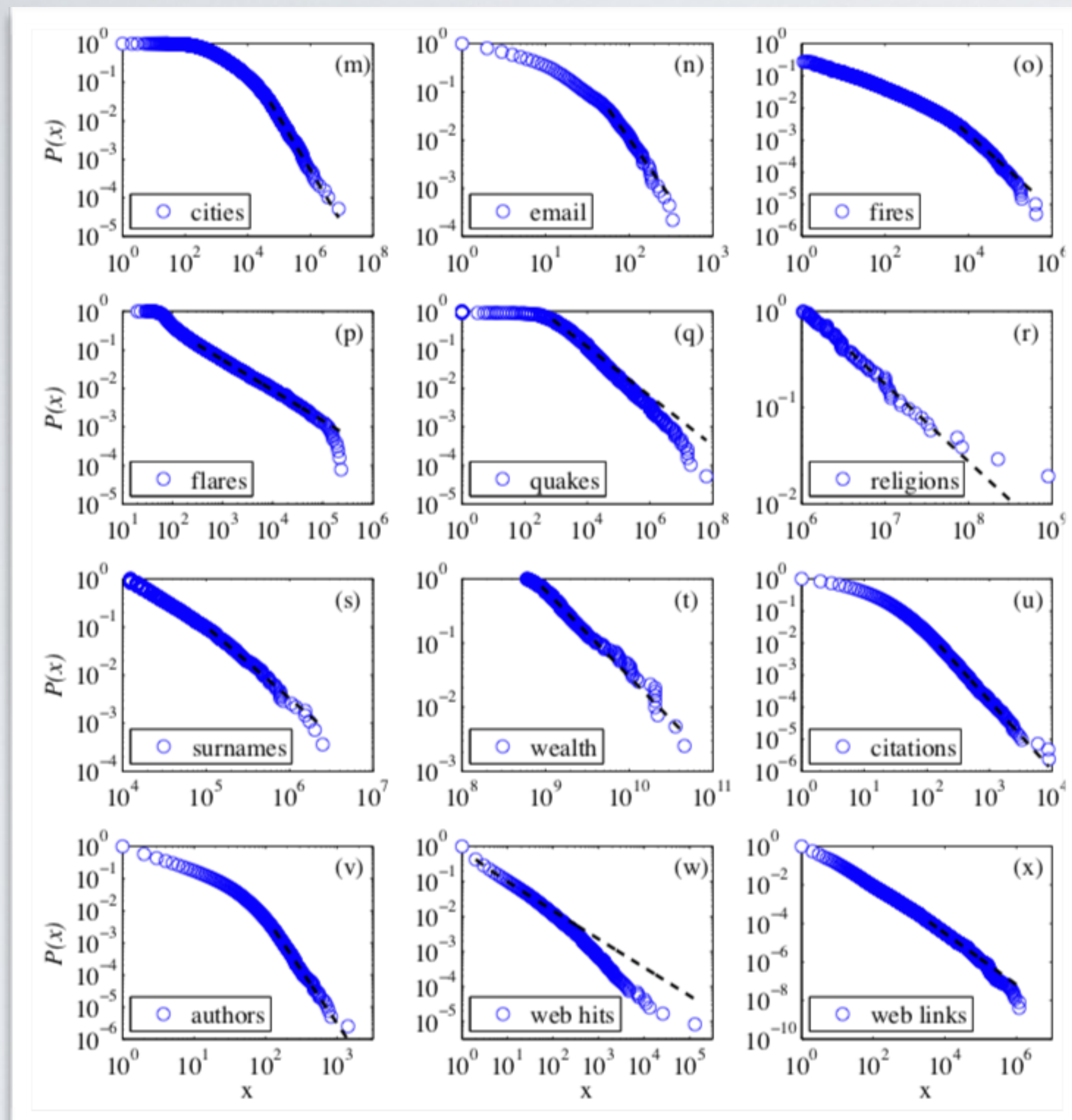
PDF (Probability Distribution Function)

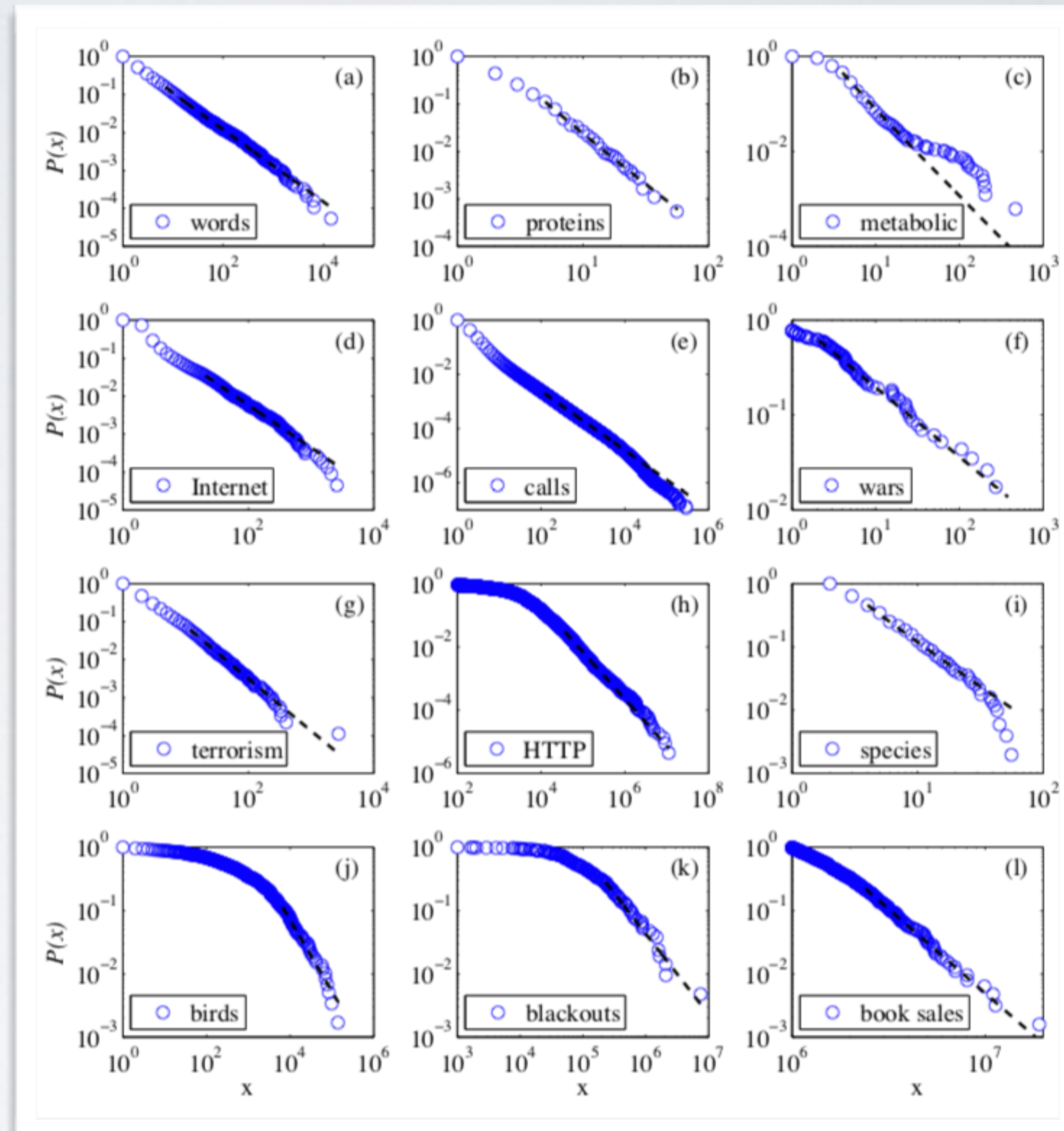
Sometimes with CDF (Cumulative Distribution Function)

DEGREE DISTRIBUTION

- In a fully random graph (Erdos-Renyi), degree distribution is a normal distribution centered on the average degree
- In real graphs, in general, it is not the case:
  - A high majority of small degree nodes
  - A small minority of nodes with very high degree (Hubs)
- Often modeled by a **power law**

DEGREE DISTRIBUTION





Power laws in empirical data (degrees and other things)

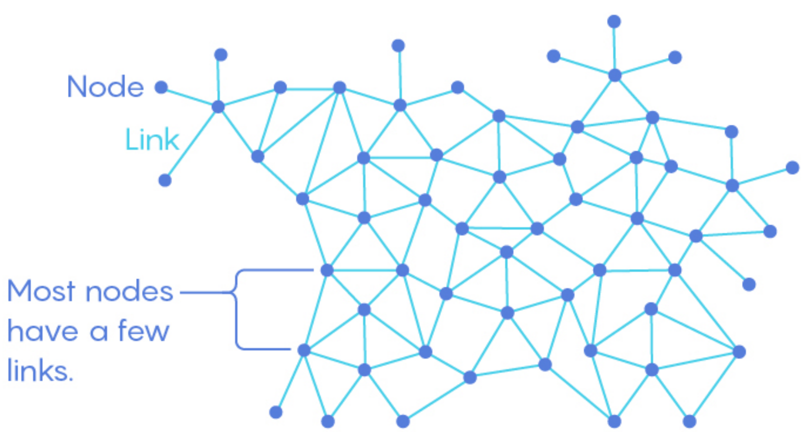
POWER-LAW DISTRIBUTIONS IN EMPIRICAL DATA [Clauset 2009]

To Be or Not to Be Scale-Free

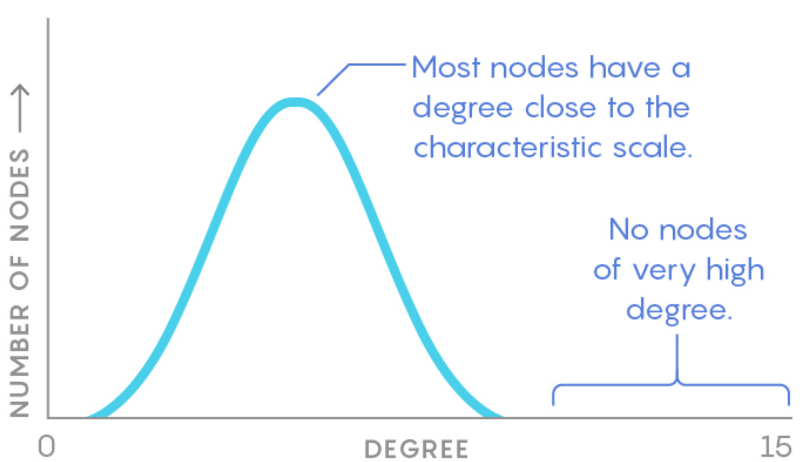
Scientists study complex networks by looking at the distribution of the number of links (or "degree") of each node. Some experts see so-called scale-free networks everywhere. But a new study suggests greater diversity in real-world networks.

Random Network

Randomly connected networks have nodes with similar degrees. There are no (or virtually no) "hubs" — nodes with many times the average number of links.

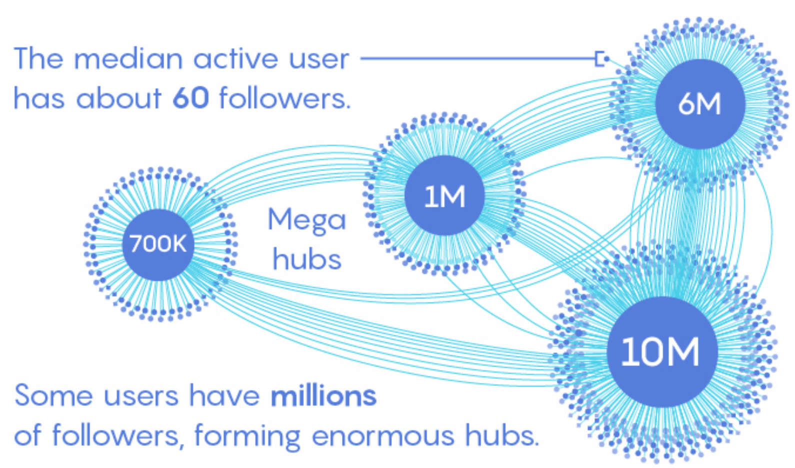


The distribution of degrees is shaped roughly like a bell curve that peaks at the network's "characteristic scale."

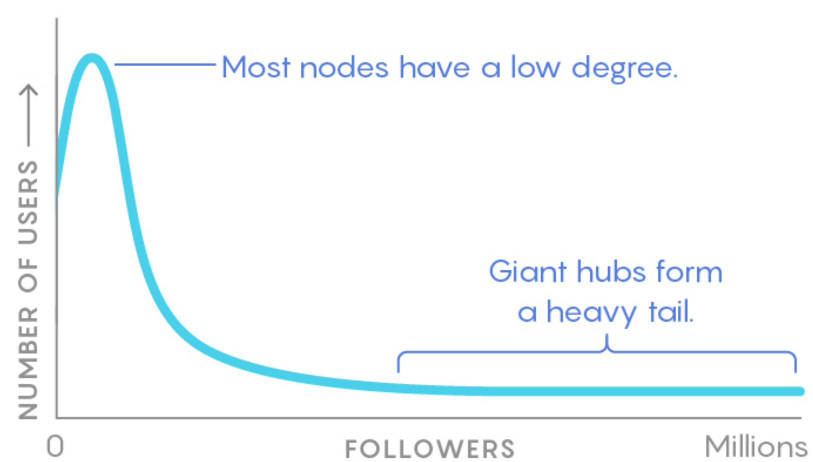


Twitter's Scale-Free Network

Most real-world networks of interest are not random. Some nonrandom networks have massive hubs with vastly higher degrees than other nodes.

