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

RESOURCES

- Website of the course:
  - http://cazabetremy.fr/Teaching/ComplexNetworks.html
  - Slides, Cheat sheets, notebooks, etc.

- Contact me: remy.cazabet@univ-lyon1.fr

- I don't have a way to contact you:
  - Please send an email to the address above with: 1)your name, 2)the master you are in (Physics, Computer science, Cognitive science, etc.)

E-LEARNING

- No live streaming (unless needed)
- Recording of classes will be available
- Discord channel, join with: https://discord.gg/vBbPDMAz
  - Ask questions that can be helpful to others, about exams, difficult points, etc.

CLASS OVERVIEW

- Network Science is multi/inter/trans/disciplinary:
  - Students from different Master:
    - Computer Science (CompSci)
    - Complex Systems (Physics, Biology) (CompSys)
    - Cognitive Science (CogSci)
- CompSys+CogSci
  - 24h lectures
  - 3*2h practicals (TD)
- CompSci
  - 32h lectures

EVALUATION

- 60%= Project.
  - In group of 2 or 3.
  - Apply class content to analyse a network of your choice
  - More details later

- 40%=Scientific article presentation
  - During a class or during the first week of January (last class for CompSys)

LECTURES

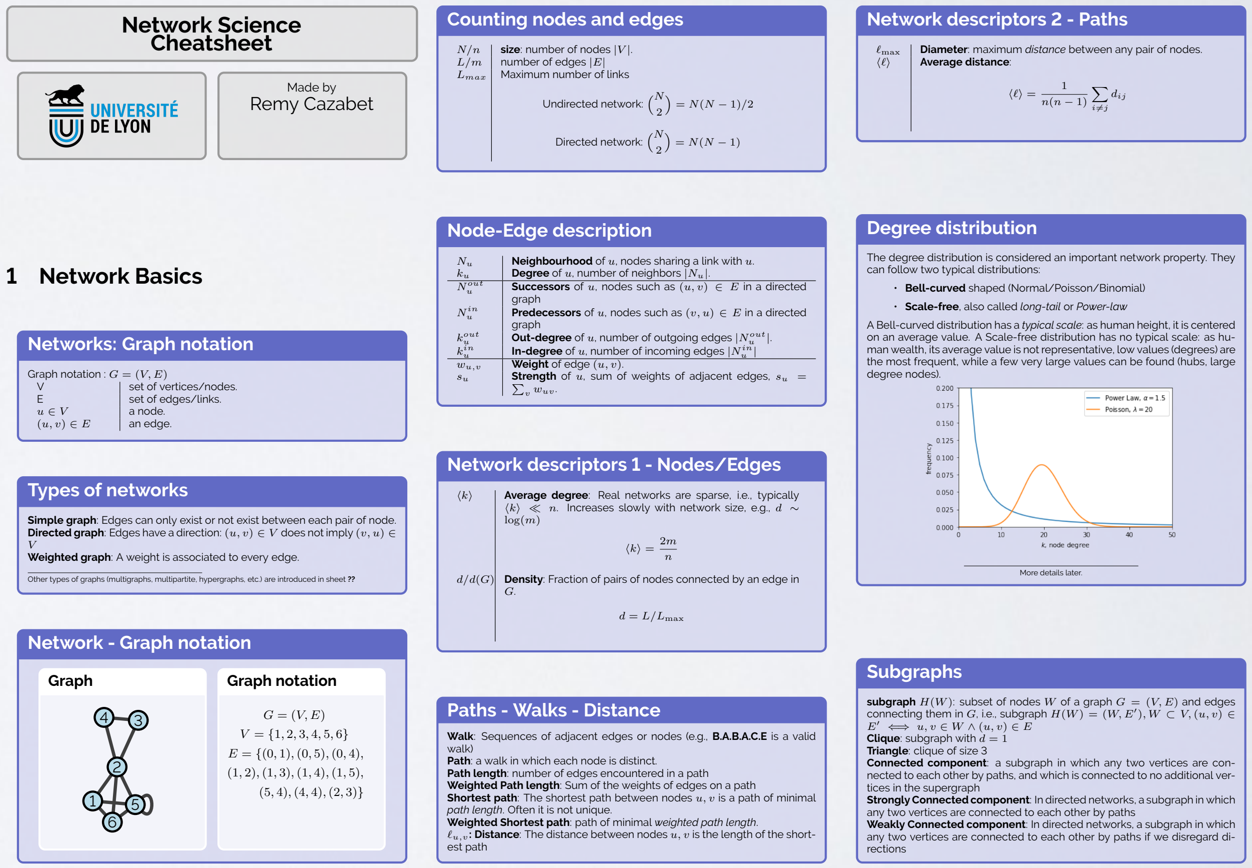
- Most lectures with me. Some lectures with Christophe Crespelle.

- From next session, lectures with me:
  - 1st half: Theory, me talking on slides
  - 2nd half: You experimenting on computers
    - Please try to bring a computer with battery,
  - Please install on your computer:
    - Gephi: https://gephi.org software to manipulate and visualize networks
    - Python, and some libraries: networkx, sklearn, seaborn (for now) cdlib, tnetwork (for later)
    - Also for python: Jupyter notebook.
    - In case of problems with your computer, all the python work can also be done using google colab (https://colab.research.google.com) an online python notebook.

LECTURES

- No need to write down definitions, etc.
  - Slides, Cheatsheet

- Questions welcomed



COMPLEX NETWORKS

(NETWORK SCIENCE)

WHAT?
WHY?
WHY NOW?
WHAT FOR?

SCIENCE

- Science: understanding how things work
  - The human body, the motion/characteristics of objects, societies, etc.
- Step 1: understand properties of things and rules applying to them
  - Fall of objects, classifications of species, etc.
  - Macro-scale properties: temperature, pression

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

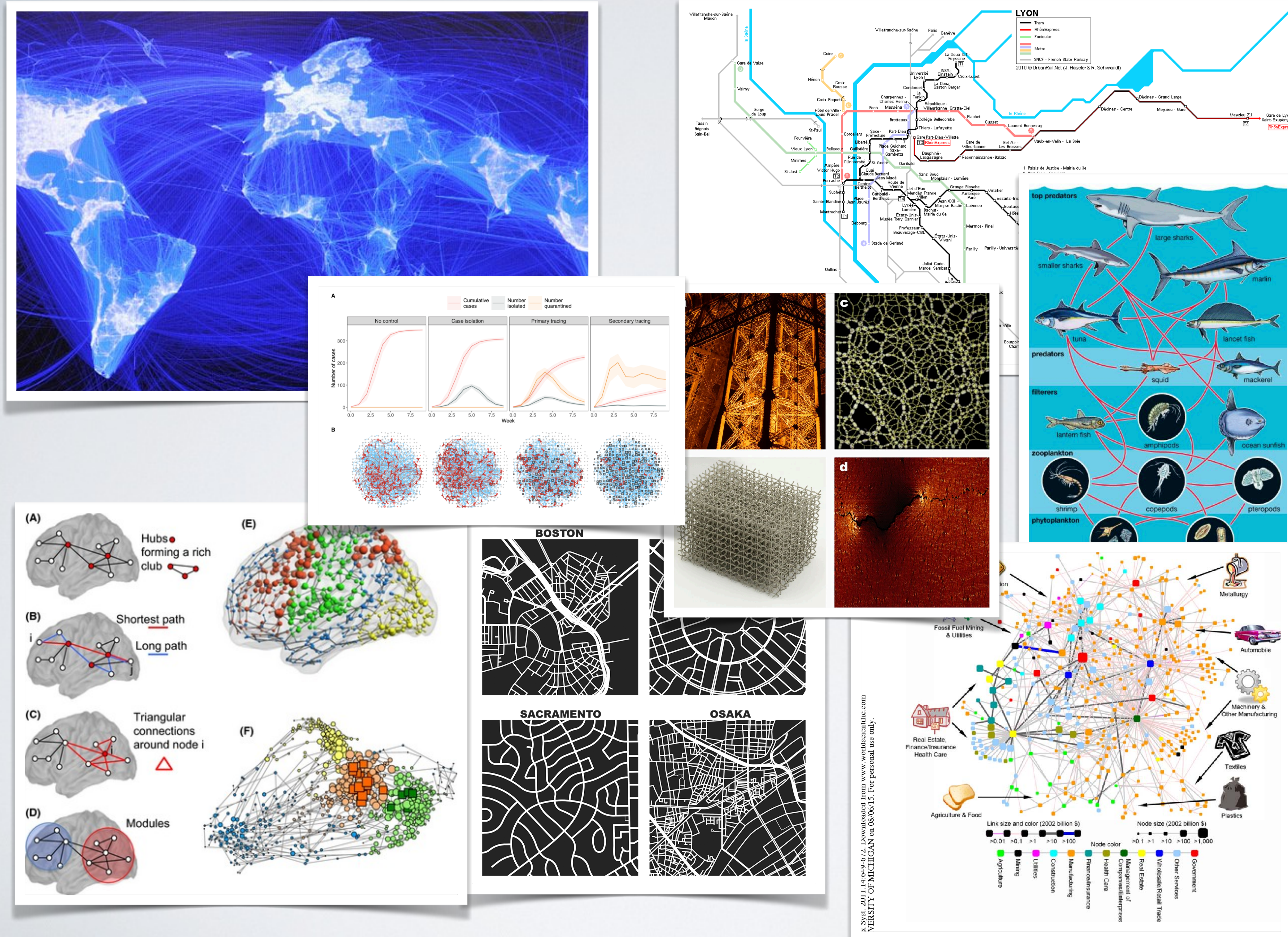
- 3)Two situations:
  - The system is **homogeneous** and/or has a **regular** structure
    - => You can explain it with a bunch of equations
  - The system is **heterogeneous** and/or **has a complex structure**
    - => Understanding each component is not enough to understand the system
    - Understanding each neuron 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



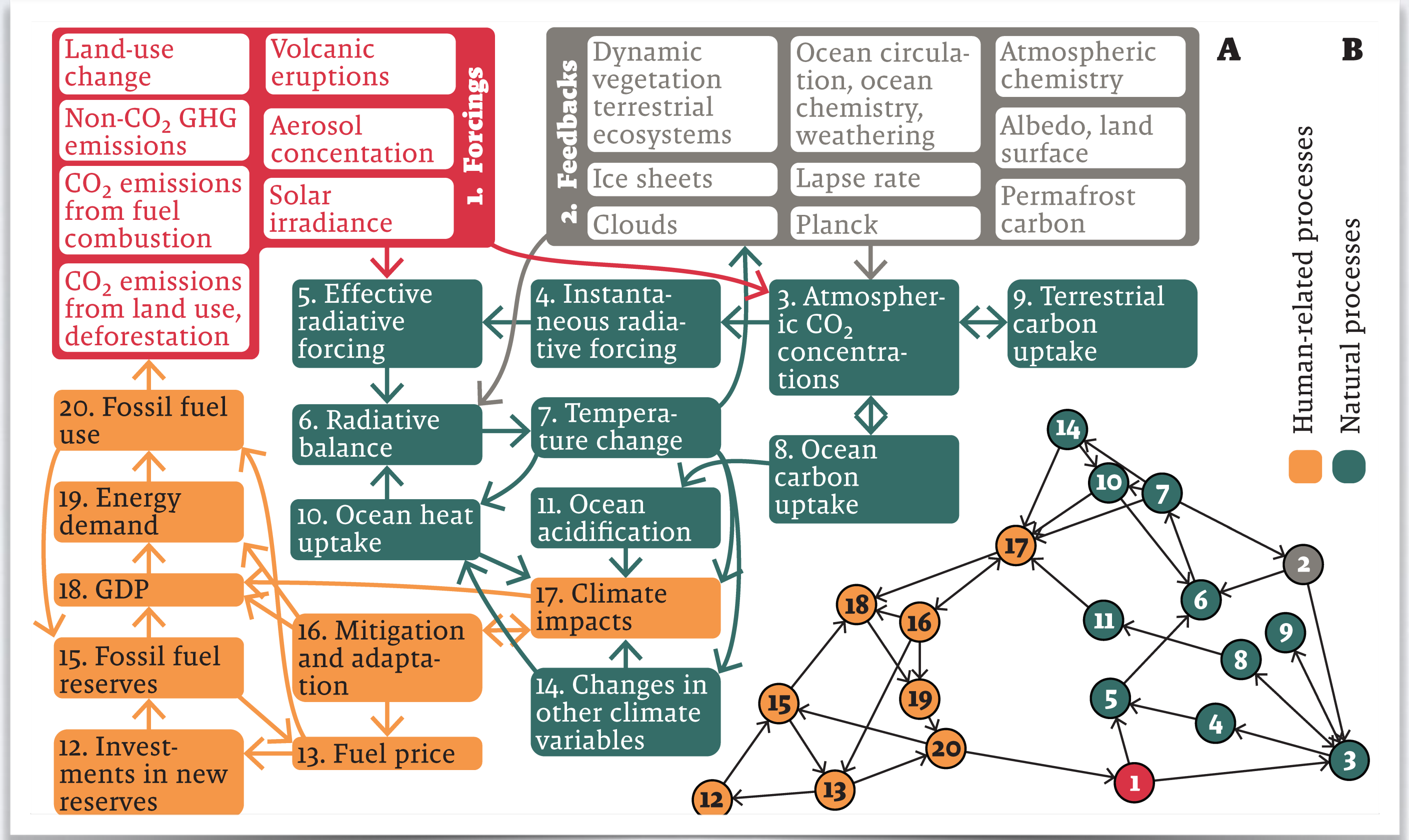
LYON
Tram
RhônExpress
Funicular
Metro
SNCF - French State Railway
2010 © UrbanRail.Net (J. Häseler & R. Schwandl)
top predators
large sharks
smaller sharks
marlin
tuna
lancet fish
predators
squid
mackerel
filterers
lantern fish
amphipods
ocean sunfish
zooplankton
shrimp
copepods
pteropods
phytoplankton
A
Cumulative cases
Number isolated
Number quarantined
No control
Case isolation
Primary tracing
Secondary tracing
Number of cases
Week
B
c
d
(A)
Hubs forming a rich club
(B)
Shortest path
Long path
(C)
Triangular connections around node i
(D)
Modules
(E)
(F)
BOSTON
SACRAMENTO
OSAKA
Metallurgy
Fossil Fuel Mining & Utilities
Automobile
Machinery & Other Manufacturing
Real Estate, Finance/Insurance Health Care
Textiles
Plastics
Agriculture & Food
Link size and color (2002 billion $)
Node size (2002 billion $)
Node color
Agriculture
Mining
Utilities
Construction
Manufacturing
Finance/Insurance
Health Care
Management of Companies/Enterprises
Real Estate
Wholesale/Retail Trade
Other Services
Government

2021 Nobel Prize in physics:

Syukuro Manabe, Klaus Hasselmann, and Giorgio Parisi

For the discovery of the interplay of disorder and fluctuations in physical systems from atomic to planetary scales.

For the physical modelling of Earth's climate, quantifying variability and reliably predicting global warming



https://petterhol.me/2021/10/06/faraway-so-close-nobel-prize-to-complex-systems/

WHO ?

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

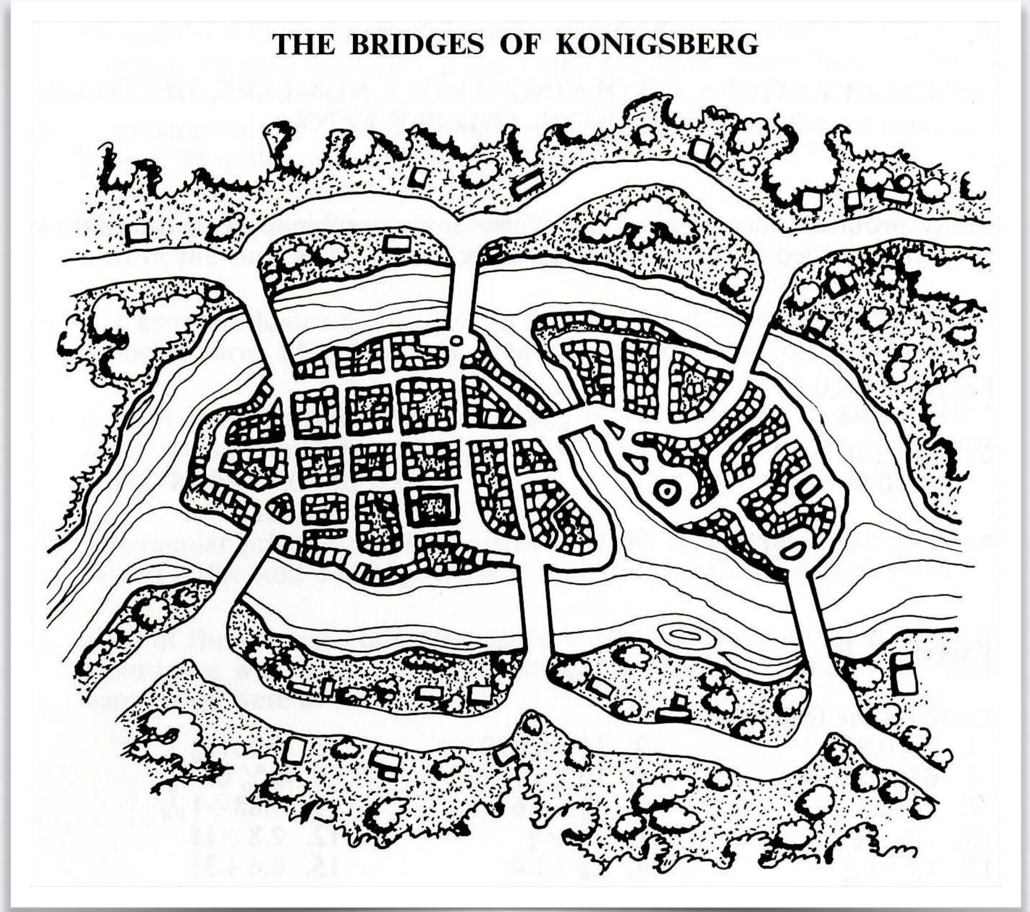
TO CONCLUDE

- 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*.

