- “Donat el caràcter i la finalitat exclusivament docent i eminentment il·lustrativa de les explicacions a classe d'aquesta presentació, l’autor s’acull a l’article 32 de la Llei de propietat intel·lectual vigent respecte de l'ús parcial d'obres alienes com ara imatges, gràfics o altre material contingudes en les diferents diapositives”
- “Dado el carácter y la finalidad exclusivamente docente y eminentemente ilustrativa de las explicaciones en clase de esta presentación, el autor se acoge al artículo 32 de la Ley de Propiedad Intelectual vigente respecto al uso parcial de obras ajenas como imágenes, gráficos u otro material contenidos en las diferentes diapositivas”.

Big Data ⟷ Big Networks

Albert Díaz Guilera

http://diaz-guilera.net

@anduviera

COMPLEXITAT

complexity lab barcelona

Institute of Complex Systems
UNIVERSITAT DE BARCELONA

Statistical Physics

- Early 20th century

Mechanics

10^23 differential eqs. of motion

Thermodinamics

macroscopic variables

$$S(E, N, V) = k_B \log(\Omega)$$

Mid 20th century

- Applications to matter: physics and chemistry
- Phase transitions: concept of universality

Late 20th century

- Applications to biology
- Concept of **complexity**
- There were other stories too: chaos, synergetis, …..

Complexity

- The whole is not the sum of its contituents
- Emergent behaviors from individual behaviors of the units that form it
- Applications to social sciences

Interdisciplinariety

- Physics
- Chemistry
- Geology
- Biology
- Mathematics
- Computation
- Social sciences

Early 21st century

- Complex networks
- How this affects previous knowledge?

Complex networks

- Many examples

What do they have in common?

- Nodes: entities
- Links: real connections (physically wired)

Data

- Not always the data has the structure of a network
- We generate relational or functional patterns
- Structure vs. functionality

Big Data: origin

- Life sciences: -omics
- Social sciences:
  - mobility
  - Communication, with machines and other humans
  - Economical transactions
- We have an important role in the experiment. Do we affect the result itself?

How we get the data?

- Data from Big Networks:
  - Facebook
  - Twitter
  - Google

https://dev.twitter.com/
https://developers.google.com/maps/
https://developers.google.com/+/api/
http://developers.facebook.com/

Complex networks

- Not always links correspond to real wired connections
- Structural networks vs. Functional networks
  - Brain networks: structure & function

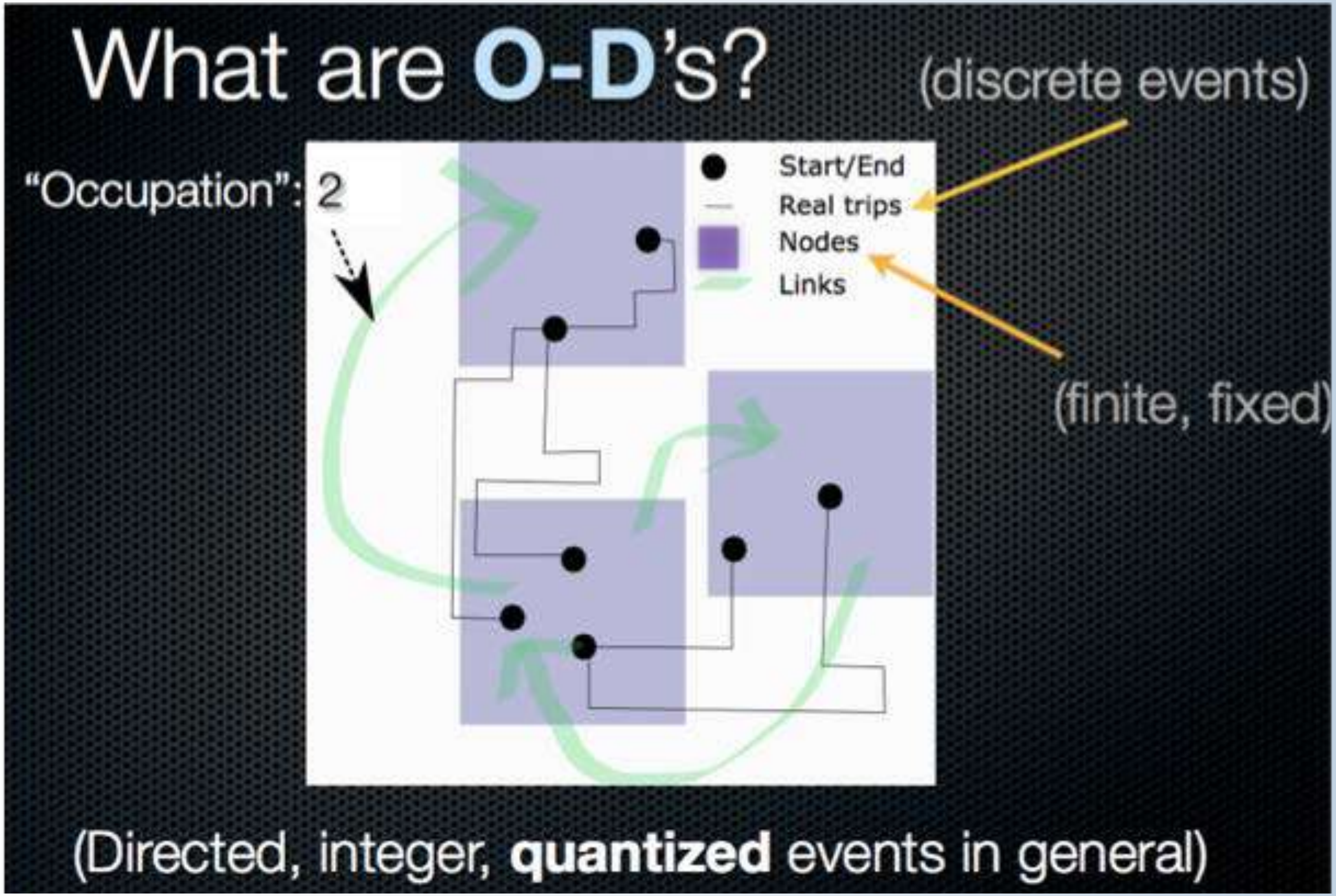
Transportation network vs. Mobility network

- Static data from maps
- Dynamic data from records:
  - GPS
  - Origin-destination polls
  - Phone calls

Mobility network

- Table with origins and destinations?
- Can we go further?

OD networks


What are O-D's?
(discrete events)
"Occupation": 2
Start/End
Real trips
Nodes
Links
(finite, fixed)
(Directed, integer, quantized events in general)

What do we have?

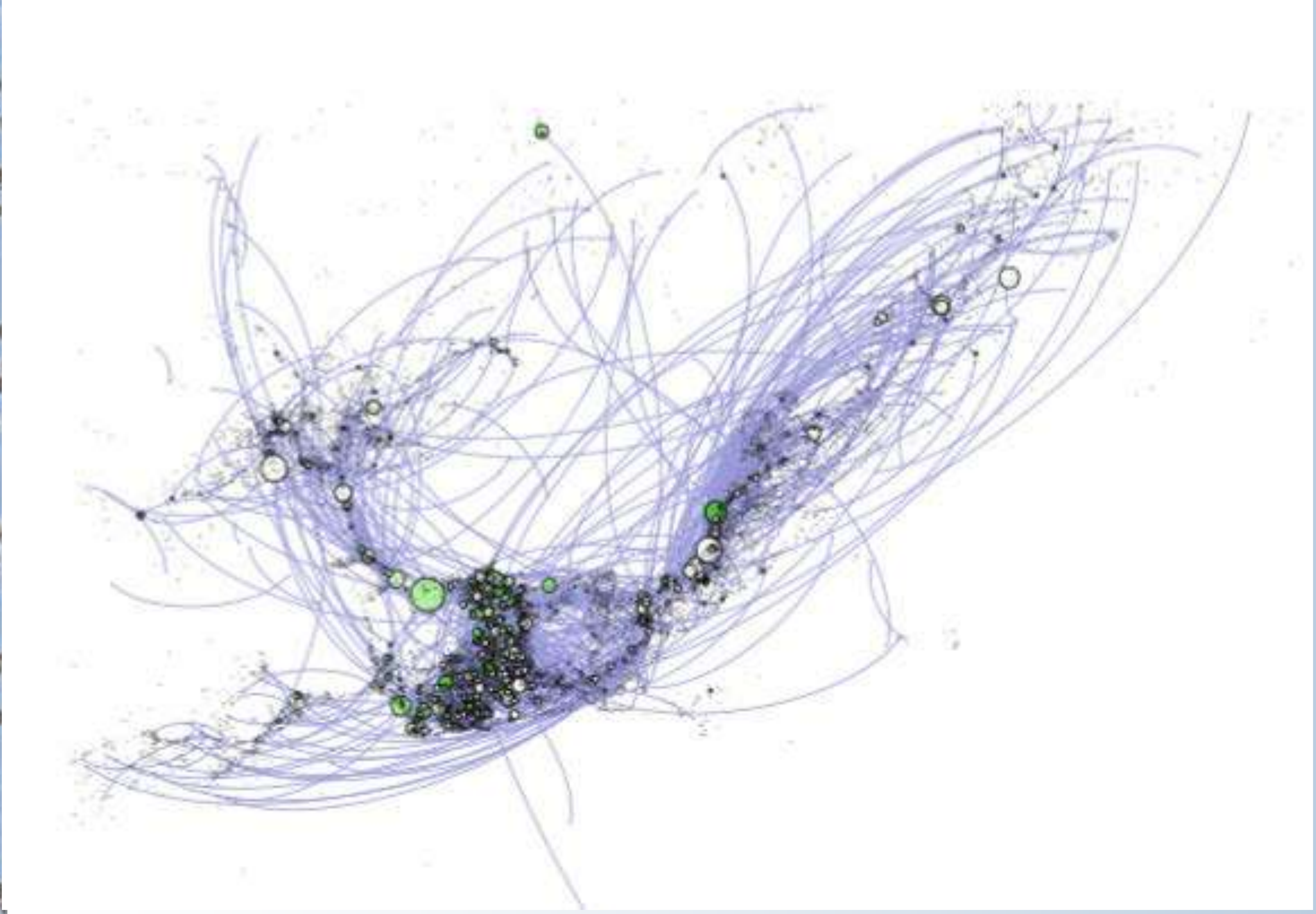
- Set of locations: **Nodes** [N]
- Set of possible trips to choose from: **States** [Nx(N-1)]
- Set of trips to distribute: **Events** [T]

We finally obtain our observables:

- Set of trips between locations ("flows") **Occupation Numbers** [{t_{ij} }]

Multi-Edge networks?

"Networks formed by a discrete set of nodes which generate a group of states which can be populated by none, one or more than one distinguishable events."



Relevance of networks

- Representation
- Table versus network

Network helps to visualize

- Redes UB

Also to characterize

- Centrality measures
  - local: who collaborates with whom
  - global: who is closer to the rest?
  - Why Google became so famous? Google Page Rank: centrality measure (I am more popular if my neighbors are also popular)

Roles of different actors

- Individual by individual
- Who is who in the structure and in the functionality

At global level

- How can we characterize a network?
- Forget individual roles
- Statistics

Statistics

- Distributions: power law distributions far from Poisson-like
- Divergencies of some features of the distributions

Complexity

- Normal statistics: a few and they are not very interesting
- Long tails: power laws and all its implications

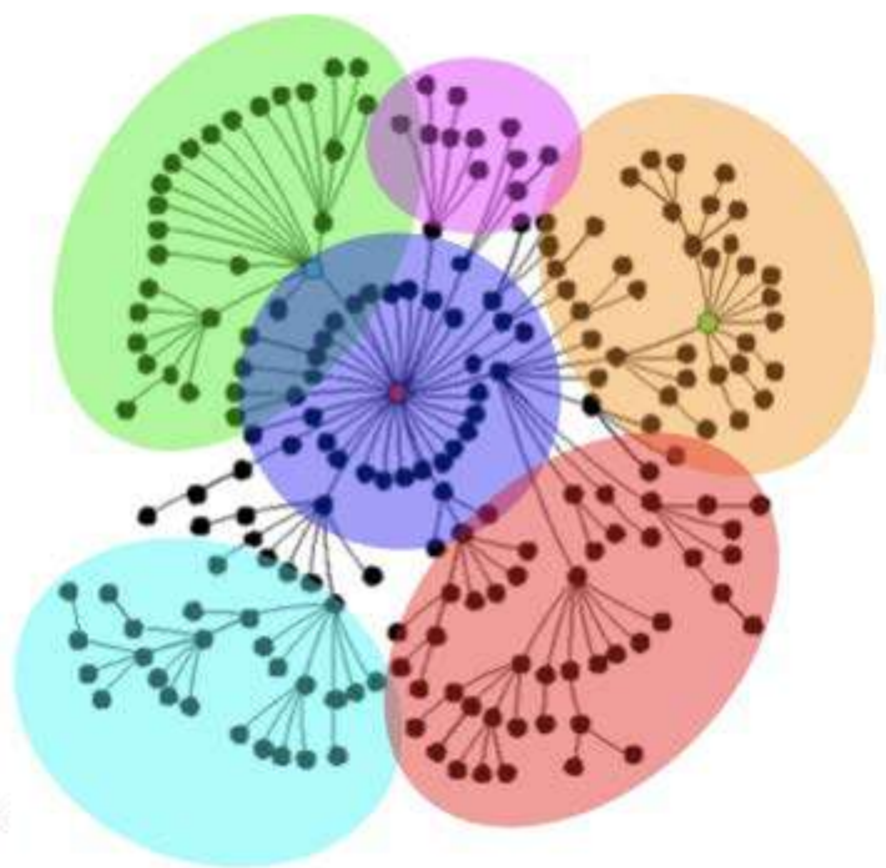
Scales

All scales are relevant (simultaneously)

- Local: microscale
- Global: macroscale
- Mid: mesoscale

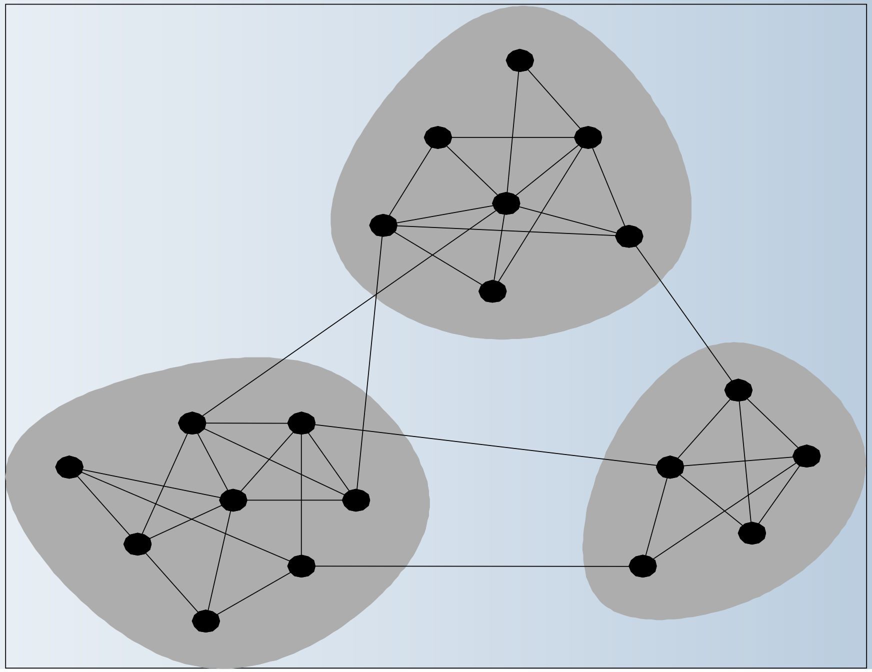
Communities

- Groups that have more internal links that with the rest of the nodes (on average)



Objectives

- Existence of communities or modules in networks
- Technical issue: finding the best partition
- Management issue: finding meaningful partitions



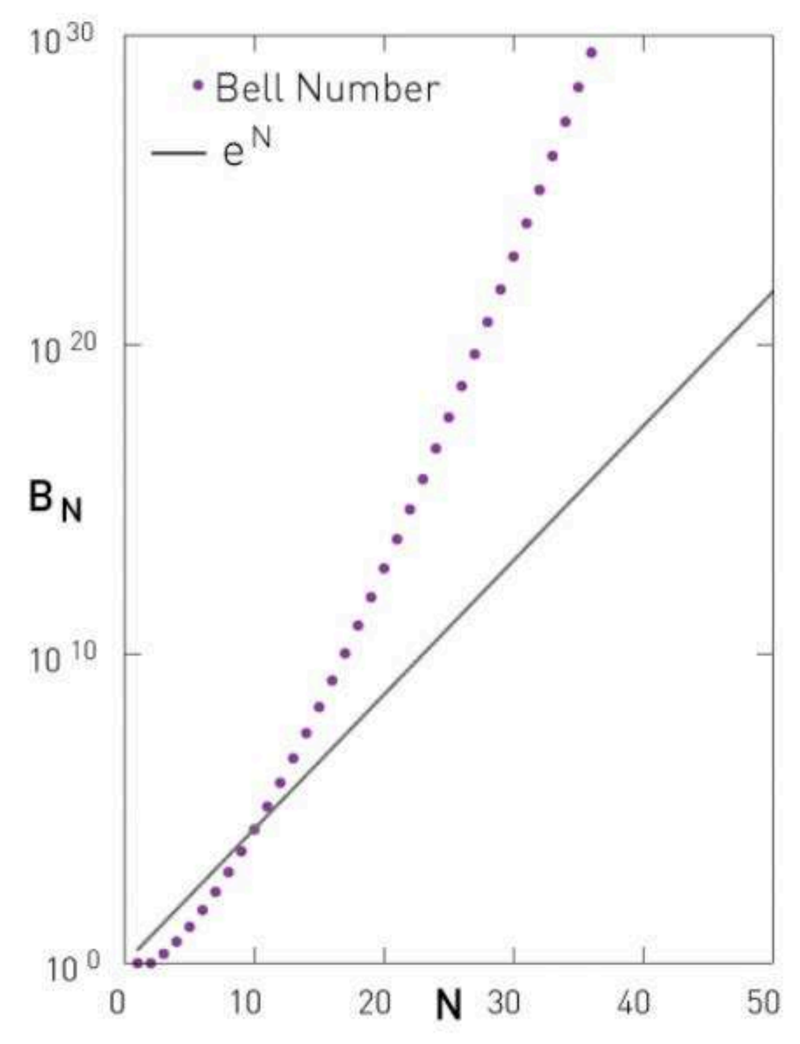
Technical issue

- We have to identify the communities
- How many possible partitions into communities?
- NP problem to find the best one

Bell number

From Wikipedia, the free encyclopedia

In combinatorial mathematics, the **Bell numbers** count the number of partitions of a set. These numbers have been studied by mathematicians since the 19th century, and their roots go back to medieval Japan, but they are named after Eric Temple Bell, who wrote about them in the 1930s.