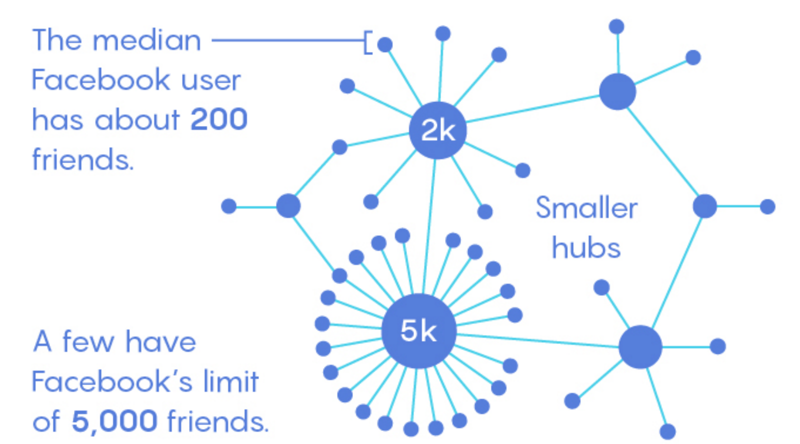


The degrees roughly follow a power law distribution that has a "heavy tail." The distribution has no characteristic scale, making it scale-free.

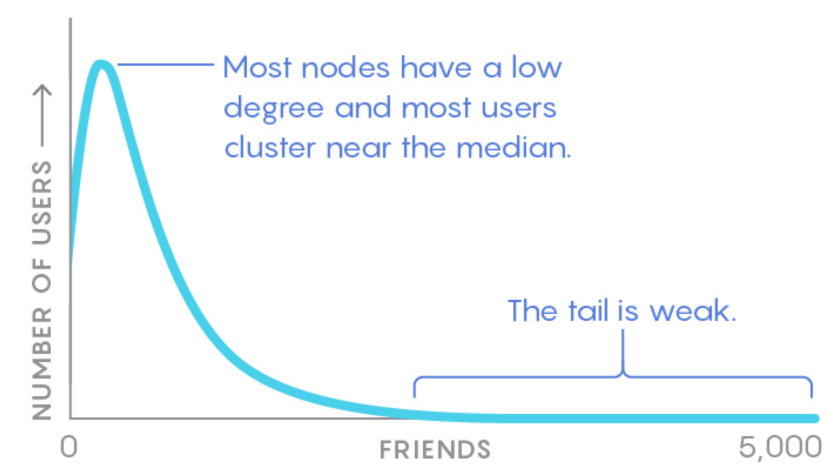


Facebook's In-Between Network

Researchers have found that most nonrandom networks are not strictly scale-free. Many have a weak heavy tail and a rough characteristic scale.



This network has fewer and smaller hubs than in a scale-free network. The distribution of nodes has a scale and does not follow a pure power law.



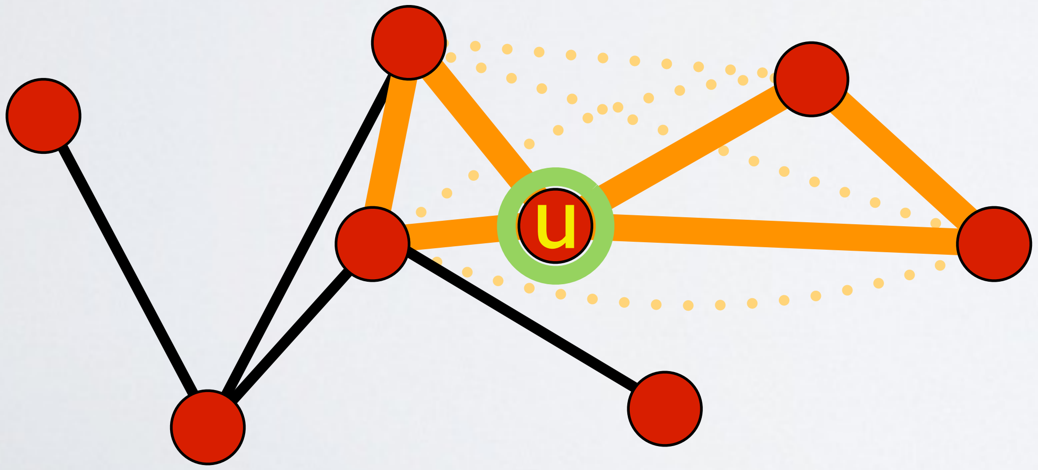
[Quanta magazine 2018]

Node clustering coefficient

- Measure of interconnectivity
- What portion of neighbours of a node are connected to each other?

Global clustering coefficient

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}} = \frac{\text{number of closed triplets}}{\text{number of connected triples of vertices}}.$$



C = 9/18 = 1/2

$C_u = (2\text{x}2)/(4\text{x}3) = 1/3$

Local clustering coefficient

$$C_u = \frac{2e_u}{k_u(k_u - 1)}$$

- e_u - number of links between the neighbours of node u
- $(k_u(k_u\text{-}1))/2$ - maximum number of triangles

Average local clustering coefficient

$$C = \frac{1}{N}\sum_u C_u$$

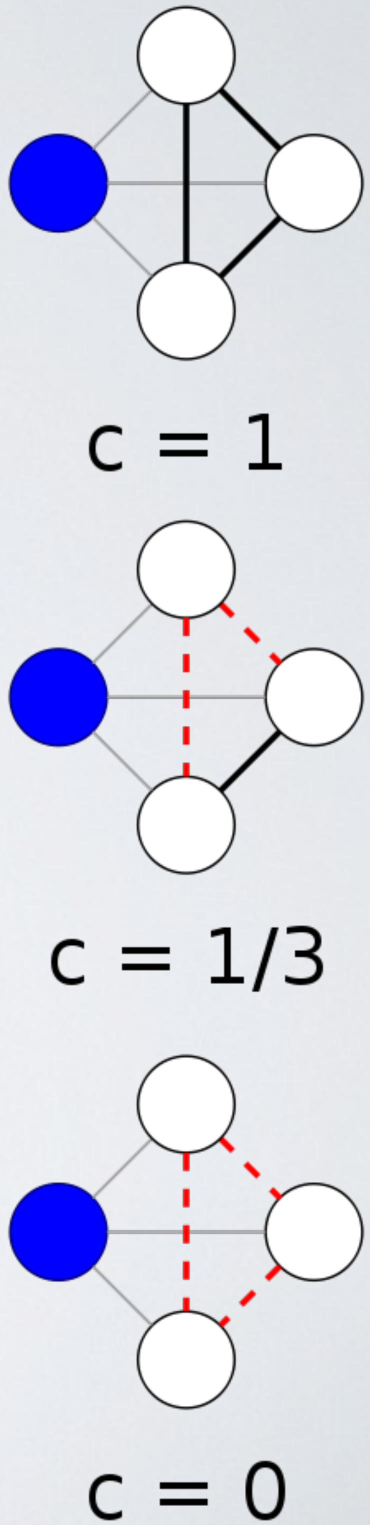
Definition: Watts and Strogatz 2002

CLUSTERING COEFFICIENT

The higher the value,
the more **locally dense** is the network.

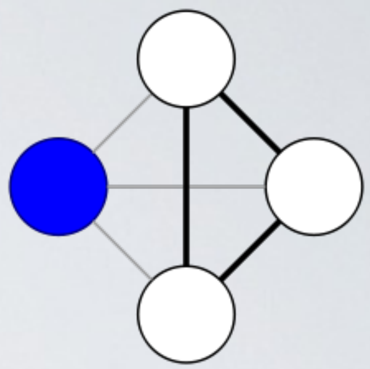
"Friends of my friends are my friends"

Higher in real networks than random

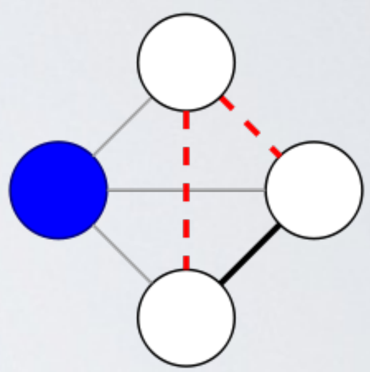


CLUSTERING COEFFICIENT

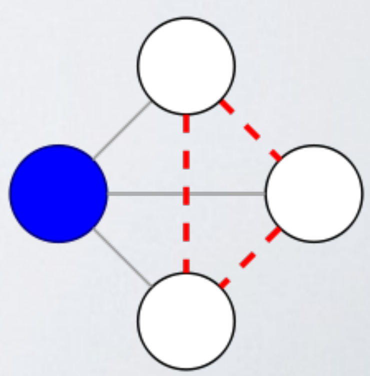
- Global CC:
  - Random (ER): =density: very small for large graphs
  - Facebook ego-networks: 0.6
  - Twitter lists: 0.56
  - California Road networks: 0.04



$c = 1$



$c = 1/3$



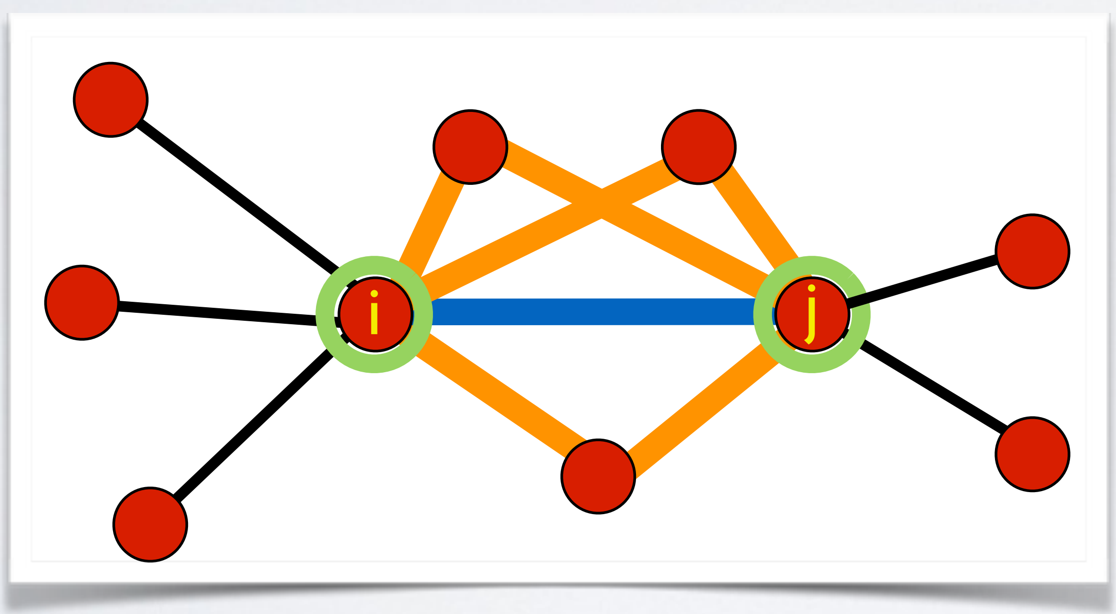
$c = 0$

Link clustering coefficient: Overlap

- Link property
- Fraction of common neighbours of a connected pair
- Jaccard index of common neighbours

$$O_{ij} = \frac{n_{ij}}{(k_i - 1) + (k_i - 1) - n_{ij}}$$

- n_i - number of common neighbours of nodes i and j
- $(k_i\text{-}1)+(k_j\text{-}1)\text{-}n_{ij}$ maximum number possible triangles between nodes i and j



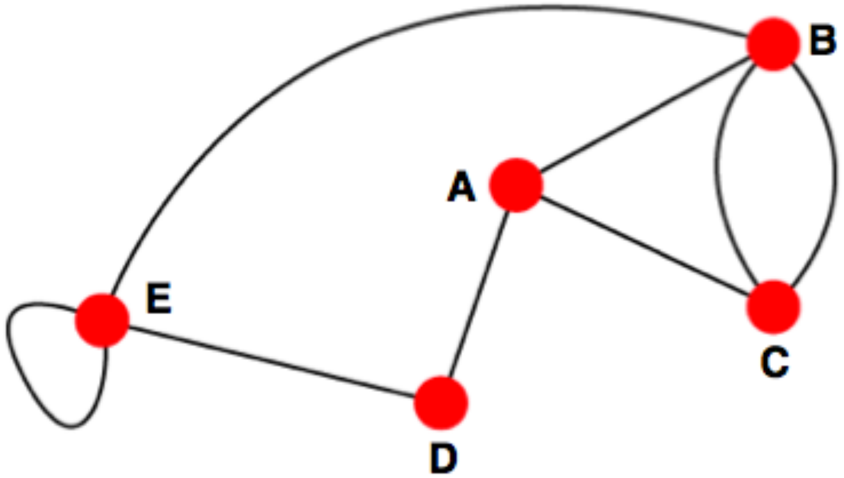
$\boxed{O_{ij} = 3/(6+5-3) = 3/8}$

Path length

A ***path is*** a sequence of nodes in which each node is adjacent to the next one

$\boldsymbol{P_{i0,in}}$ of length $\boldsymbol{n}$ between nodes i_0 and i_n is an ordered collection of ***n+1*** nodes and ***n*** links

$$P_n = \{i_0, i_1, i_2, \ldots, i_n\} \qquad P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)\}$$

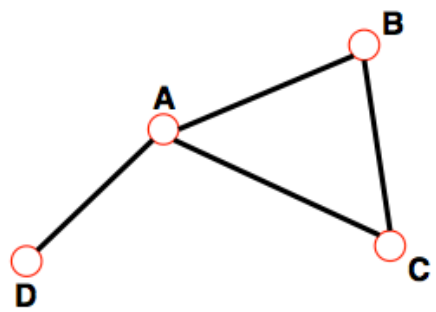


•A path can intersect itself and pass through the same link repeatedly. Each time a link is crossed, it is counted separately

•A legitimate path on the graph on the right:
ABCBCADEEBA

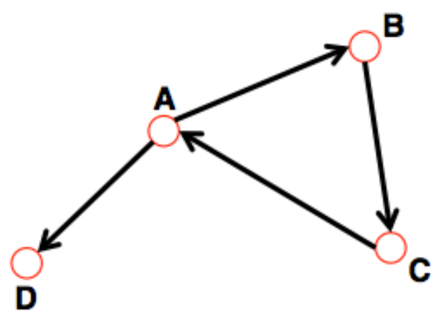
• In a directed network, the path can follow only the direction of an arrow.