A BRIEF HISTORY

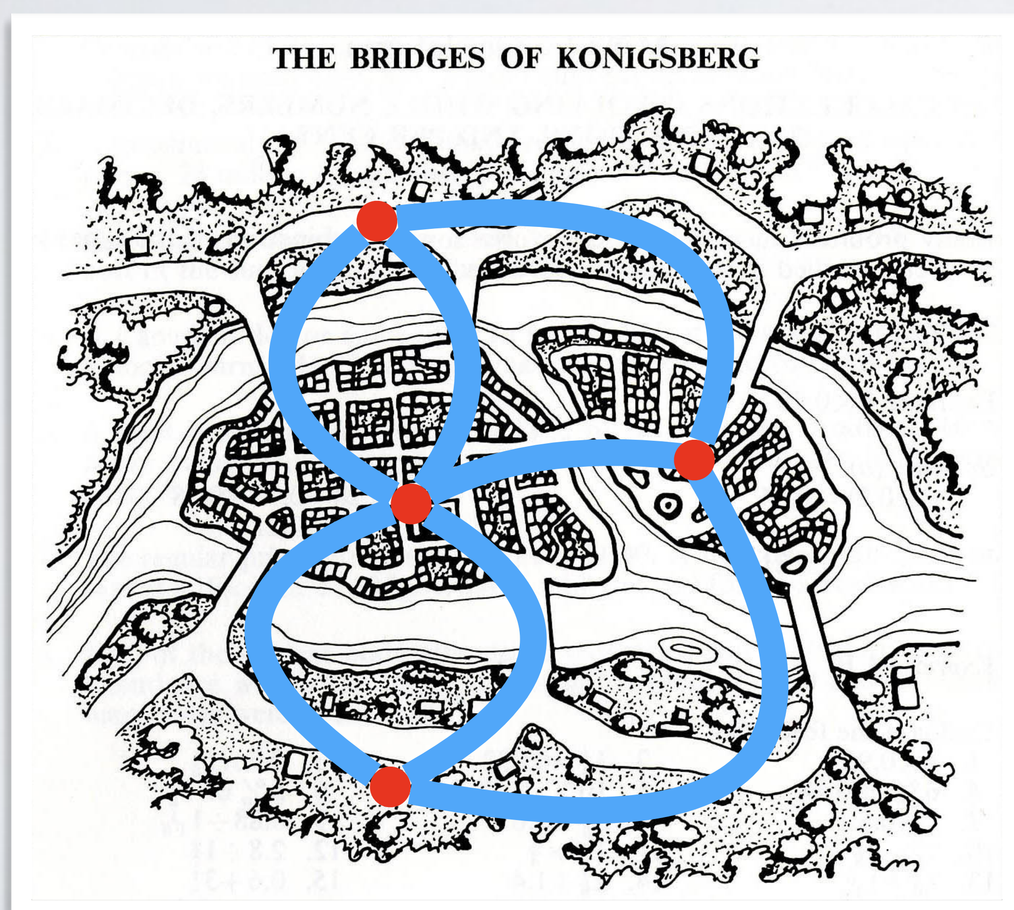
A BRIEF HISTORY

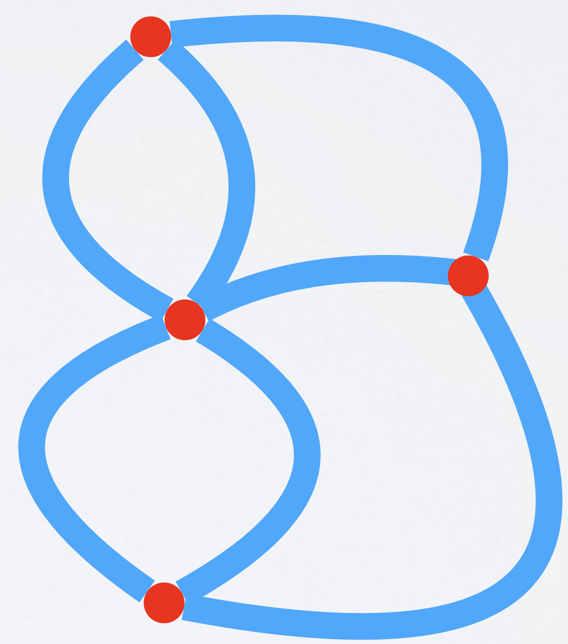
- Graph theory: 1736 - Euler and the bridges of konigsberg



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

A BRIEF HISTORY





Answer: **No**

A BRIEF HISTORY

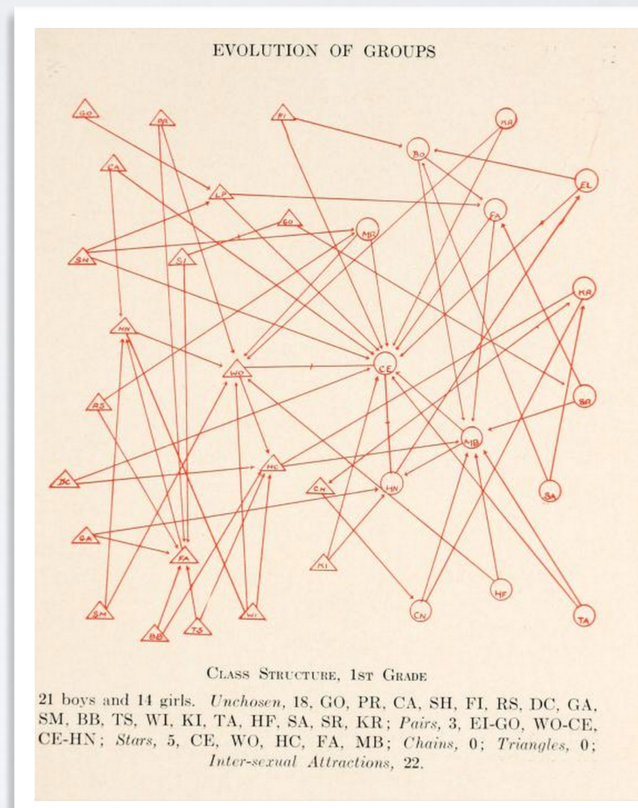
- Social networks: 1934 - Jacob Moreno

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

Sociograms

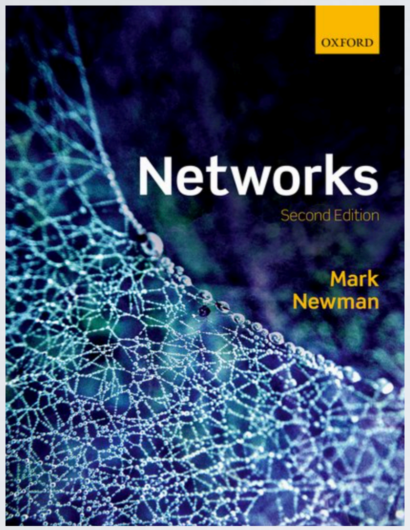
KEY PUBLICATIONS

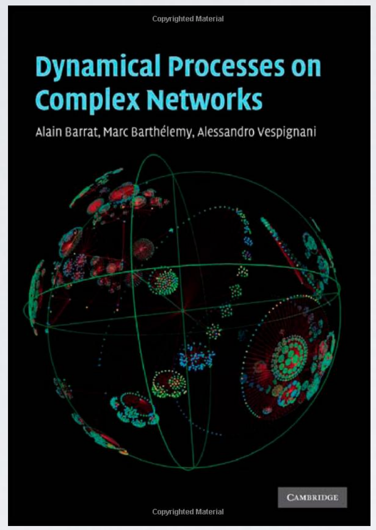
- 1998: Watts & Strogatz - Small-World:
  - 2nd Most cited paper of the year in Nature
- 1999: Barabasi & Albert - scale-free networks:
  - Most cited paper of the year in Science
- 2002: Girvan & Newman - Community detection:
  - Most cited paper of the year in PNAS
- 2004: Barabasi & Oltvai - Network 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
- …

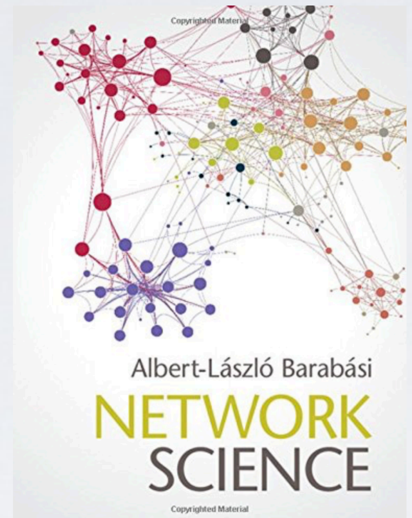
(As of 2020)

Materials

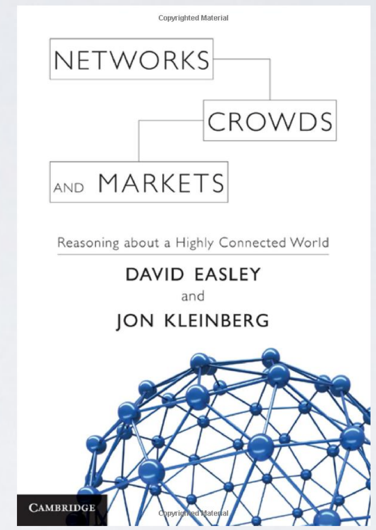
Lecture books



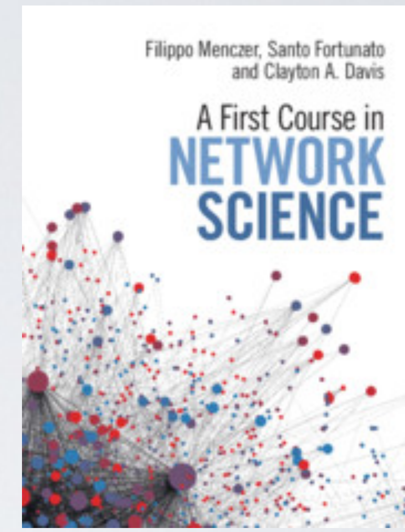




available free online



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
Physics Reports
journal homepage: www.elsevier.com/locate/physrep
ELSEVIER

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
Physics Reports
journal homepage: www.elsevier.com/locate/physrep
ELSEVIER

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
Physics Reports
journal homepage: www.elsevier.com/locate/physrep
ELSEVIER

Spatial networks

Marc Barthélemy *

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

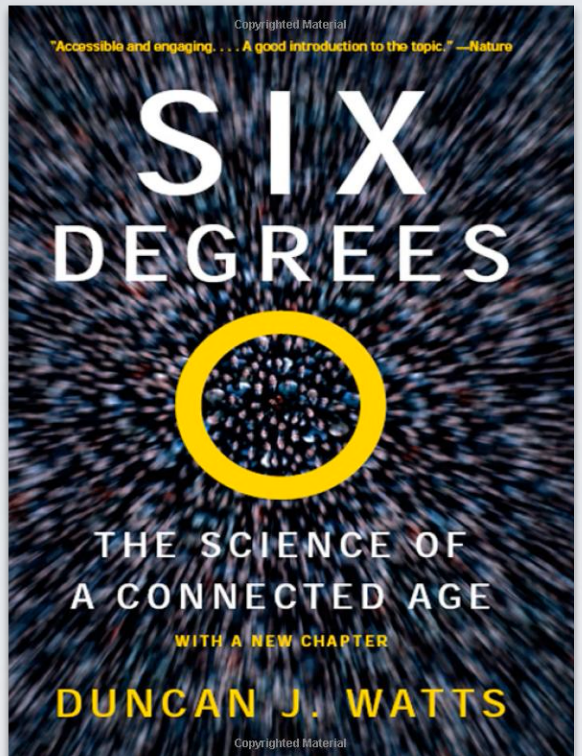
The structure and dynamics of multilayer networks

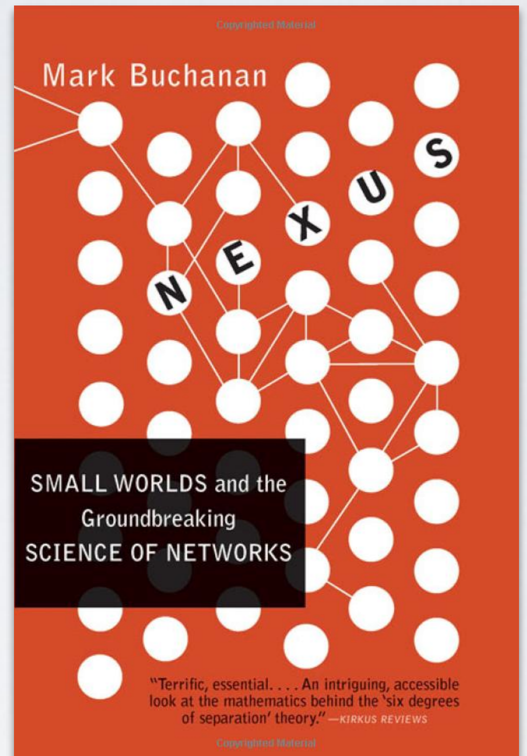
S. Boccaletti [a,b,*], G. Bianconi [c], R. Criado [d,e], C.I. del Genio [f,g,h], J. Gómez-Gardeñes [i], M. Romance [d,e], I. Sendiña-Nadal [j,e], Z. Wang [k,l], M. Zanin [m,n]

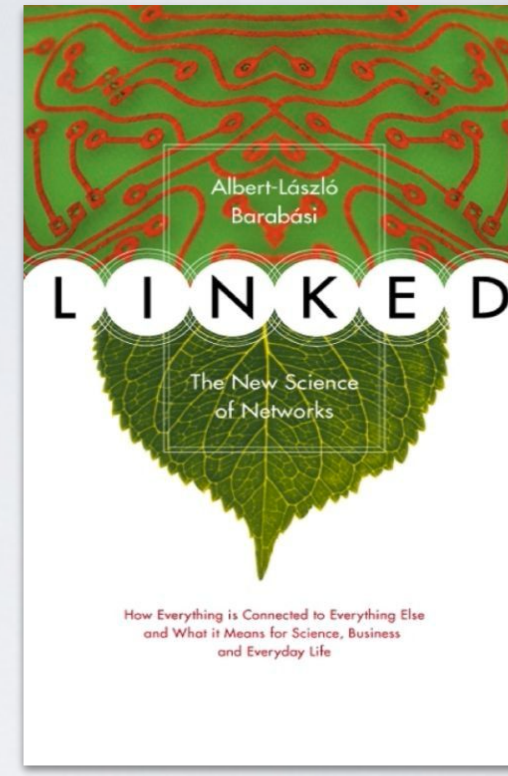
…and many more…all of them on arXiv.org!