Number of Partitions

The number of partitions of a network of size *N* is provided by the Bell number (9.6). The figure compares the Bell number to an exponential function, illustrating that the number of possible partitions grows faster than exponentially. Given that there are over 1040 partitions for a network of size *N*=50, brute-force approaches that aim to identify communities by inspecting all possible partitions are computationally infeasible.

10^{40}

Graph Partitioning

We can solve the graph bisection problem by inspecting all possible divisions into two groups and choosing the one with the smallest cut size. To determine the computational cost of this brute force approach we note that the number of distinct ways we can partition a network of N nodes into groups of N_1 and N_2 nodes is

$$\frac{N!}{N_1!N_2!} \qquad (9.3)$$

Using Stirling's formula

$$n! \simeq \sqrt{2\pi n}(n/e)^n$$

we can write (9.3) as

$$\frac{N!}{N_1!N_2!} \simeq \frac{\sqrt{2\pi N}(N/e)^N}{\sqrt{2\pi N_1}(N_1/e)^{N_1}\sqrt{2\pi N_2}(N_2/e)^{N_2}} \sim \frac{N^{N+1/2}}{N_1^{N_1+1/2}N_2^{N_2+1/2}} \qquad (9.4)$$

To simplify the problem let us set the goal of dividing the network into two equal sizes $N_1 = N_2 = N/2$. In this case (9.4) becomes

$$\frac{2^{N+1}}{\sqrt{N}} = e^{(N+1)\ln 2 - \frac{1}{2}\ln N} \qquad (9.5)$$

indicating that the number of bisections increases exponentially with the size of the network.

Maximum cliques

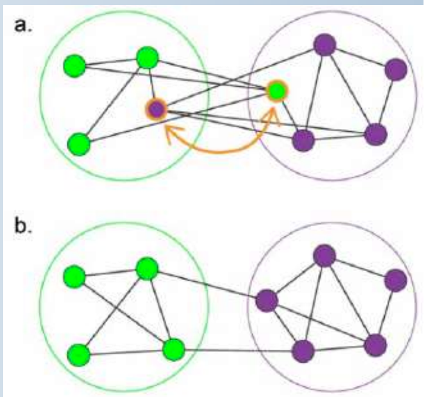
- Define a community as group of individuals whose members all know each other. In graph theoretic terms this means that a community is a complete subgraph, or a **clique**.
- A clique is a connected subgraph with maximal link density.
- But:
  - Triangles define clustering and are quite common
  - Squares, ... very very rare
  - Too restrictive definition

Strong and weak

- ***Strong community***: C is a strong community if each node within C has more links within the community than with the rest of the graph
- ***Weak community***: C is a weak community if the total internal degree of a subgraph exceeds its total external degree

Partition

- A partition is a division of the network into groups, communities or clusters
- The question is: Which of all possible partitions is the best?
- NP problem
- Community detection:
  - From computer scientists (graph partitioning, Kernighan-Lin)
  - To statistical physicists (Girvan-Newman, PNAS 99, 7821, 2002)



Quantifying a partition

- **Modularity:**

$$Q = \sum_i (e_{ii} - a_i^2)$$

- **e_{ij}: fraction of total links starting at a node in partition i and ending at a node in partition j**
- **a_i: fraction of links connected to i**
- **a_i^2: number of intracommunity links**

Modularity

Consider a network with N nodes and L links and a partition into n_c communities, each community having N_c nodes connected to each other by L_c links, where c=1,...,n_c. If L_c is larger than the expected number of links between the N_c nodes given the network's degree sequence, then the nodes of the subgraph C_c could indeed be part of a true community, as expected based on the Density Hypothesis H2 (Image 9.2). We therefore measure the difference between the network's real wiring diagram (A_{ij}) and the expected number of links between i and j if the network is randomly wired (p_{ij}),

$$M_c = \frac{1}{2L} \sum_{(i,j)\in C_c} (A_{ij} - p_{ij}) \qquad (9.9)$$

Here p_{ij} can be determined by randomizing the original network, while keeping the expected degree of each nod unchanged. Using the degree preserving null model (7.1) we have

$$p_{ij} = \frac{k_i k_j}{2L} \qquad (9.10)$$

If M_c is positive, then the subgraph C_c has more links than expected by chance, hence it represents a potential community. If M_c is zero then the connectivity between the N_c nodes is random, fully explained by the degree distribution. Finally, if M_c is negative, then the nodes of C_c do not form a community.

Using (9.10) we can derive a simpler form for the modularity (9.9) (ADVANCED TOPICS 9.B)

$$M_c = \frac{L_c}{L} - \left(\frac{k_c}{2L}\right)^2 \qquad (9.11)$$

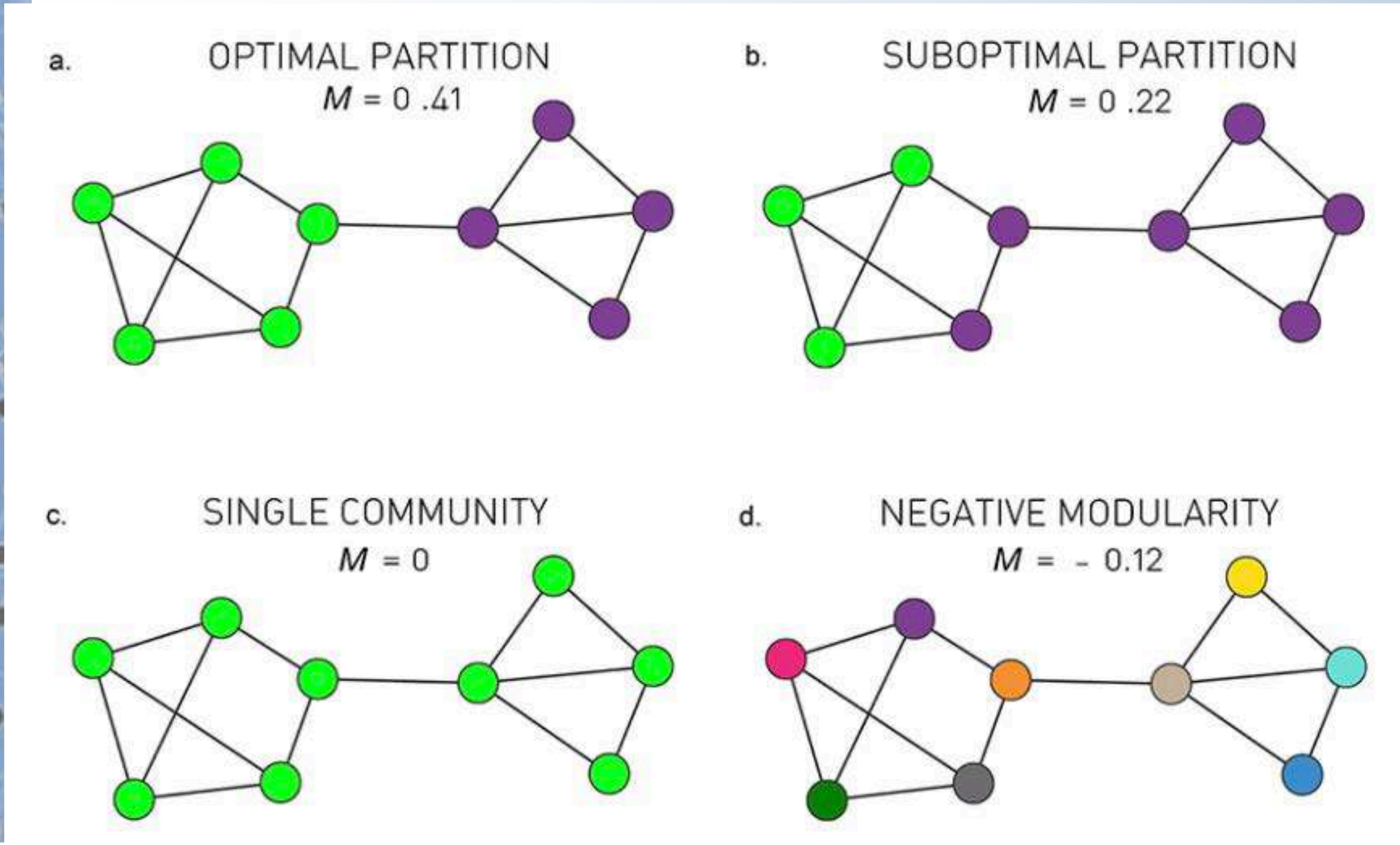
where L_c is the total number of links within the community C_c and k_c is the total degree of the nodes in this community.

To generalize these ideas to a full network consider the complete partition that breaks the network into n_c communities. To see if the local link density of the subgraphs defined by this partition differs from the expected density in a randomly wired network, we define the partition's *modularity* by summing (9.11) over all n_c communities [23]

$$M = \sum_{c=1}^{n_c} \left[\frac{L_c}{L} - \left(\frac{k_c}{2L}\right)^2\right] \qquad (9.12)$$

Modularity

- Higher modularity better partition
- Zero modularity (1 or N)



Methods of community identification

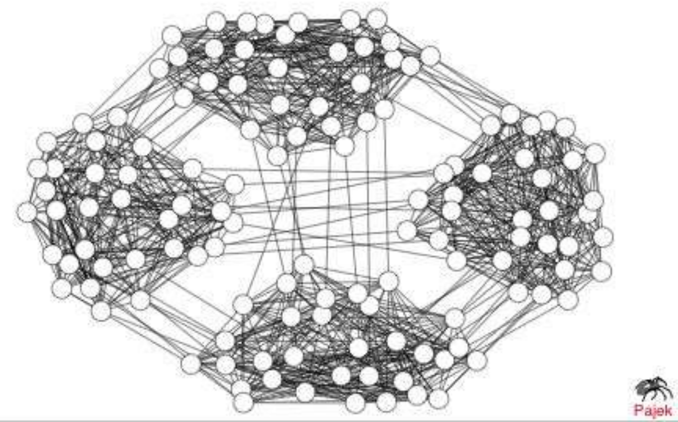
- L. Danon, J. Duch, A.D-G, A. Arenas J. Stat. Mech. (2005) P09008
  - Link removal methods
  - Agglomerative methods
  - Maximizing modularity
  - Spectral analysis methods
  - Based on physics: resistor networks, q-Potts model
- More recent reviews:
  - S. Fortunato, *Community detection in graphs* (Phys. Rep. 486, 75-174, 2010)

Computational costs

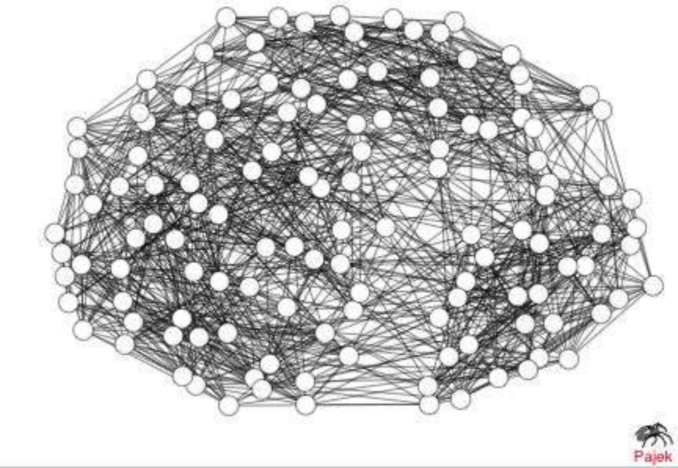
Reference	Alias	Order
(Newman and Girvan, 2004)	NG	$O(m^2 n)$
(Girvan and Newman, 2002)	GN	$O(n^2 m)$
(Fortunato et al., 2004)	FLM	$O(n^4)$
(Radicchi et al., 2004)	RCCLP	$O(n^2)$
(Newman, 2004b)	NF	$O(n \log^2 n)$
(Donetti and Muñoz, 2004),	DMSA	$O(n^3)$
(Donetti and Muñoz, 2004),	DMCA	$O(n^3)$
(Eckmann and Moses, 2002)	EM	$O(m\langle k^2\rangle)$
(Zhou and Lipowsky, 2005)	ZL	$O(n^3)$
(Reichardt and Bornholdt, 2004)	RB	unknown
(Bagrow and Bollt, 2004)	BB	$O(n^3)$
(Duch and Arenas, 2005)	DA	$O(n^2 \log n)$
(Capocci et al., 2004)	CSCC	$O(n^2)$
(Wu and Huberman, 2004)	WH	$O(n+m)$

Comparing algorithms

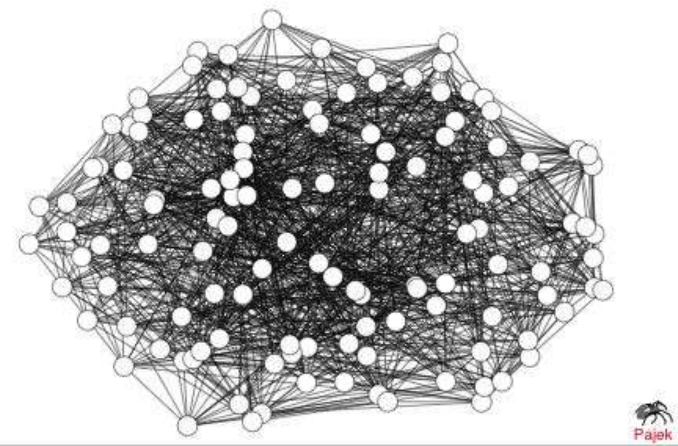
- *ad-hoc* networks (Newman-Girvan, PRE 69, 026113, 2004)
  - 128 nodes
  - 4 communities of 32 nodes each
  - Each node has 16 links:
    - z_{in} internal nodes within the community
    - z_{out} nodes out of its community