Path length



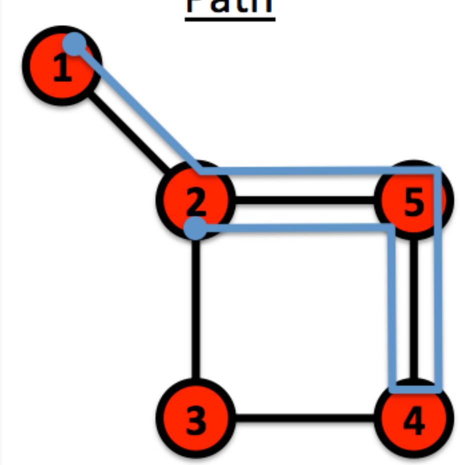
The ***distance (shortest path, geodesic path)*** between two nodes is defined as the number of edges along the shortest path connecting them.

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



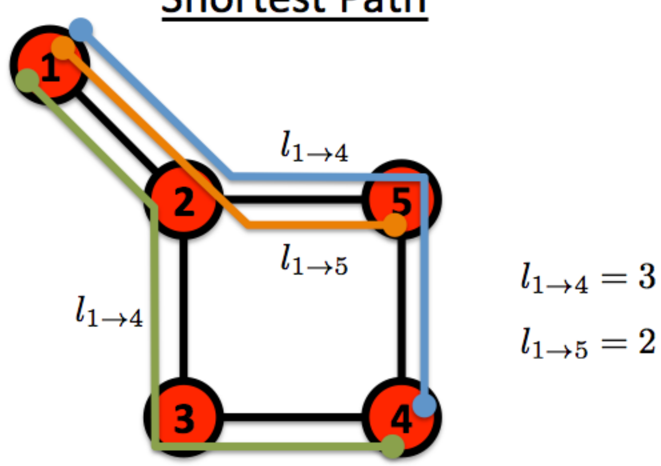
In directed graphs each path needs to follow the direction of the arrows.
Thus in a digraph the distance from node A to B (on an AB path) is generally different from the distance from node B to A (on a BCA path).

Path



A sequence of nodes such that each node is connected to the next node along the path by a link.

Shortest Path



The path with the shortest length between two nodes (distance).

Path length

N_{ij},number of paths between any two nodes *i* and *j*:

Length n=1: If there is a link between ***i*** and ***j***, then $A_{ij}=1$ and $A_{ij}=0$ otherwise.

Length n=2: If there is a path of length two between ***i*** and ***j***, then $A_{ik}A_{kj}=1$, and $A_{ik}A_{kj}=0$ otherwise.
The number of paths of length 2:

$$N_{ij}^{(2)} = \sum_{k=1}^{N} A_{ik}A_{kj} = [A^2]_{ij}$$

Length n: In general, if there is a path of length ***n*** between ***i*** and ***j***, then $A_{ik}\ldots A_{lj}=1$ and $A_{ik}\ldots A_{lj}=0$ otherwise.
The number of paths of length ***n*** between ***i*** and ***j*** is*

$$N_{ij}^{(n)} = [A^n]_{ij}$$

****holds for both directed and undirected networks.***

Path length

- d_{max} diameter- the maximum distance between any pairs of nodes

- $\langle d \rangle$ average path length - for directed graphs

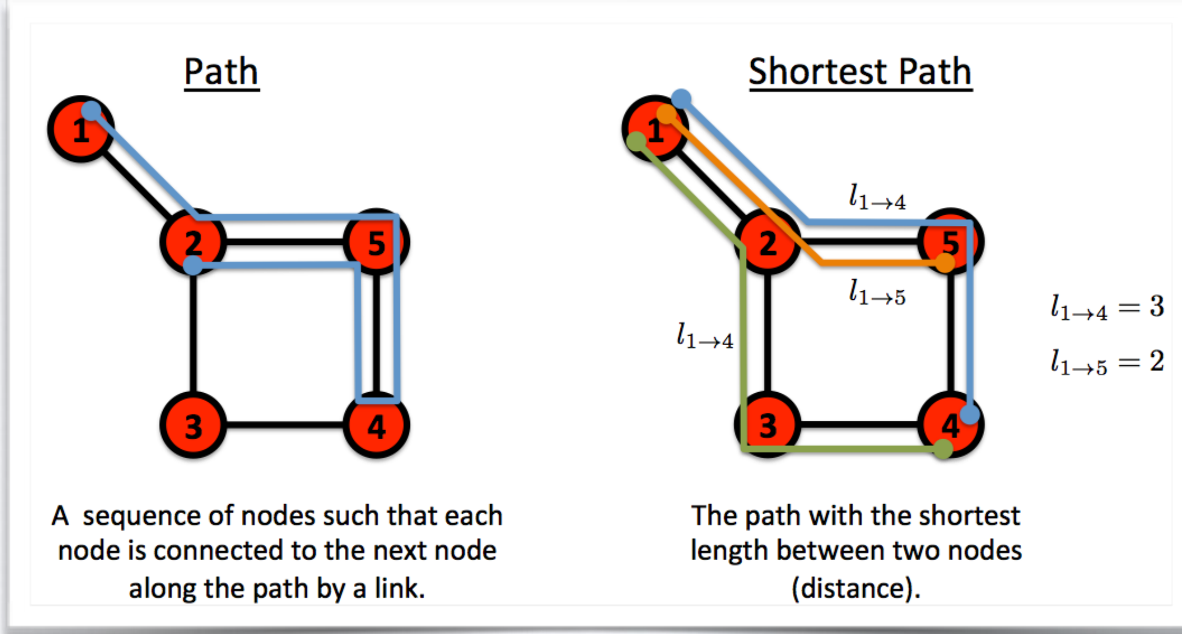
$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}$$

- where d_{ij} is the shortest distance between nodes i and j
- multiplicative is *(2 x max number of links)*
- distance between unconnected nodes is 0

- $\langle d \rangle$ average path length - for un-directed graphs

$$\langle d \rangle = \frac{2}{N(N-1)} \sum_{i < j} d_{ij}$$

- since $d_{ij} = d_{ji}$
- multiplicative is *(max number of links)*



AVERAGE PATH LENGTH

- The famous 6 degrees of separation (Milgram experiment)
  - In fact 6 hops
  - (More on that next slide)
- Not too sensible to noise
- Tells you if the network is "stretched" or "hairball" like

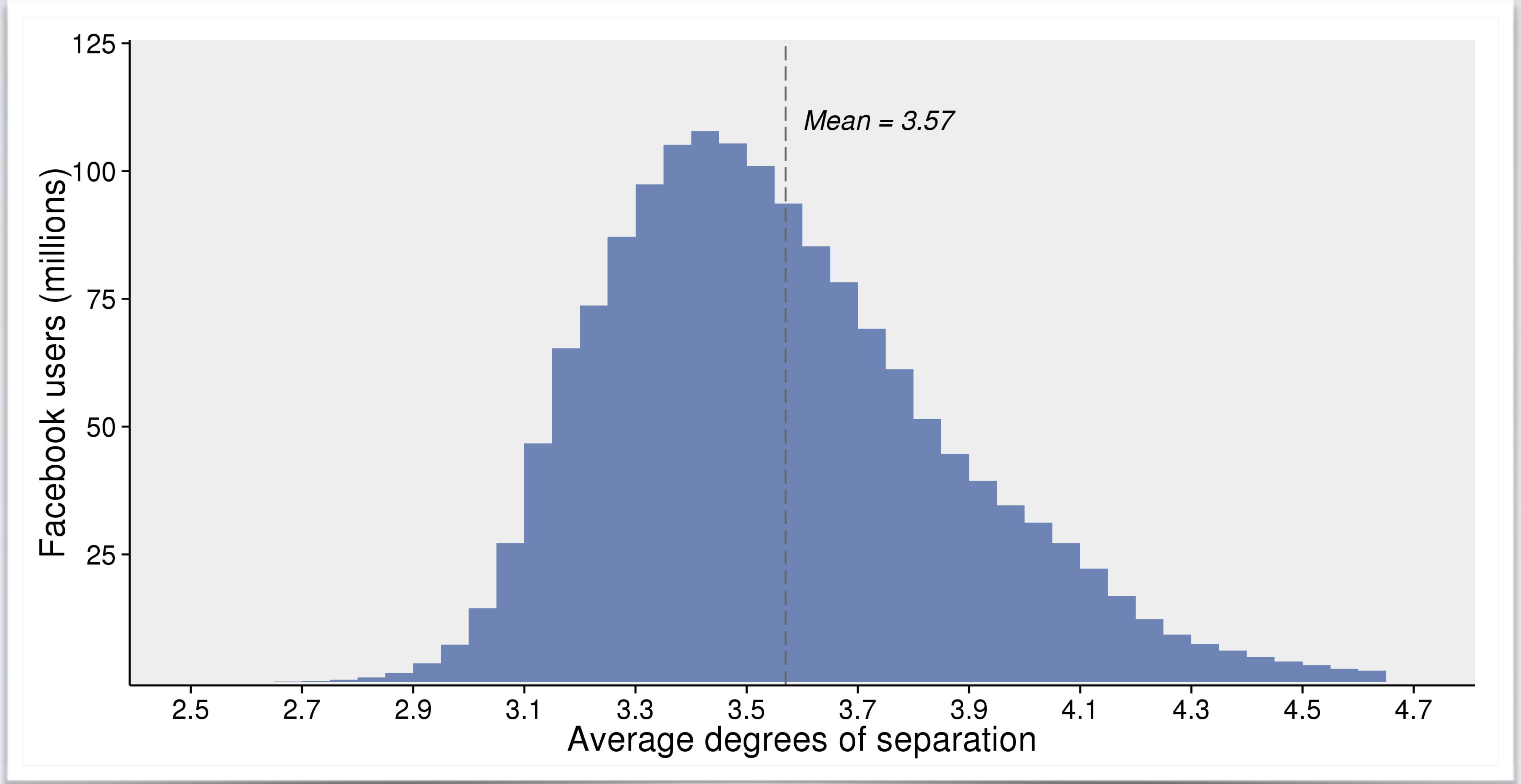
SIDE-STORY: MILGRAM EXPERIMENT

- Small world experiment (60's)
  - Give a (physical) mail to random people
  - Ask them to send to someone they don't know
    - They know his city, job
  - They send to their most relevant contact
- Results: In average, 6 hops to arrive

SIDE-STORY: MILGRAM EXPERIMENT

- Many criticism on the experiment itself:
  - Some mails did not arrive
  - Small sample
  - …
- Checked on “real” complete graphs (giant component):
  - MSN messenger
  - Facebook
  - The world wide web
  - …

SIDE-STORY: MILGRAM EXPERIMENT



Facebook

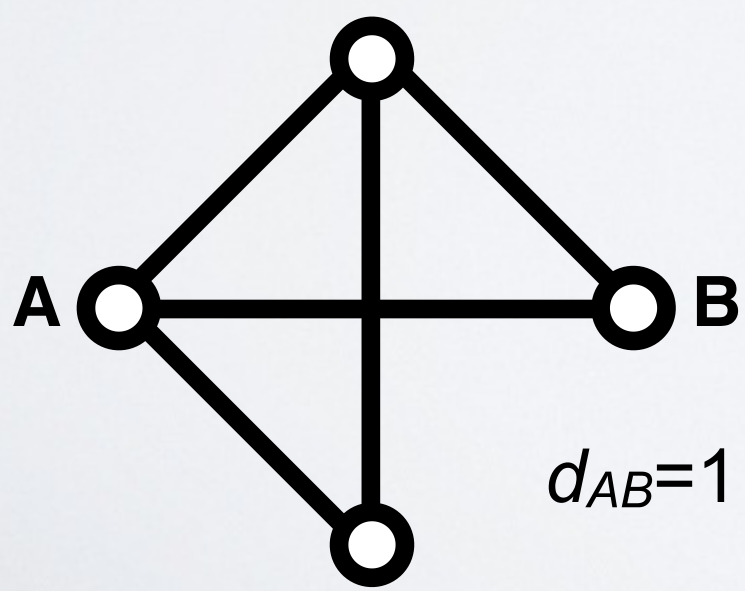
Weighted path length

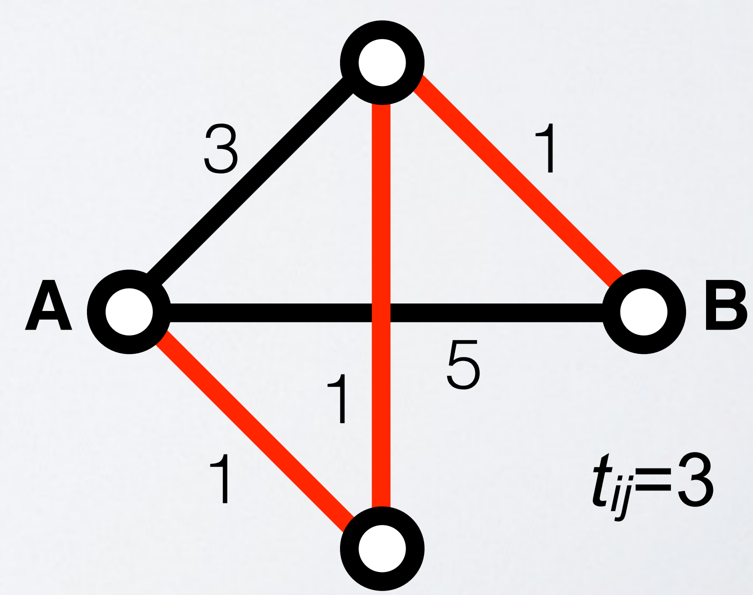
length of a shortest path $P(i \to j)$ ≠ length of a weighted shortest path $P(i \to j)$