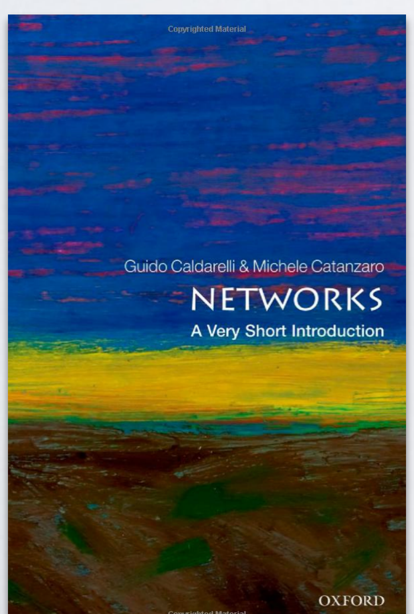
Materials

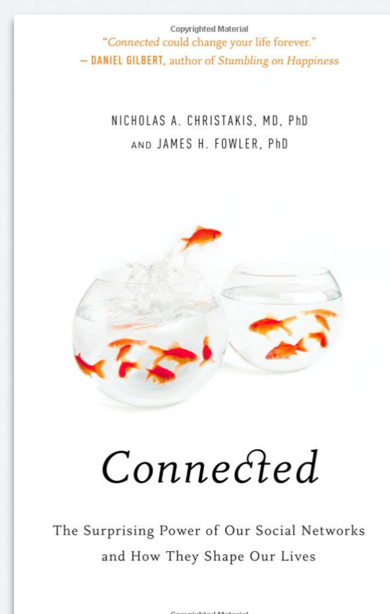
Pop-science books

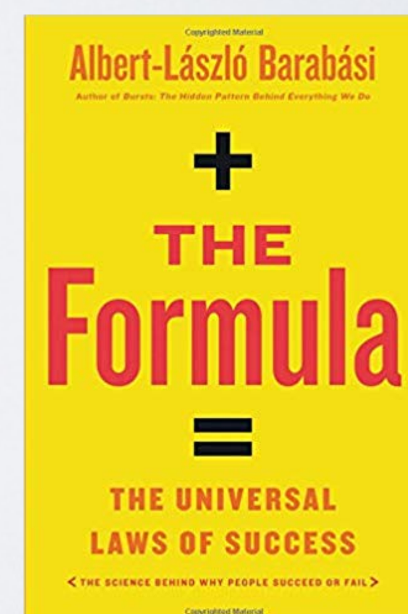








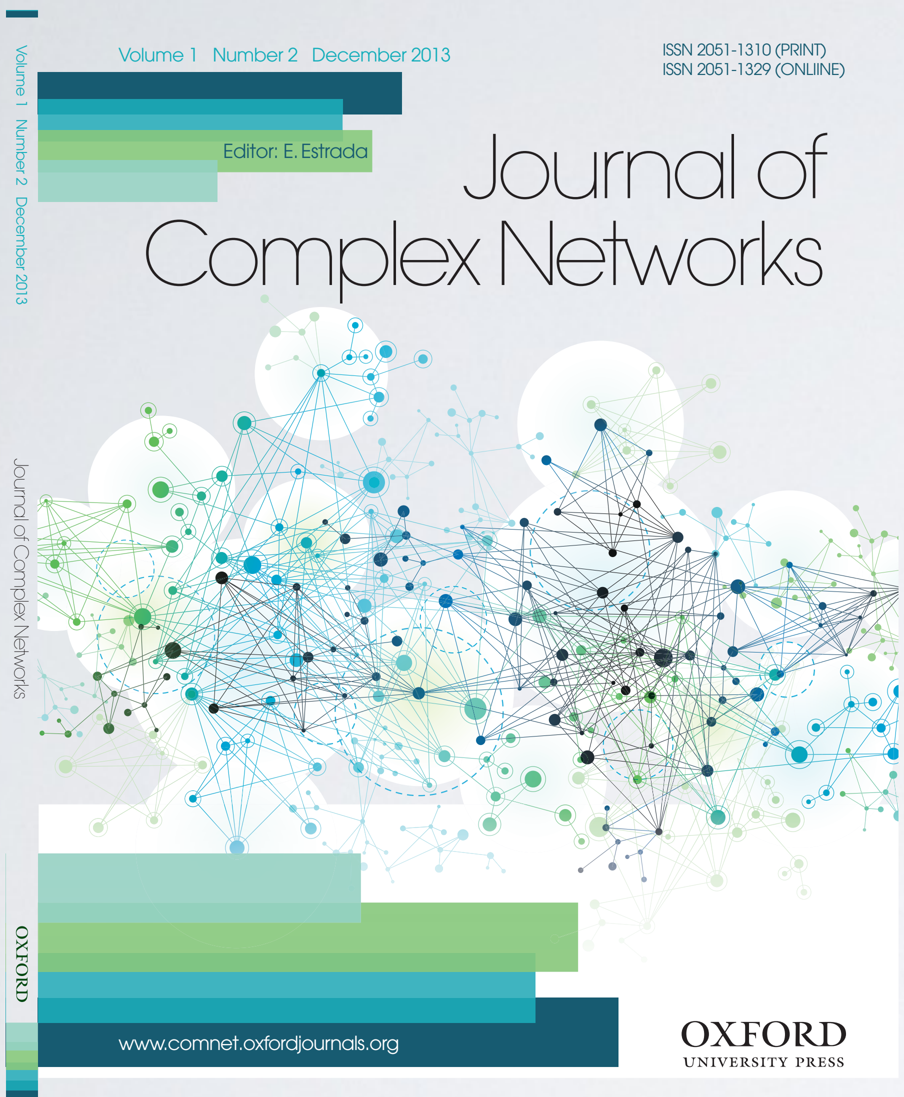


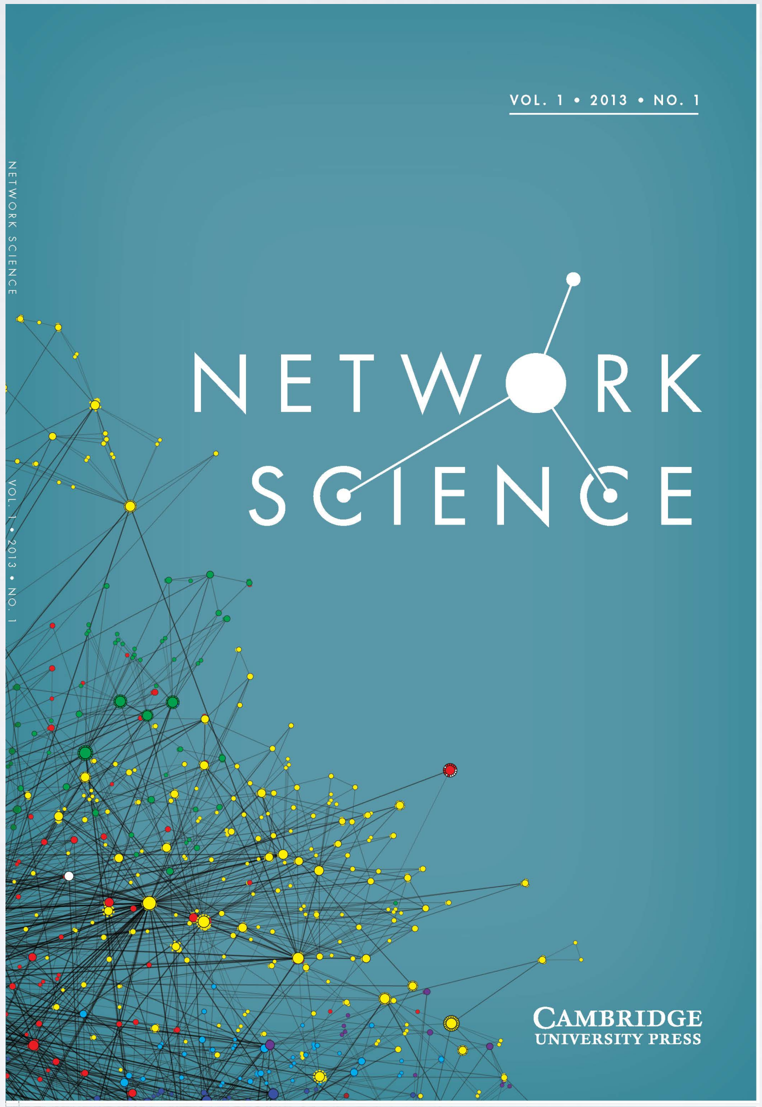


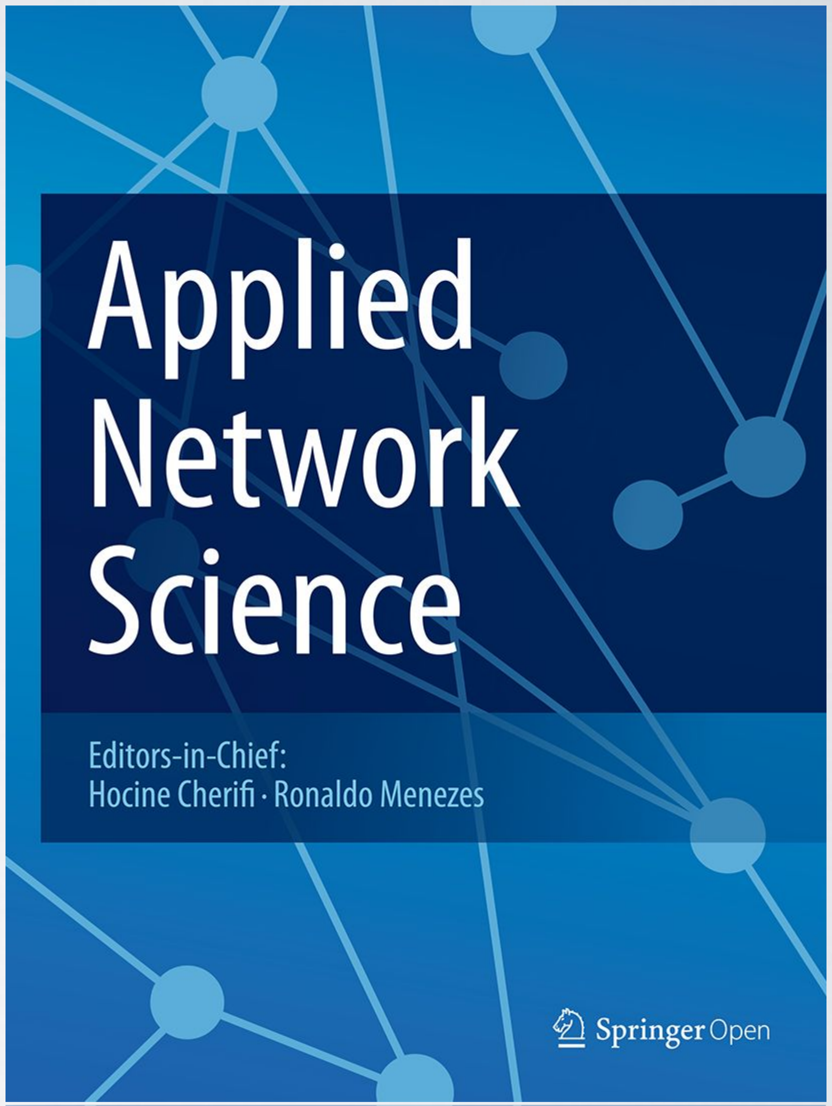
I have a copy I can lend

Materials

Specific Journals







CONFERENCES

- NetSci, NetSci X - The Network Science Society (Since 2006)
- International Conference on Complex Networks and their Applications (Since 2011)
- CompleNet - International Conference on Complex Networks (Since 2009)
- France:
  - MARAMI (Modèles & Analyse des Réseaux : Approches Mathématiques & Informatiques) (Since 2009)

PROGRAM

Day	Time	Room	Group	Topic	Resources
Tuesday Nov.16	8h00-10h00	B	All	Introduction, Describing Networks	
Thursday Nov. 18	10h15-12h15	C	All	Centralities, Gephi, networkx intro	
Tuesday Nov. 23	08h00-10h00	B	All	**Teacher: Christophe Crespelle**. Phase transition in ER random graphs	
Thursday Nov. 25	8h00-10h00	C	**CS only**	(practicals)Data to Network: Scientometric Networks PDF	
Thursday Nov. 25	10h15-12h15	C	All	Random Graph Models II, Community Structure	
Tuesday Nov. 30	08h00-10h00	B	All	**Teacher: Christophe Crespelle**. Community detection algorithms.	
Thursday Dec. 2	8h00-10h00	C	**CS only**	(practicals)Data to Network: Movies PDF	
Thursday Dec. 2	10:15-12:15	C	All	Community Evaluation, Hypergraphs, Multigraphs, etc.	
Tuesday Dec. 7	08h00-10h00	B	All	Visualization - Assortativity	
Thursday Dec. 9	8h00-10h00	C	**CS only**	(practicals)Data to Network: Project	
Thursday Dec. 9	10h15-12h15	C	All	Dynamic Networks	
Tuesday Dec. 14	8h00-10h00	B	All	Spatial Networks	
Thursday Dec. 16	8h00-10h00	C	**CS only**	(practicals)Data to Network: Project + Optional	
Thursday Dec. 16	10h15-12h15	C	All	Spreading Processes	
Tuesday Jan. 4	8h00-10h00	B	All	Machine Learning on graphs (Link Prediction, Node Classification)	
Thursday Jan. 6	10h15-12h15	C	All	End of article presentations	
Tuesday Jan. 11	8h00-10h00	B	**Info only**	**Teacher: Christophe Crespelle**. Betweenness centrality and graph editing	
Thursday Jan. 13	10h15-12h15	B1	**Info only**	Graph Embedding	
Tuesday Jan. 18	8h00-10h00	B	**Info only**	Graph Convolutional Networks	
Thursday Jan. 20	10h15-12h15	B1	**Info only**	TBA	

INTERNSHIPS

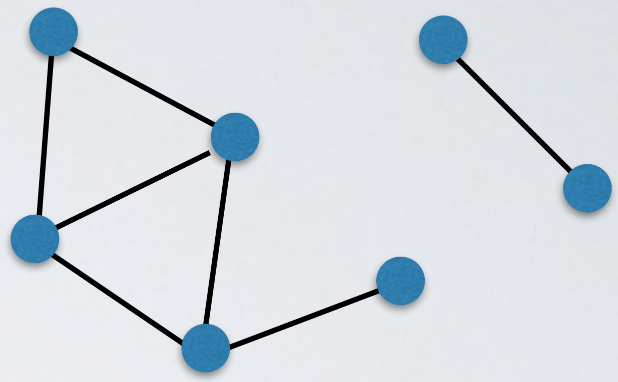
Graph Analysis for illegal activity tracking in Bitcoin transaction network

http://cazabetremy.fr/rRessources/Bitcoin_Internship.pdf

Contact me before the end of the week !

GRAPHS & NETWORKS

GRAPHS & NETWORKS



Network often refers to real systems

- www,
- social network
- metabolic network.
- Language: (Network, node, link)

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

Graph is the mathematical representation of a network

- Language: (Graph, vertex, edge)

In most cases we will use the two terms interchangeably.

GRAPH REPRESENTATION

NETWORK REPRESENTATIONS

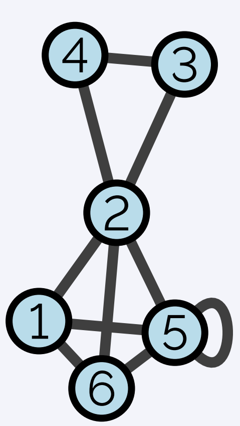
Networks: Graph notation

Graph notation : $G = (V, E)$

V	set of vertices/nodes.
E	set of edges/links.
$u \in V$	a node.
$(u, v) \in E$	an edge.

Network - Graph notation

Graph



Graph notation

$$G = (V, E)$$
$$V = \{1, 2, 3, 4, 5, 6\}$$
$$E = \{(1, 2), (1, 6), (1, 5), (2, 4), (2, 3), (2, 5), (2, 6), (6, 5), (5, 5), (4, 3)\}$$

NETWORK REPRESENTATIONS

- $G = (V, E)$
  - Often encoded as **edge list** or **adjacency list**
- Software: custom data structure and manipulation
  - add_nodes([i,j]), add_edge(i,j), …
- Libraries in many languages
  - Networkx (python)
  - igraph (python, C, R)
  - Graph-tools (python, C)

```
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
```

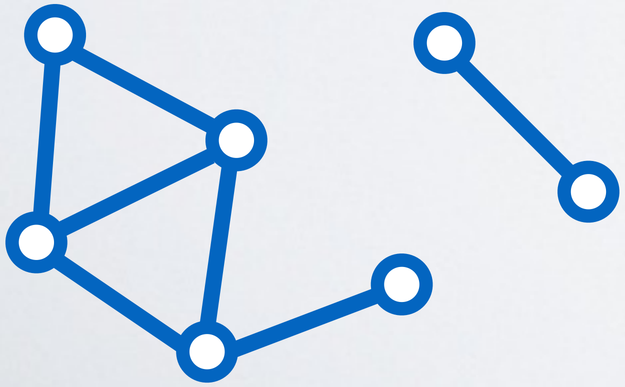
Types of Networks

Undirected networks

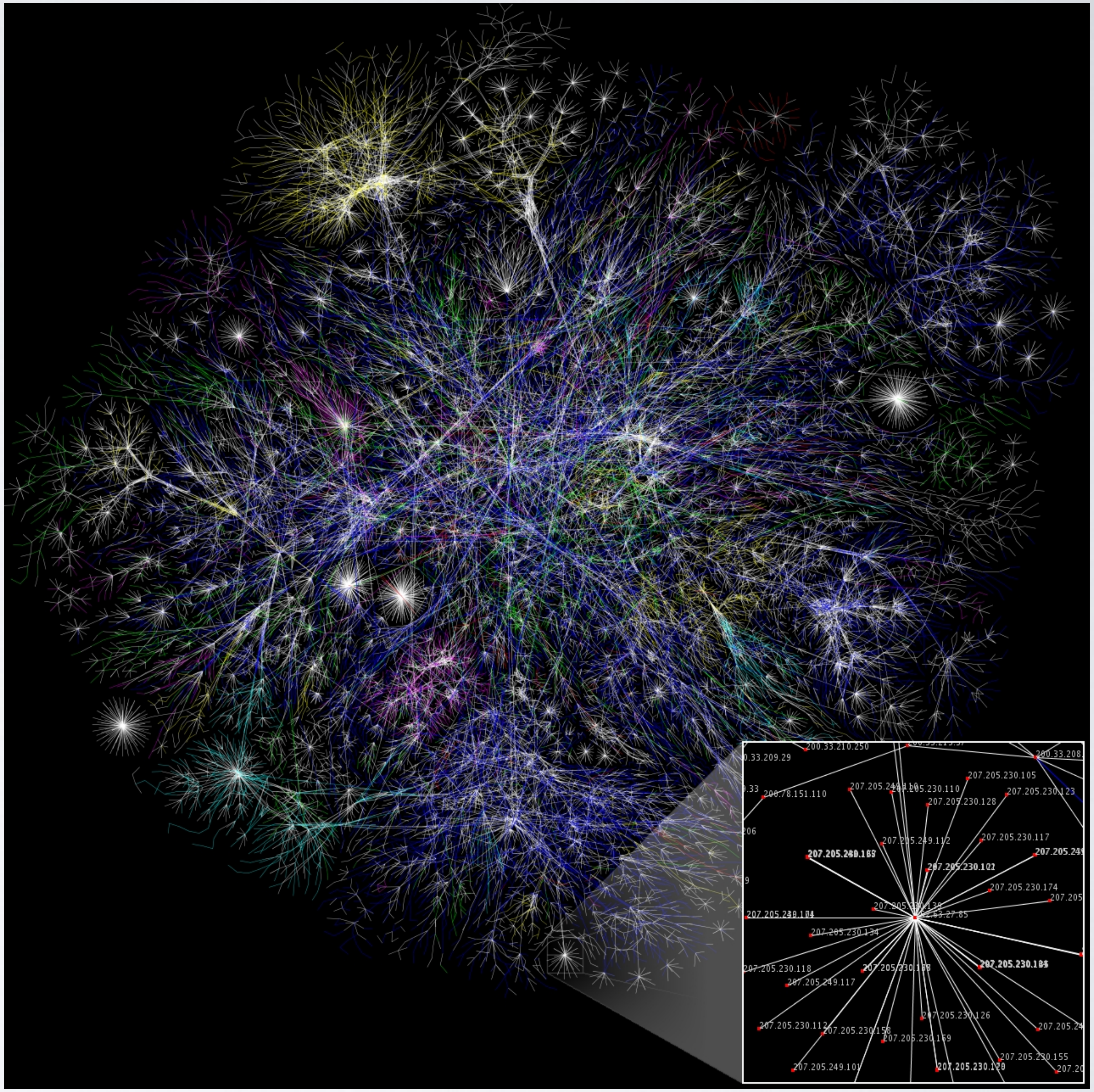
$G=(V, E)$

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

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



Opte project



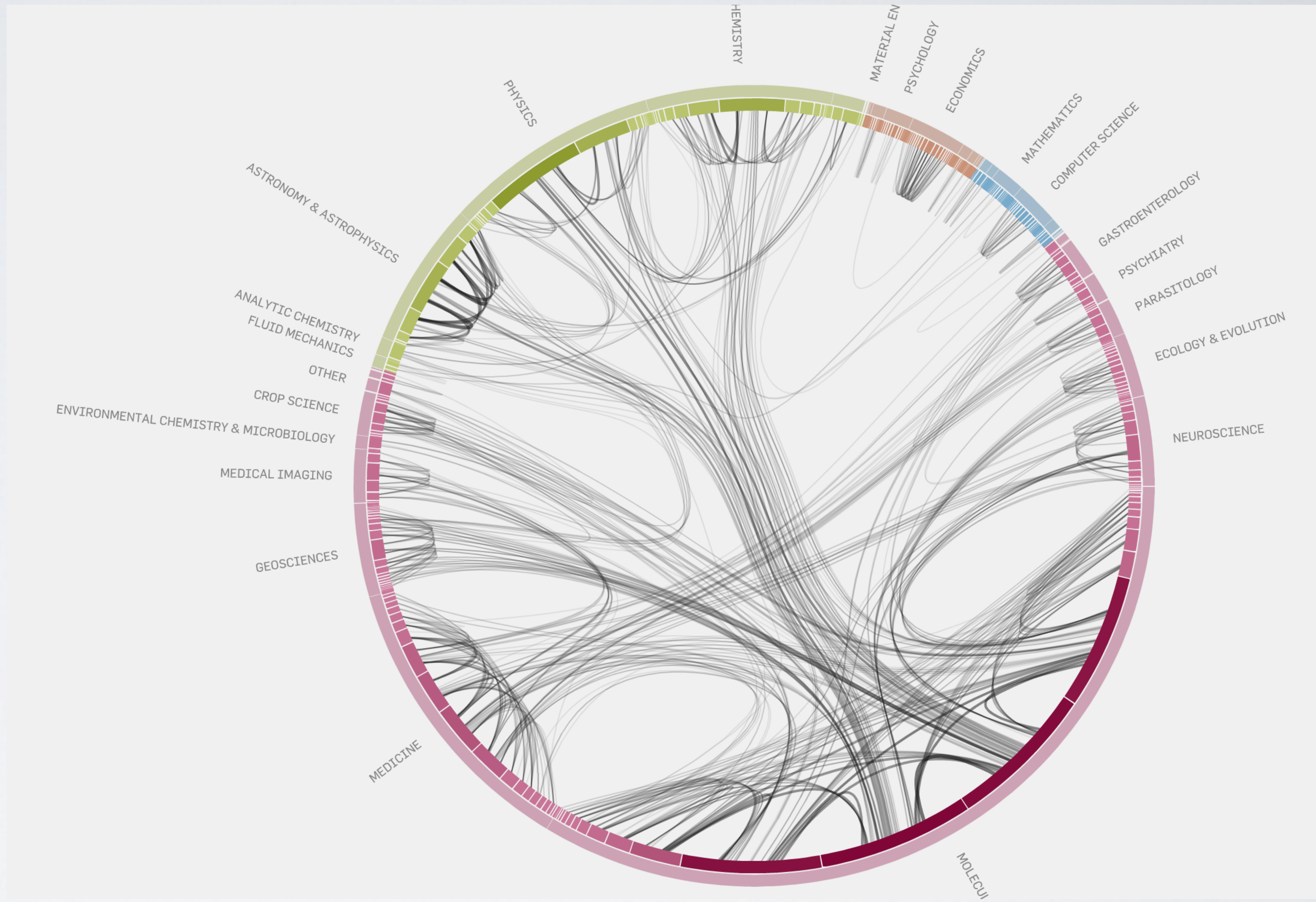
The Internet: Nodes - routers, Links - physical wires