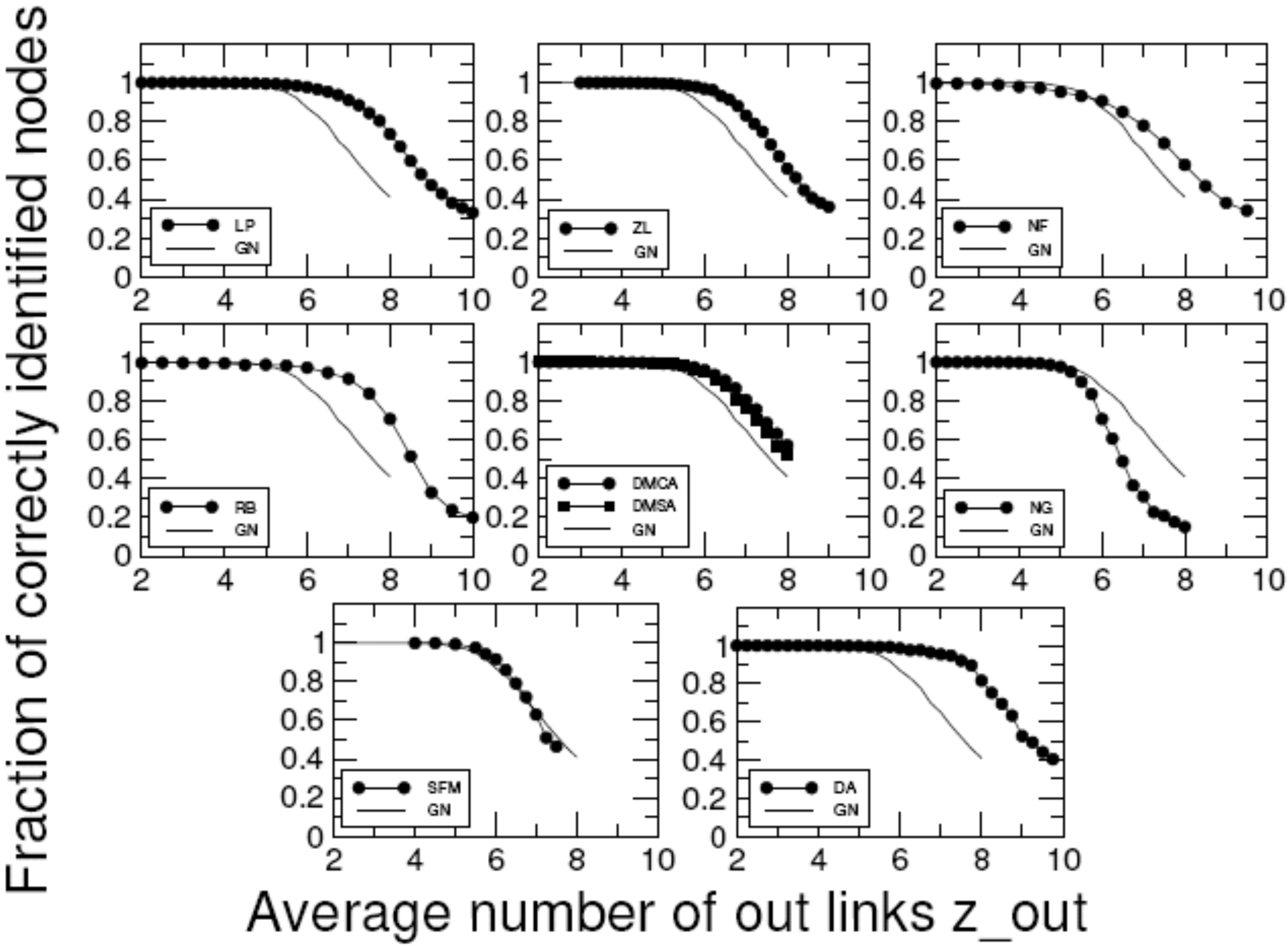
z_{in}=15



z_{in}=12



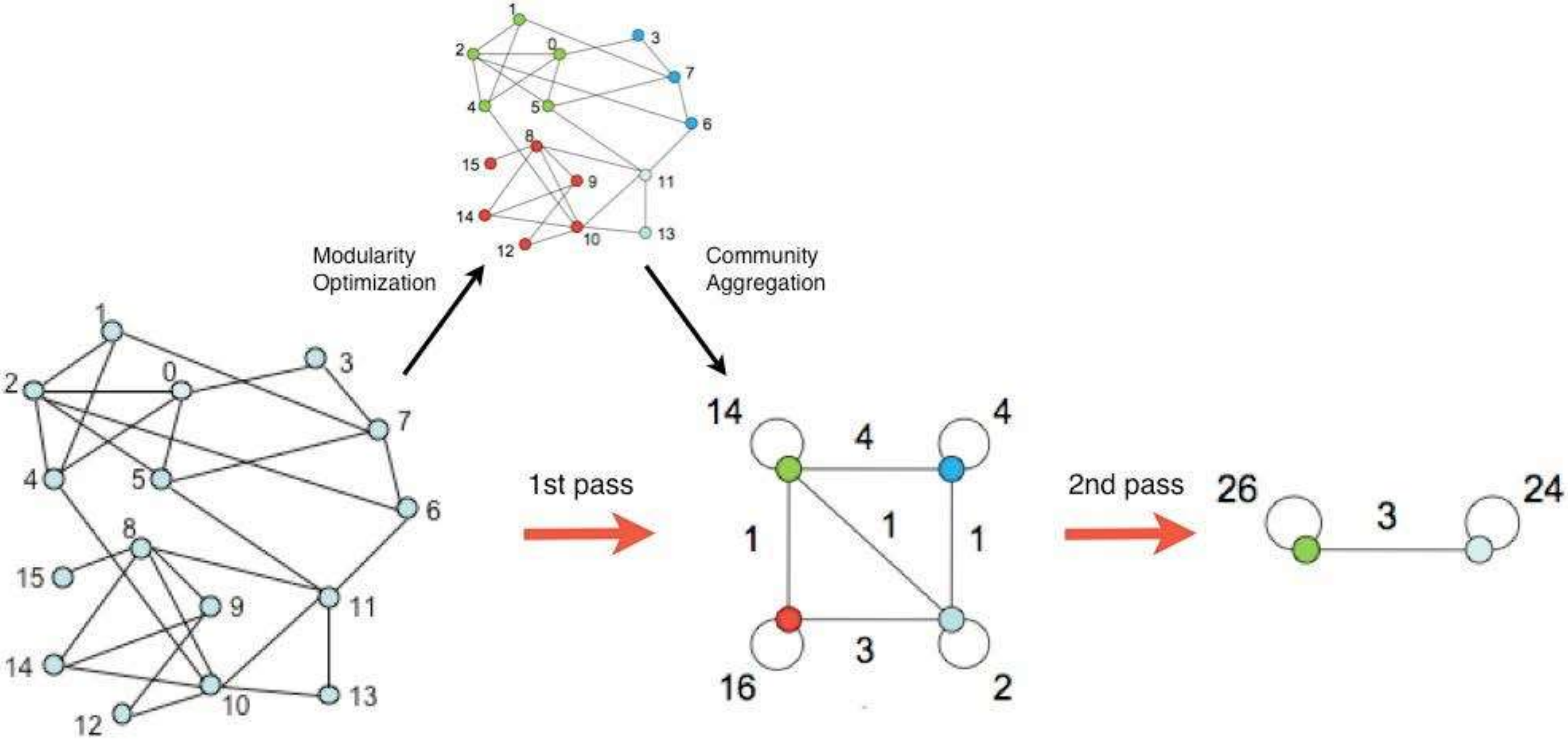
z_{in}=8


Fraction of correctly identified nodes
Average number of out links z_out
LP
GN
ZL
NF
RB
DMCA
DMSA
NG
SFM
DA

Optimal?????

- For a given network the partition with **maximum modularity** corresponds to the **optimal** community structure.
- The hypothesis is supported by the inspection of small networks, for which the **maximum M agrees with the expected** communities

Louvain algorithm



Community detection for NetworkX's documentation

This module implements community detection.

It uses the louvain method described in Fast unfolding of communities in large networks, Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Renaud Lefebvre, Journal of Statistical Mechanics: Theory and Experiment 2008(10), P10008 (12pp)

It depends on Networkx to handle graph operations : http://networkx.lanl.gov/

The program can be found in a repository where you can also report bugs :

https://bitbucket.org/taynaud/python-louvain

```
import community
import networkx as nx
import matplotlib.pyplot as plt

#better with karate_graph() as defined in networkx example.
#erdos renyi don't have true community structure
G = nx.erdos_renyi_graph(30, 0.05)

#first compute the best partition
partition = community.best_partition(G)

#drawing
size = float(len(set(partition.values())))
pos = nx.spring_layout(G)
count = 0.
for com in set(partition.values()) :
    count = count + 1.
    list_nodes = [nodes for nodes in partition.keys()
                                if partition[nodes] == com]
    nx.draw_networkx_nodes(G, pos, list_nodes, node_size = 20,
                                node_color = str(count / size)


nx.draw_networkx_edges(G,pos, alpha=0.5)
plt.show()
```

Community Detection with the Karate Club Network

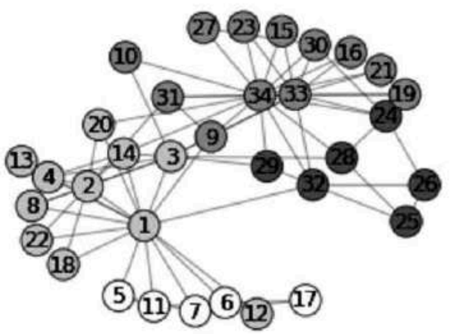
In [25]:
```
import community

G=nx.read_edgelist("./data/karate.dat")

#first compute the best partition
partition = community.best_partition(G)

#plot the network
size = float(len(set(partition.values())))
pos = nx.spring_layout(G)
count = 0.
plt.axis('off')
for com in set(partition.values()) :
    count = count + 1.
    list_nodes = [nodes for nodes in partition.keys() \
                  if partition[nodes] == com]
    nx.draw_networkx_nodes(G, pos, list_nodes, node_size = 300, \
                           node_color = str(count / size))
    nx.draw_networkx_labels(G,pos)

nx.draw_networkx_edges(G,pos, alpha=0.5,width=1)
savefig('./data/karate_community.png',dpi=600)
```



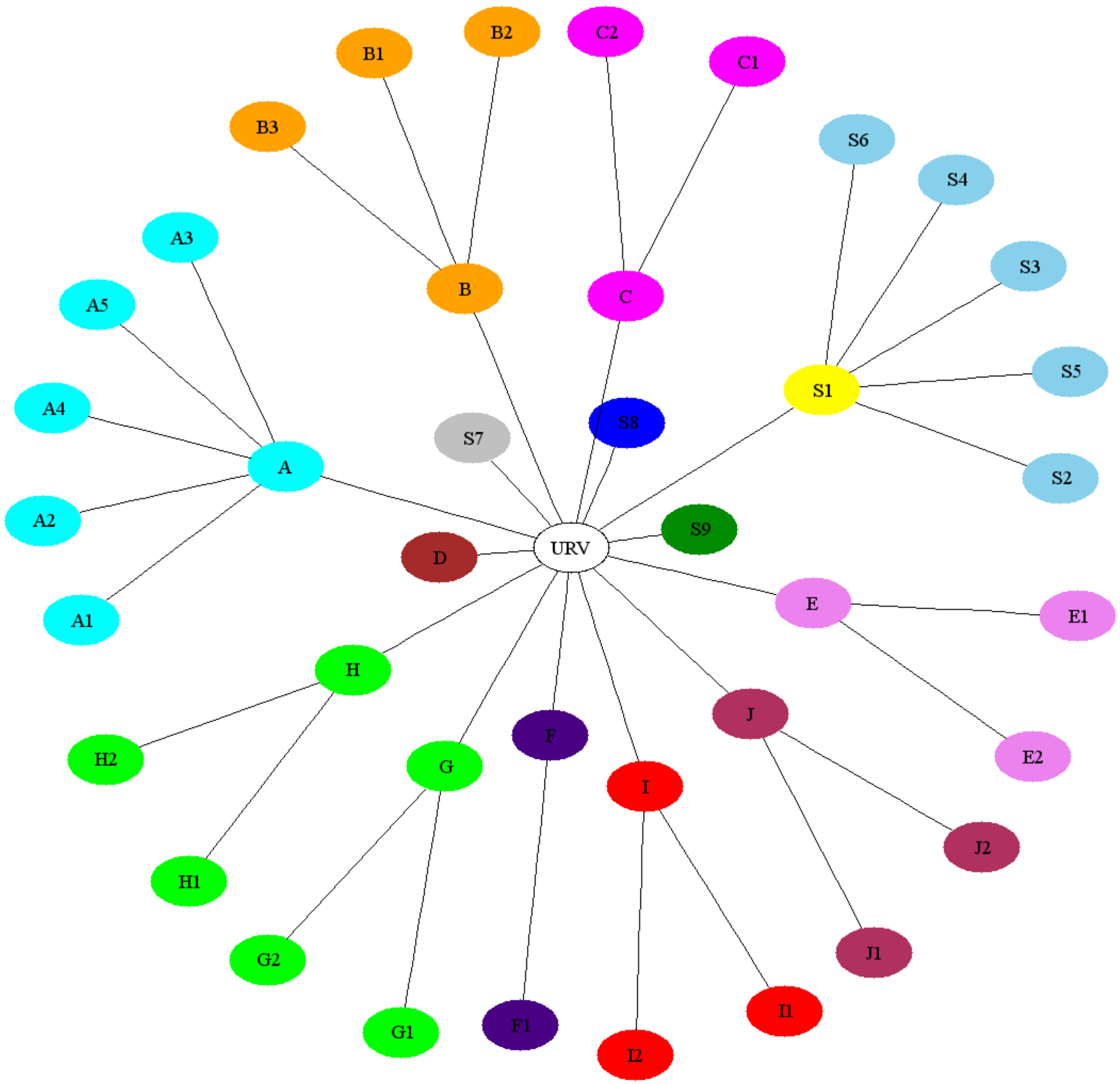
Identifying communities

- Identifying what communities are
- Managerial point of view:
  - How a company is organized
  - How powerful is the formed informal chart

Two networks

- E-mail network at Universitat Rovira i Virgili
- FisEs

URV


URV
A
A1
A2
A3
A4
A5
B
B1
B2
B3
C
C1
C2
D
E
E1
E2
F
F1
G
G1
G2
H
H1
H2
I
I1
I2
J
J1
J2
S1
S2
S3
S4
S5
S6
S7
S8
S9

Importance from management

Unravel the real (informal) organization behind the formal chart

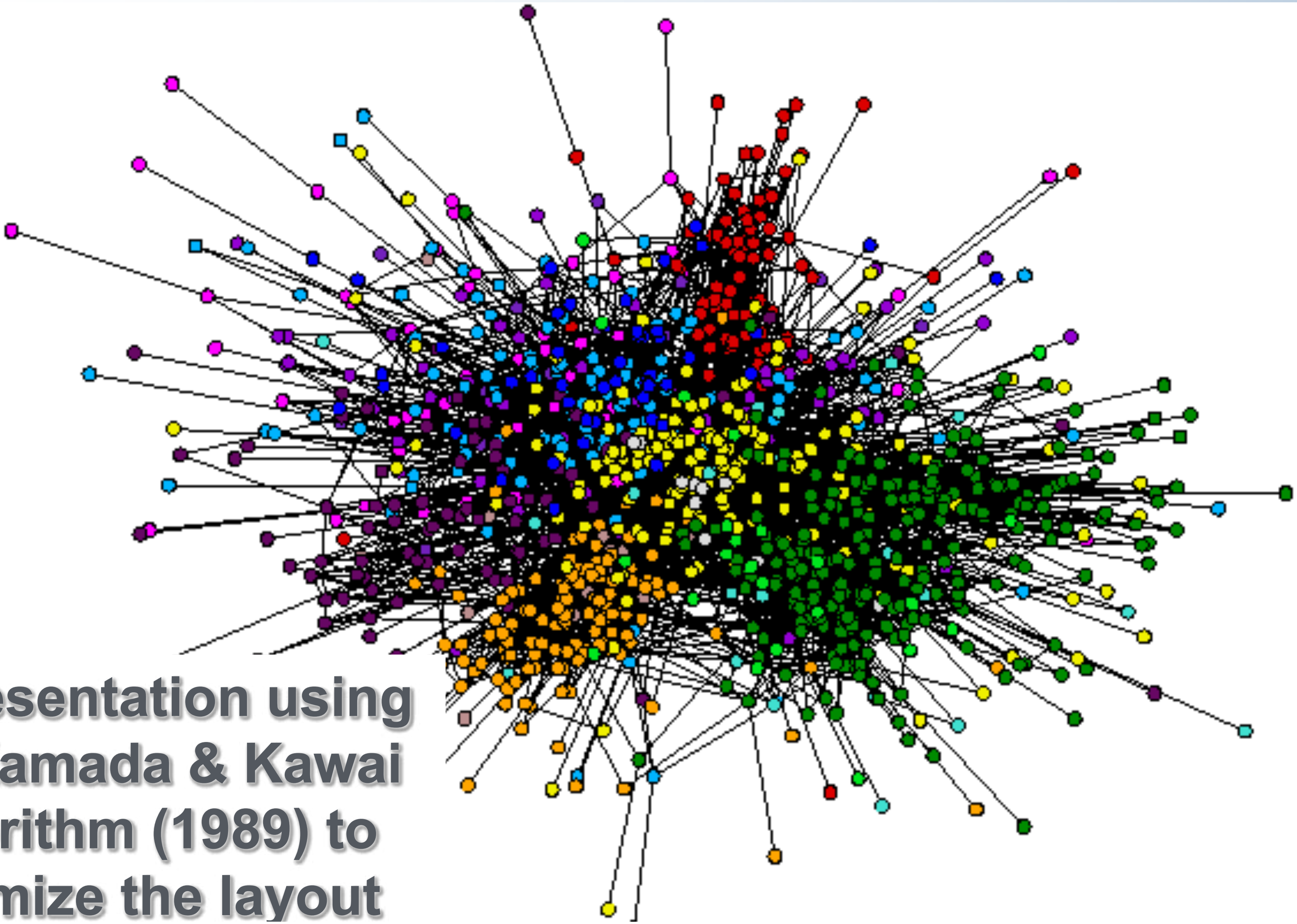
"If the formal organization is the skeleton of a company, the informal is the central nervous system... Complex webs of social ties form every time colleagues communicate and solidify over time into surprisingly stable networks."