$$d_{ij} = \sum_{e_{mn} \in P(i \to j)} A_{mn}$$

$$t_{ij} = \sum_{e_{mn} \in P(i \to j)} w_{mn}$$

Shortest path ≠ Weighted shortest path





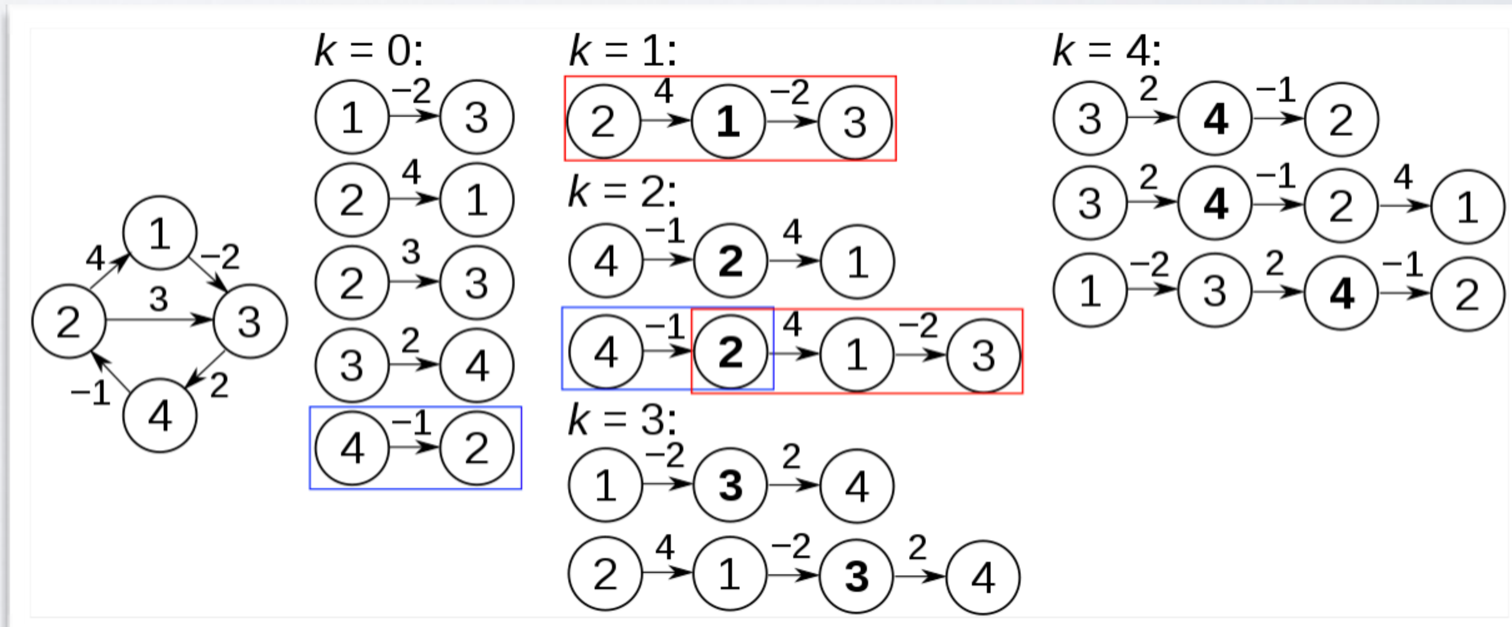
All shortest path algorithm

finding shortest paths in a **weighted graph** with **positive** or **negative edge weights** (but with no negative cycles)

```
proc FloydWarshall(G=(V,E,w))
// let dist be a |V| × |V| array of minimum distances initialized to ∞ (infinity)
for each edge (u,v)
   dist[u][v] ← w(u,v)  // the weight of the edge (u,v)
for each vertex v
   dist[v][v] ← 0
for k from 1 to |V|
   for i from 1 to |V|
      for j from 1 to |V|
         if dist[i][j] > dist[i][k] + dist[k][j]
            dist[i][j] ← dist[i][k] + dist[k][j]
         end if
```

Checking and updating all paths going through nodes k=1, 2, 3, … , N by assuming that:

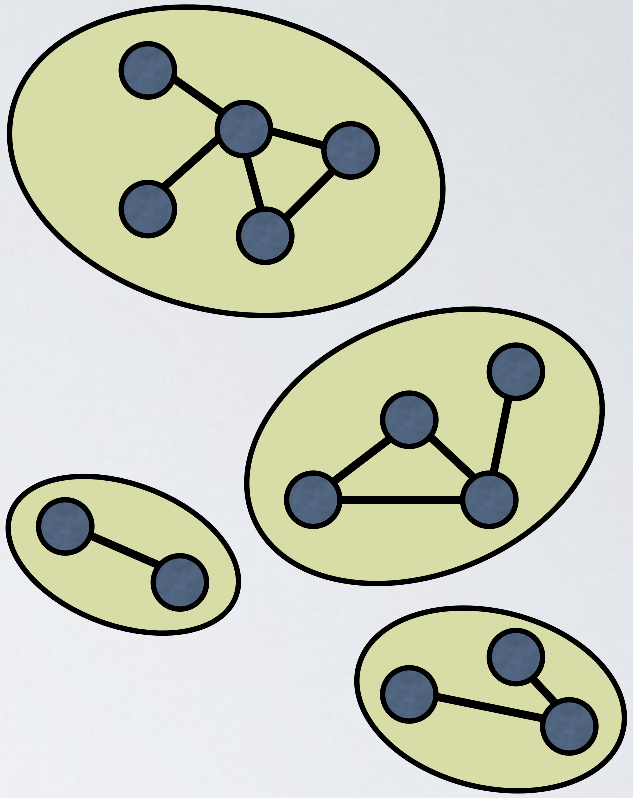
$$shp(i,j,k)= min(shp(i,j,k\text{-}1)), shp(i,k,k\text{-}1)+shp(k,j,k\text{-}1))$$



Complexity: $O(n^3)$

Connectivity and components

- A connected component is a subset of vertices with at least one path connecting each of them
- A network may consist of a single connected component (a connected network) or several of those
- Distances between nodes in disjoint components are not defined (infinite)
- Bridge: if we remove it, the graph becomes disconnected.
- The adjacency matrix of a network with several components can be written in a block-diagonal form, so that nonzero elements are confined to squares, with all other elements being zero



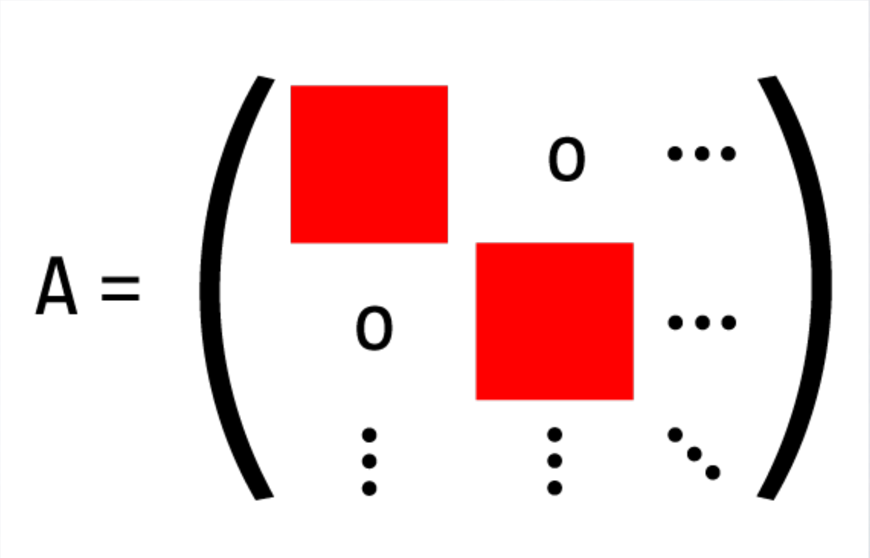


Figure after Newman, 2010

Connected components algorithm

```
proc connectedComponents(G=(V,E))
   // Mark all the vertices as not visited
   bool visited=[V]
   for v from 0 to V-1
      visited[v] ← false

   for v from 0 to V-1
      if visited[v]==false
         // print all reachable vertices from v
         DFSUtil(v, visited)
         print("\n")
      end if
```

- Compute with recursive DFS (or BFS) algorithm started from each unvisited node
- **Complexity**: *O(|V| + |E|)*

```
proc DFSUtil(v, visited[])
   // Mark the current node as visited and print it
   visited[v] ← true
   print(v)

   // Recur for all the vertices adjacent to this vertex
   for each i in adj[v]
      if visited[i]==false
         DFSUtil(i, visited)
      end if
```

- Better solution exists using **disjoint set structures**

Connectivity and components - directed networks

- Strongly connected component (SCC): has a path from each node to every other node in the component
- Weakly connected component (WCC): it is connected if we disregard the directions
- In-component: nodes that can reach the SCC
- Out-component: nodes that can be reached from SCC

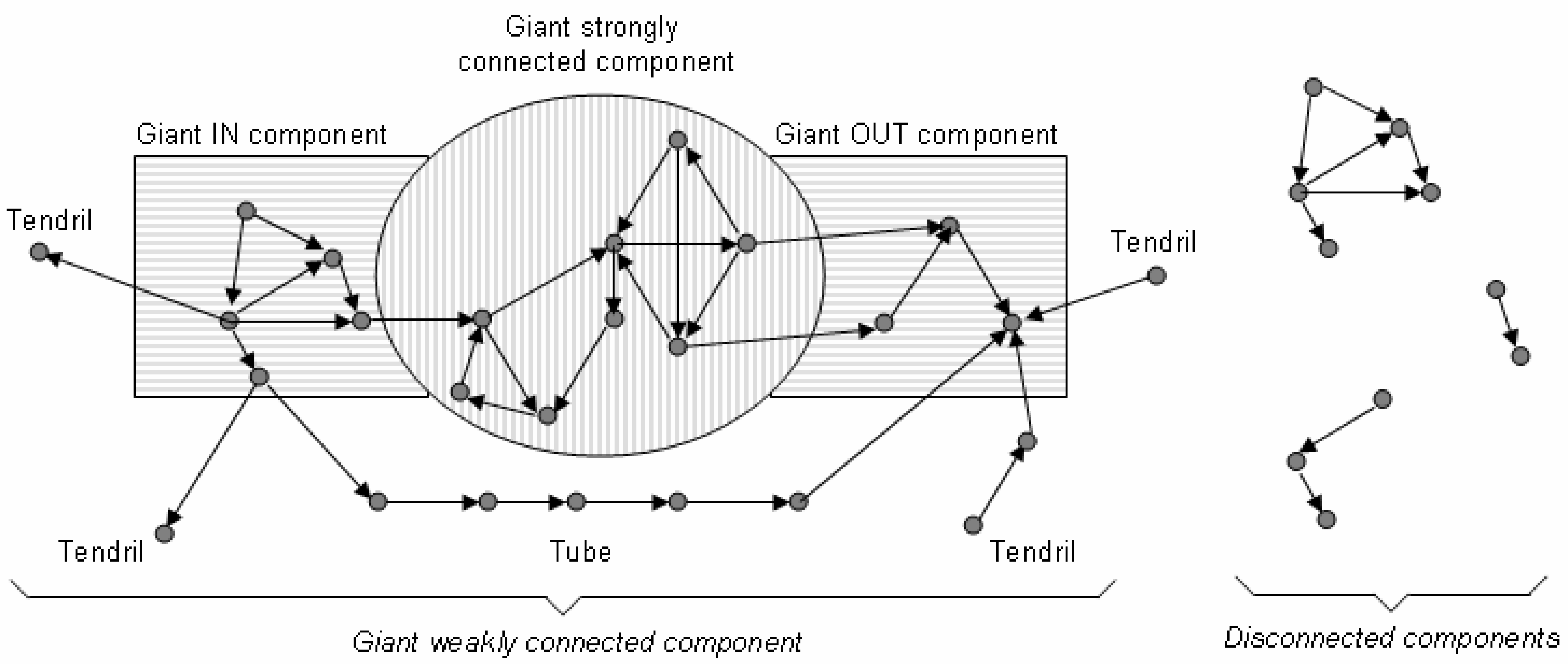


Figure from Broder et. al. (2000)

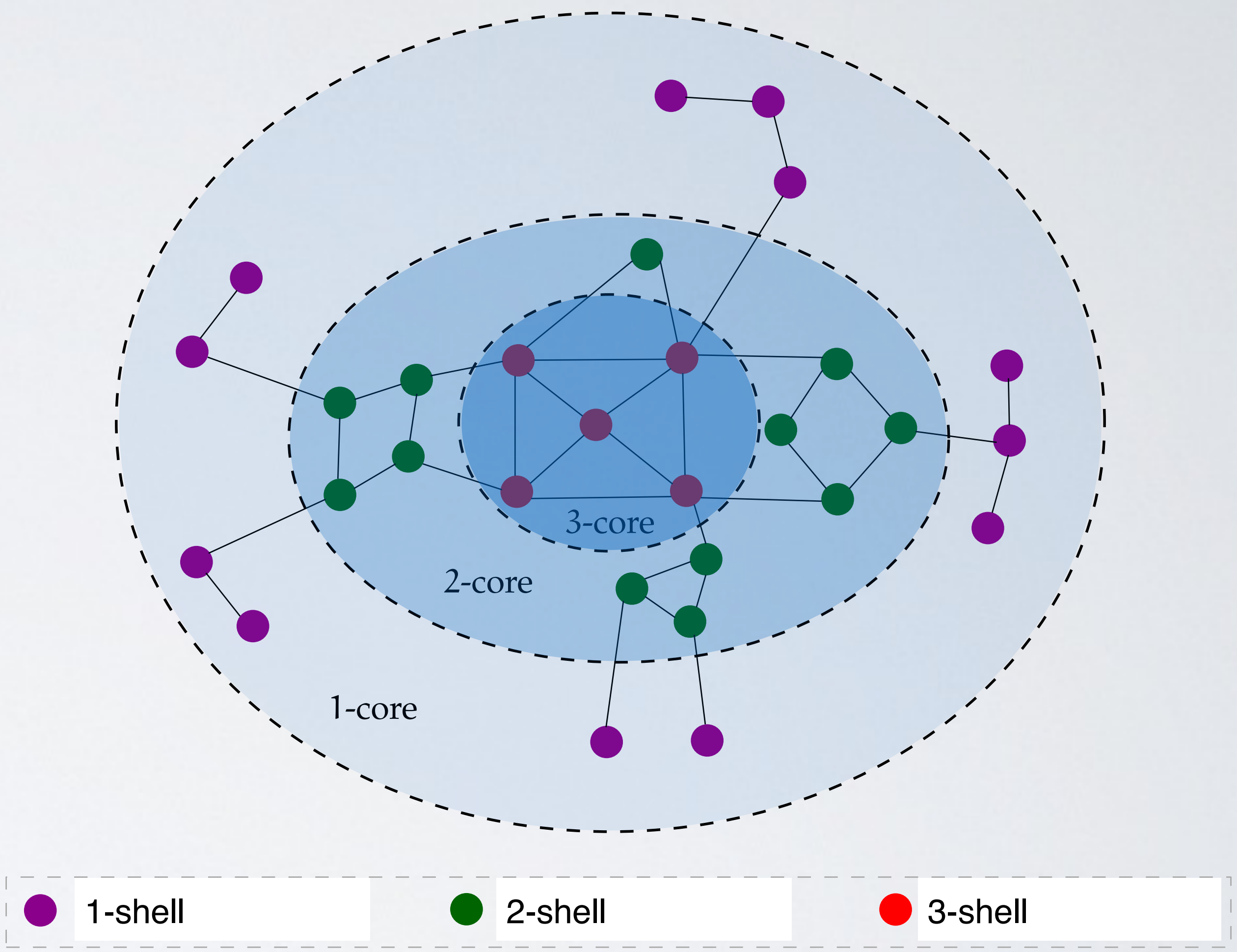
k-core decomposition

Goal: To identify dense cores of high degree nodes in networks

Given graph $G = (V, E)$

Definition: A subgraph $H = (C, E|C)$ induced by the set $C \subseteq V$ is a **k-core or a core of order k** iff $\forall v \in C : degree(H(v)) \geq k$, and H is the maximum subgraph with this property.