Directed networks

Moritz Stefaner, eigenfactor.com

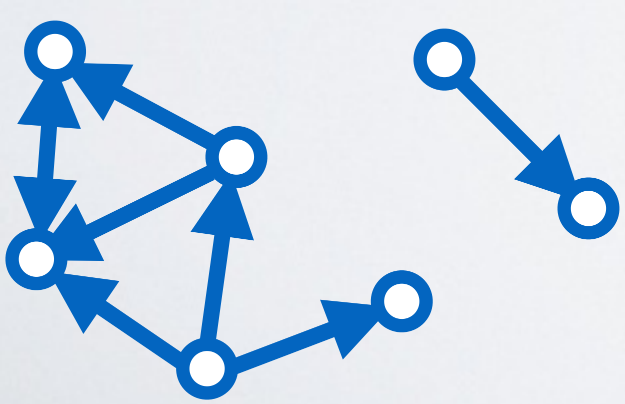
$G=(V, E)$

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



Citation network: Nodes - publications, Links - references

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

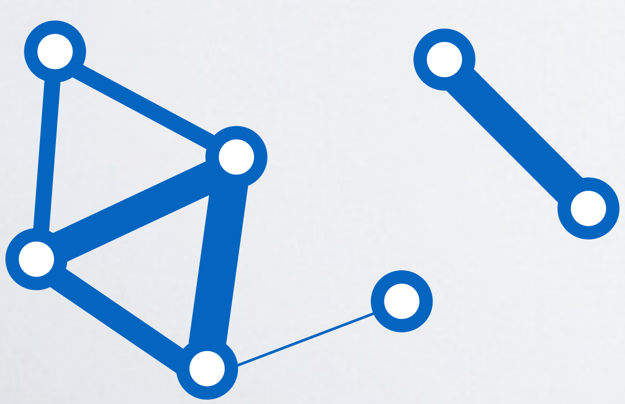


Weighted networks

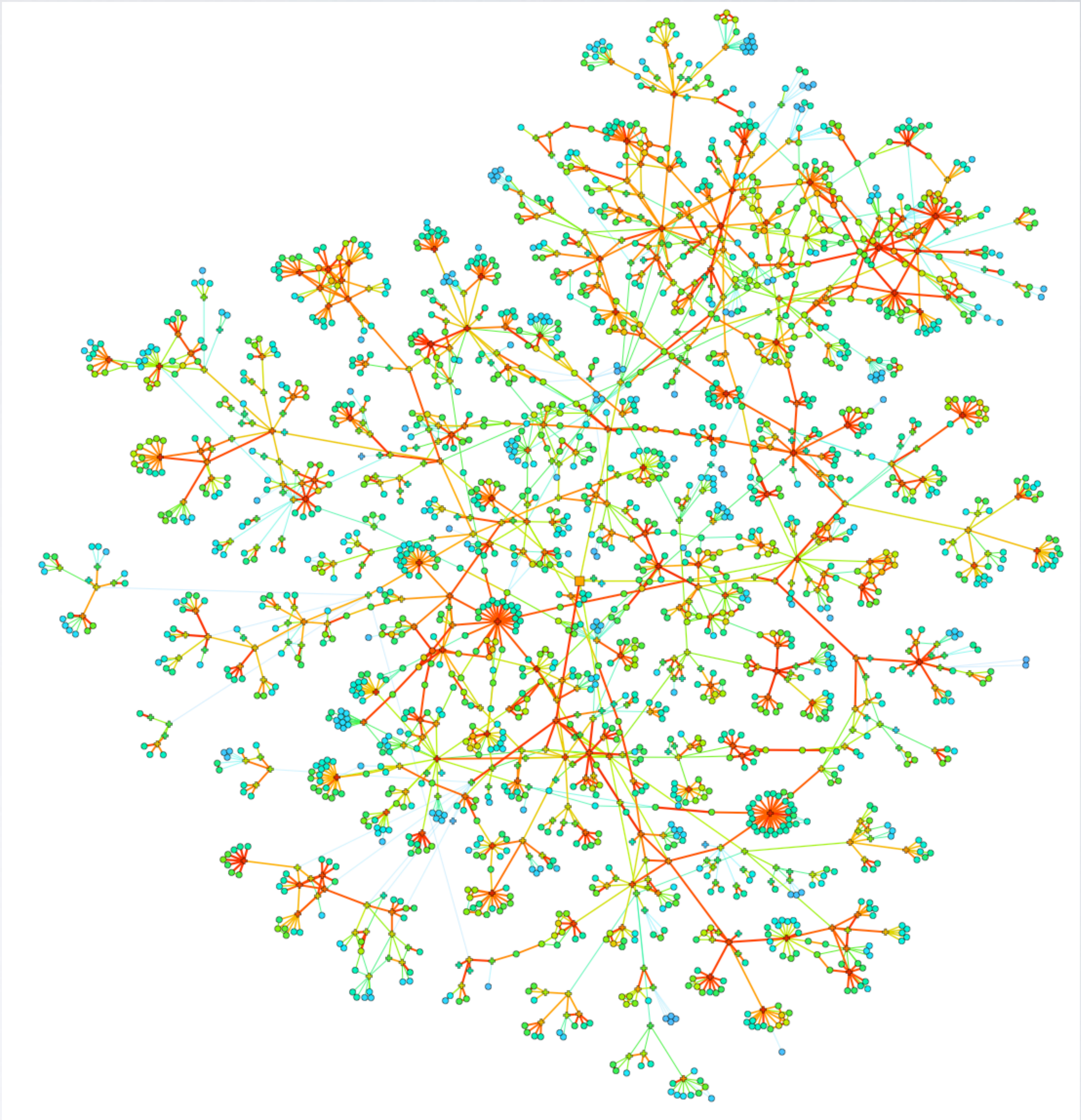
$G=(V, E, w)$

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

- Strength of interactions are assigned by the weight of links



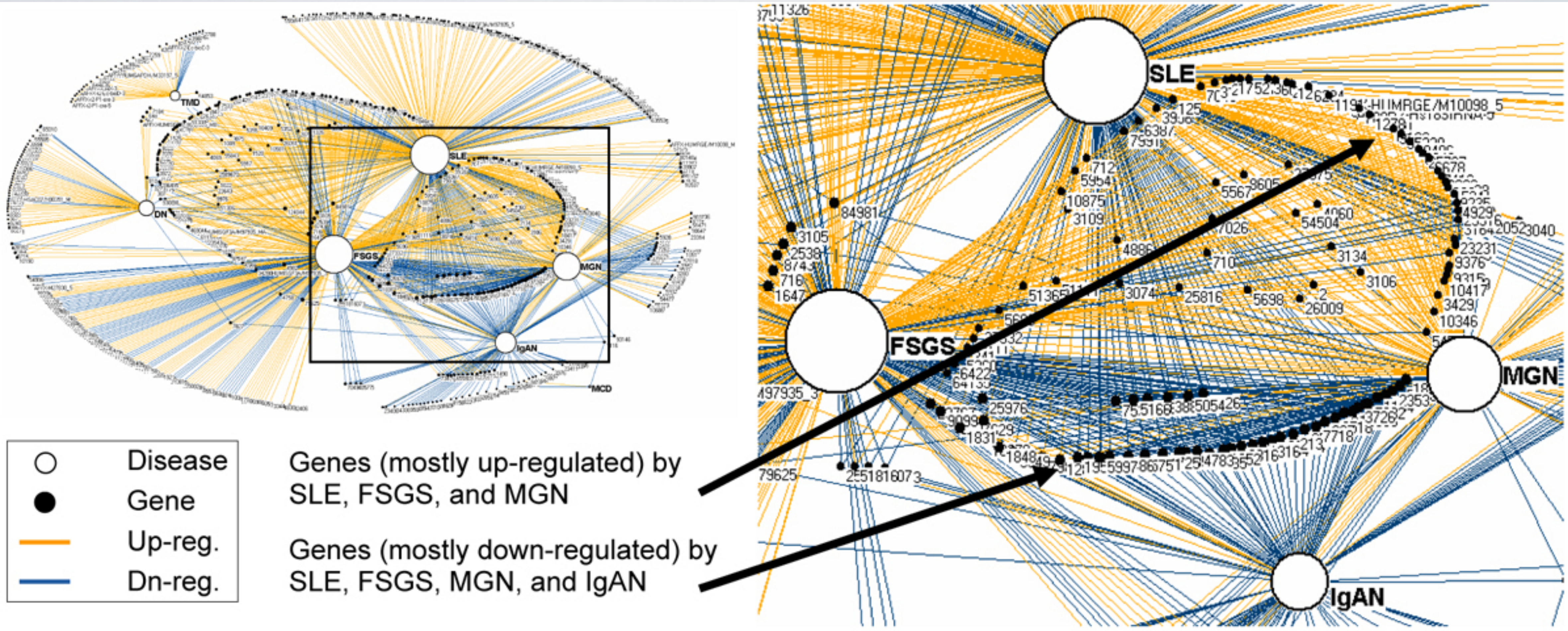
Onnela et.al. New Journal of Physics 9, 179 (2007).



Social interaction network: Nodes - individuals

Links - social interactions

Bipartite network

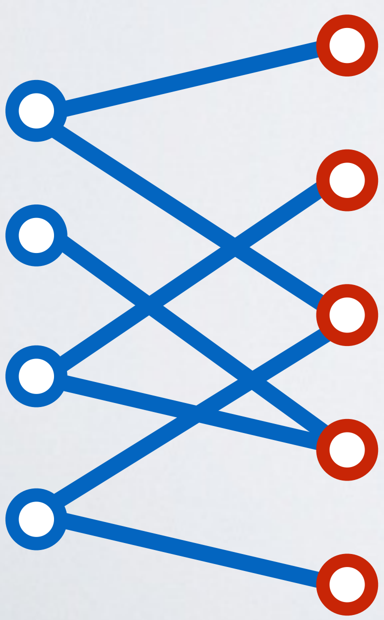


Bhavnani et.al. BMC Bioinformatics 2009, **10**(Suppl 9):S3

Gene-desease network:

Nodes - Desease (7)&Genes (747)

Links - gene-desease relationship



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

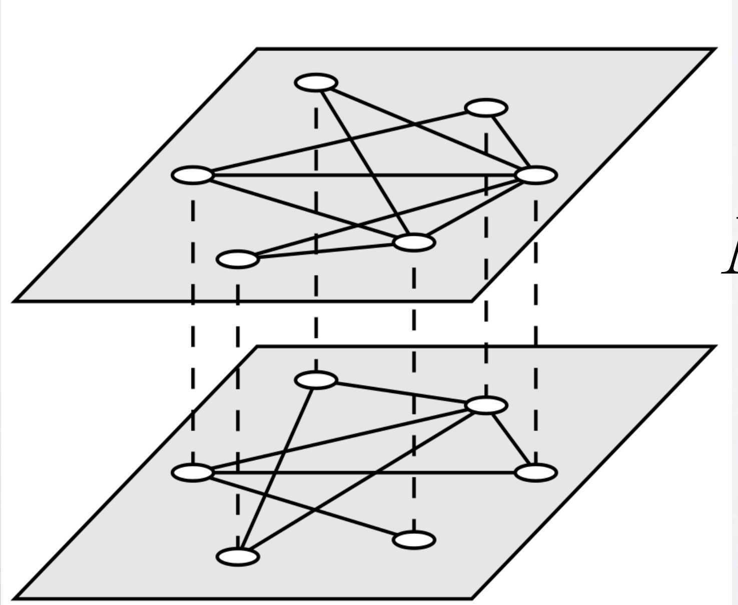
$$U \cap V = \varnothing$$

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

Multiplex and multilayer networks

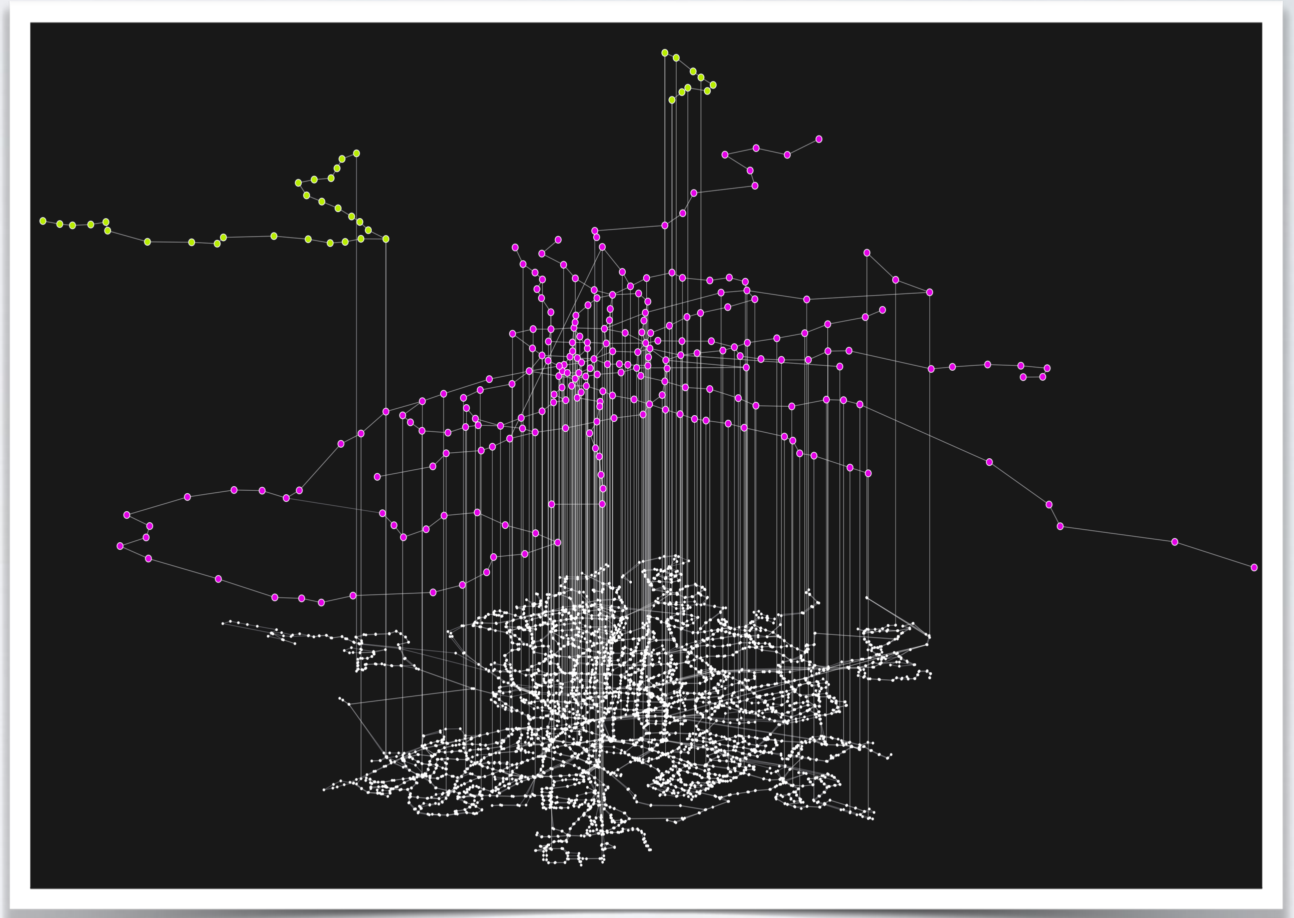
$G=(V, E_i),\ i=1\ldots M$

- 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)

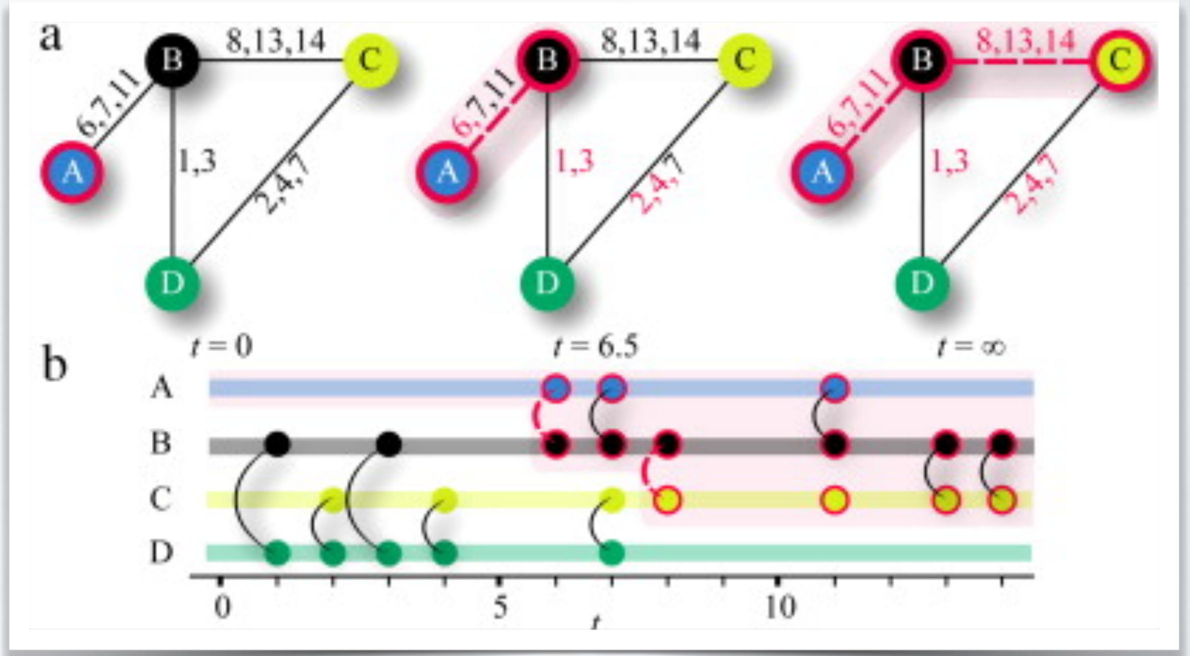
- Temporal links encode time varying interactions

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

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

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

- Dynamical nodes and links encode the evolution of the network



Mobile communication network
Nodes - individuals
Links - calls and SMS

NETWORK REPRESENTATIONS

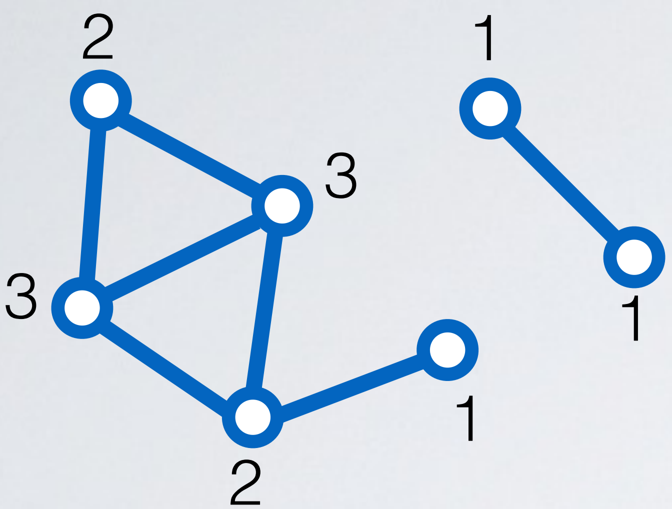
Node-Edge description

N_u	**Neighbourhood** of u, nodes sharing a link with u.
k_u	**Degree** of u, number of neighbors $\|N_u\|$.
N_u^{out}	**Successors** of u, nodes such as $(u, v) \in E$ in a directed graph
N_u^{in}	**Predecessors** of u, nodes such as $(v, u) \in E$ in a directed graph
k_u^{out}	**Out-degree** of u, number of outgoing edges $\|N_u^{out}\|$.
k_u^{in}	**In-degree** of u, number of incoming edges $\|N_u^{in}\|$
$w_{u,v}$	**Weight** of edge (u, v).
s_u	**Strength** of u, sum of weights of adjacent edges, $s_u = \sum_v w_{uv}$.