D. Krackhardt and J. R. Hanson, Harvard Business Review, 71, 104-113 (1993)

Data acquisition to construct the e-mail network of the URV

- Node => e-mail address
- Link => bidirectional e-mails between nodes (undirected graph)
- Number of users approx. 1700 (professors, technicians, administrators, graduate students)
- We consider only e-mails sent within the University during the first 3 months of 2002 (stable network)
- Non "spam" mail: (neglect >50 recipients)

Email at URV



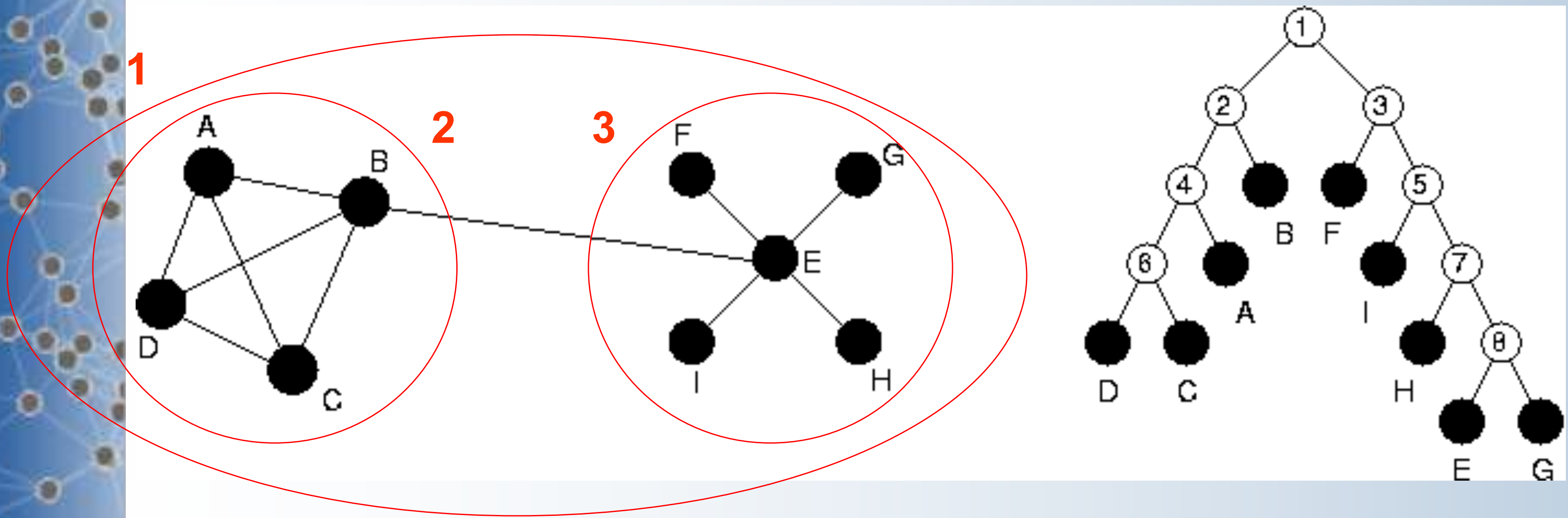
Representation using the Kamada & Kawai algorithm (1989) to optimize the layout

Community identification: the Girvan & Newman (GN) algorithm*

- **Definition**: Betweenness of a link = # minimum paths connecting pairs of nodes that go through that link
- **Idea in GN algorithm**: The links which connect highly clustered communities have a higher link betweenness. Then cut these links to separate communities.

*Girvan and Newman, PNAS USA 99, 7821-7826 (2002)

Communities

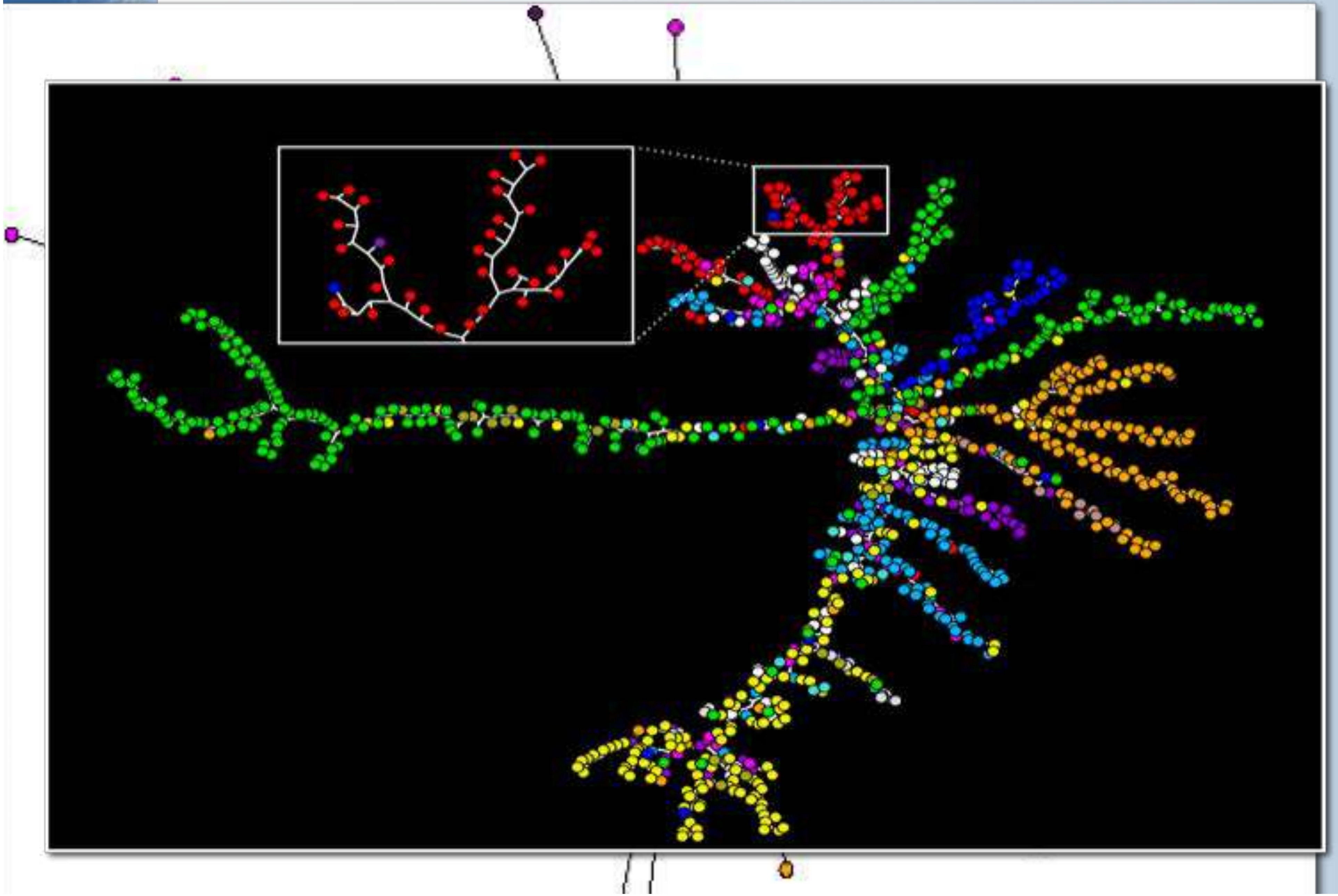


A network containing two clear communities linked by BE.

Since there is no more community structure, the rest of the nodes will be separated one by one generating a binary tree with two branches corresponding to the two communities.

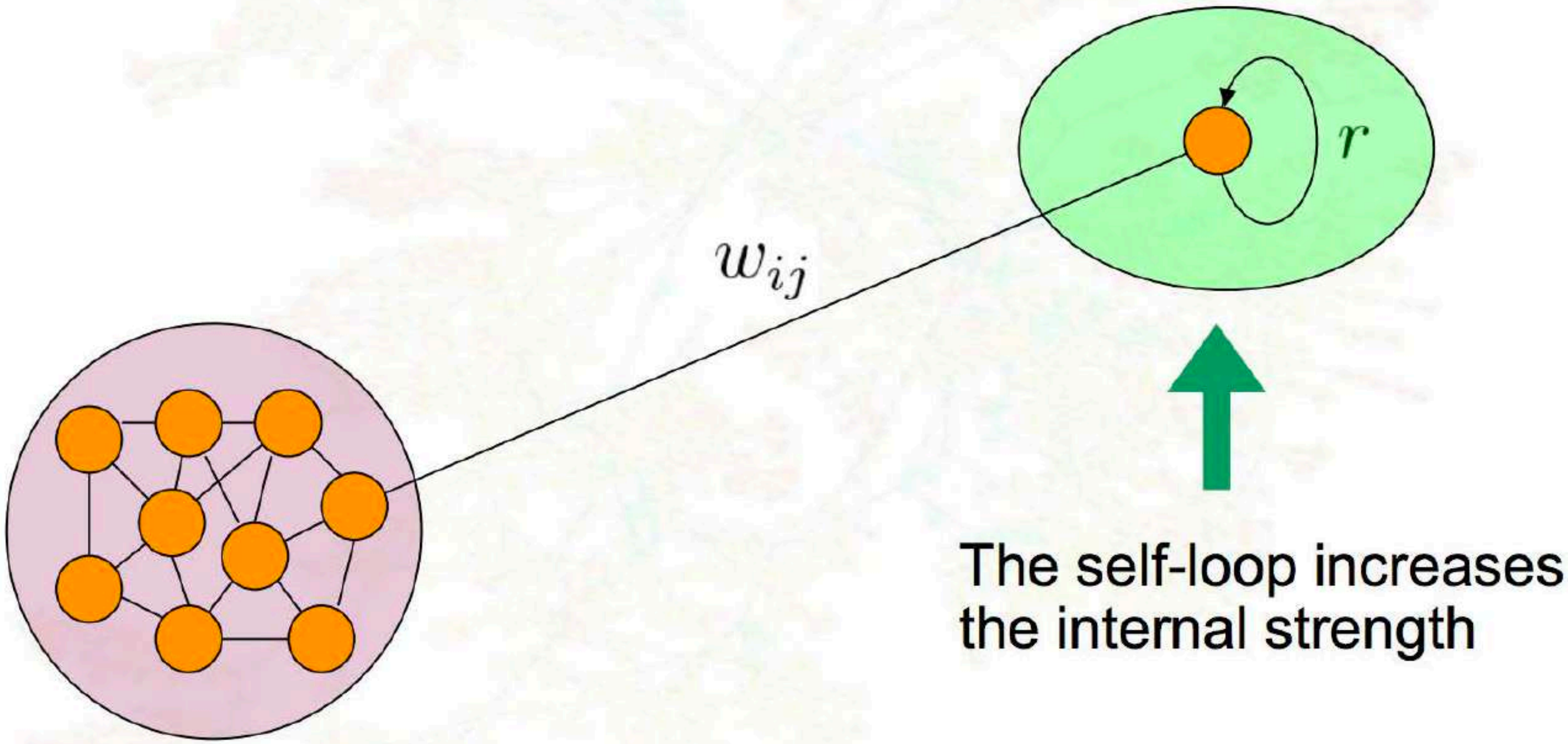
Leaders are at the tips of the branches

Email network



Communities at all levels (self-loops)

- Tune the *resistance* of nodes to join communities, adding *self-loops*



- Multiple resolution method
  (Arenas, Fernandez & Gómez (2008) New J Phys **10**, 053039)

  - Add a common resistance (self-loop) to all nodes

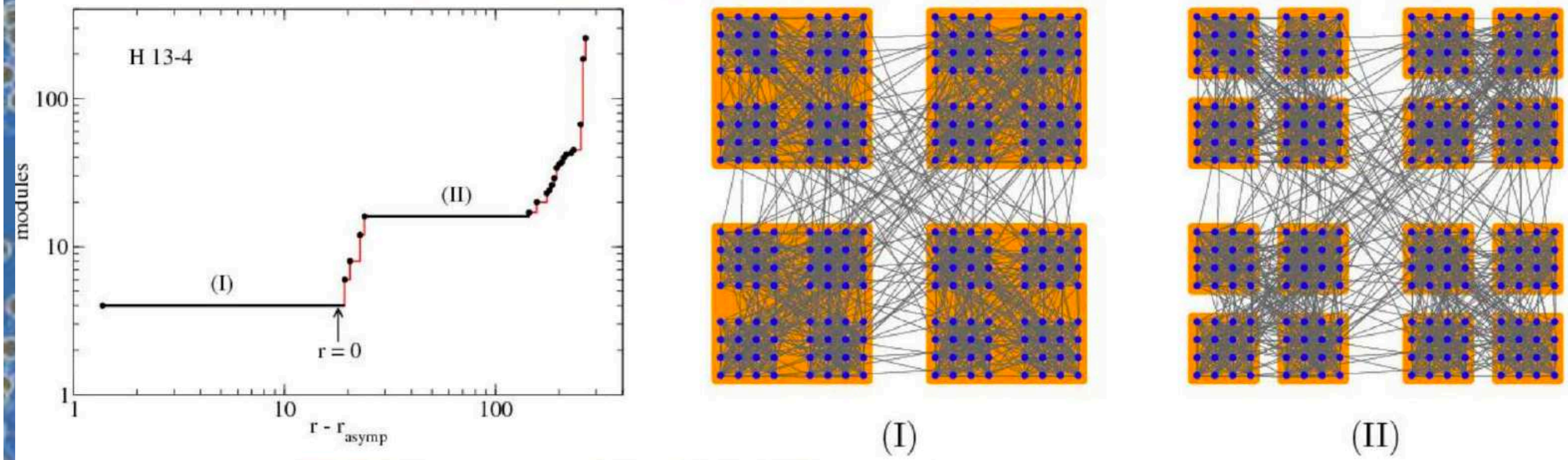
$$w'_{ij} = \begin{cases} w_{ij} & \text{if} \quad i \neq j \\ r & \text{if} \quad i = j \end{cases}$$

$$w'_i = w_i + r$$
$$2w' = 2w + Nr$$

  - Optimize modularity

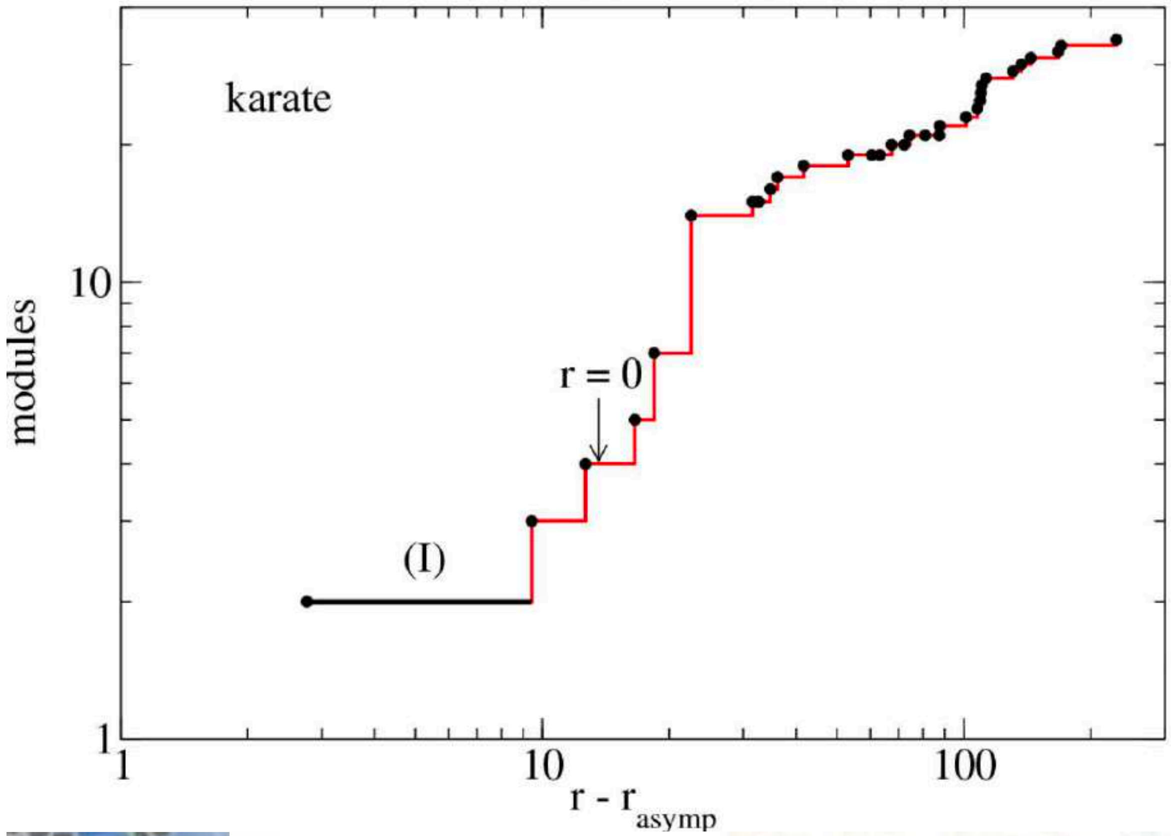
$$Q_r = \frac{1}{2w'} \sum_i \sum_j \left(w'_{ij} - \frac{w'_i w'_j}{2w'} \right) \delta(C_i, C_j)$$

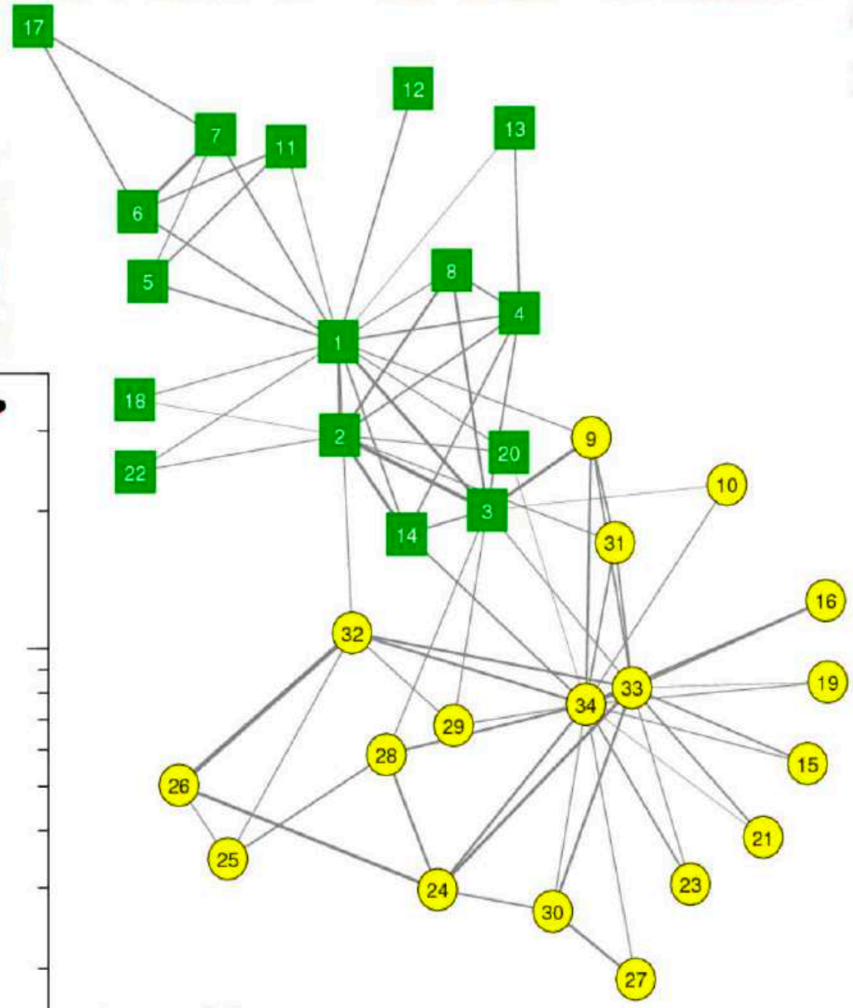
- **Homogeneous with two hierarchical levels**
  (Arenas, Diaz-Guilera & Perez-Vicente (2006) Phys Rev Lett **96**, 114102)