- A k-core of G can be obtained by recursively removing all the vertices of degree less than k, until all vertices in the remaining graph have at least degree k.



Definition: A vertex i has **coreness c** if it belongs to the c-$core$ but not to $(c + 1)$-$core$.

Definition: A **c-shell** is composed by all the vertices whose coreness is c. The k-core is thus the union of all shells with $c \geq k$.

core decomposition

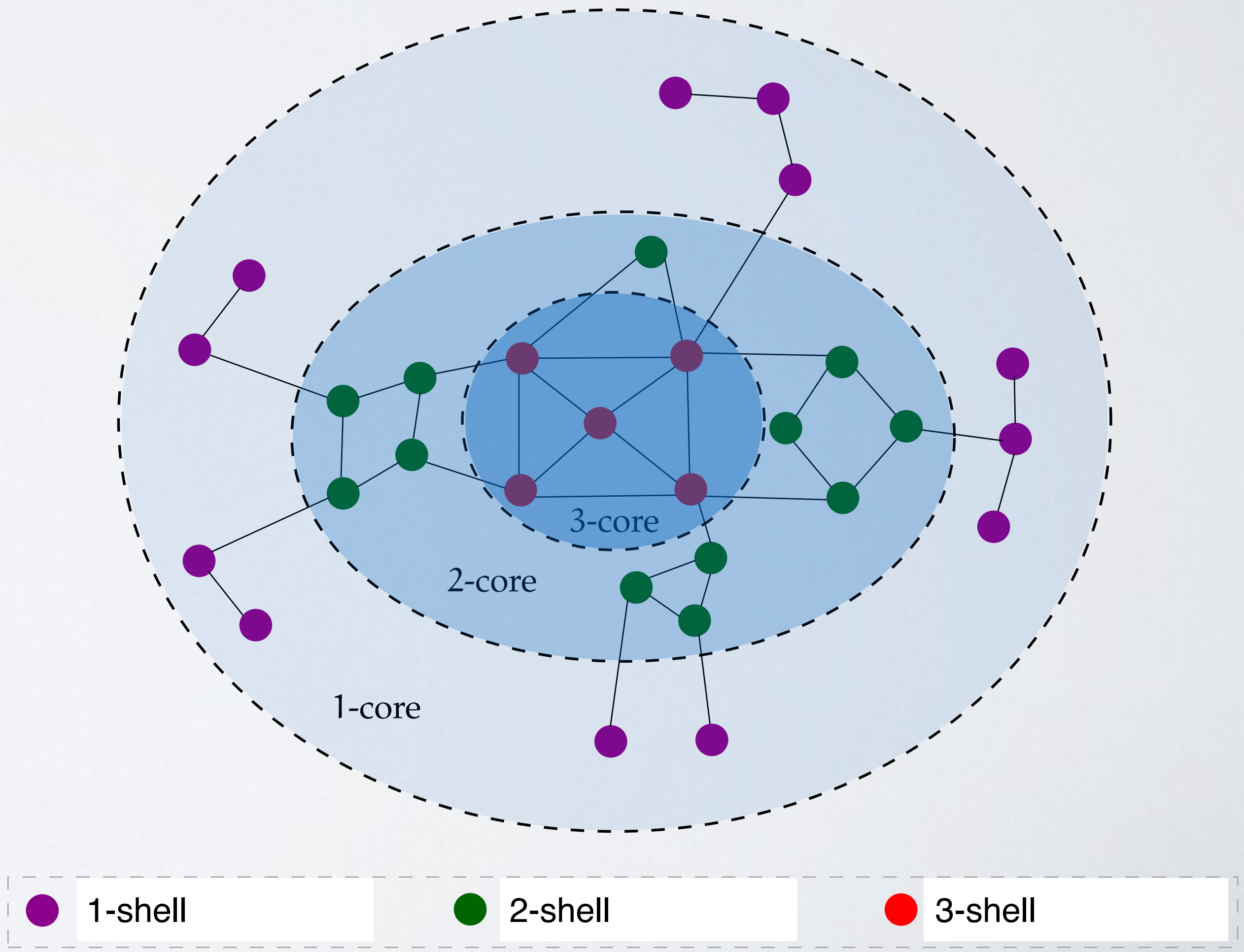
Intuitive algorithm

1. Take a directed or undirected network
2. Remove nodes with degree $k(=1)$ and all of those which degree became $k(=1)$ because of the removal process
3. Repeat step 2 for $k=2,3,\ldots$ until no node can be removed

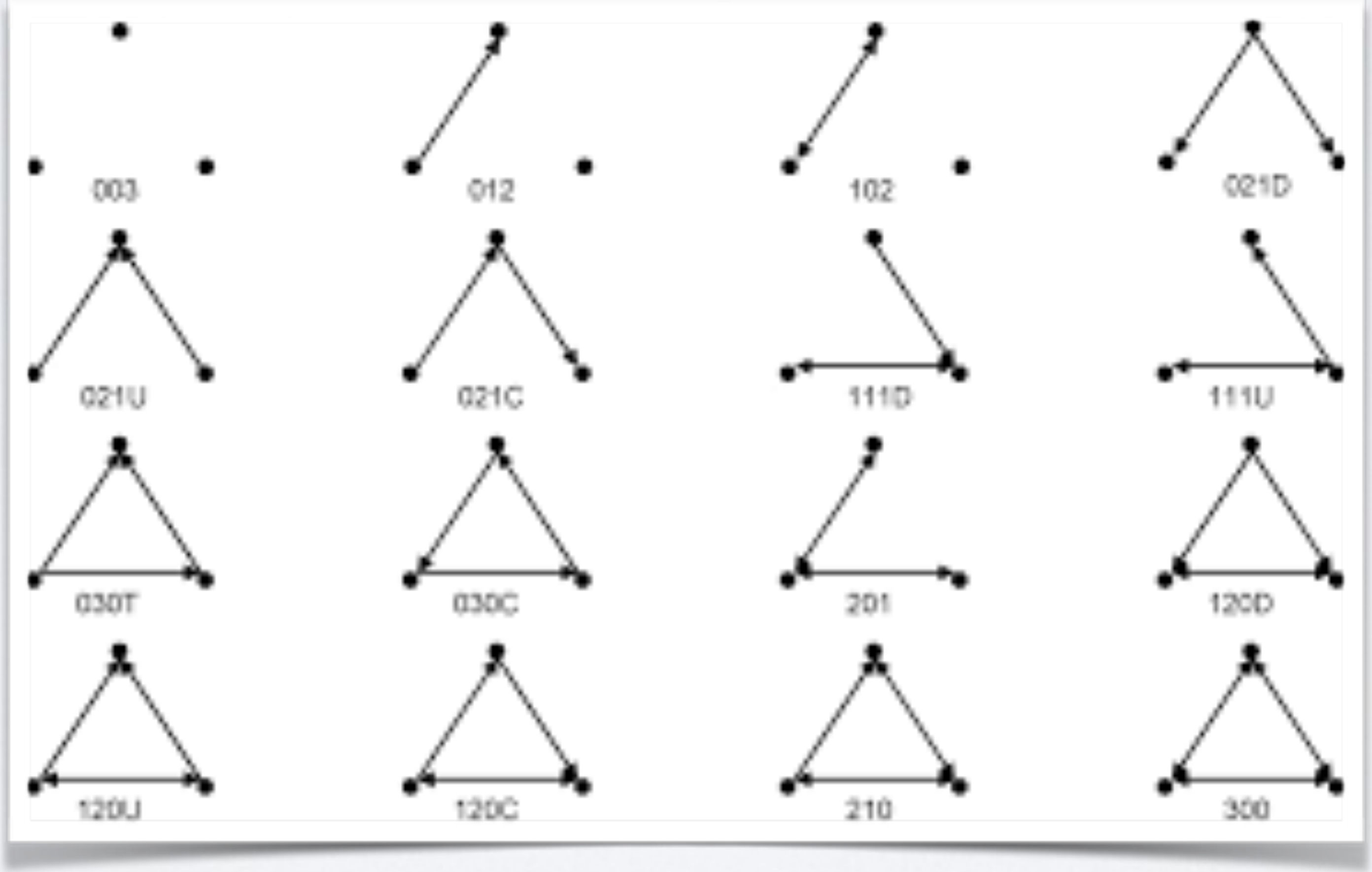
- Nodes removed in the kth turn are in the k-shell and the remaining nodes form the k-core

```
proc CoreDecomposition(G=(V,E))
   compute the degrees of vertices
   order v∈V in increasing degree order
   core[V]=0
   for each v∈V in the order
      core[v] := degree[v];
      for each u ∈ adj(v) do
         if deg[u] > deg[v] then
            degree[u] := degree[u] - 1;
            reorder V
         end if
```

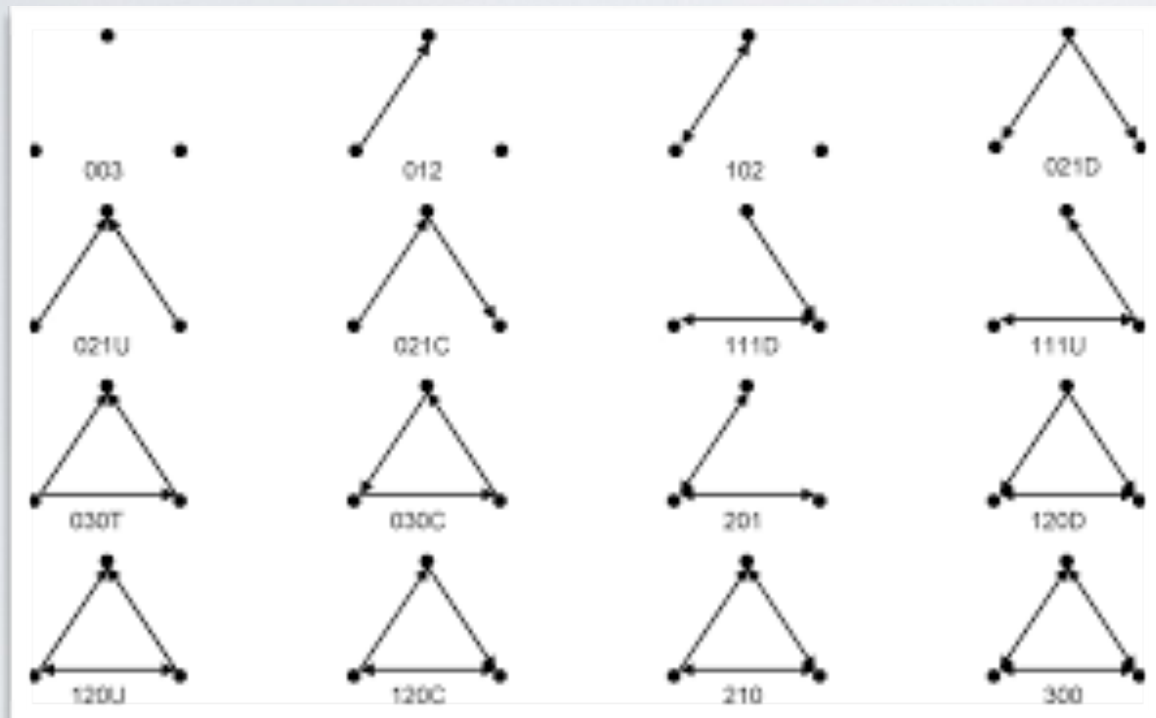
Batagelj, Zversnik (2002)