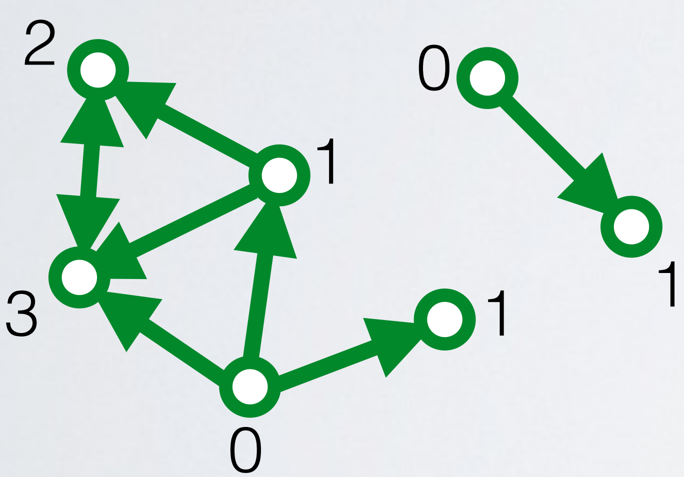
Node degree

Number of connections of a node

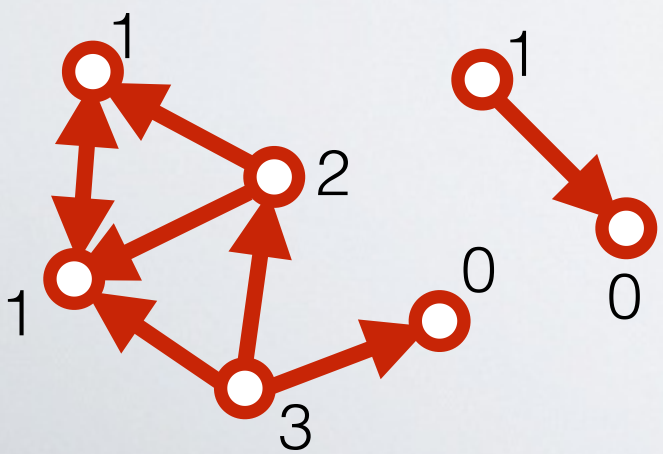
- Undirected network



- Directed network

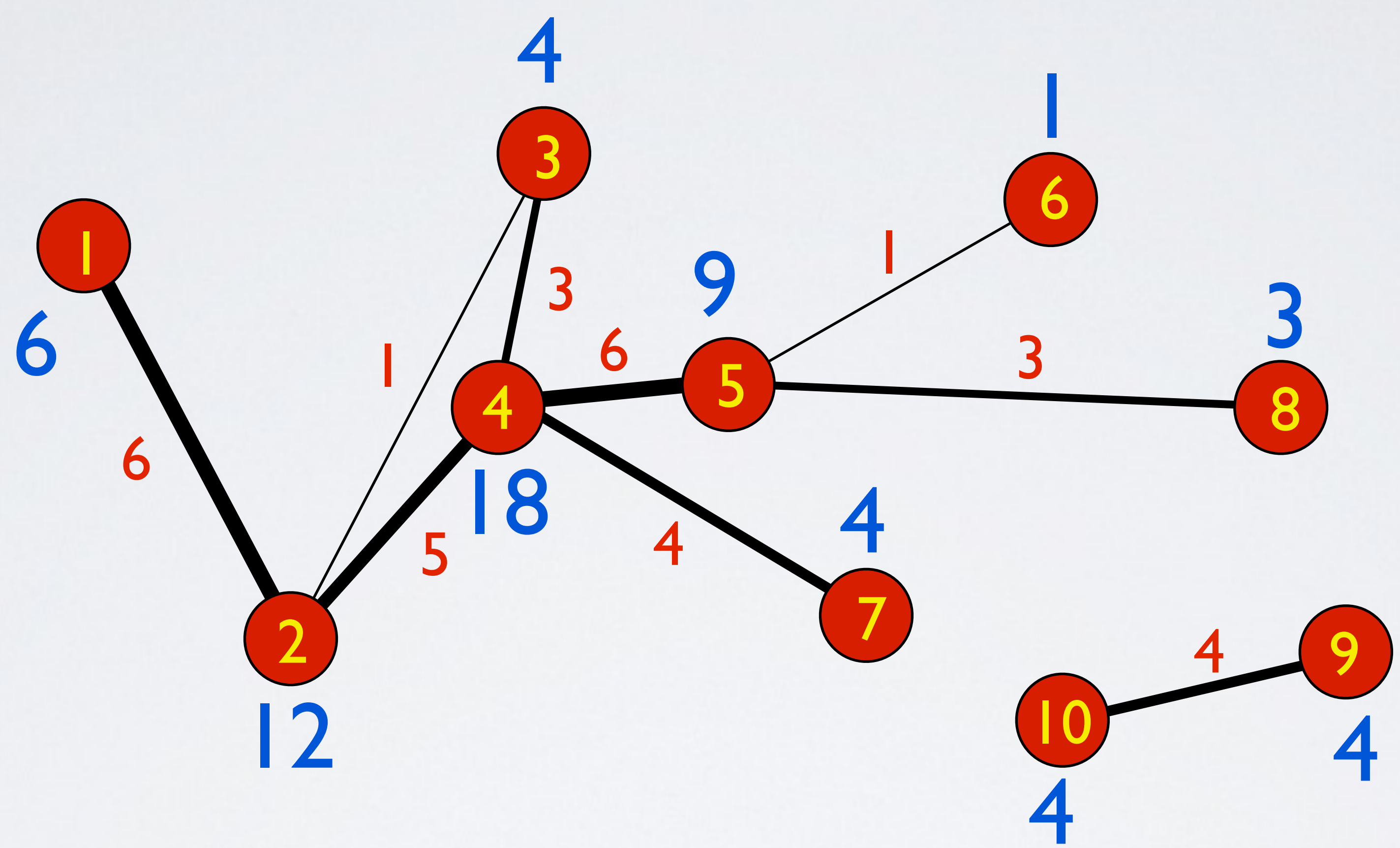


In degree



Out degree

Weighted degree: strength



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

Counting nodes and edges

N/n	**size**: number of nodes $\|V\|$.
L/m	number of edges $\|E\|$
L_{max}	Maximum number of links

Undirected network: $\binom{N}{2} = N(N-1)/2$

Directed network: $\binom{N}{2} = N(N-1)$

SIZE

	#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

Network descriptors - Nodes/Edges

$\langle k \rangle$	**Average degree**: Real networks are sparse, i.e., typically $\langle k \rangle \ll n$. Increases slowly with network size, e.g., $\langle k \rangle \sim \log(m)$[a] $$\langle k \rangle = \frac{2m}{n}$$
$d/d(G)$	**Density**: Fraction of pairs of nodes connected by an edge in G. $$d = L/L_{\max}$$

[a] Leskovec, Kleinberg, and Faloutsos 2005.

DENSITY

	#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

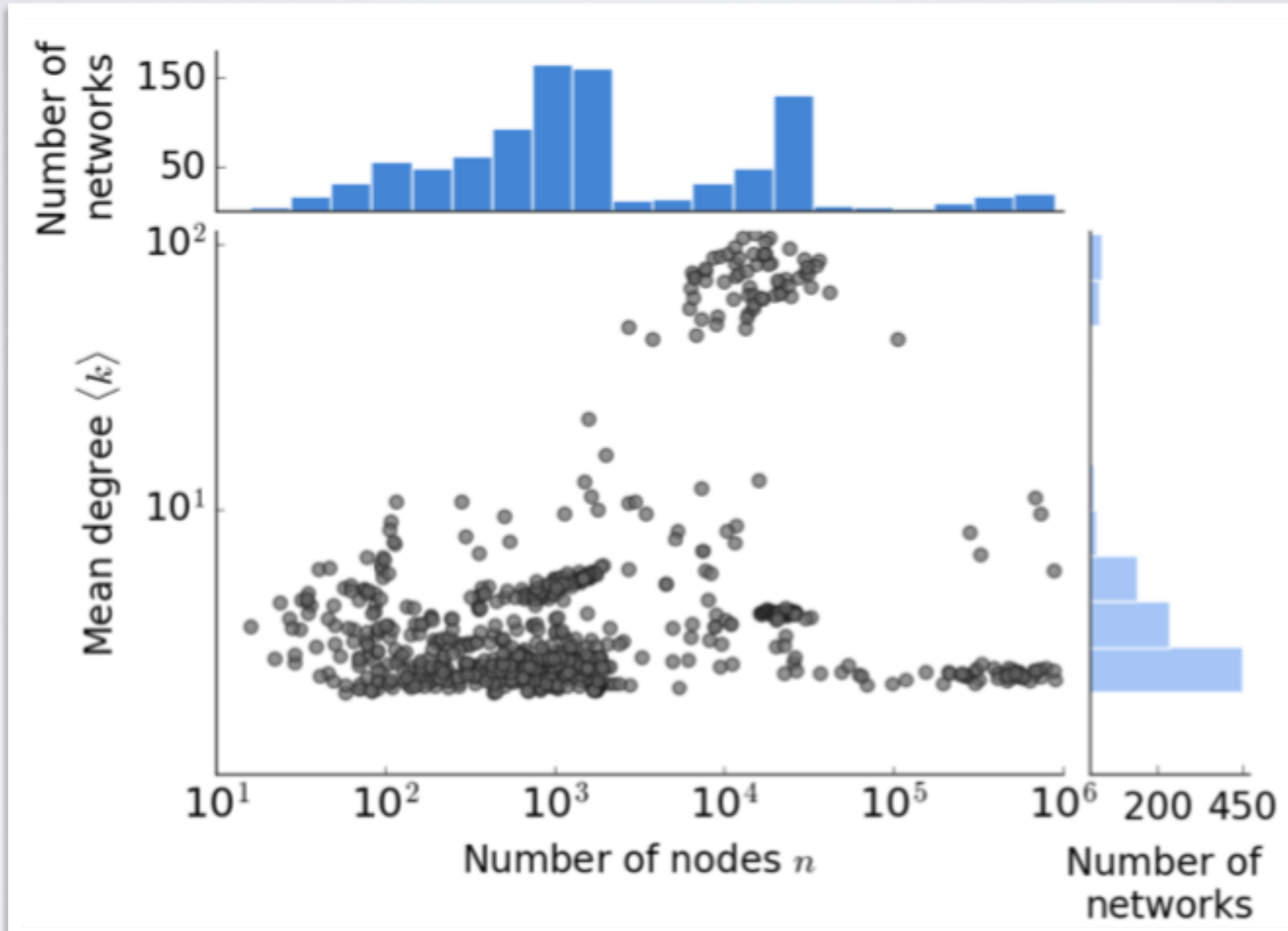
Beware: density hard to compare between graphs of different sizes

DENSITY

- It has been observed that:
  - When graphs increase in size, the average degree increases
    - (Density on the contrary, decreases)
  - This increase is very slow
- Think of friends in a social network

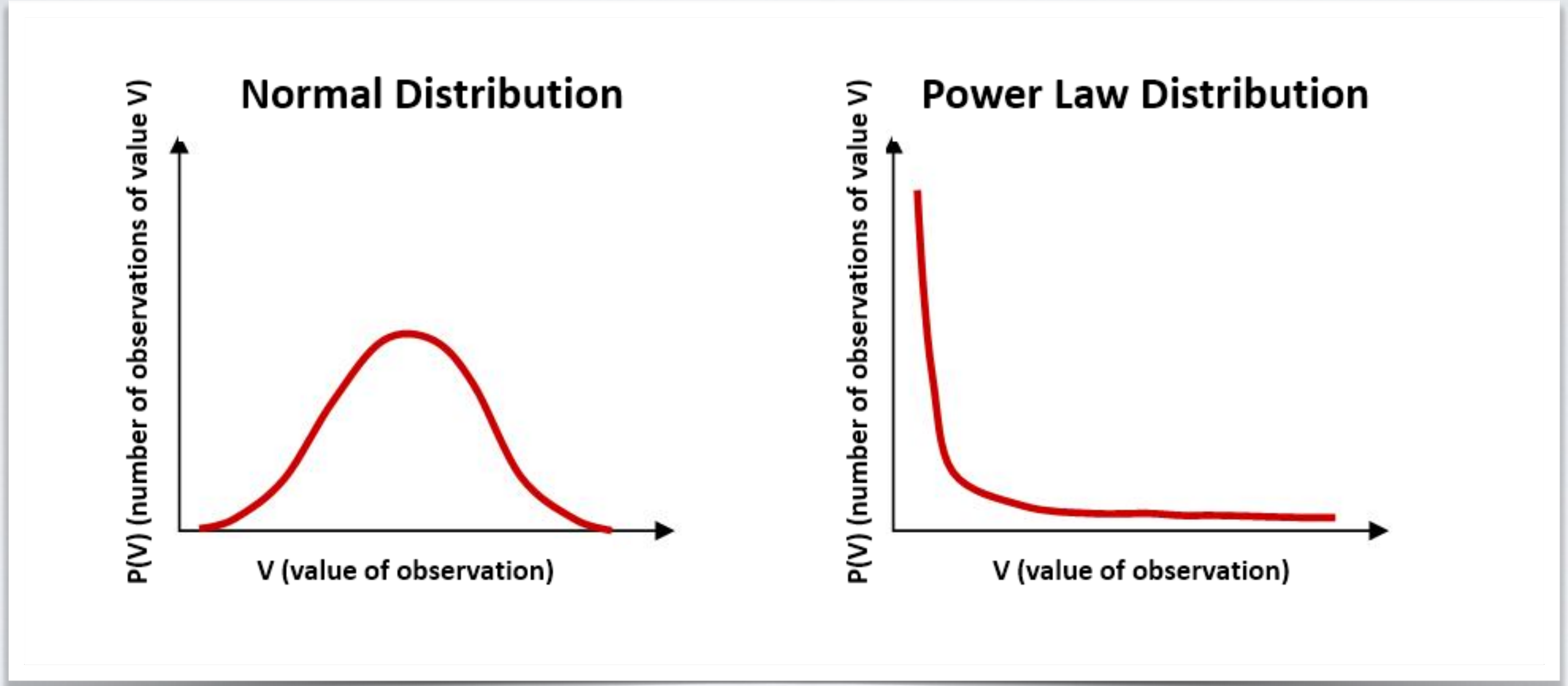
Leskovec, Jure, Jon Kleinberg, and Christos Faloutsos. "Graphs over time: densification laws, shrinking diameters and possible explanations." *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. 2005.

DENSITY



Broido, Anna D., and Aaron Clauset. "Scale-free networks are rare." *Nature communications* 10.1 (2019): 1-10.

DEGREE DISTRIBUTION



PDF (Probability Distribution Function)

DEGREE DISTRIBUTION

- In a fully random graph (Erdos-Renyi), degree distribution is (close to) 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**
  - More details later in the course

SUBGRAPHS

Subgraphs

Subgraph $H(W)$ (induced subgraph): subset of nodes W of a graph $G = (V, E)$ and edges connecting them in G, i.e., subgraph $H(W) = (W, E'), W \subset V, (u, v) \in E' \iff u, v \in W \wedge (u, v) \in E$
Clique: subgraph with $d = 1$
Triangle: clique of size 3
Connected component: a subgraph in which any two vertices are connected to each other by paths, and which is connected to no additional vertices in the supergraph
Strongly Connected component: In directed networks, a subgraph in which any two vertices are connected to each other by paths
Weakly Connected component: In directed networks, a subgraph in which any two vertices are connected to each other by paths if we disregard directions

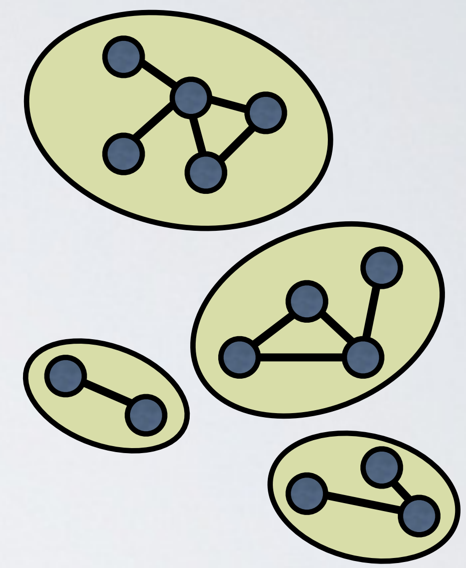
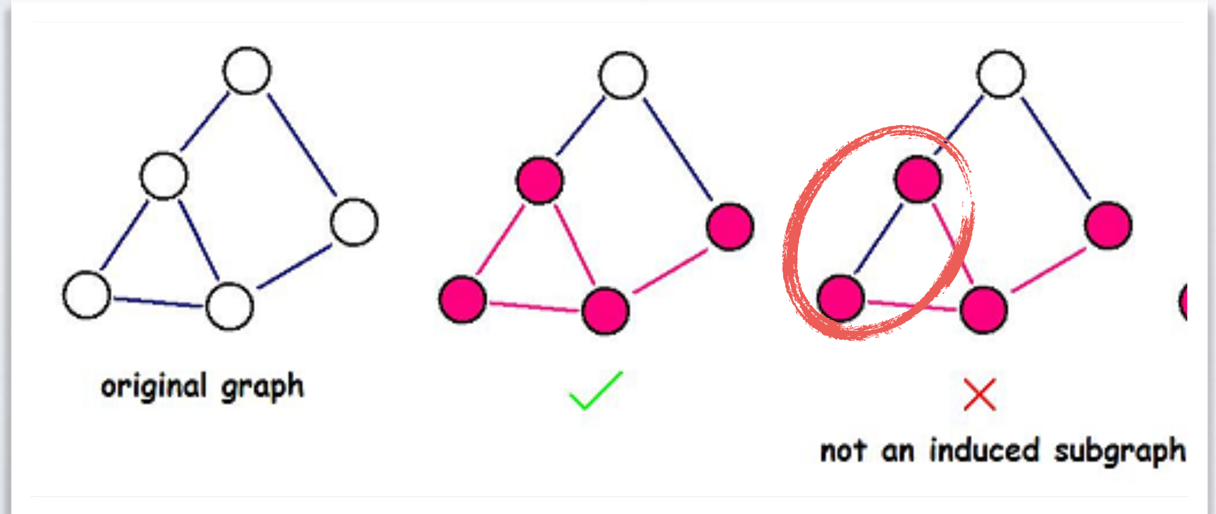
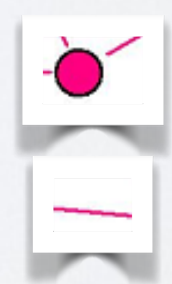


Figure after Newman, 2010



After "A. DZY Loves Physics"



Nodes/Edges in the subgraph

CLUSTERING COEFFICIENT

- **Clustering coefficient** or **triadic closure**
- Triangles are considered important in real networks
  - Think of social networks: *friends of friends are my friends*
  - # triangles is a big difference between real and random networks

CLUSTERING COEFFICIENT

Triangles counting

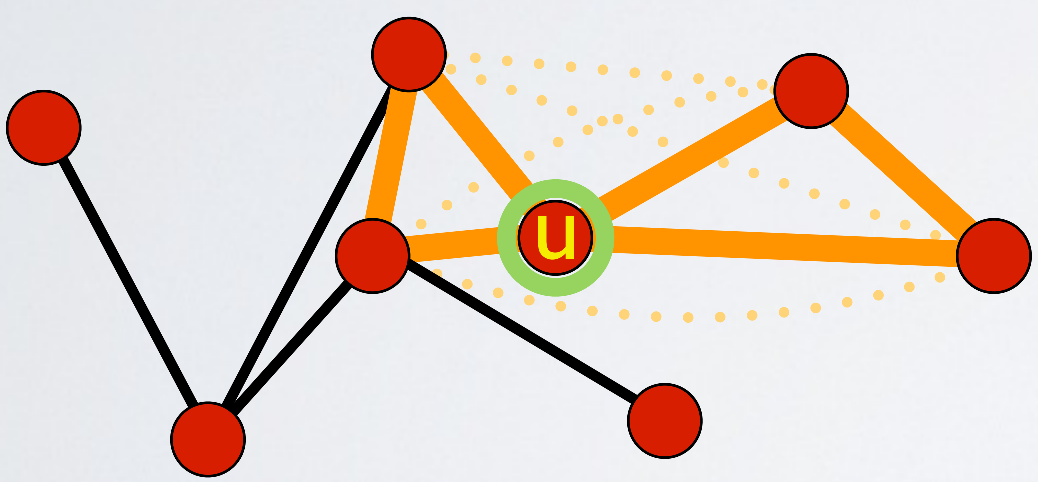
δ_u - **triads of u:** number of triangles containing node u
Δ - **number of triangles in the graph** total number of triangles in the graph, $\Delta = \frac{1}{3}\sum_{u \in V} \delta_u$.