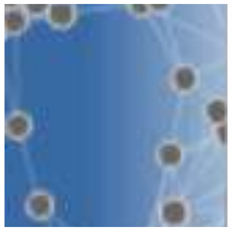


Social networks

- Zachary karate club
  (Zachary (1977) J Anthropol Res **33**, 452)

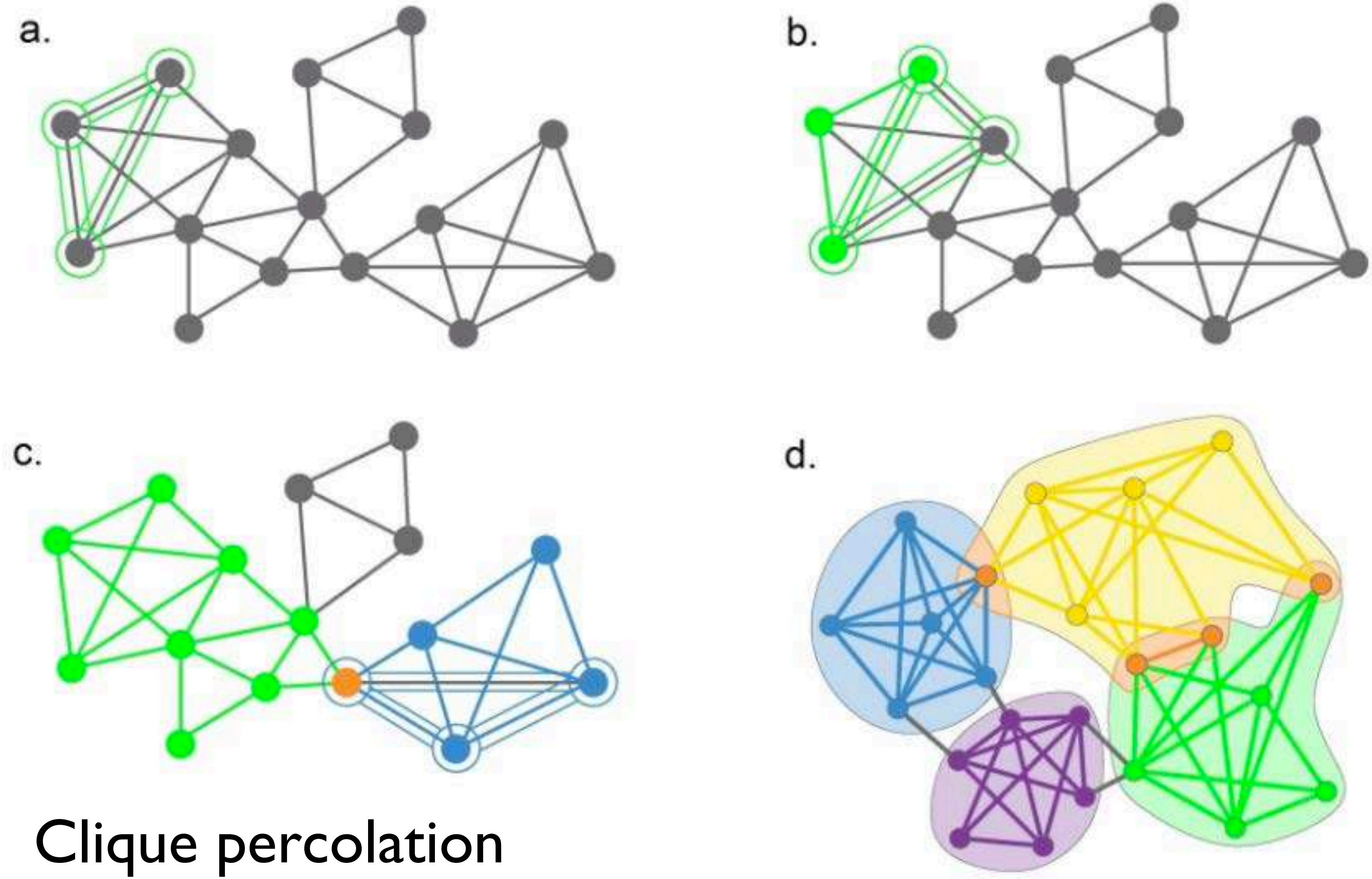






- Software: Radatools
  - http://deim.urv.cat/~sgomez/radatools.php

Overlapping communities



Clique percolation

From data to models

- Statistical models
- Physical models

Network models

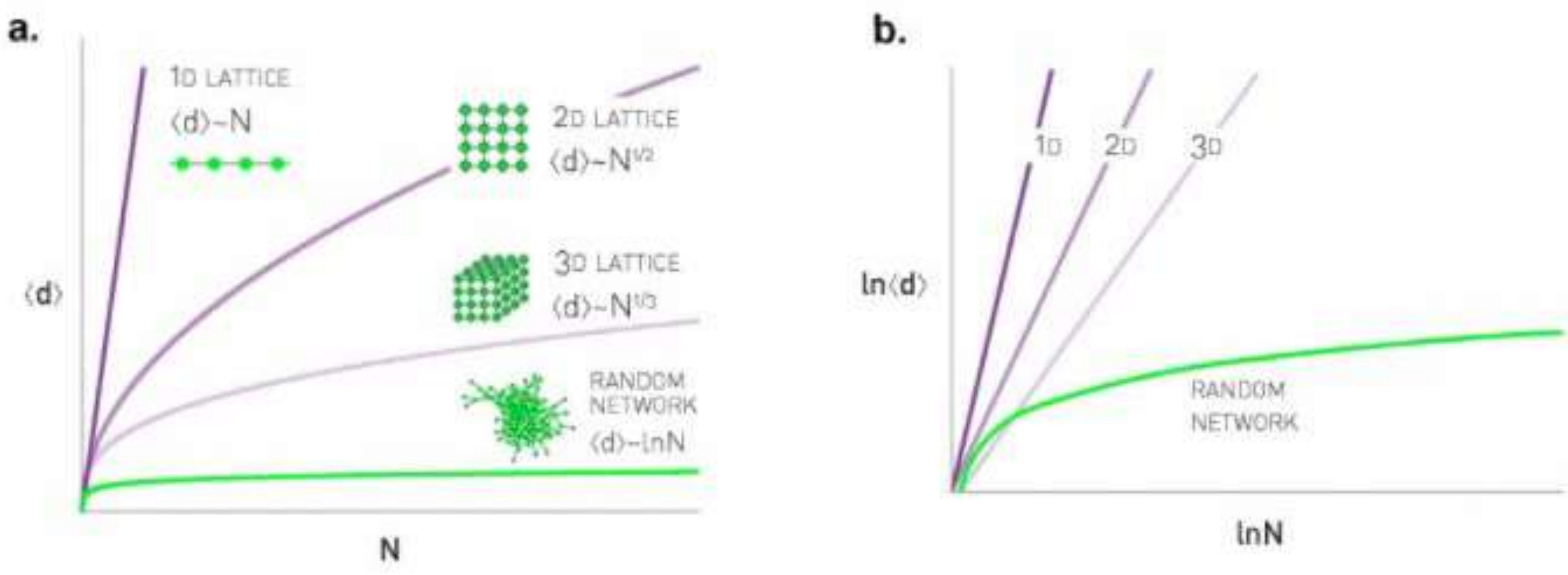
- Erdös-Reny: Random graphs
- Watts-Strogatz: small worlds
- Barabasi-Albert: scale free networks

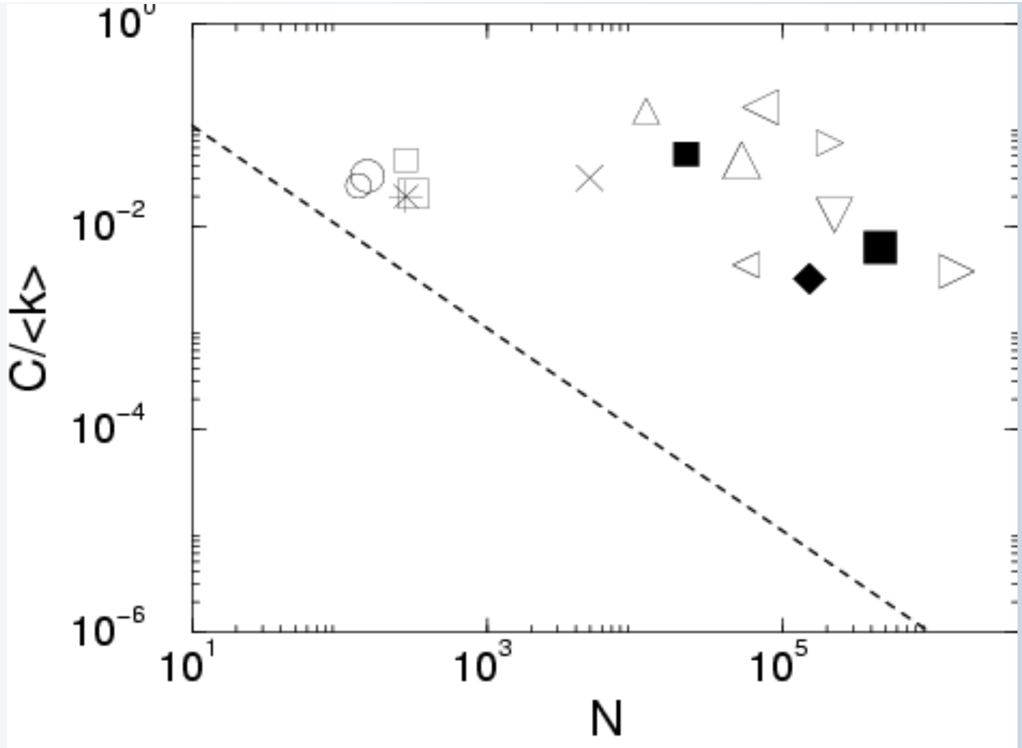
Universality

- What do have in common networks in so different "worlds"?
- Minimize the number of relevant variables
- We propose mechanisms
- Particular cases are not so important

1 Erdos-Renyi: random graph model

- Definition: N labeled nodes connected by n links which are chosen randomly from the N(N-1)/2 possible links
- There are $\binom{N(N-1)/2}{n}$ graphs with N nodes and n links


a.
1D LATTICE
⟨d⟩~N
2D LATTICE
⟨d⟩~N^1/2
3D LATTICE
⟨d⟩~N^1/3
RANDOM
NETWORK
⟨d⟩~lnN
⟨d⟩
N
b.
1D
2D
3D
ln⟨d⟩
RANDOM
NETWORK
lnN


10^0
10^-2
10^-4
10^-6
C/<k>
10^1
10^3
10^5
N

Random Graphs

Generators for random graphs.

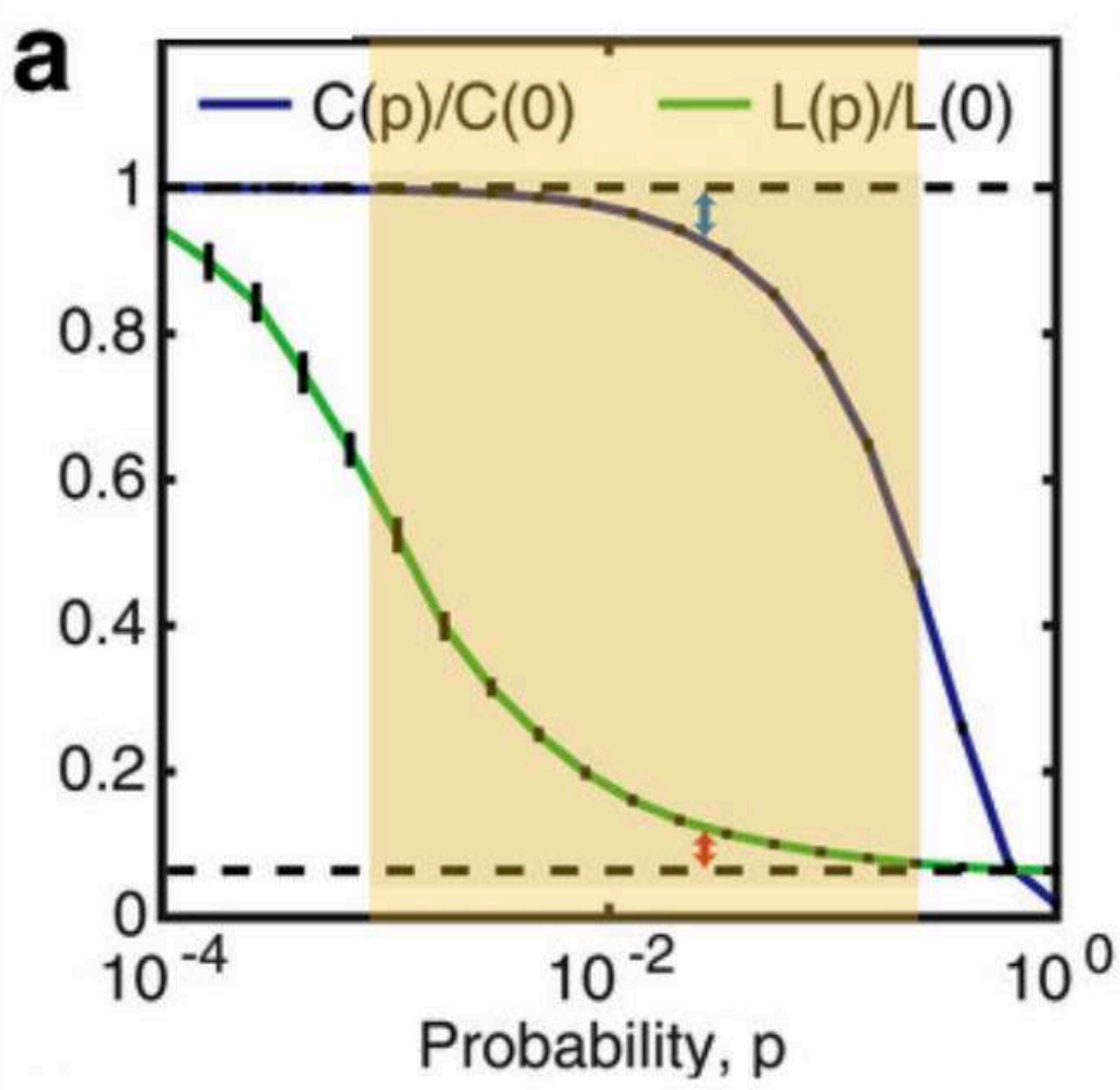
fast_gnp_random_graph(n, p[, seed, directed])	Return a random graph G_{n,p} (Erdős-Rényi graph, binomial graph).
gnp_random_graph(n, p[, seed, directed])	Return a random graph G_{n,p} (Erdős-Rényi graph, binomial graph).
dense_gnm_random_graph(n, m[, seed])	Return the random graph G_{n,m}.
gnm_random_graph(n, m[, seed, directed])	Return the random graph G_{n,m}.
erdos_renyi_graph(n, p[, seed, directed])	Return a random graph G_{n,p} (Erdős-Rényi graph, binomial graph).
binomial_graph(n, p[, seed, directed])	Return a random graph G_{n,p} (Erdős-Rényi graph, binomial graph).
newman_watts_strogatz_graph(n, k, p[, seed])	Return a Newman-Watts-Strogatz small world graph.
watts_strogatz_graph(n, k, p[, seed])	Return a Watts-Strogatz small-world graph.
connected_watts_strogatz_graph(n, k, p[, ...])	Return a connected Watts-Strogatz small-world graph.
random_regular_graph(d, n[, seed])	Return a random regular graph of n nodes each with degree d.
barabasi_albert_graph(n, m[, seed])	Return random graph using Barabási-Albert preferential attachment model.
powerlaw_cluster_graph(n, m, p[, seed])	Holme and Kim algorithm for growing graphs with powerlaw
random_lobster(n, p1, p2[, seed])	Return a random lobster.
random_shell_graph(constructor[, seed])	Return a random shell graph for the constructor given.
random_powerlaw_tree(n[, gamma, seed, tries])	Return a tree with a powerlaw degree distribution.
random_powerlaw_tree_sequence(n[, gamma, ...])	Return a degree sequence for a tree with a powerlaw distribution.