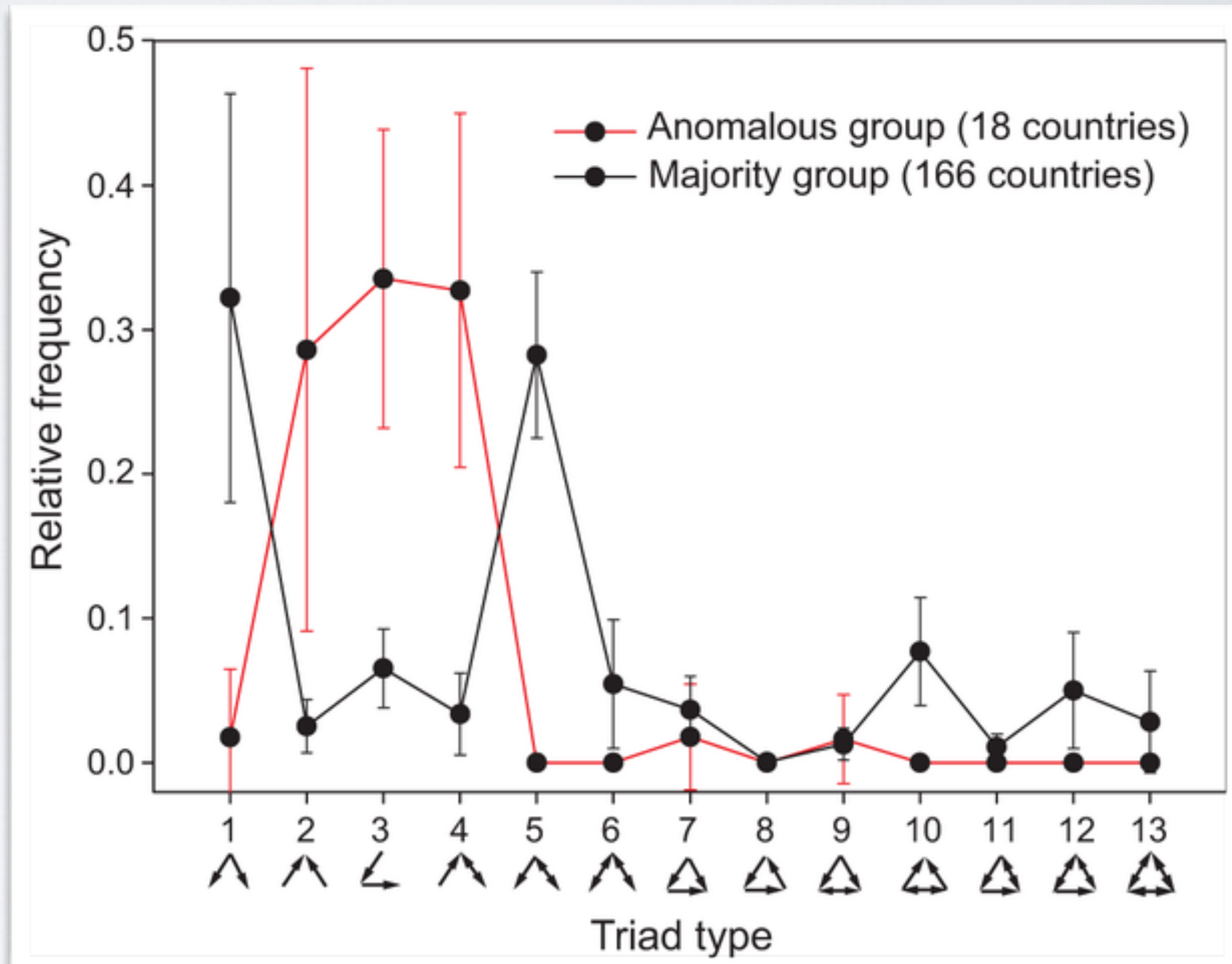


TRIADS COUNTING

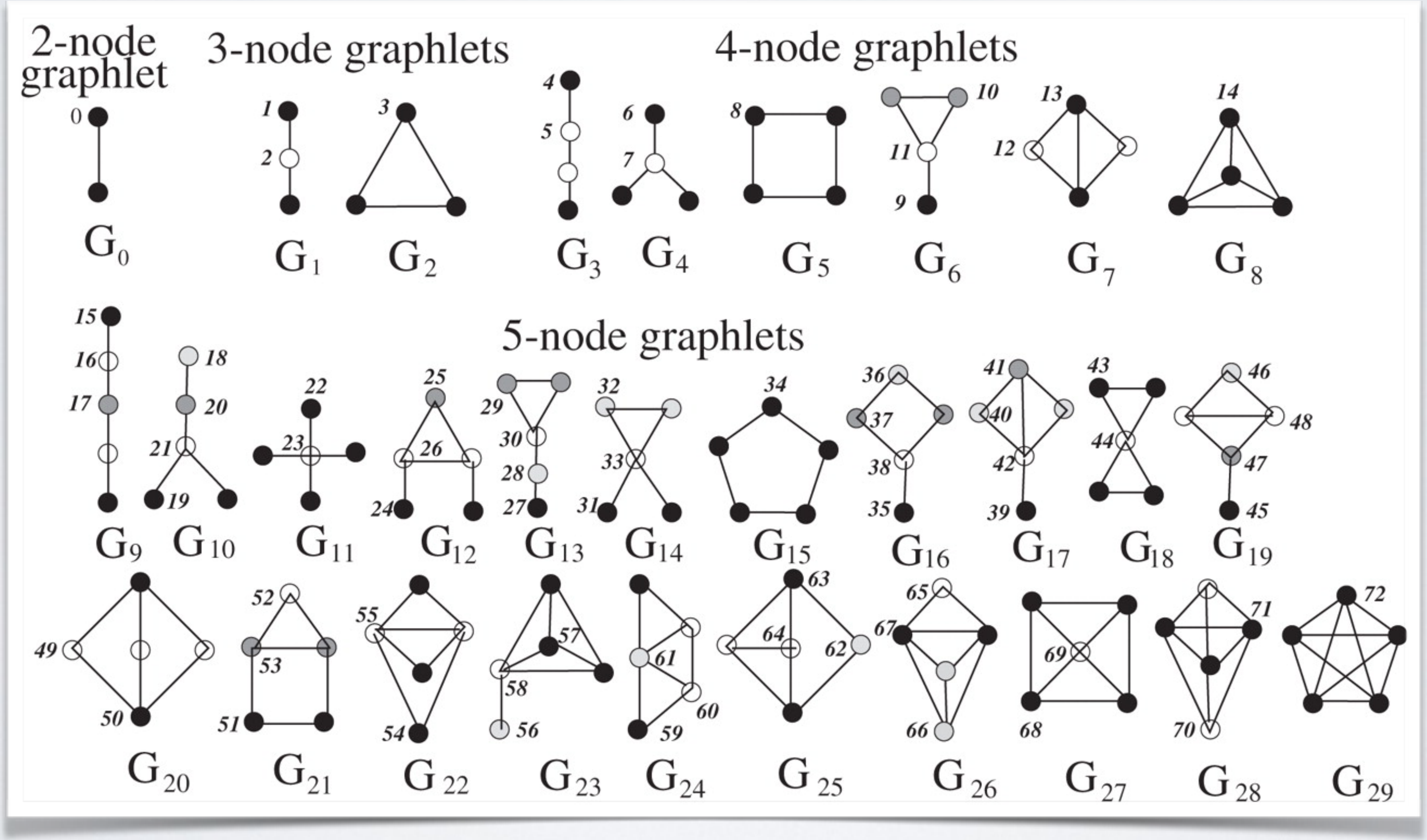


TRIADS COUNTING



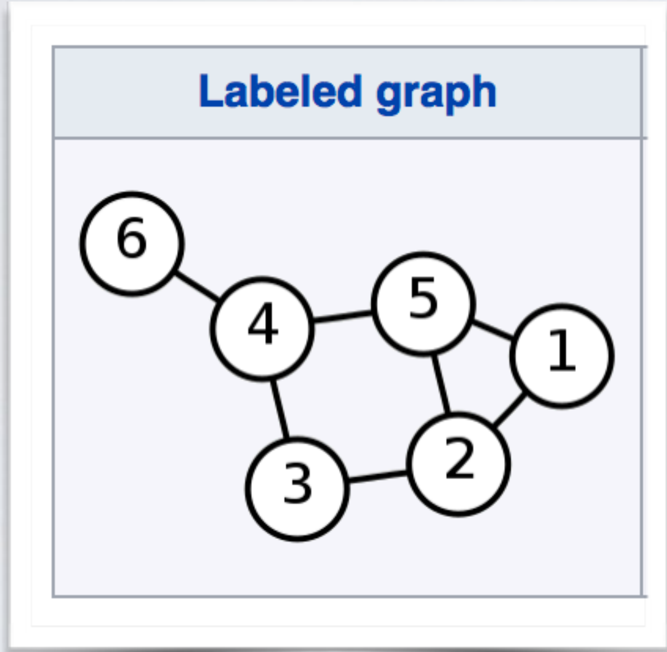


GRAPHLETS



2-node graphlet
3-node graphlets
4-node graphlets
5-node graphlets
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
G0
G1
G2
G3
G4
G5
G6
G7
G8
G9
G10
G11
G12
G13
G14
G15
G16
G17
G18
G19
G20
G21
G22
G23
G24
G25
G26
G27
G28
G29

MATRIX PROPERTIES



Adjacency matrix

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

- What is a Matrix?
  - <u>Not</u> a 2D data table
  - It describes a *linear transformation*, or *linear function*
  - Said differently, it represents a *set of equations*

MATRIX PROPERTIES

$$\begin{array}{c|c} & \begin{matrix} x1 & x2 & x3 & x4 & x5 & x6 \end{matrix} \\ \begin{matrix} x1' \\ x2' \\ x3' \\ x4' \\ x5' \\ x6' \end{matrix} & \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

$$x1' = 0x_1 + 1x_2 + 0x_3 + 0x_4 + 1x_5 + 0x_6$$
$$x2' = x_1 + x_3 + x_5$$
$$x3' = x_2 + x_4$$
$$\cdots$$

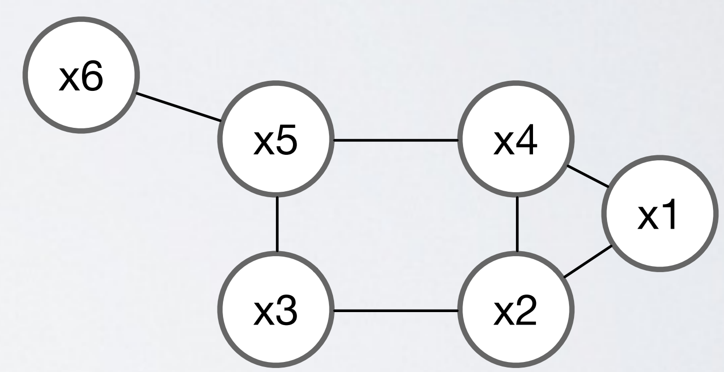
MATRIX PROPERTIES

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \qquad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_5 \\ x_5 \\ x_6 \end{pmatrix}$$

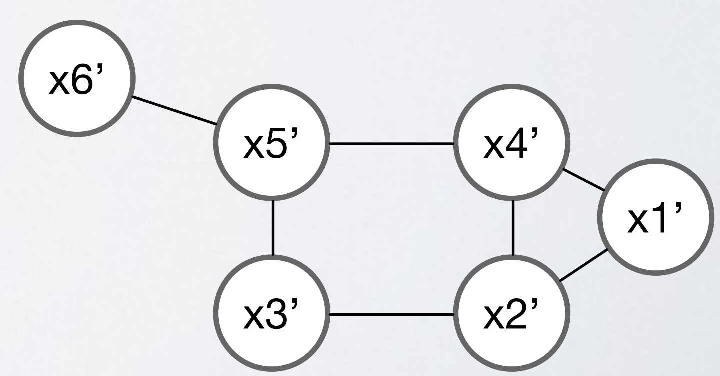
$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_5 \\ x_5 \end{pmatrix}$$

$$Ax = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_2 + x_4 \\ x_1 + x_3 + x_5 \\ x_2 + x_4 \\ x_3 + x_5 + x_6 \\ x_1 + x_2 + x_4 \\ x_4 \end{pmatrix}$$

A, x =



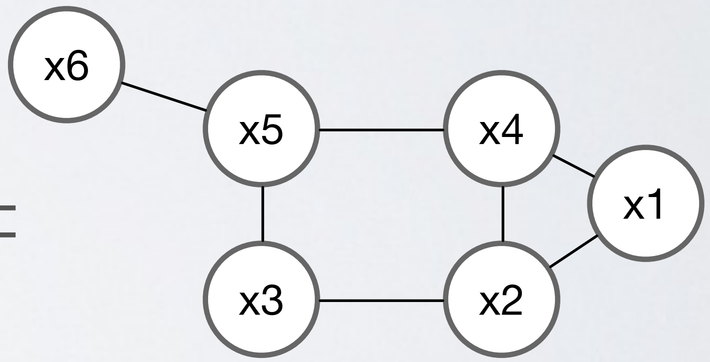
A, Ax=



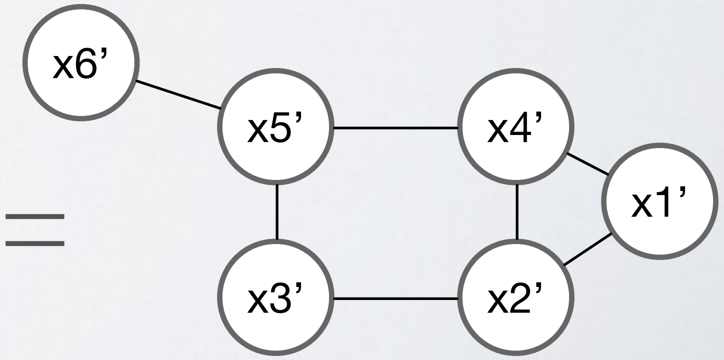
MATRIX PROPERTIES

- Question: What is the result of Ax if
  - x1=x2=x3=x4=x5=x6=1 ?