Each triangle in the graph is counted as a triad once by each of its nodes.

$\delta_u^{\max}$ - **triads potential of u:** maximum number of triangles that could exist around node u, given its degree: $\delta_u^{\max} = \tau(u) = \binom{k_i}{2}$
$\Delta^{\max}$ - **triangles potential of G:** maximum number of triangles that could exist in the graph, given its degree distribution: $\Delta^{\max} = \frac{1}{3}\sum_{u \in V} \delta^{\max}(u)$

CLUSTERING COEFFICIENT

C_u - **Node clustering coefficient:** density of the subgraph induced by the neighborhood of u, $C_u = d(H(N_u))$. Also interpreted as the fraction of all possible triangles in N_u that exist, $\frac{\delta_u}{\delta_u^{\max}}$



Triangles=2

Possible triangles=$\binom{4}{2}$=6

C_u=2/6=1/3

Edges: 2

Max edges: 4*3/2=6

C_u=2/6=1/3

CLUSTERING COEFFICIENT

$\langle C \rangle$ - **Average clustering coefficient:** Average clustering coefficient of all nodes in the graph, $\bar{C} = \frac{1}{N}\sum_{u \in V} C_u$.

Be careful when interpreting this value, since all nodes contributes equally, irrespectively of their degree, and that low degree nodes tend to be much more frequent than hubs, and their C value is very sensitive, i.e., for a node u of degree 2, $C_u \in 0, 1$, while nodes of higher degrees tend to have more contrasted scores.

C^g - **Global clustering coefficient:** Fraction of all possible triangles in the graph that do exist, $C^g = \frac{3\Delta}{\Delta^{\max}}$

CLUSTERING COEFFICIENT

- Global CC:
  - In random networks, GCC = density
    - =>very small for large graphs

Network	Size	$\langle k \rangle$	C	C_{rand}	Reference
WWW, site level, undir.	153 127	35.21	0.1078	0.00023	Adamic, 1999
Internet, domain level	3015–6209	3.52–4.11	0.18–0.3	0.001	Yook *et al.*, 2001a, Pastor-Satorras *et al.*, 2001
Movie actors	225 226	61	0.79	0.00027	Watts and Strogatz, 1998
LANL co-authorship	52 909	9.7	0.43	1.8×10^{-4}	Newman, 2001a, 2001b, 2001c
MEDLINE co-authorship	1 520 251	18.1	0.066	1.1×10^{-5}	Newman, 2001a, 2001b, 2001c
SPIRES co-authorship	56 627	173	0.726	0.003	Newman, 2001a, 2001b, 2001c
NCSTRL co-authorship	11 994	3.59	0.496	3×10^{-4}	Newman, 2001a, 2001b, 2001c
Math. co-authorship	70 975	3.9	0.59	5.4×10^{-5}	Barabási *et al.*, 2001
Neurosci. co-authorship	209 293	11.5	0.76	5.5×10^{-5}	Barabási *et al.*, 2001
E. coli, substrate graph	282	7.35	0.32	0.026	Wagner and Fell, 2000
E. coli, reaction graph	315	28.3	0.59	0.09	Wagner and Fell, 2000
Ythan estuary food web	134	8.7	0.22	0.06	Montoya and Solé, 2000
Silwood Park food web	154	4.75	0.15	0.03	Montoya and Solé, 2000
Words, co-occurrence	460.902	70.13	0.437	0.0001	Ferrer i Cancho and Solé, 2001
Words, synonyms	22 311	13.48	0.7	0.0006	Yook *et al.*, 2001b
Power grid	4941	2.67	0.08	0.005	Watts and Strogatz, 1998
C. Elegans	282	14	0.28	0.05	Watts and Strogatz, 1998

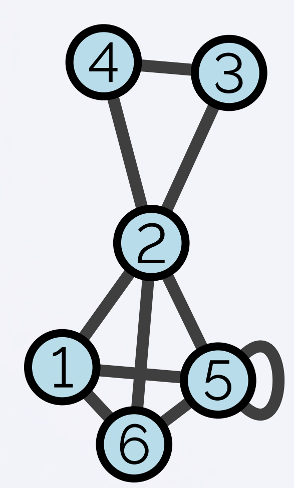
Albert, R. et.al. Rev. Mod. Phy. (2002)

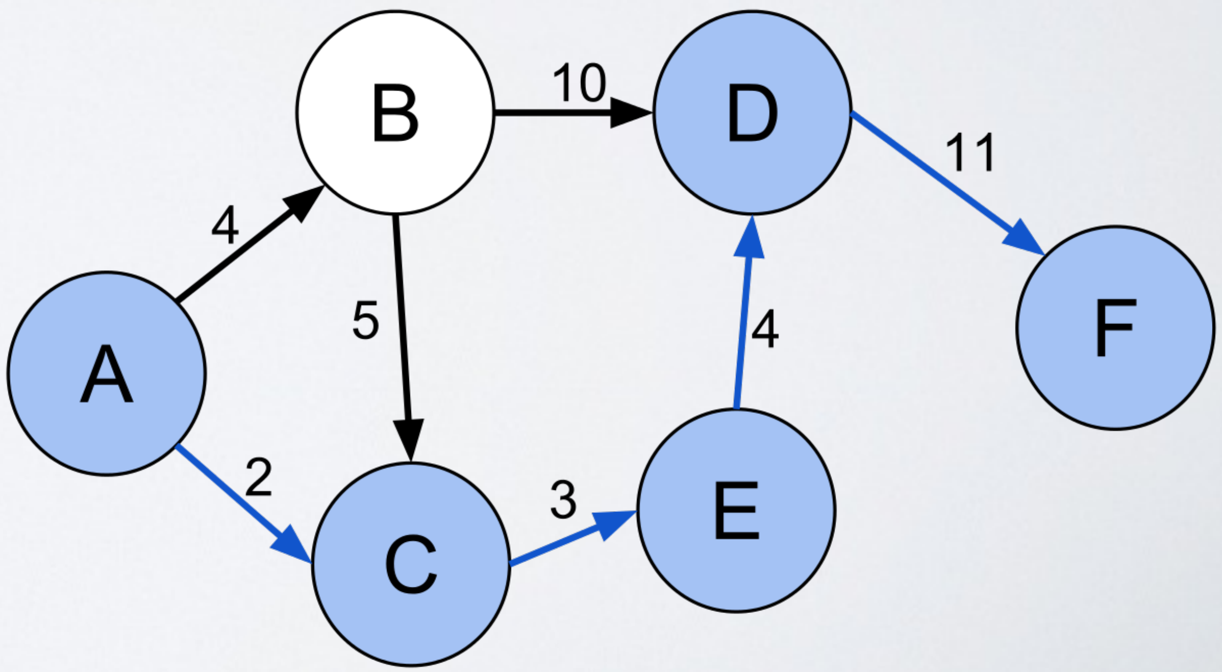
PATH RELATED SCORES

Paths - Walks - Distance

Walk: Sequences of adjacent edges or nodes (e.g., **1.2.1.6.5** is a valid walk)
Path: a walk in which each node is distinct.
Path length: number of edges encountered in a path
Weighted Path length: Sum of the weights of edges on a path
Shortest path: The shortest path between nodes u, v is a path of minimal *path length*. Often it is not unique.
Weighted Shortest path: path of minimal *weighted path length*.
$\ell_{u,v}$**: Distance**: The distance between nodes u, v is the length of the shortest path