2 Watts-Strogatz: small-world model

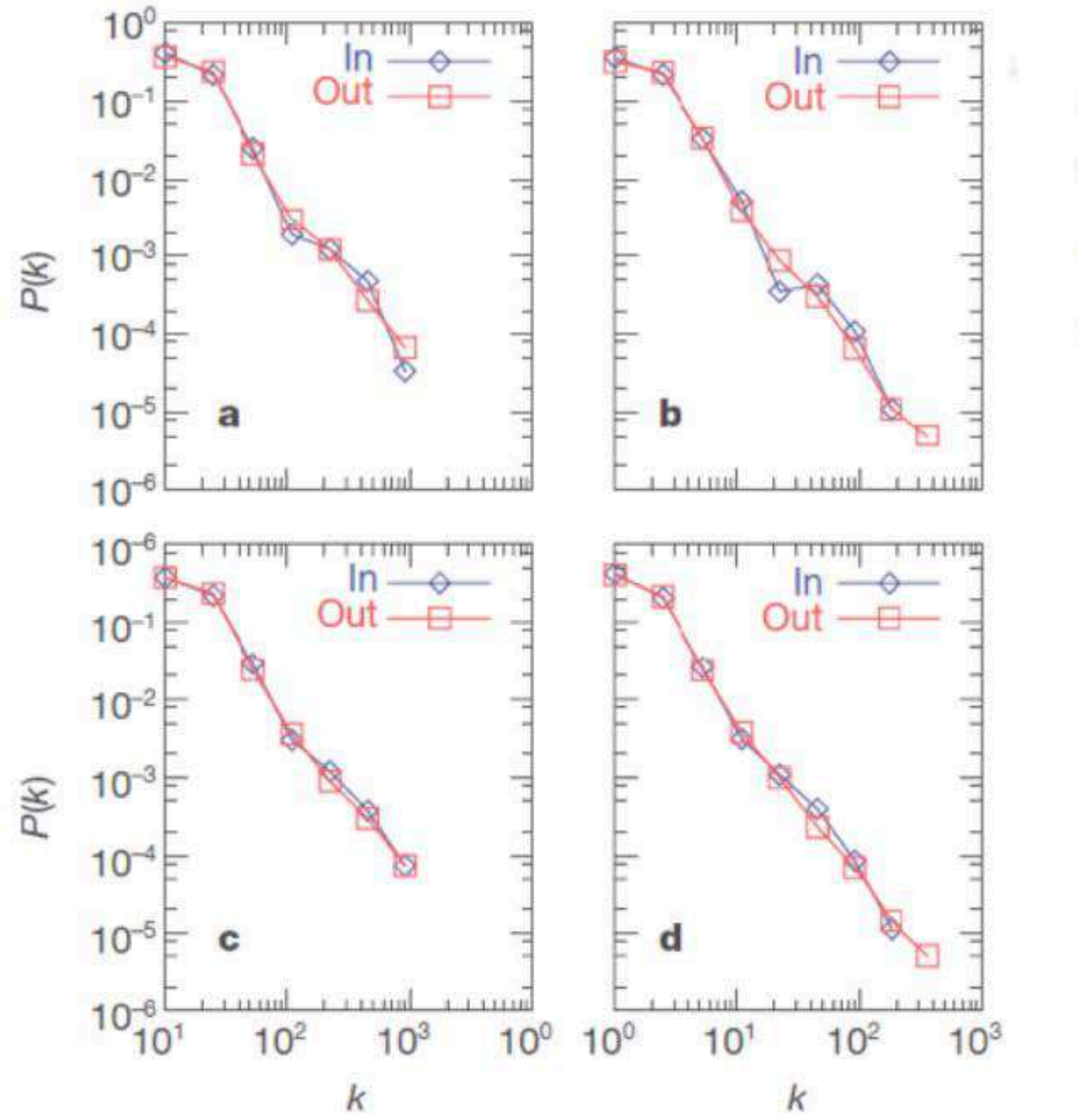
- Small world: the average shortest path length in a real network is small
- Six degrees of separation (Milgram, 1967)
- Local neighborhood + long-range friends
- A random graph is a small world

Model proposed

- Crossover from regular lattices to random graphs
- Tunable
- Small world network with (simultaneously):
  - Small average shortest path
  - Large clustering coefficient (not obeyed by RG)


a
C(p)/C(0)
L(p)/L(0)
1
0.8
0.6
0.4
0.2
0
10 -4
10 -2
10 0
Probability, p

3 Barabasi-Albert: scale-free network



JEONG ET AL. NATURE 407, 651 (2000)

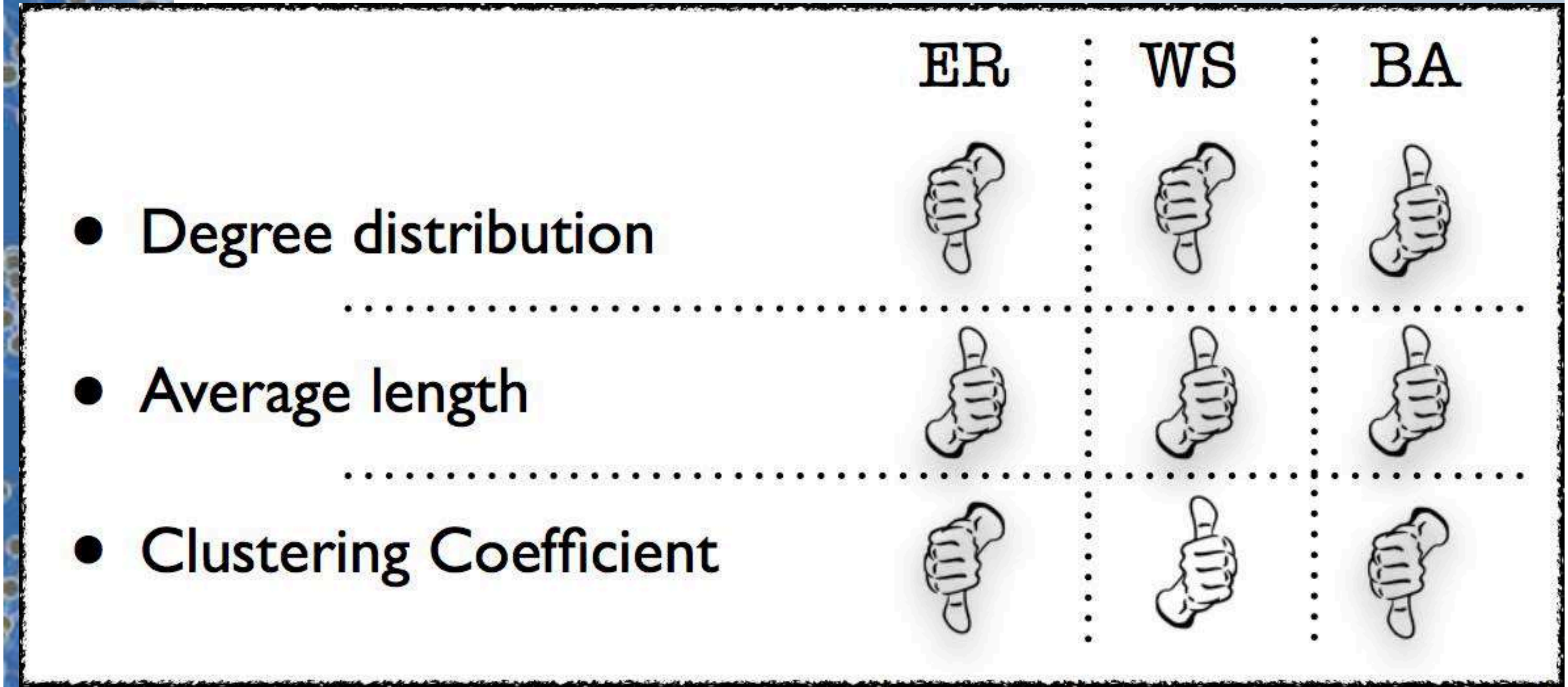
- Many large networks are scale free
- The degree distribution has a power-law behavior for large k (far from a Poisson distribution)
- Random graph theory and the Watts-Strogatz model cannot reproduce this feature

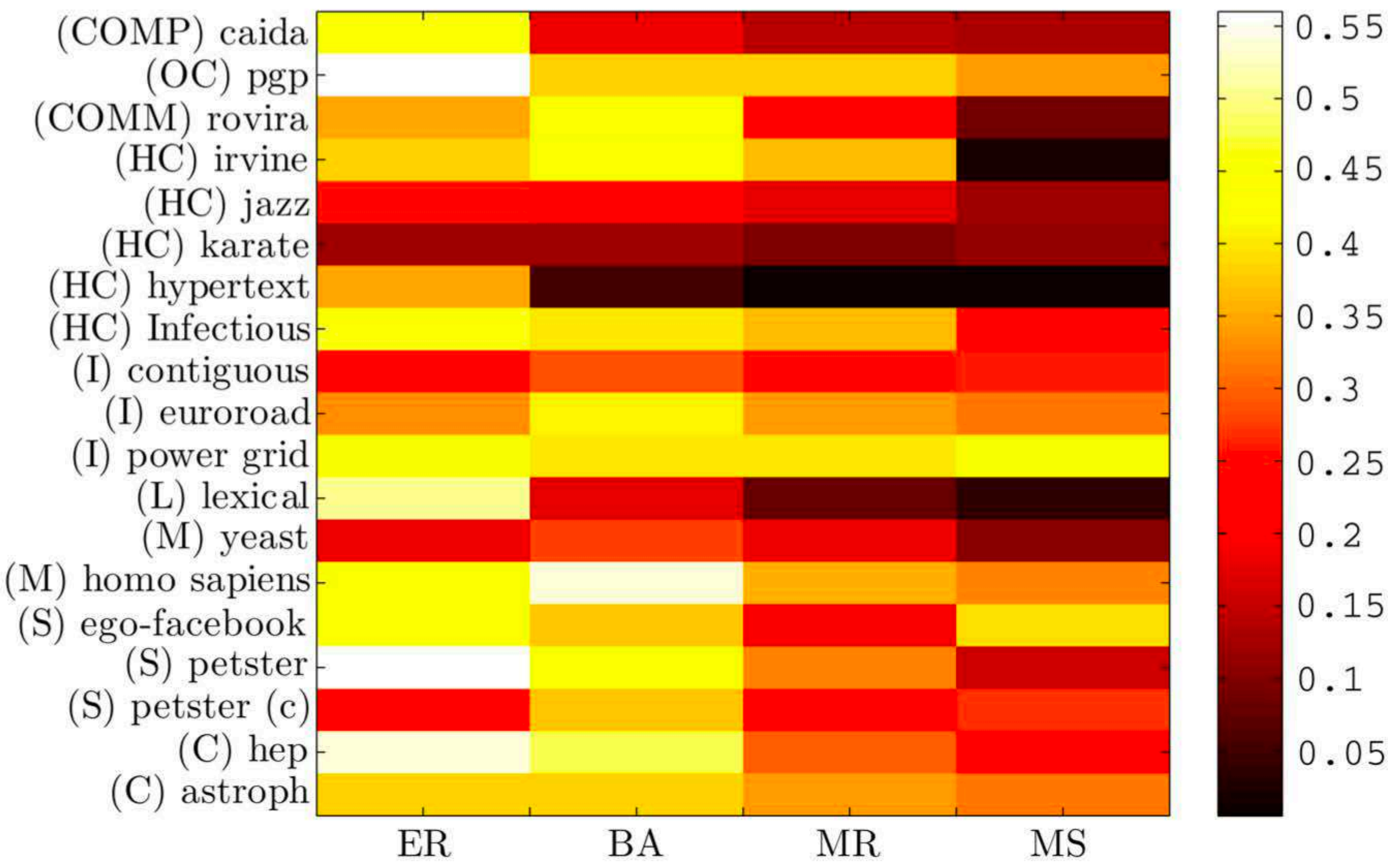
Preferential attachment: rich gets richer

- When choosing the nodes to which the new connects, the probability Π that a new node will be connected to node *i* depends on the degree k_i of node *i*

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

Linear attachment (more general models)
Sum over all existing nodes


ER
WS
BA
• Degree distribution
• Average length
• Clustering Coefficient

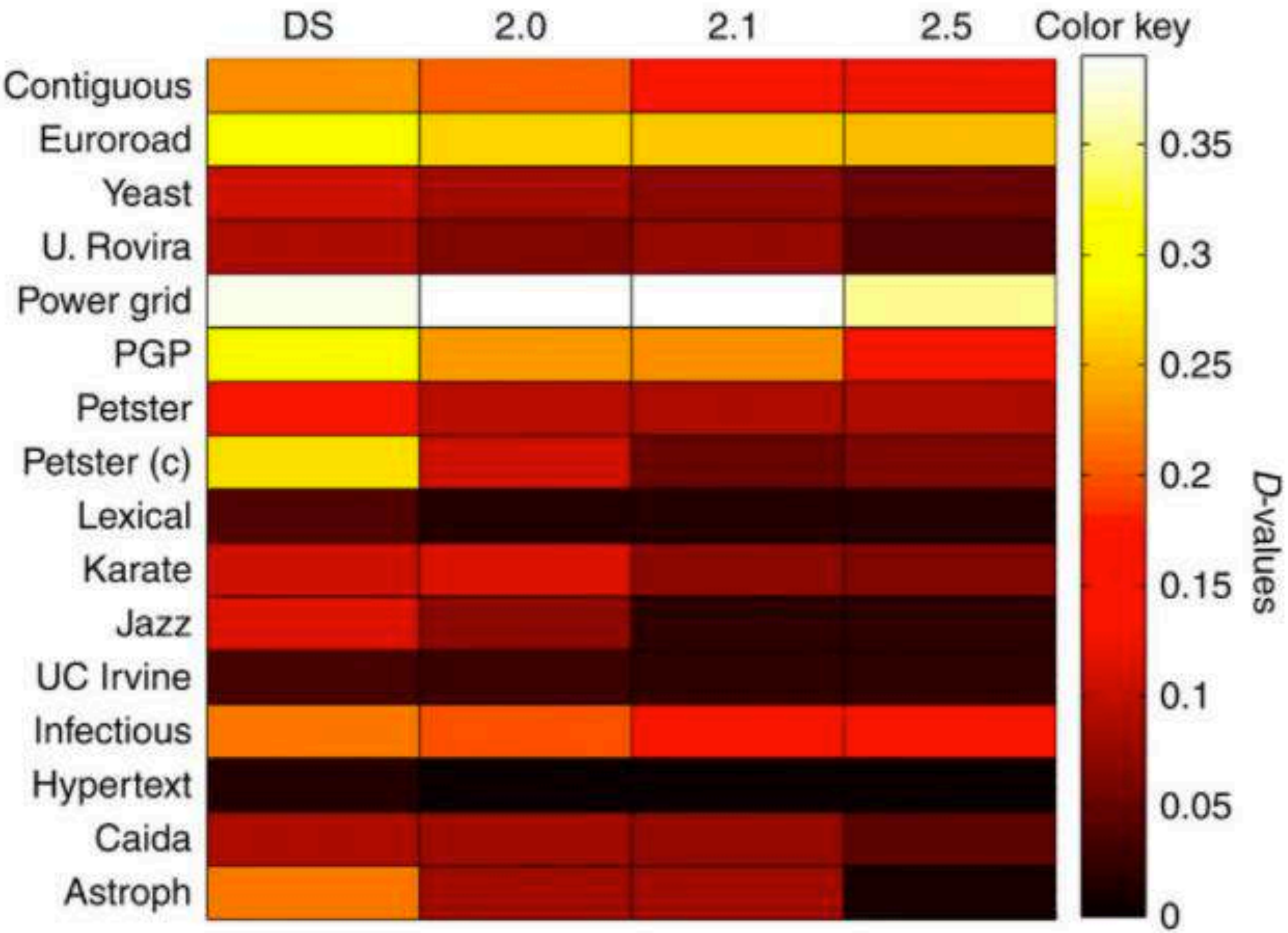

(COMP) caida
(OC) pgp
(COMM) rovira
(HC) irvine
(HC) jazz
(HC) karate
(HC) hypertext
(HC) Infectious
(I) contiguous
(I) euroroad
(I) power grid
(L) lexical
(M) yeast
(M) homo sapiens
(S) ego-facebook
(S) petster
(S) petster (c)
(C) hep
(C) astroph
ER
BA
MR
MS
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05

Random Network Generator

RandNetGen

Random Network Generator

- http://polcolomer.github.io/RandNetGen/



Dissimilarity values between real-world networks and four null models. From left to right we report averaged results after 30 independent runs; DS, degree sequence (MS, MR and k=1.0), generates equivalent results and the last three columns are obtained using the dk model for different k-values (2.0, 2.1 and 2.5).

Why dynamics?

- Nodes are entities that have dynamical properties
- Links represent the interaction patterns

nature Vol 457 | 19 February 2009 | doi:10.1038/nature07634

LETTERS

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg[1], Matthew H. Mohebbi[1], Rajan S. Patel[1], Lynnette Brammer[2], Mark S. Smolinski[1] & Larry Brilliant[1]

Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year[1]. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities[2]. Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza[3,4]. One way to improve early detection is to monitor health-seeking behaviour in the form of queries to online search engines, which are submitted by millions of users around the world each day. Here we present a method of analysing large numbers of Google search queries to track influenza-like illness in a population. Because the relative frequency of certain queries is highly correlated with the percentage of physician visits in which a patient presents with influenza-like symptoms, we can accurately estimate the current level of weekly influenza activity in each region of the United States, with a reporting lag of about one day. This approach may make it possible to use search queries to detect influenza epidemics in areas with a large population of web search users.