A, x =



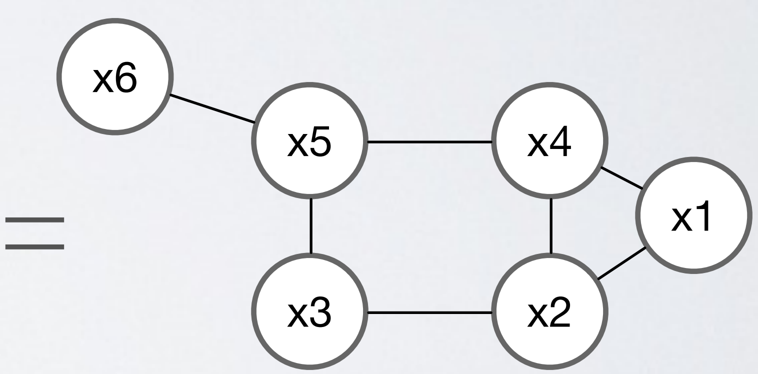
A, Ax=



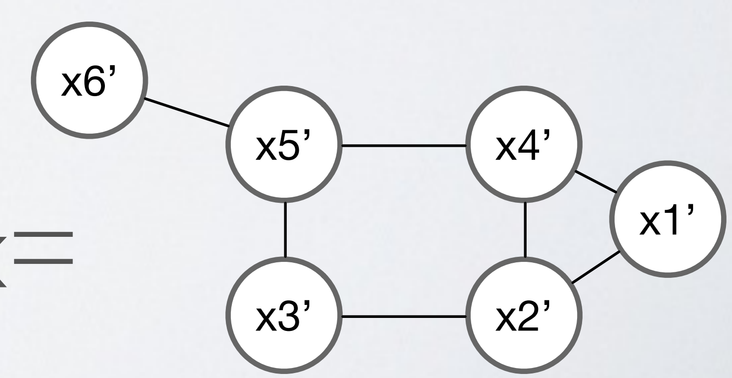
MATRIX PROPERTIES

- Question: What is the result of Ax if
  - x1=x2=x3=x4=x5=x6=1 ?
  - =>New values are degrees

A, x =



A, Ax=



MATRIX PROPERTIES

- What about A^2 ?

- Define a new *function*
  - A encodes the *number of paths of lengths exactly* ***1*** *between pairs of nodes*
  - A^2 encodes the *number of paths of lengths exactly* ***2*** *between pairs of nodes*
  - A^3 encodes the *number of paths of lengths exactly* ***3*** *between pairs of nodes*
  - …

- Graph matrices operations can be interpreted as:
  - Diffusion phenomenons
  - Random walks

Graph Spectral properties

Adjacency matrix

Given a simple graph $G = (V, E)$ with an adjacency matrix $\boldsymbol{A}$

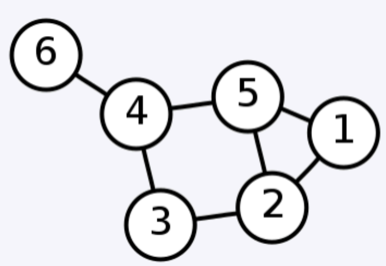
- if G is undirected it has a complete set of real eigenvalues
- Set of eigenvalues define the **spectrum** of G

- Interesting properties:
  - The largest eigenvalue λ_0 of a graph G lies between the average and maximum degrees
  - The number of closed walks of length k in G equals $\sum_{i=0}^{n} \lambda_i^k$
  - A graph is bipartite if and only if its spectrum is symmetric (ie if λ is an eigenvalue, then so is $-\lambda$, and with the same multiplicity).
  - If G is connected, then the diameter of G is strictly less than its number of distinct eigenvalues

Graph Spectral properties

Graph Laplacian $\boldsymbol{L_{(NxN)}=D\text{-}A}$ where D is the degree matrix of G

$$L_{i,j} := \begin{cases} \deg(v_i) & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and } v_i \text{ is adjacent to } v_j \\ 0 & \text{otherwise} \end{cases}$$

Labeled graph	Degree matrix	Adjacency matrix	Laplacian matrix
	$\begin{pmatrix} 2&0&0&0&0&0\\ 0&3&0&0&0&0\\ 0&0&2&0&0&0\\ 0&0&0&3&0&0\\ 0&0&0&0&3&0\\ 0&0&0&0&0&1 \end{pmatrix}$	$\begin{pmatrix} 0&1&0&0&1&0\\ 1&0&1&0&1&0\\ 0&1&0&1&0&0\\ 0&0&1&0&1&1\\ 1&1&0&1&0&0\\ 0&0&0&1&0&0 \end{pmatrix}$	$\begin{pmatrix} 2&-1&0&0&-1&0\\ -1&3&-1&0&-1&0\\ 0&-1&2&-1&0&0\\ 0&0&-1&3&-1&-1\\ -1&-1&0&-1&3&0\\ 0&0&0&-1&0&1 \end{pmatrix}$

- Interesting properties (assuming G is undirected with eigenvalues $\lambda_0 \leq \lambda_1 \leq \ldots \lambda_{n-1}$)
  - L is symmetric and positive definite ($\lambda_i \geq 0$ for all i)
  - $\lambda_0 = 0$ and the number of 0 eigenvalues gives the number of connected components in G

- If G has multiple connected components, **L is a block diagonal matrix**, where each block is the respective Laplacian matrix for each component

Graph Spectral properties

- Graph Spectral Analysis is a whole field of research
- We will introduce more of it in later parts of the course
  - Centralities
  - Community Detection
  - embedding
  - …

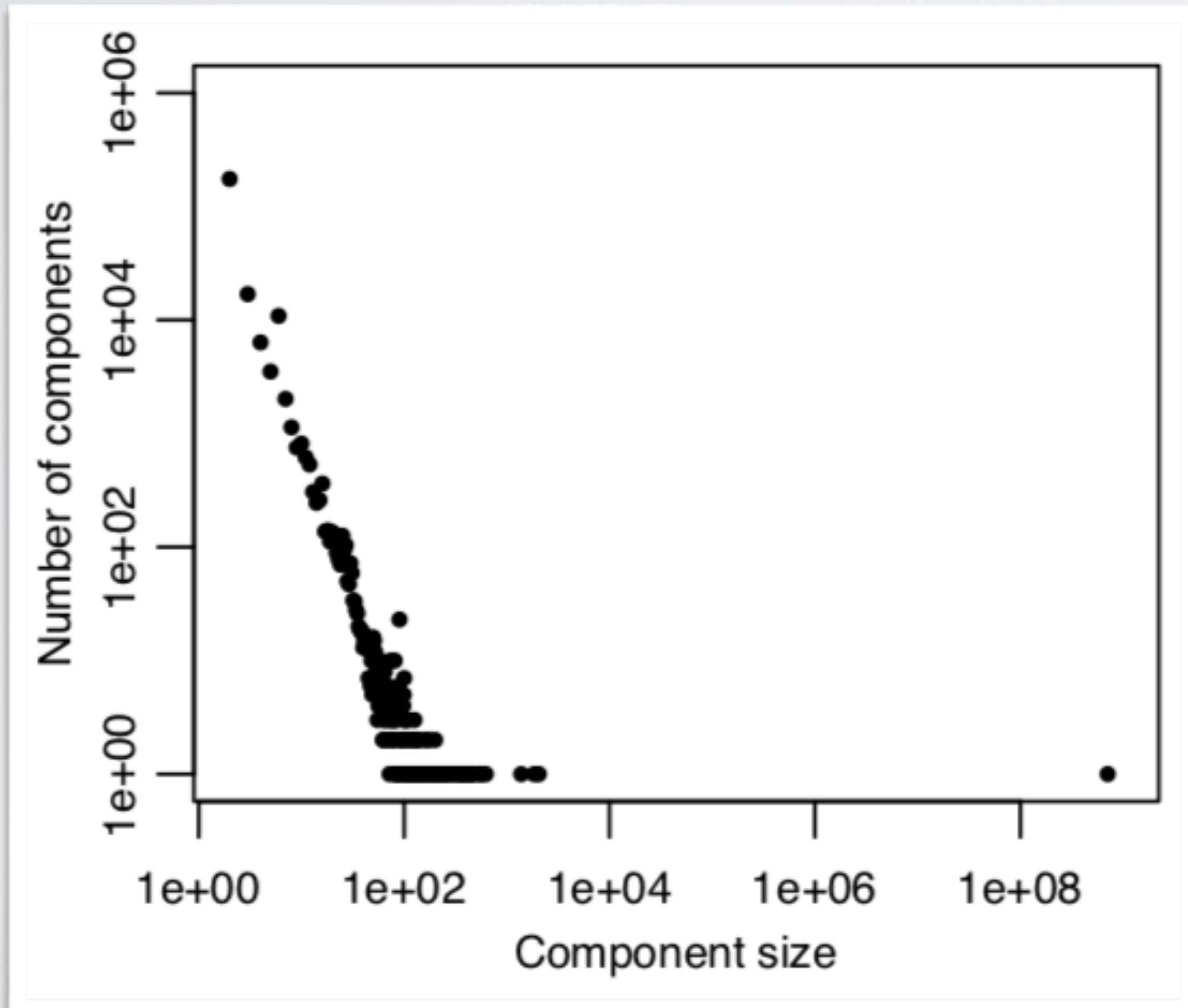
EXEMPLE OF GRAPH ANALYSIS

- Source: [The Anatomy of the Facebook Social Graph, Ugander et al. 2011]
- The Facebook friendship network in 2011

EXEMPLE OF GRAPH ANALYSIS

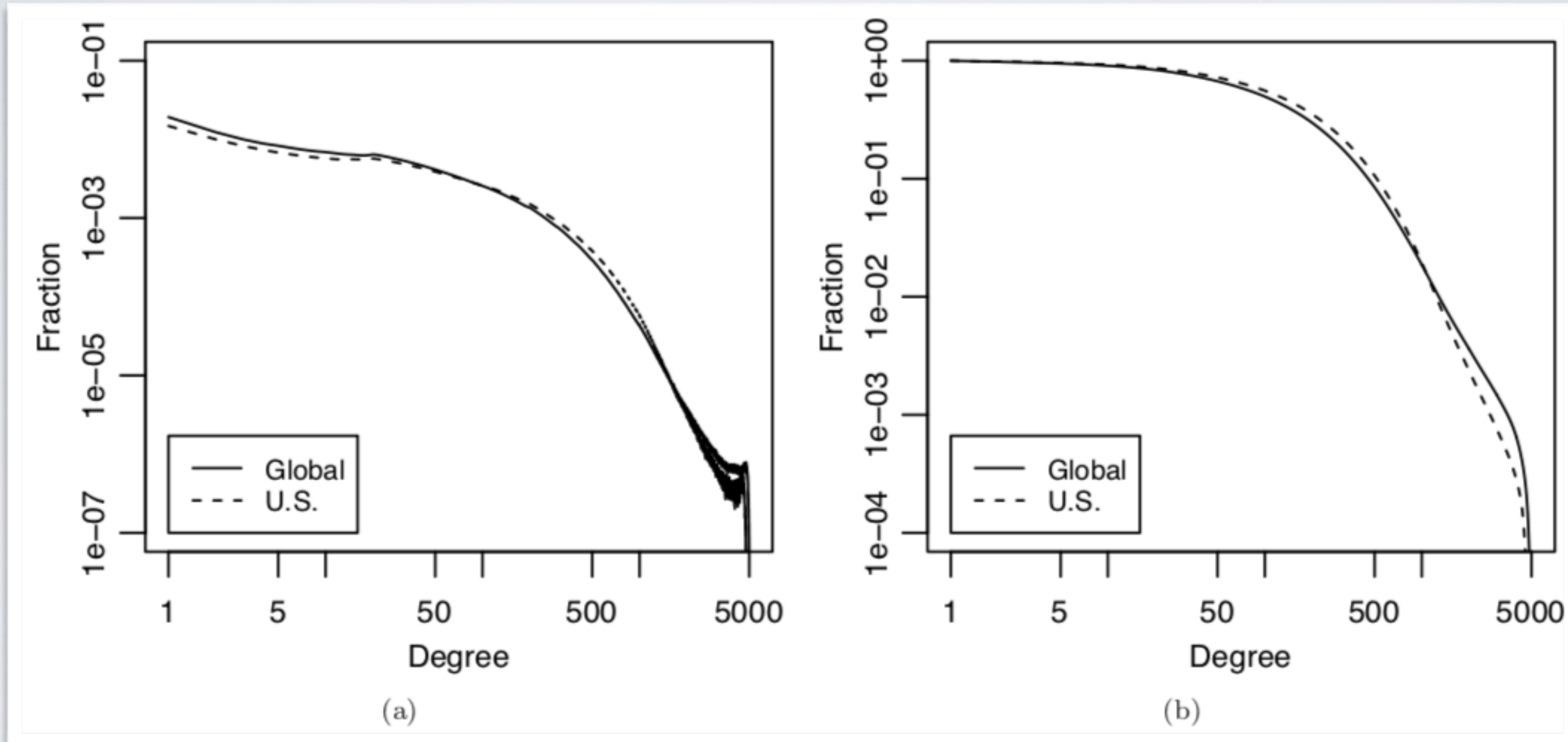
- 721M users (nodes) (active in the last 28 days)
- 68B edges
- Average degree: 190 (average # friends)
- Median degree: 99
- Connected component: 99.91%