Graph





All shortest path algorithm

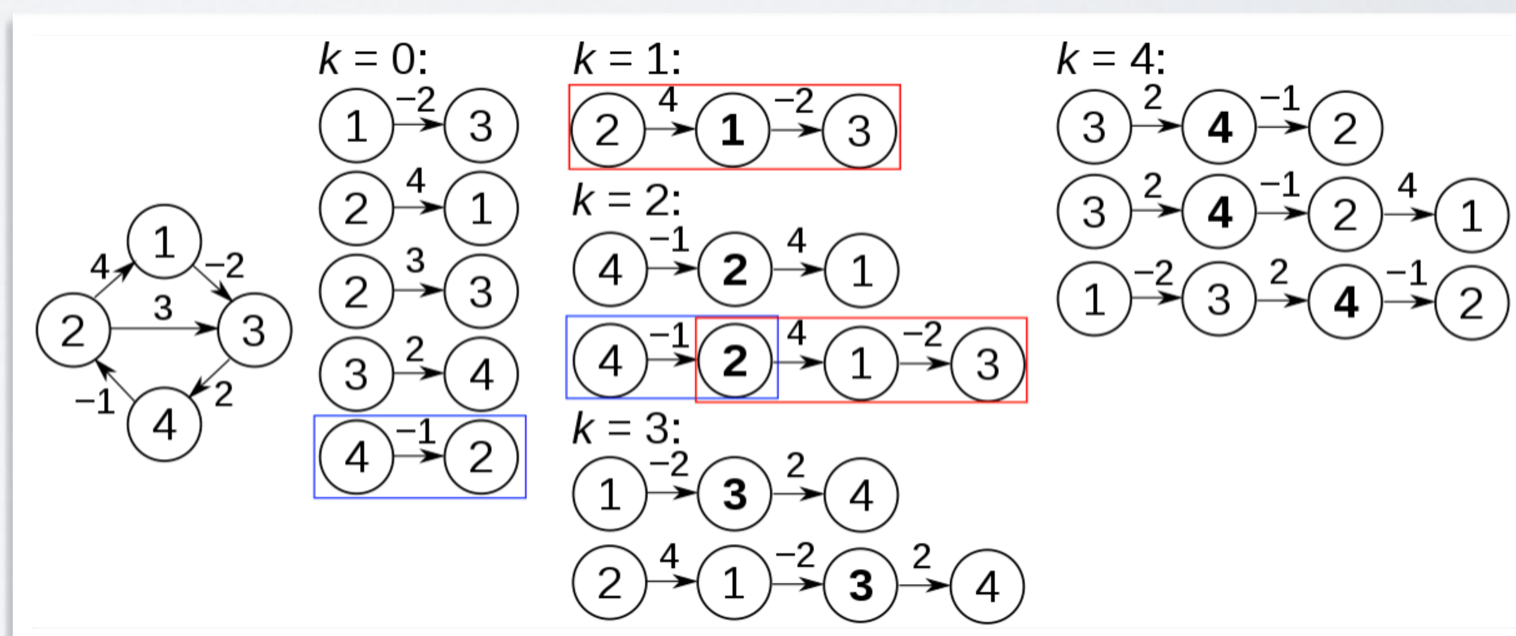
finding shortest paths in a **weighted graph** with **positive** or **negative edge weights** (but with no negative cycles)

```
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:

$$shp(i,j,k) = min(shp(i,j,k-1)), shp(i,k,k-1)+shp(k,j,k-1))$$

Complexity: $O(n^3)$



PATH RELATED SCORES

Network descriptors 2 - Paths

ℓ_{max} | **Diameter**: maximum *distance* between any pair of nodes.

$\langle \ell \rangle$ | **Average distance**:

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

AVERAGE PATH LENGTH

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

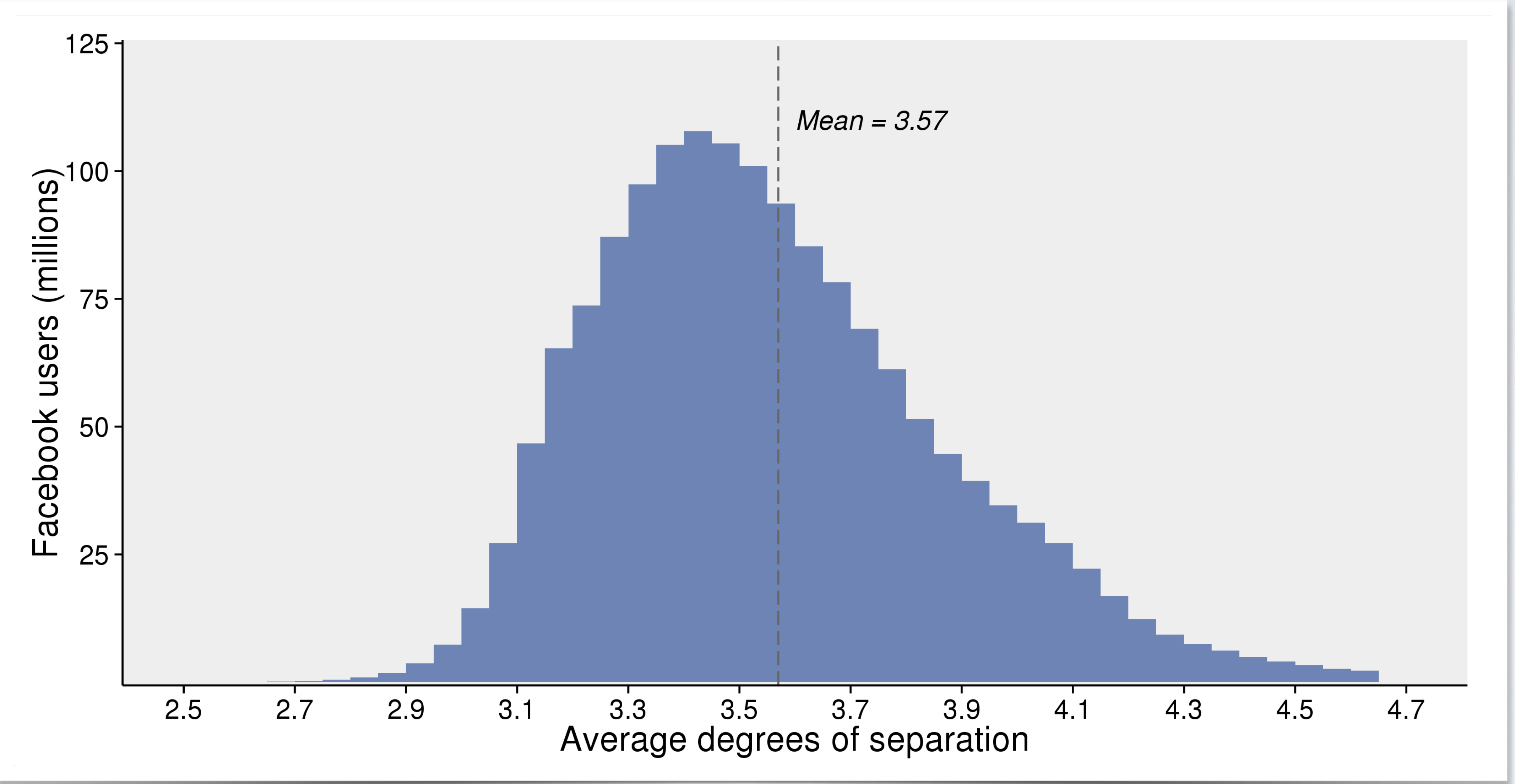
SIDE-STORY: MILGRAM EXPERIMENT

- 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

SMALL WORLD

Small World Network

A network is said to have the **small world** property when it has some structural properties. The notion is not quantitatively defined, but two properties are required:

- Average distance must be short, i.e., $\langle \ell \rangle \approx \log(N)$
- Clustering coefficient must be high, i.e., much larger than in a random network , e.g., $C^g \gg d$, with d the network density

More on this during the random network class

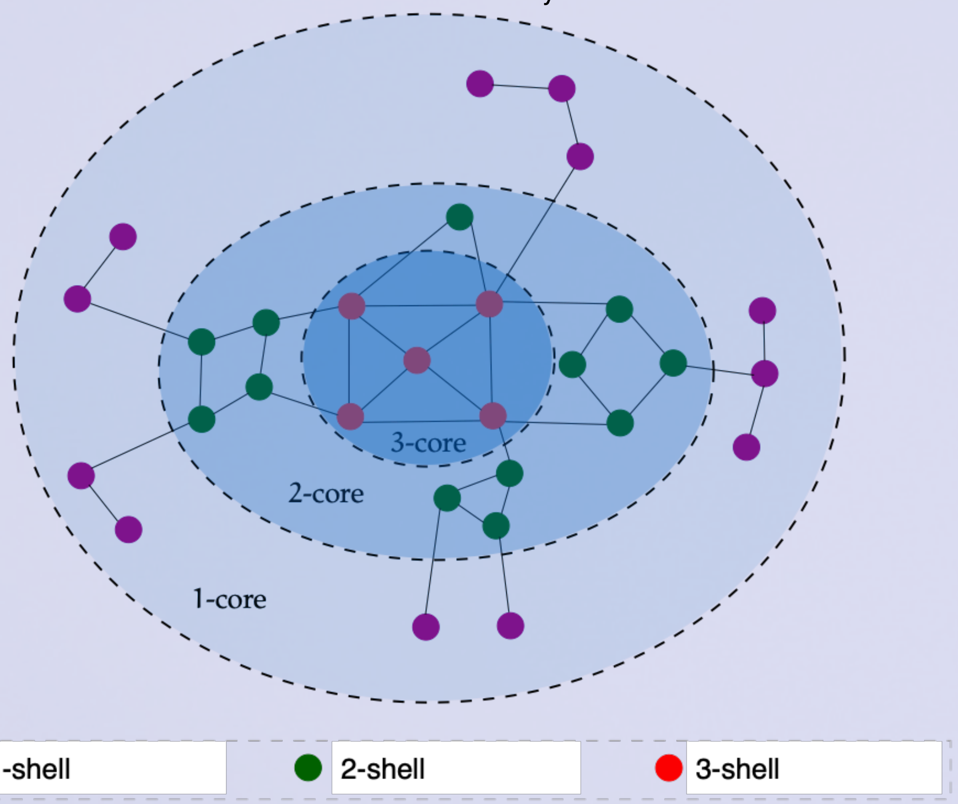
CORE-PERIPHERY : CORENESS

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

Cores and Shells

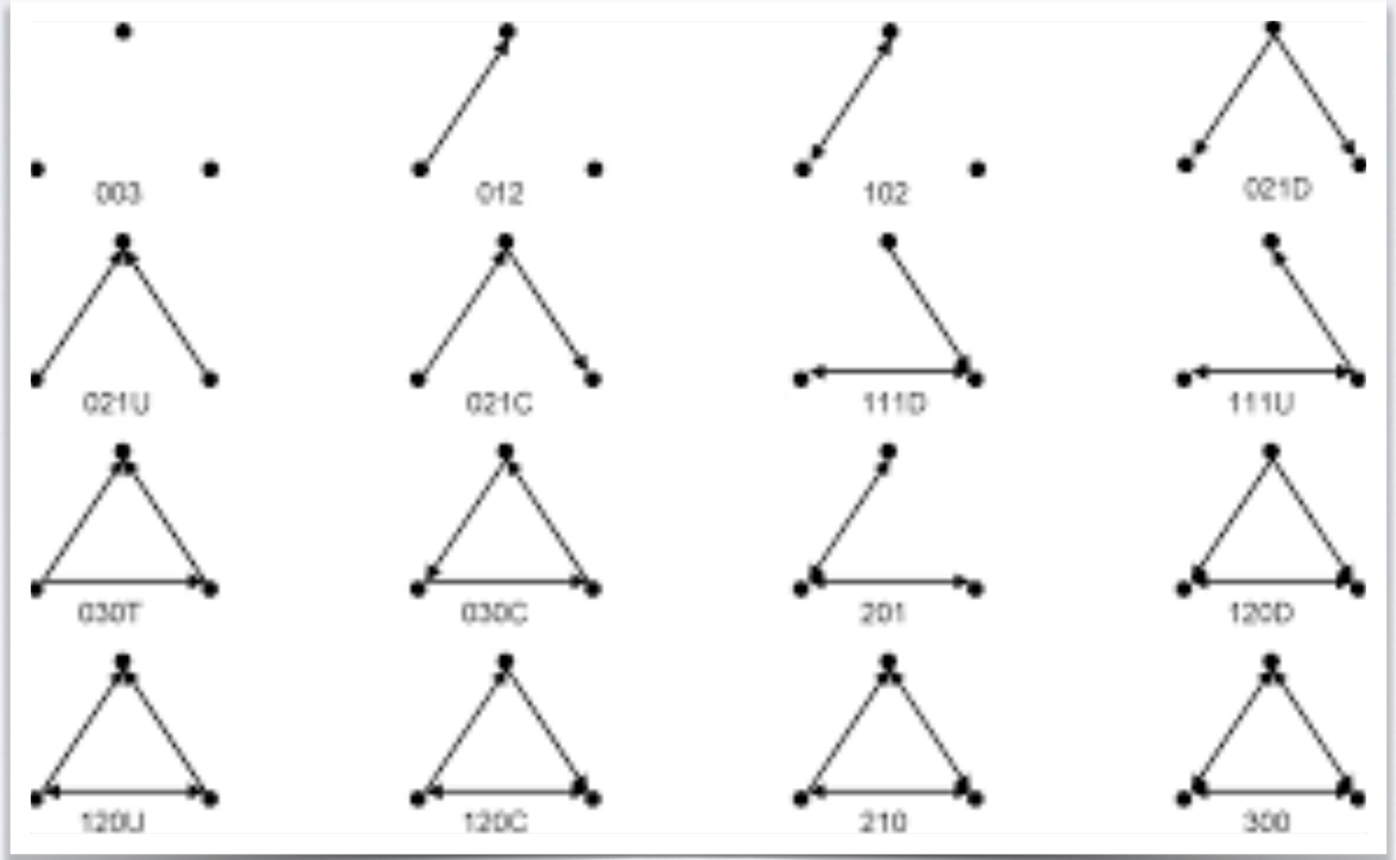
Many real networks are known to have a **core-periphery** structure, i.e., there is a densely connected core at its center and a more peripheral zone in which nodes are loosely connected between them and to the core.

k-core: The k-core (core of order k) of $G(V, E)$ is the largest subgraph $H(C)$ such as all nodes have at least a degree k, i.e., $\forall u \in C, k_u^H \leq k$, with k_u^H the degree of node u in subgraph H.
coreness: A vertex u has coreness k if it belongs to the k-core but not to the $k+1$-core.
c-shell: all vertices whose coreness is exactly c.

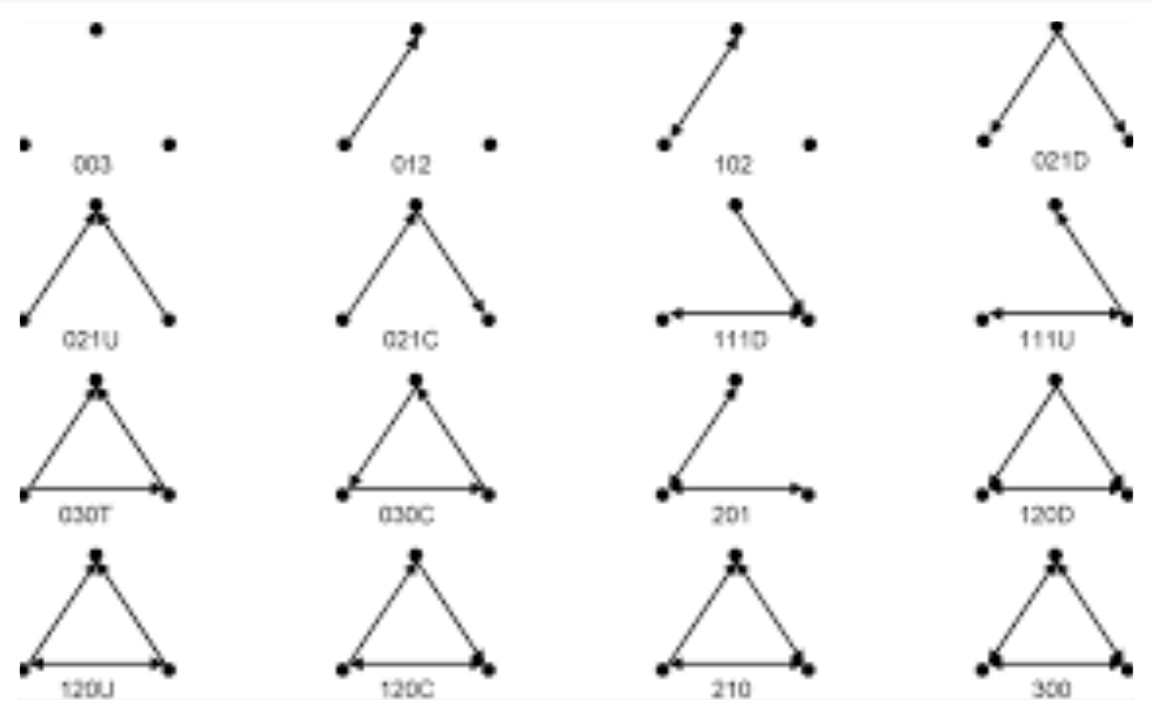


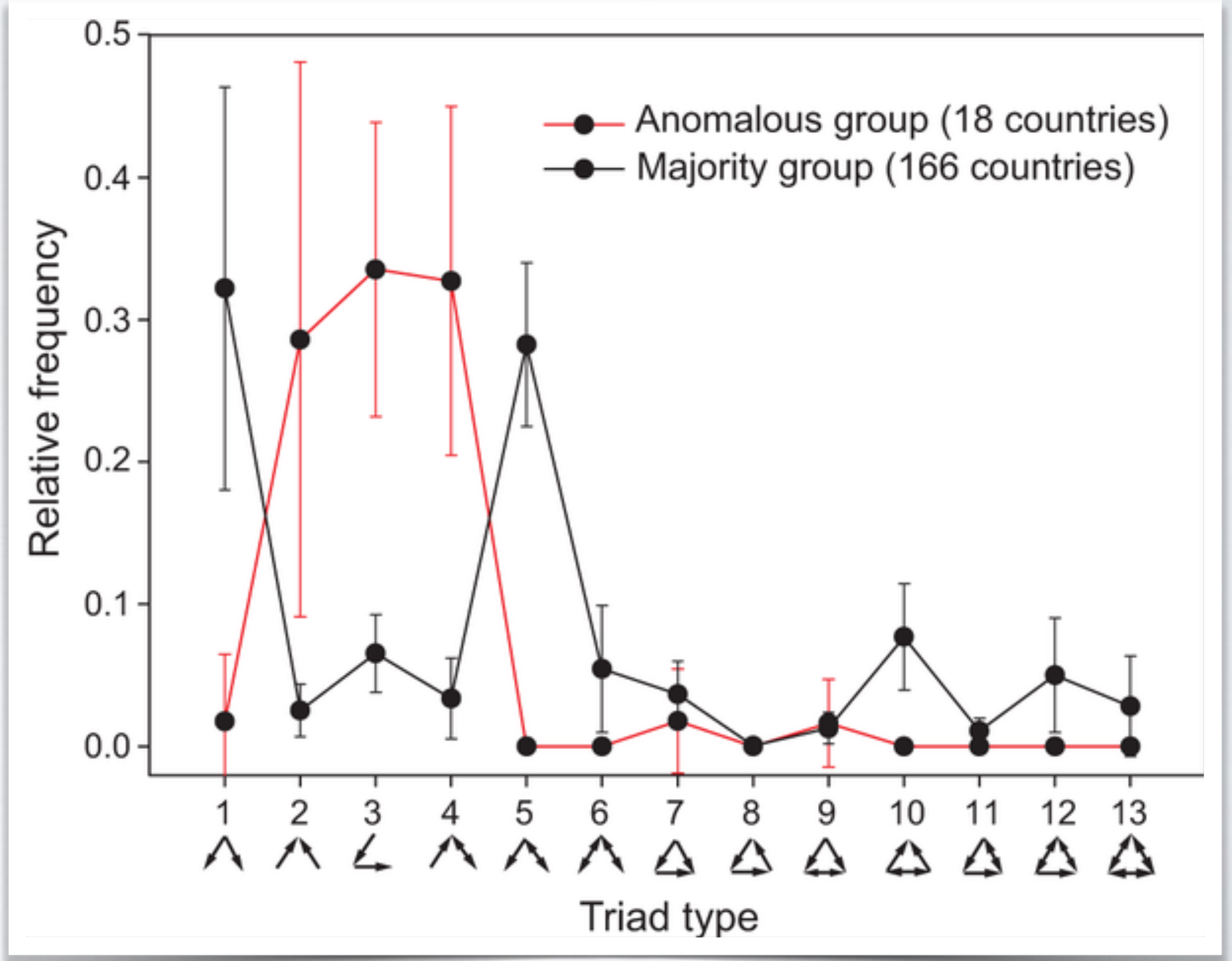
- A k-core of G can be obtained by recursively removing all the vertices of degree less than k, until all vertices in the remaining graph have at least degree k.

TRIADS COUNTING

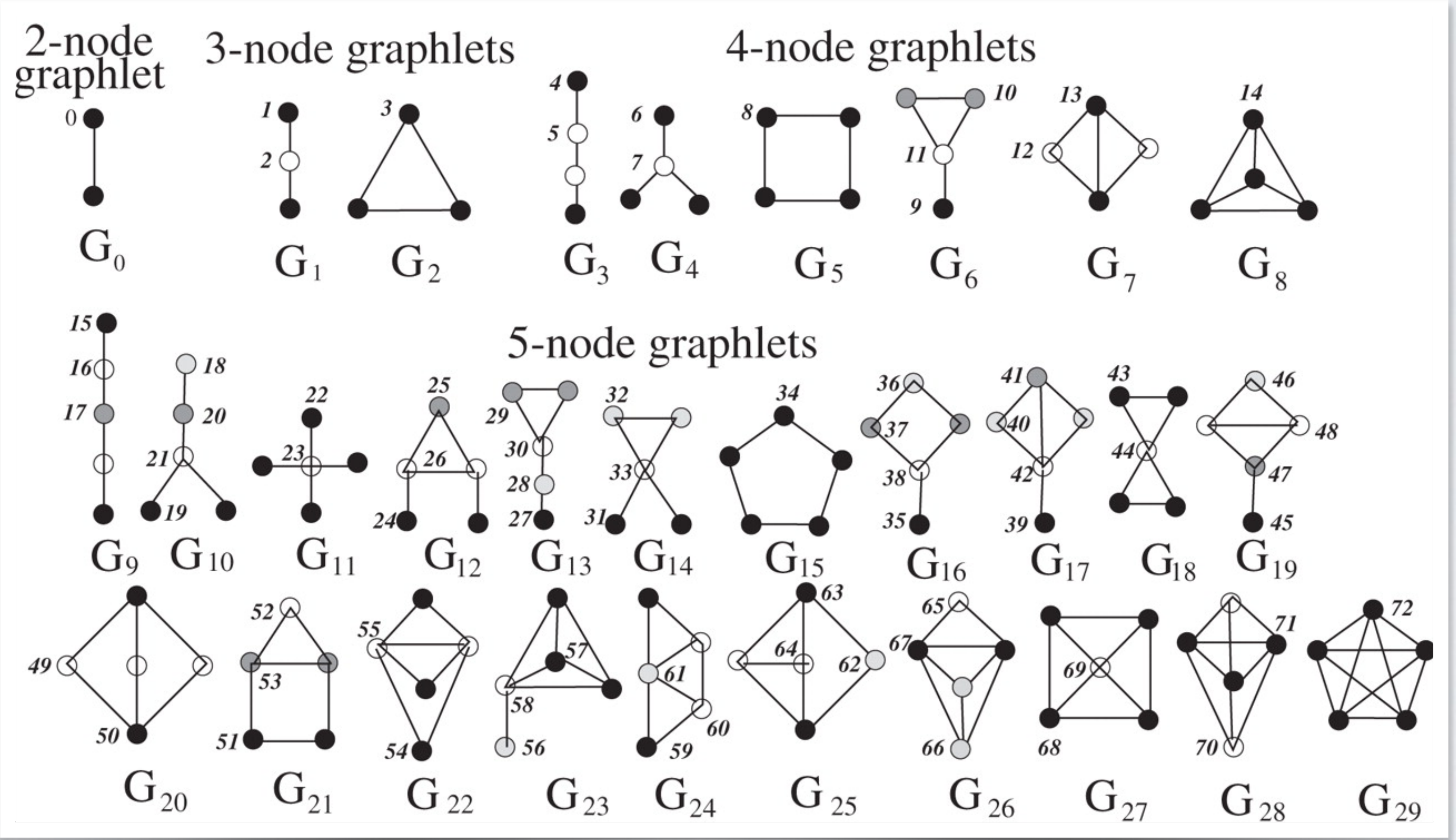


TRIADS COUNTING





GRAPHLETS



2-node graphlet
3-node graphlets
4-node graphlets
5-node graphlets
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

GRAPHS AS MATRICES

Matrices in short

Matrices are mathematical objects that can be thought as *tables* of numbers. The size of a matrix is expressed as $m \times n$, for a matrix with m rows and n columns. **The order (row/column) is important**.
M_{ij} is a notation representing the element on **row** m and **column** j.

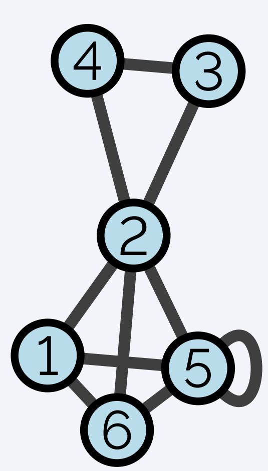
ADJACENCY MATRIX

A - Adjacency matrix

The most natural way to represent a graph as a matrix is called the Adjacency matrix A. It is defined as a square matrix, such as the number of rows (and the number of columns) is equal to the number of nodes N in the graph. Nodes of the graph are numbered from 1 to N, and there is an edge between nodes i and j if the corresponding position of the matrix A_{ij} is not 0.

- A value on the diagonal means that the corresponding node has a **self-loop**
- the graph is **undirected**, the matrix is **symmetric**: $A_{ij} = A_{ji}$ for any i, j.
- In an **unweighted** network, and edge is represented by the value 1.
- In a **weighted** network, the value A_{ij} represents the **weight** of the edge (i, j)

Graph



A - Adjacency Mat.

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

ADJACENCY MATRIX

Typical operations on A

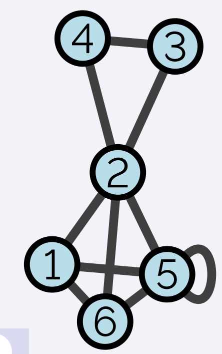
Some operations on Adjacency matrices have straightforward interpretations and are frequently used

---

Multiplying A by **itself** allows to know the number of walks of a given length that exist between any pair of nodes: A^2_{ij} corresponds to the number of walks of length 2 from node i to node j, A^3_{ij} to the number of walks of length 3, etc.

Multiplying A by a **column vector** W of length $1 \times N$ can be thought as setting the i th value of the vector to the ith node, and each node *sending* its value to its neighbors (for undirected graphs). The result is a column vector with N elements, the ith element corresponding to the sum of the values of its neighbors in W. This is convenient when working with **random walks** or **diffusion** phenomenon.