By aggregating historical logs of online web search queries submitted between 2003 and 2008, we computed a time series of weekly counts for 50 million of the most common search queries in the United States. Separate aggregate weekly counts were kept for every query in each state. No information about the identity of any user was retained. Each time series was normalized by dividing the count for each query in a particular week by the total number of online search queries submitted in that location during the week, resulting in a query fraction (Supplementary Fig. 1).

We sought to develop a simple model that estimates the probability that a random physician visit in a particular region is related to an ILI; this is equivalent to the percentage of ILI-related physician visits. A single explanatory variable was used: the probability that a random search query submitted from the same region is ILI-related, as determined by an automated method described below. We fit a linear model using the log-odds of an ILI physician visit and the log-odds of an ILI-related search query: $\text{logit}(I(t)) = \alpha\,\text{logit}(Q(t)) + \varepsilon$, where $I(t)$ is the percentage of ILI physician visits, $Q(t)$ is the ILI-related query fraction at time t, α is the multiplicative coefficient, and ε is the error term. $\text{logit}(p)$ is simply $\ln(p/(1-p))$.

Publicly available historical data from the CDC's US Influenza era

La 20th Century Fox se disculpa por propagar noticias falsas sobre Putin y Trump

Publicado: 18 feb 2017 00:55 GMT | Última actualización: 18 feb 2017 09:06 GMT

2.5K G+ 244

En el marco de una campaña publicitaria, la compañía creó portales de noticias ficticios que supuestamente ofrecían información de actualidad. Algunas notas involucraban incluso a personalidades mundiales.

Efecto Trump: Zuckerberg hará esto para evitar propagación de noticias falsas en Facebook

En vista a las fuertes críticas que sufrió Facebook después de saberse que las noticias falsas fueron más comunes que las reales en la red social. El CEO de la misma, Mark Zuckerberg ha anunciado medidas para contrarrestar este grave problema.

Fundamental Question

?

@gomezgardenes

Main dynamical processes

- Cascades (Failures and Attacks)
- Simple diffusion & random walks
- Synchronization
- Contagion processes
- Evolutionary games
- Chaotic dynamics
- ………….

Spreading

Phenomena	Agent	Network
Venereal Disease	Pathogens	Sexual Network
Rumor Spreading	Information, Memes	Communication Network
Diffusion of Innovations	Ideas, Knowledge	Communication Network
Computer Viruses	Malwares, Digital viruses	Internet
Mobile Phone Virus	Mobile Viruses	Social Network/Proximity Network
Bedbugs	Parasitic Insects	Hotel - Traveler Network
Malaria	Plasmodium	Mosquito - Human network

Brockman Lab

Research on Complex Systems

Brockmann Lab

Home Research Projects Publications Contact Interactive Tea

2014 Ebola Outbreak: Worldwide Air-Transportation, Relative Import Risk and Most Probable Spreading Routes

An interactive network analysis

August 4th, 2014

Dirk Brockmann[1,2,3,*], Lars Schaade[1], Luzie Verbeek[1]

Social models: agent based models (ABM)

- Voter model
- Opinion dynamics
- Language competition
- Cultural dynamics
- Naming game
- Evolutionary games

Big data => more complex networks

- New paradigms:
  - Networks of networks
  - Time dependent networks
  - Interconnected networks
  - Networks in multiple layers (multiplex)

Interconnected networks

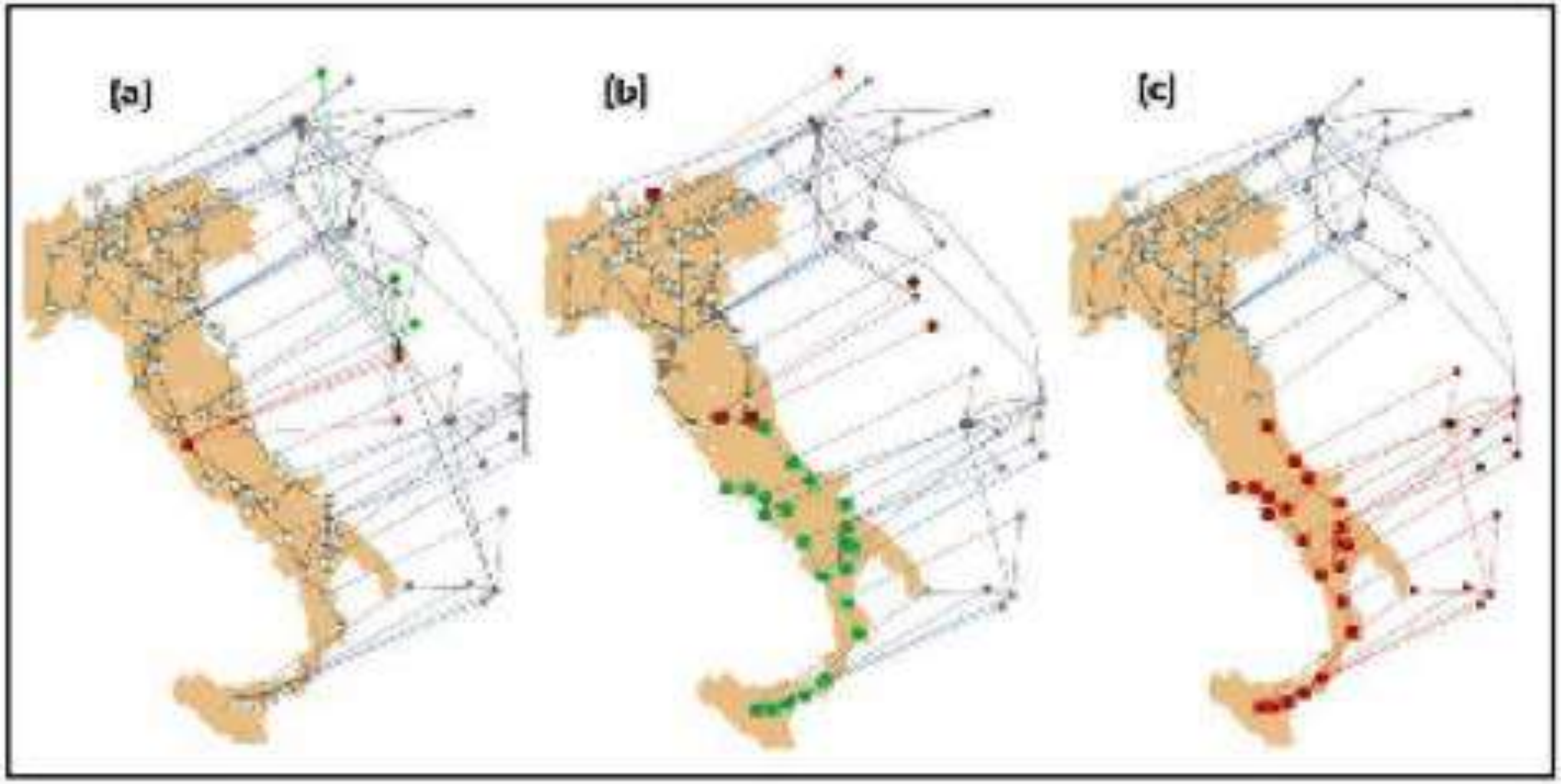
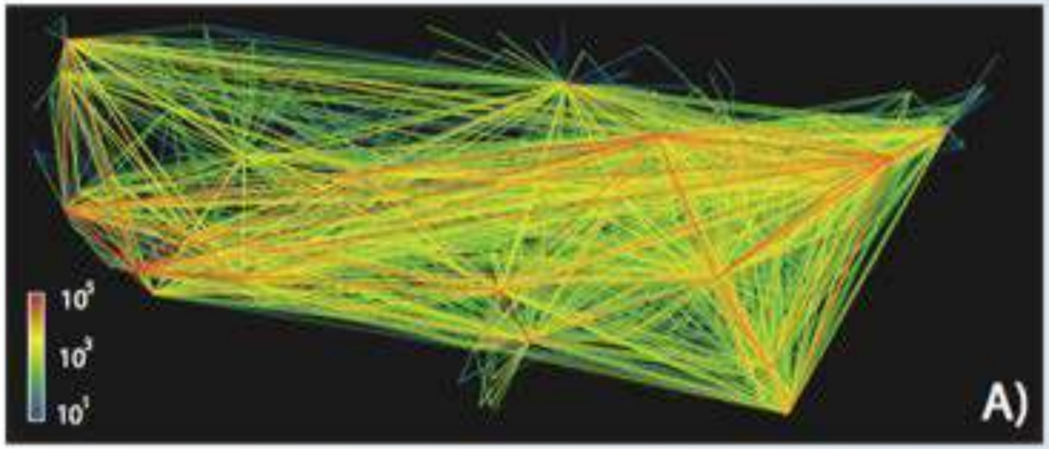
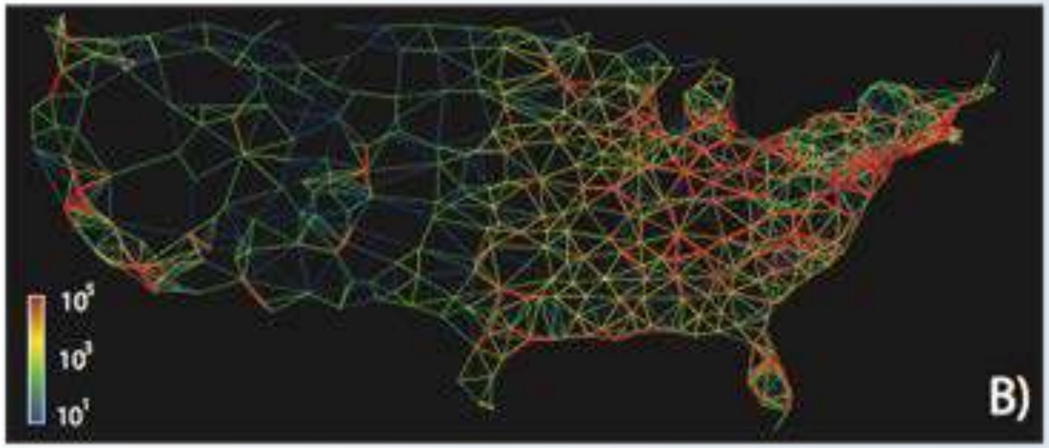


Fig. 2. – Cartoon of a typical cascade obtained by implementing the described model on the rea coupled system in Italy. Over the map is the network of the Italian power network and, slightl shifted to the top, is the communication network. Every server was considered to be connecte to the geographically nearest power station. (After Buldyrev *et al.* [15])

Networks of networks



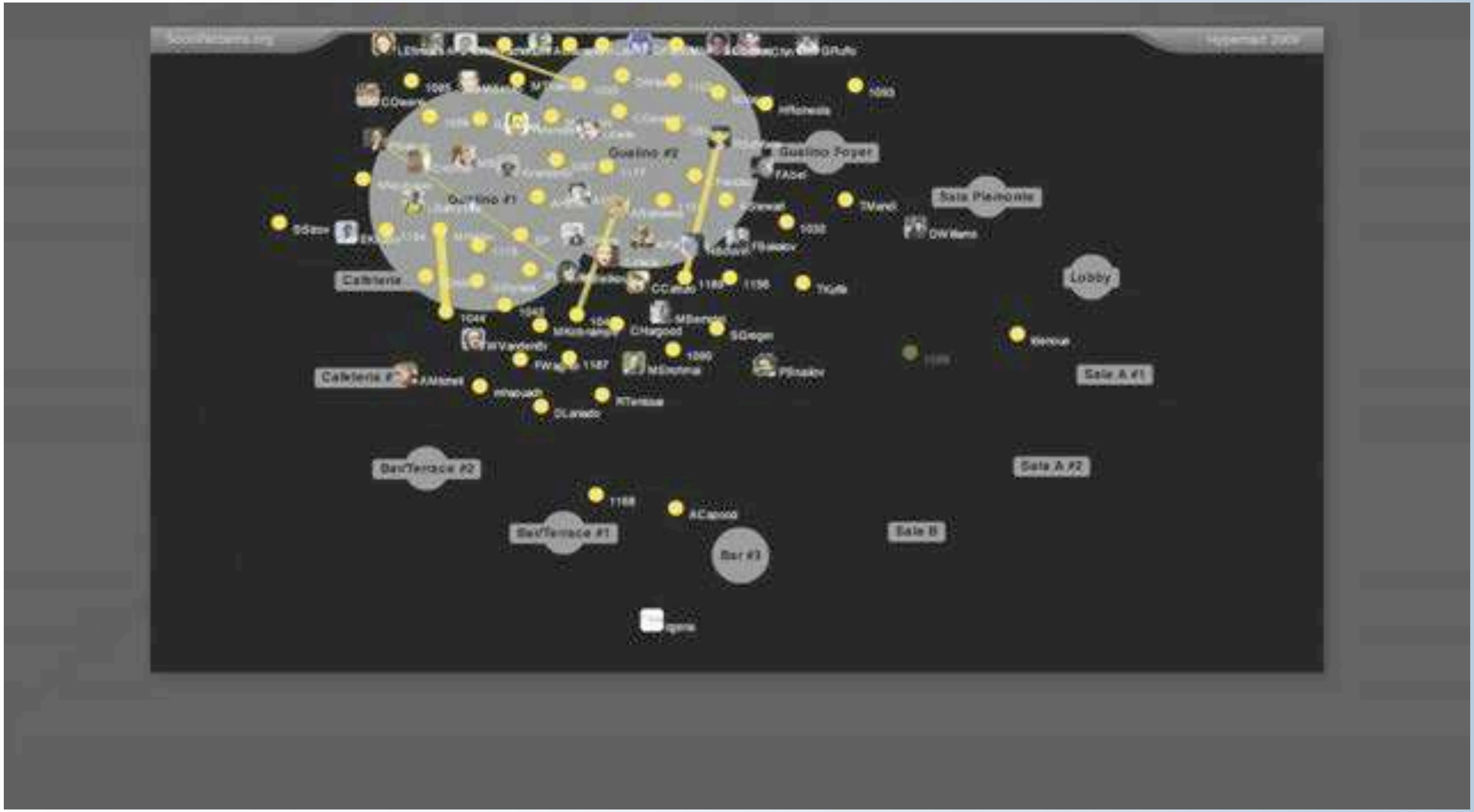
airline transportation network



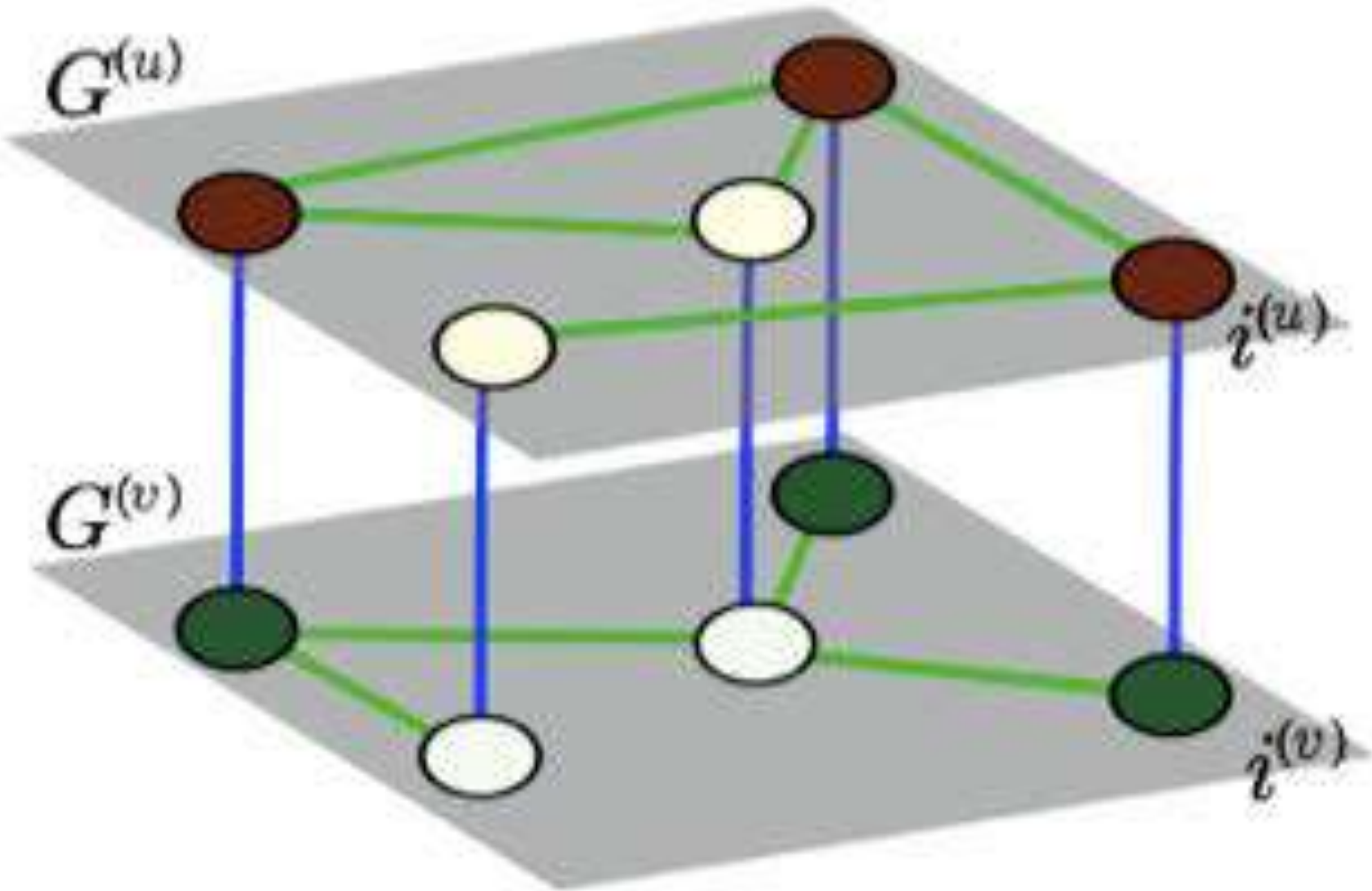
commuting network

Balcan et al., PNAS 196, 21484 (2009)