EXEMPLE OF GRAPH ANALYSIS



Component size
Distribution

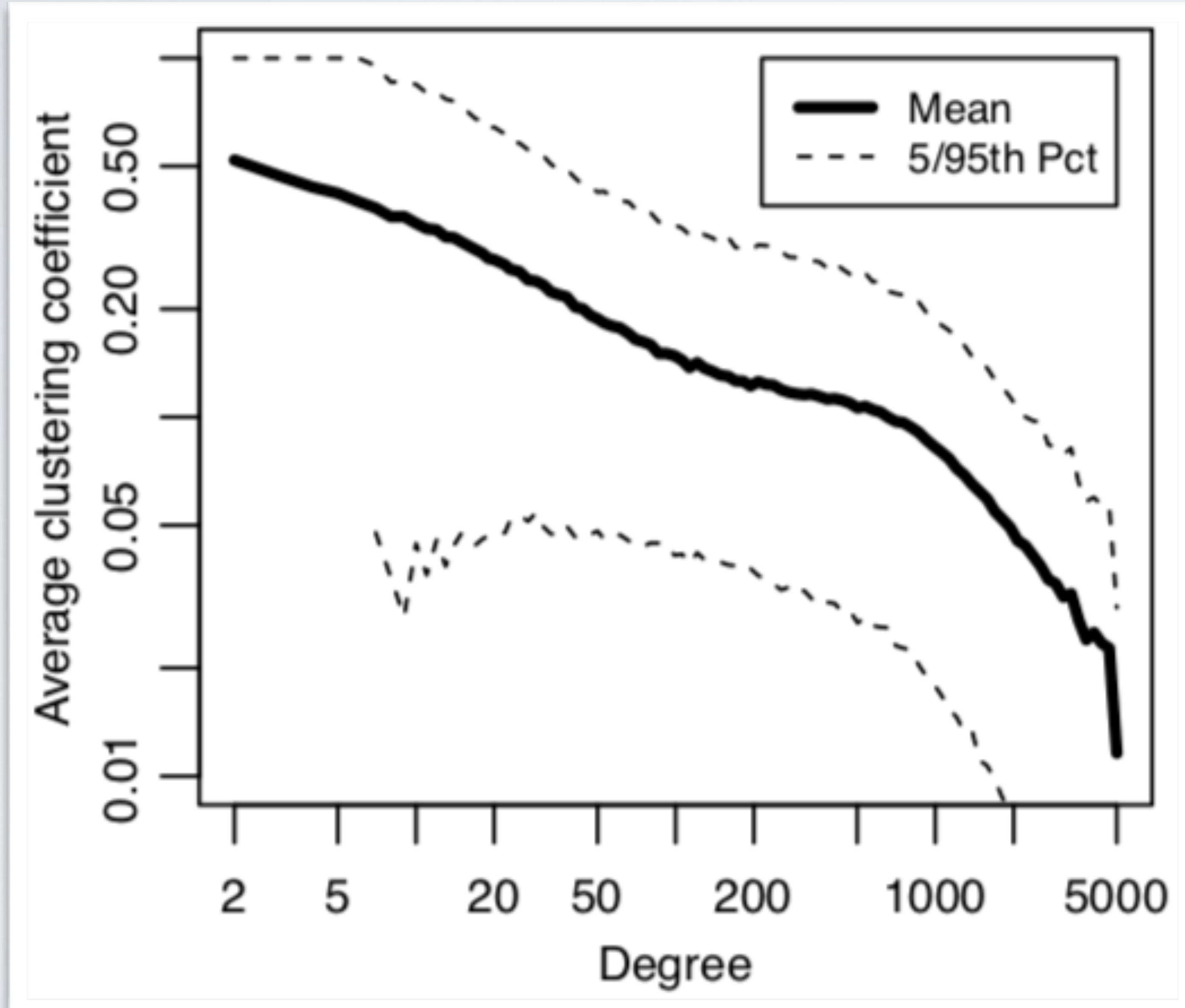
EXEMPLE OF GRAPH ANALYSIS



Cumulative

Degree distribution

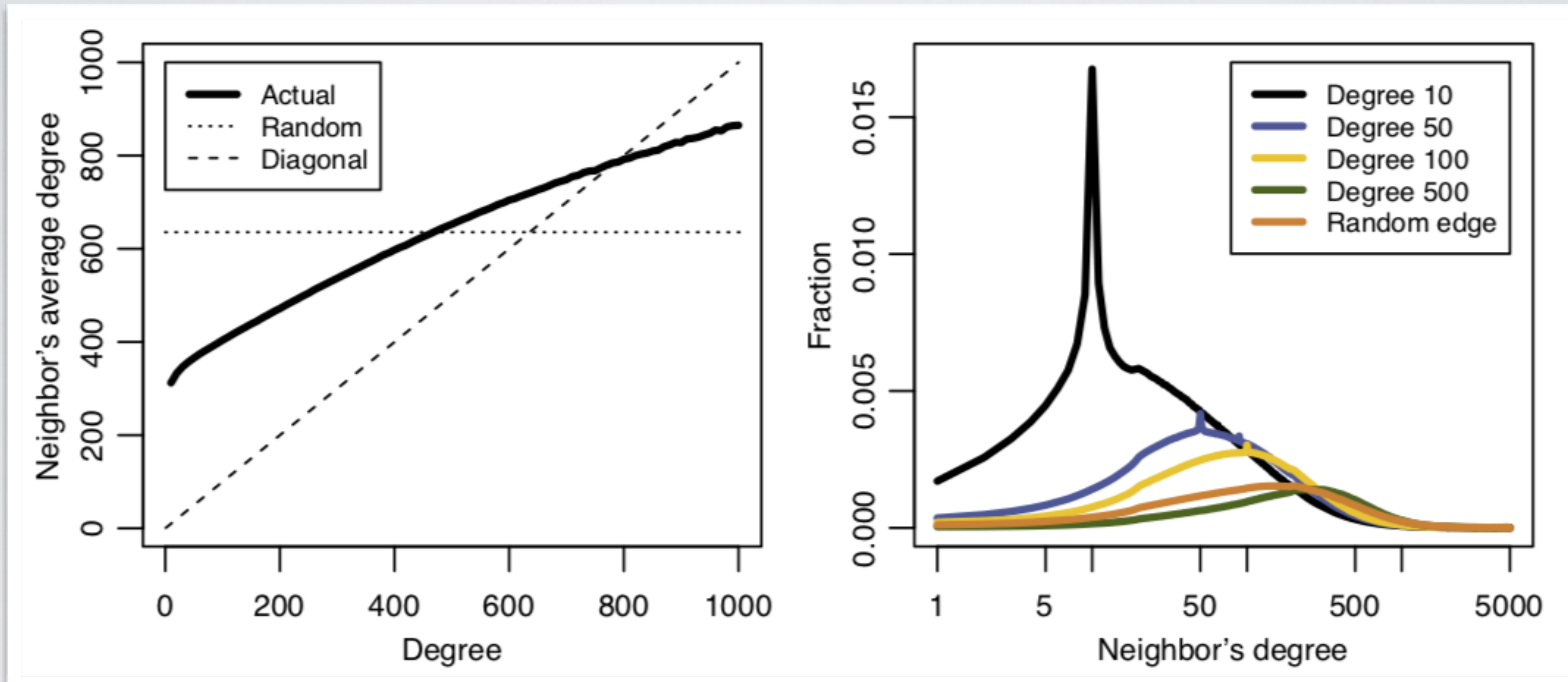
EXEMPLE OF GRAPH ANALYSIS



Clustering coefficient
By degree

Median user: 0.14:
14% of users with a common friend are friends

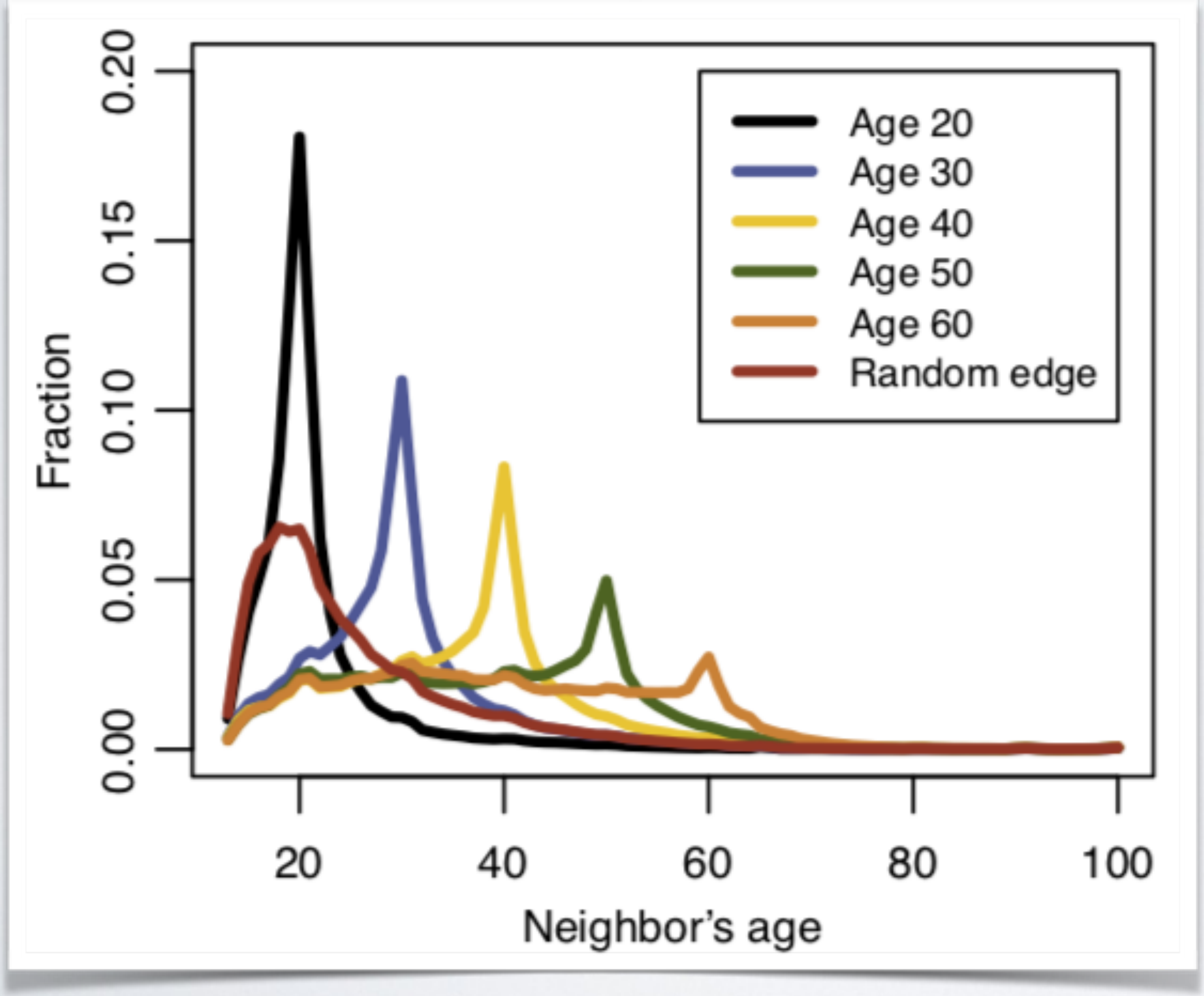
EXEMPLE OF GRAPH ANALYSIS



My friends have more Friends than me!

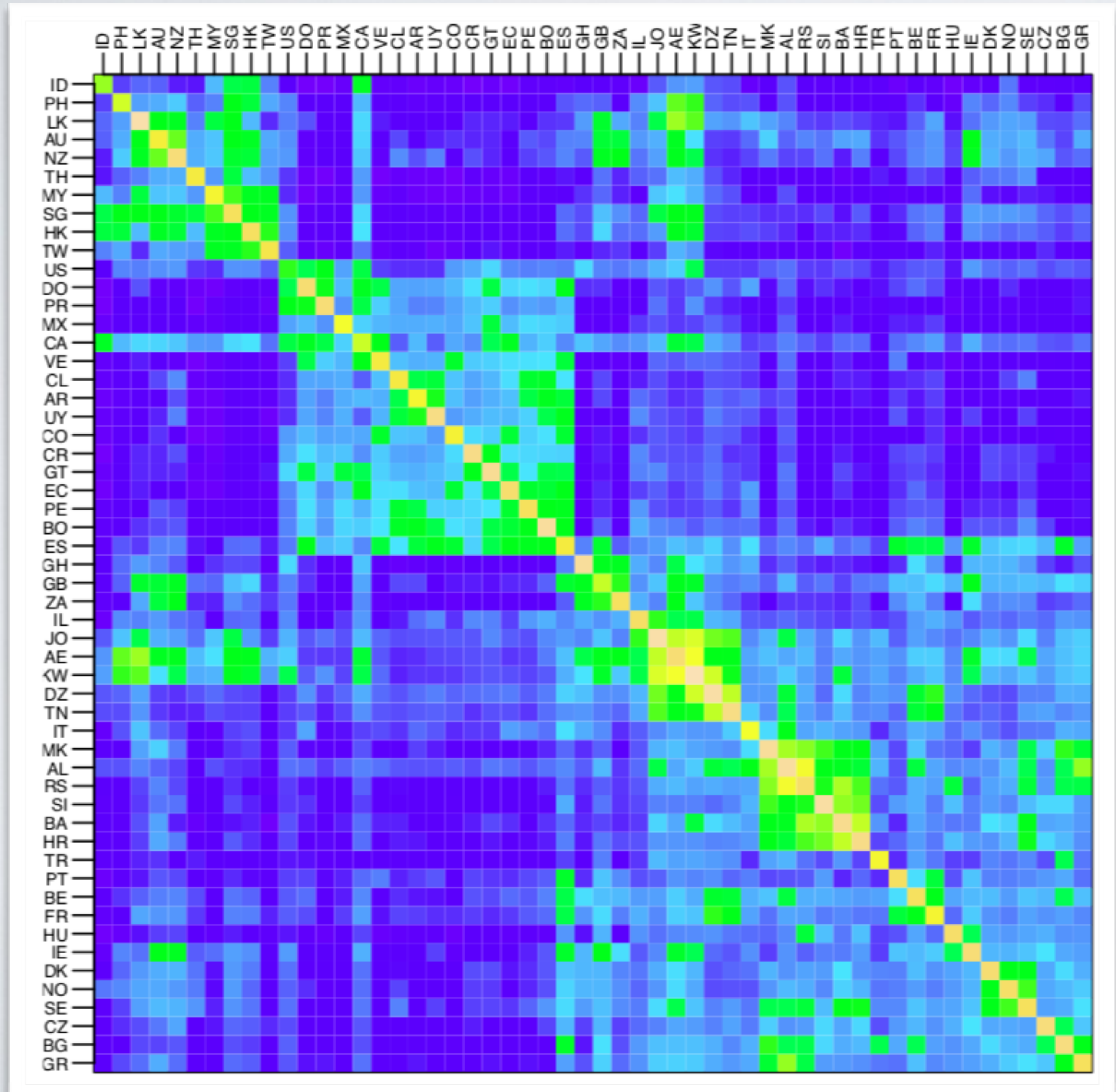
Many of my friends have the Same # of friends than me!

EXEMPLE OF GRAPH ANALYSIS



Age homophily

EXEMPLE OF GRAPH ANALYSIS



Country similarity

84.2% percent of edges are within countries

(More in the community detection class)