Graph



A **- Adjacency Mat.**

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

A^2

$$\begin{pmatrix} 3 & 2 & 1 & 1 & 3 & 2 \\ 2 & 5 & 1 & 1 & 3 & 2 \\ 1 & 1 & 2 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 1 & 1 \\ 3 & 3 & 1 & 1 & 4 & 3 \\ 2 & 2 & 1 & 1 & 3 & 3 \end{pmatrix}$$

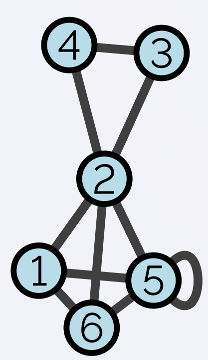
LAPLACIAN

Graph Laplacian

The **Graph Laplacian**, or **Laplacian Matrix** of a graph is a variant of the Adjacency matrix, often used in *Graph theory* and *Spectral Graph Theory*.
It is defined as $D - A$, with D the *Degree matrix* of the graph, defined as a $N \times N$ matrix with $D_{ii} = k_i$ and zeros everywhere else.

Intuitively, Laplace operator is a generalization of the second derivative, and is defined in discrete situations, for each value, as the sum of differences between the value and its "neighbors". e.g., in time, the 2$^{\text{nd}}$ derivative *acceleration* is the difference between current speed and previous speed. In a B&W picture, it's the difference between the greylevel on current pixel and the greylevel of 4 or 8 closest pixels, and perform *edge detection*. On a graph, with W a column vector representing values on nodes, LW computes for each node the difference to neighbors.

Graph



A - Adjacency Mat.

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

D - Degree Matrix

$$\begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

L - Laplacian

$$\begin{pmatrix} 3 & -1 & 0 & 0 & -1 & -1 \\ -1 & 5 & -1 & -1 & -1 & -1 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & -1 & 2 & 0 & 0 \\ -1 & -1 & 0 & 0 & 4 & -1 \\ -1 & -1 & 0 & 0 & -1 & 3 \end{pmatrix}$$

SPECTRAL GRAPH THEORY

Spectral properties of A

Spectral Graph Theory is a whole field in itself, and beyond the scope of this class. A few elements for those with a *linear algebra* background:

- The adjacency matrix of an undirected simple graph is symmetric, and therefore has a complete set of real eigenvalues and an orthogonal eigenvector basis.
- The set of eigenvalues of a graph is the spectrum of the graph.
- Eigenvalues are denoted as $\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \lambda_n$
- The largest eigenvalue λ_0 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 (i.e., if λ is an eigenvalue, then so is $-\lambda$
- If G is connected, then the diameter of G is strictly less than its number of distinct eigenvalues

SPECTRAL GRAPH THEORY

Spectral properties of L

Eigenvalues of the Laplacian have many applications, such as *spectral clsutering*, *graph matching*, *embedding*, etc. Assuming G undirected with eigenvalues $\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \lambda_n$, here are some interesting properties:

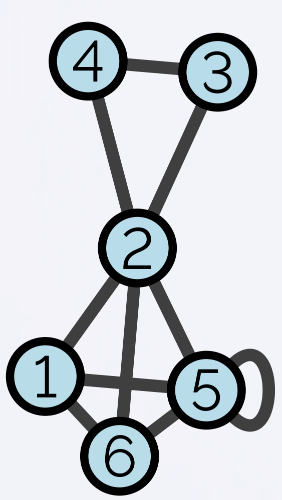
- The smallest eigenvalue λ_i equals 0
- The number of 0 eigenvalues gives the number of connected components

RANDOM WALK MATRIX

Random Walk matrix

Another useful matrix of a graph is the **Random Walk Transition Matrix** R. It is the column normalized version of the adjacency matrix. R_{ij} can be understood as the probability for a random walker located on node i to move to j.

Graph



A - Adjacency Mat.

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

Random W. mat.

$$\begin{pmatrix} 0 & \frac{1}{5} & 0 & 0 & \frac{1}{4} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{4} & \frac{1}{3} \\ 0 & \frac{1}{5} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{5} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{5} & 0 & 0 & \frac{1}{4} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{5} & 0 & 0 & \frac{1}{4} & 0 \end{pmatrix}$$

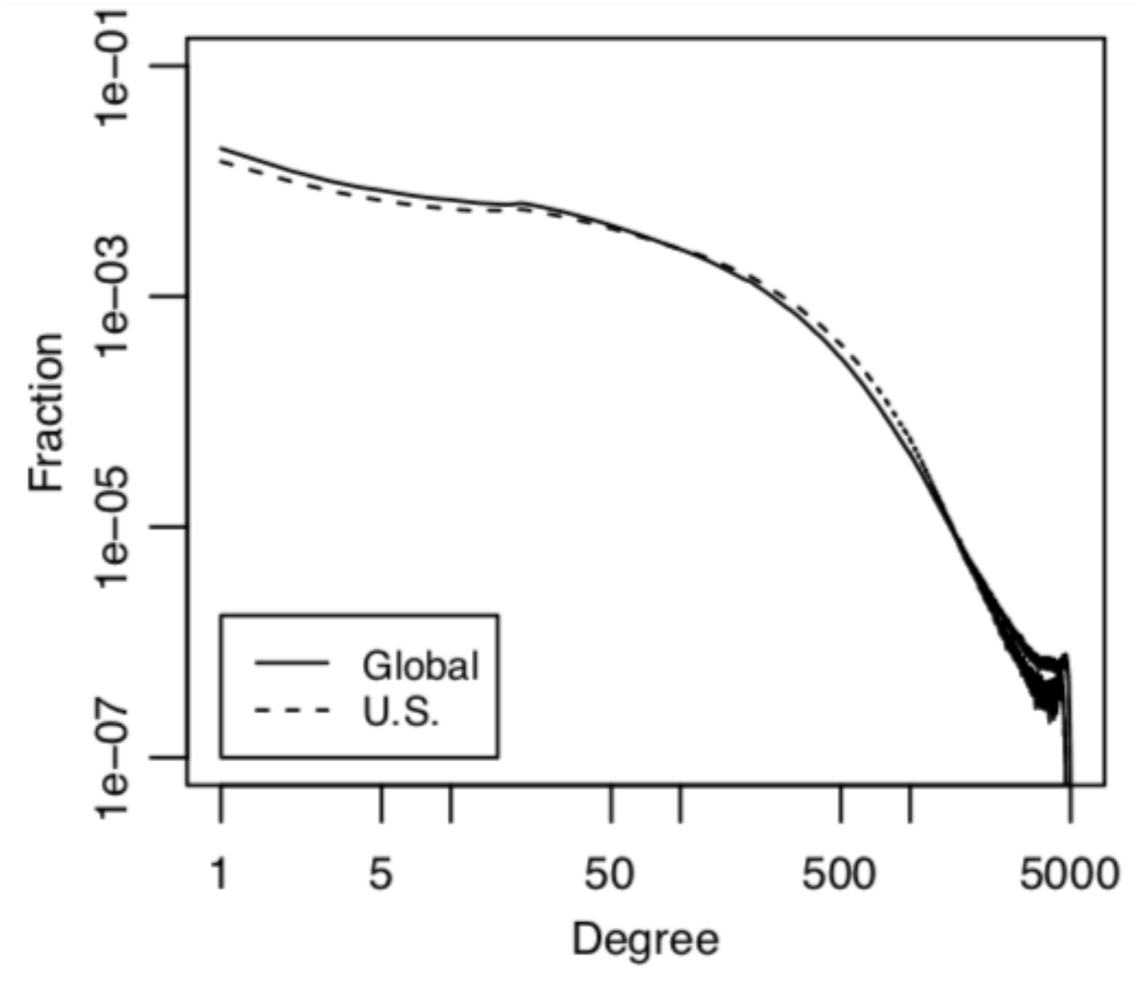
EXEMPLE OF GRAPH ANALYSIS

- 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%

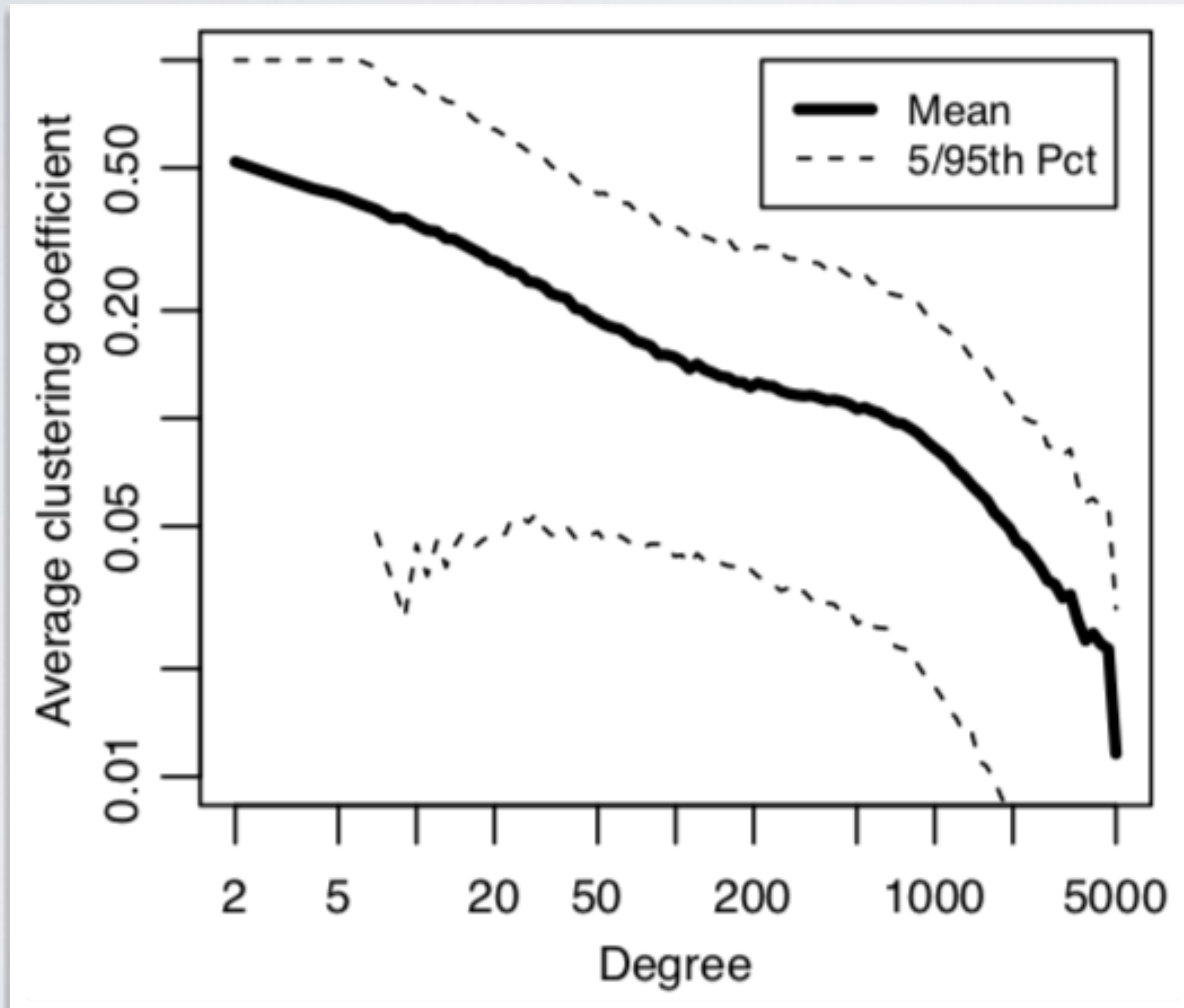
EXEMPLE OF GRAPH ANALYSIS



(a)

Degree distribution

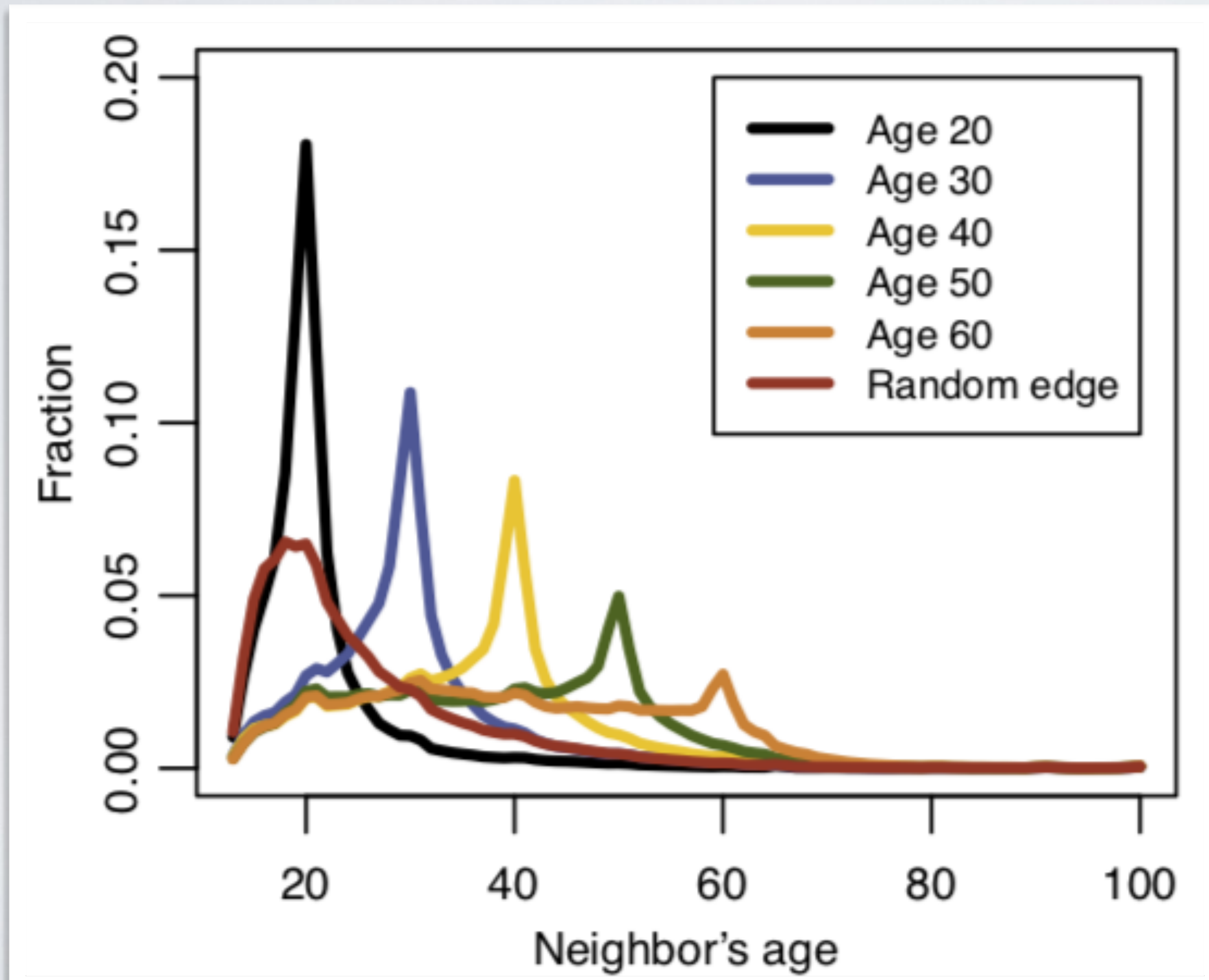
EXEMPLE OF GRAPH ANALYSIS



Clustering coefficient
By degree

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

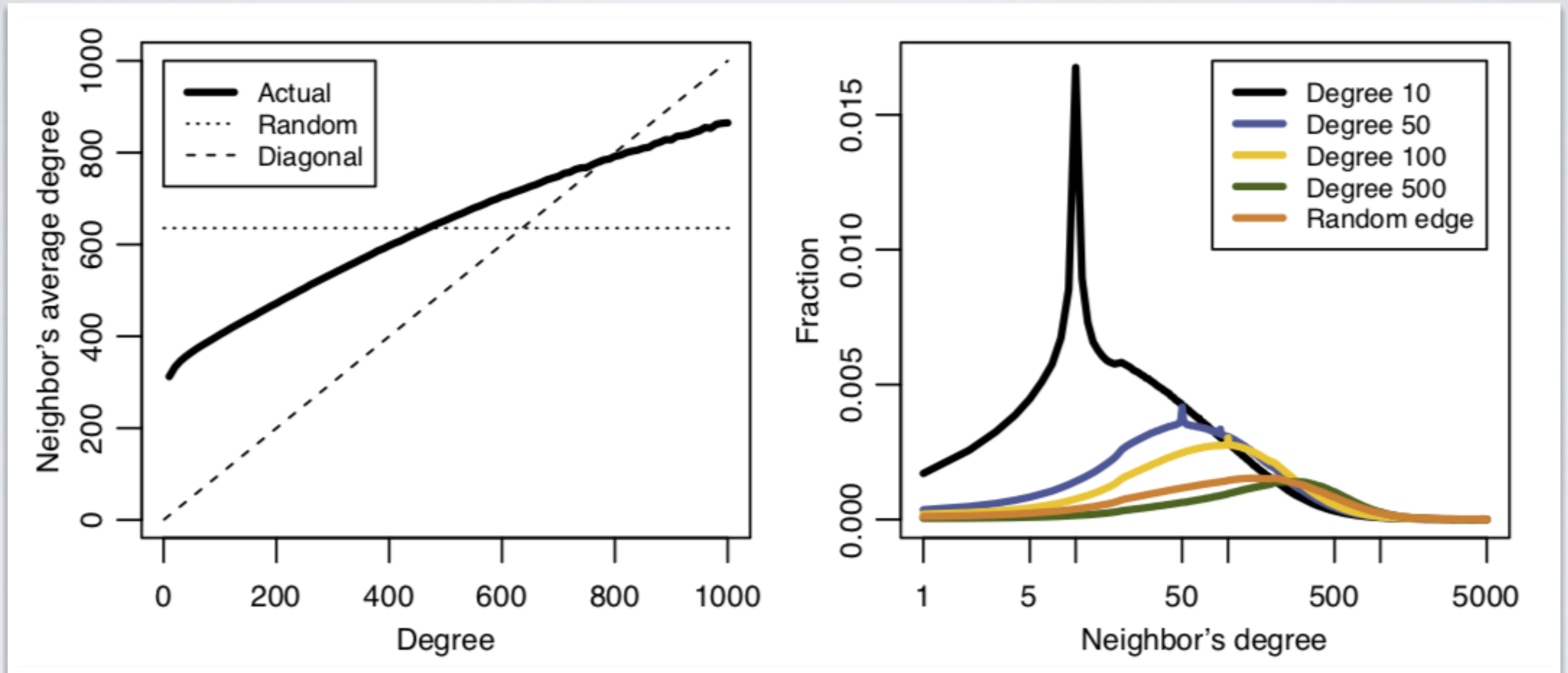
EXEMPLE OF GRAPH ANALYSIS



Age homophily

(More next class)

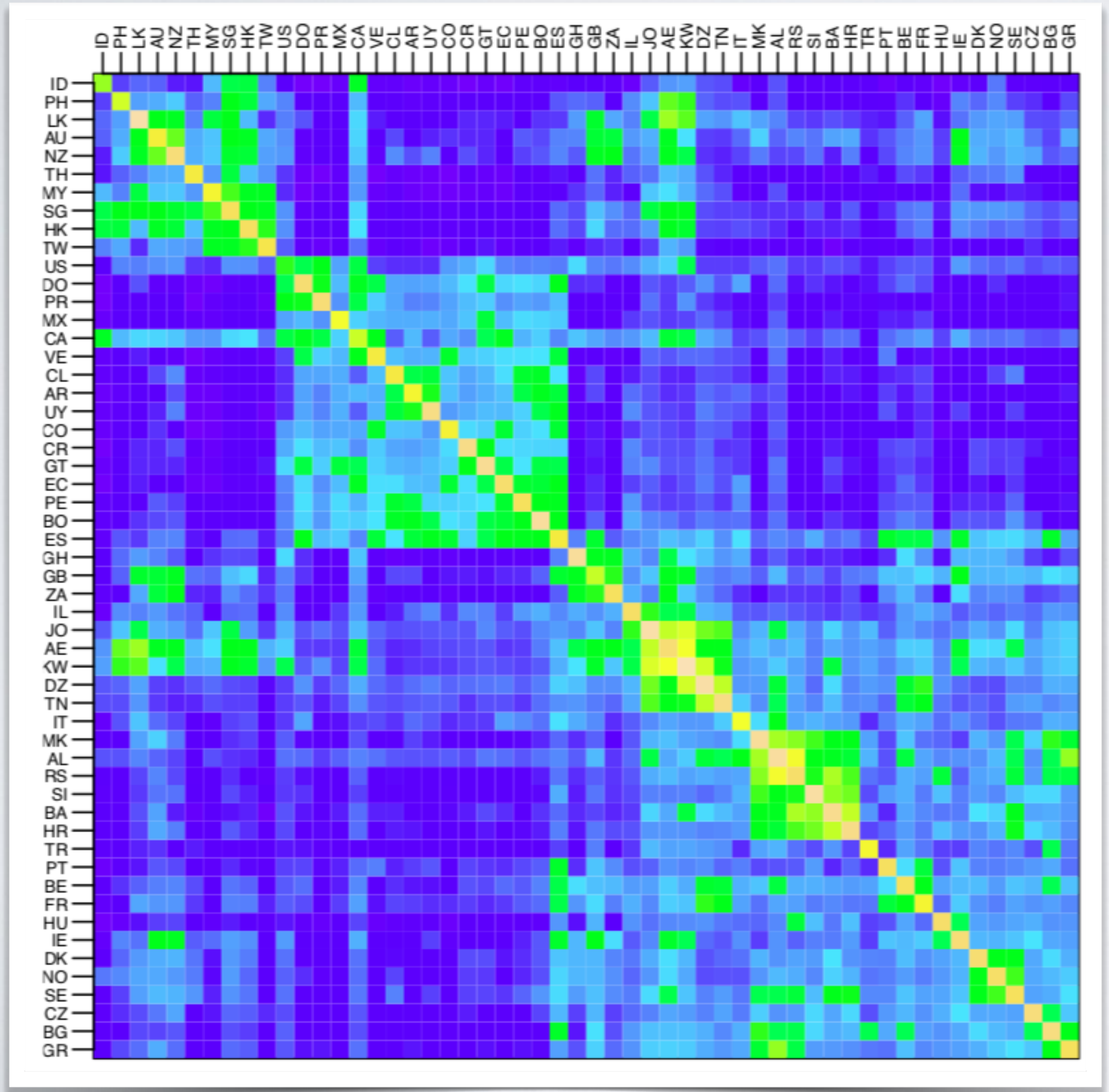
EXEMPLE OF GRAPH ANALYSIS



My friends have more Friends than me!

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

EXEMPLE OF GRAPH ANALYSIS



Country similarity

84.2% percent of edges are within countries

(More in the community detection class)