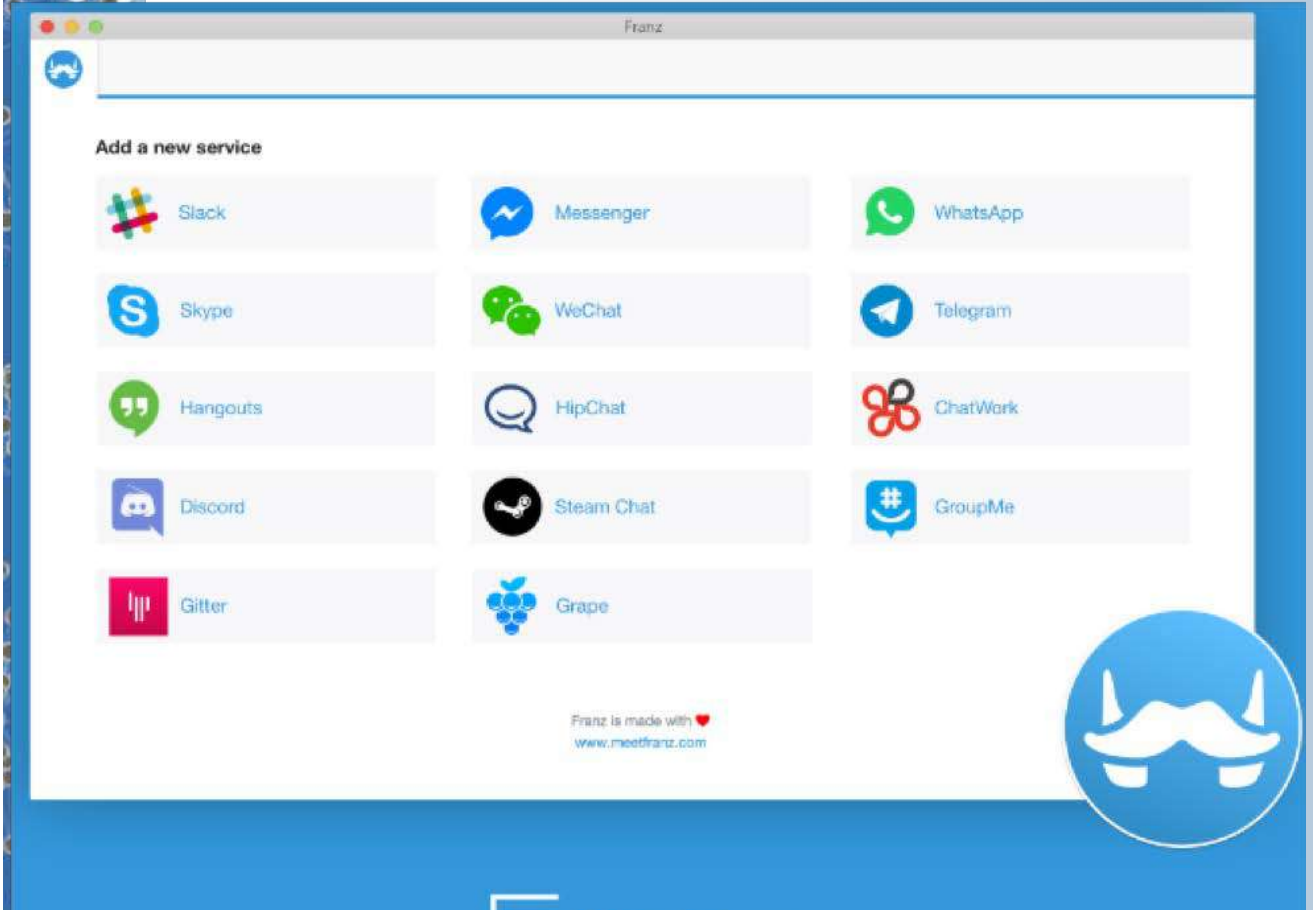
Contact networks



Multiplex



Big big networks


Franz
Add a new service
Slack
Messenger
WhatsApp
Skype
WeChat
Telegram
Hangouts
HipChat
ChatWork
Discord
Steam Chat
GroupMe
Gitter
Grape
Franz is made with
www.meetfranz.com

KONECT: The Koblenz Network Connection

- http://konect.uni-koblenz.de/

The Koblenz Network Collection

KONECT (the Koblenz Network Collection) is a project to collect large network datasets of all types in order to perform research in network science and related fields, collected by the Institute of Web Science and Technologies at the University of Koblenz–Landau.

KONECT contains several hundred network datasets of various types, including directed, undirected, bipartite, weighted, unweighted, signed and rating networks. The networks of KONECT cover many diverse areas such as social networks, hyperlink networks, authorship networks, physical networks, interaction networks, and communication networks. The KONECT project has developed free software network analysis tools which are used to compute network statistics, to draw plots and to implement various link prediction algorithms. The result of these analyses are presented on these pages. Whenever we are allowed to do so, we provide a download of the networks.

KONECT currently holds 261 networks, of which

- 63 are undirected,
- 107 are directed,
- 91 are bipartite,
- 125 are unweighted,
- 90 allow multiple edges,
- 6 have signed edges,
- 10 have ratings as edges,
- 3 allow multiple weighted edges,
- 18 allow positive weighted edges,
- and 89 have edge arrival times.

The Colorado Index of Complex Networks

- https://icon.colorado.edu/#!/

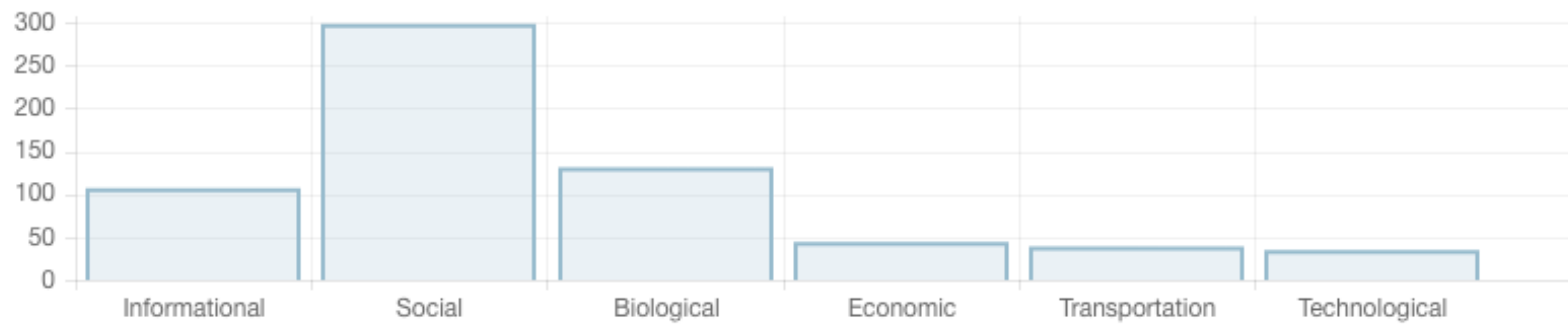
The Colorado Index of Complex Networks (ICON)

ICON is a comprehensive index of research-quality network data sets from all domains of network science, including social, web, information, biological, ecological, connectome, transportation, and technological networks.

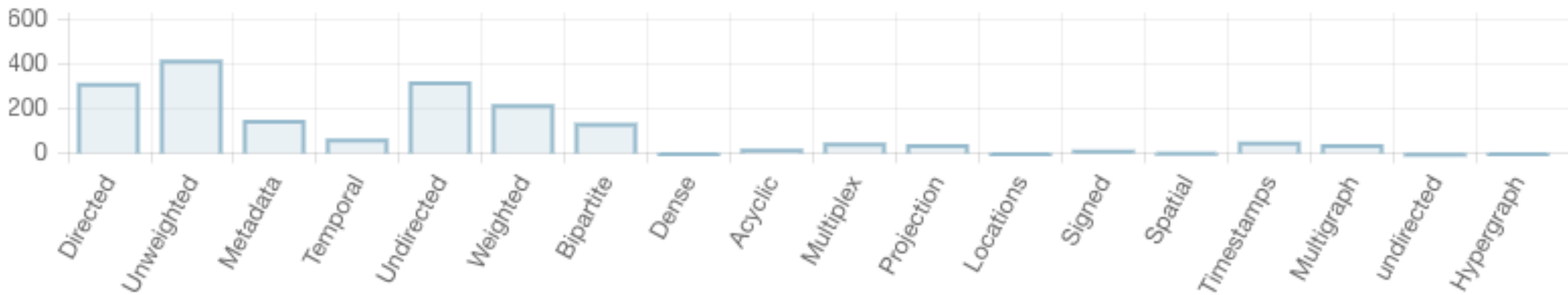
Each network record in the index is annotated with and searchable or browsable by its graph properties, description, size, etc., and many records include links to multiple networks. The contents of ICON are curated by volunteer experts from Prof. Aaron Clauset's research group at the University of Colorado Boulder.

Click on the NETWORKS tab above to get started.

Entries found: 654 Networks found: 5318



Property Counts



OCT
3
2015

¿Por qué me vigilan, si no soy nadie? – Marta Peirano

10 Votos

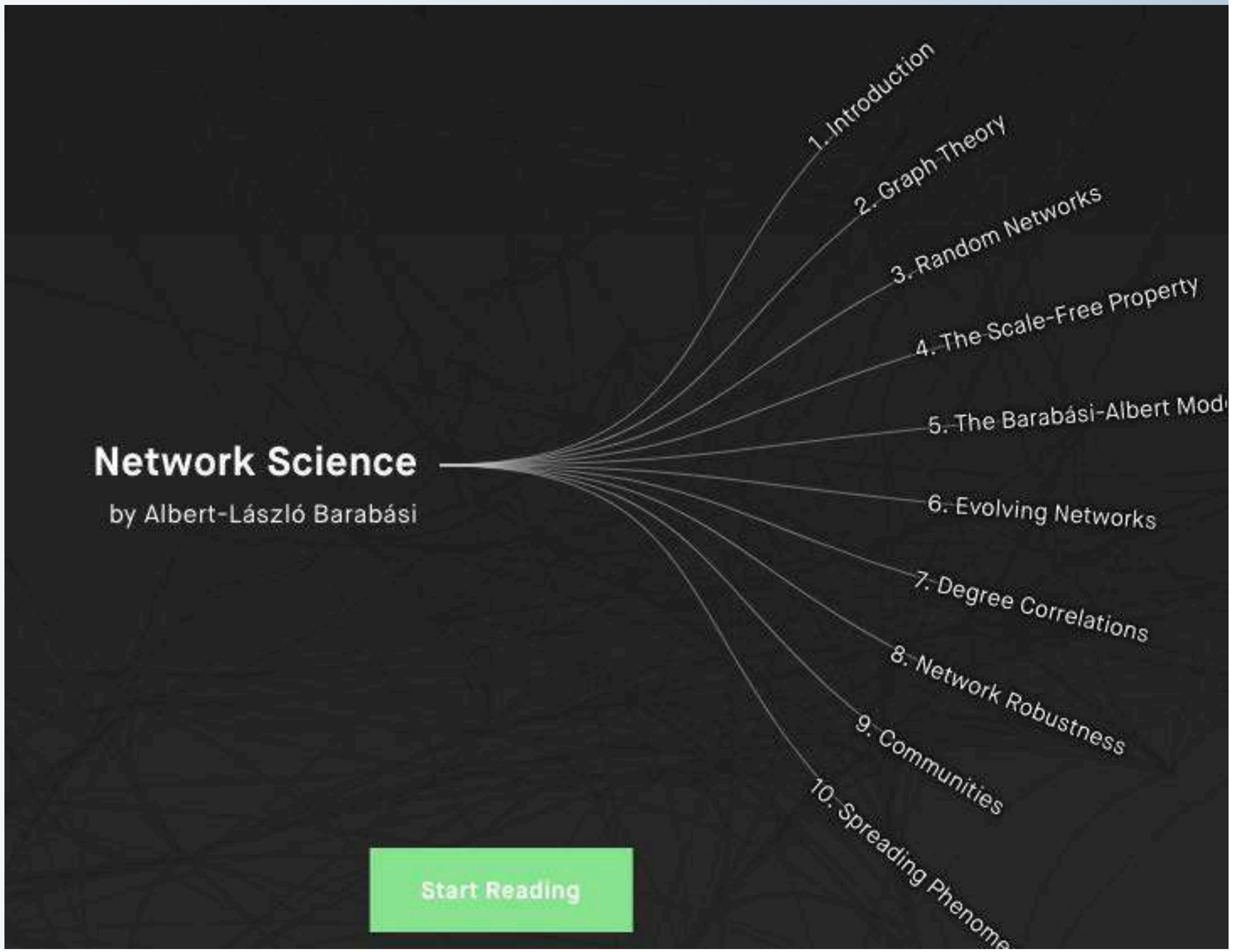
Por TEDx Talks

Books

2003 2008 2010 2011 2016

Barabasi book

- http://barabasi.com/networksciencebook/



https://textbooks.opensuny.org/introduction-to-the-modeling-and-analysis-of-complex-systems/

Introduction to the Modeling and Analysis of Complex Systems

License: Attribution-NonCommercial-ShareAlike CC BY-NC-SA

Author(s): Hiroki Sayama

Keep up to date on *Introduction to Modeling and Analysis of Complex Systems* at http://bingweb.binghamton.edu/~sayama/textbook/!

Introduction to the Modeling and Analysis of Complex Systems introduces students to mathematical/computational modeling and analysis developed in the emerging interdisciplinary field of Complex Systems Science. Complex systems are systems made of a large number of microscopic components interacting with each other in nontrivial ways. Many real-world systems can be understood as complex systems, where critically important information resides in the relationships between the parts and not necessarily within the parts themselves. This textbook offers an accessible yet technically-oriented introduction to the modeling and analysis of complex systems. The topics covered include: fundamentals of modeling, basics of dynamical systems, discrete-time models, continuous-time models, bifurcations, chaos, cellular automata, continuous field models, static networks, dynamic networks, and agent-based models. Most of these topics are discussed in two chapters, one focusing on computational modeling and the other on mathematical analysis. This unique approach provides a comprehensive view of related concepts and techniques, and allows readers and instructors to flexibly choose relevant materials based on their objectives and needs. Python sample codes are provided for each modeling example.

This textbook is available for purchase in both grayscale and color via Amazon.com and CreateSpace.com.

REVIEWS:

> *Hiroki Sayama's book "Introduction to the Modeling and Simulation of Complex Systems" is … a unique and welcome addition to any instructor's collection. What makes it valuable is that it not only presents a state-of-the-art review of the domain but also serves as a gentle guide to learning the sophisticated art of modeling complex systems. –Muaz A. Niazi, Complex Adaptive Systems Modeling 2016 4:3*

PDF DOWNLOAD
19 MB

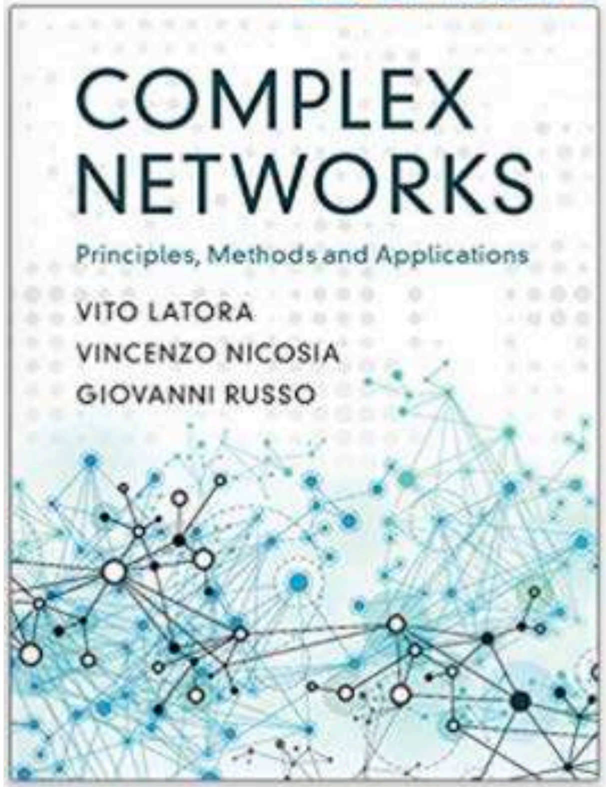
42604
Downloads from Open SUNY Textbooks

AVAILABLE NOW AT amazon.com

Publication Date: August 13, 2015

2017: Latora, Nicosia, Russo

Echa un vistazo



Complex Networks: Principles, Methods and Applications

(Inglés) Tapa dura – 28 sep 2017

de Vito Latora (Autor), Vincenzo Nicosia (Autor), Giovanni Russo (Autor)

Sé el primero en opinar sobre este producto

Ver los 2 formatos y ediciones

Versión Kindle	Tapa dura
EUR 37,01	EUR 66,75 prime
Leer con nuestra App gratuita	8 Nuevo desde EUR 62,08

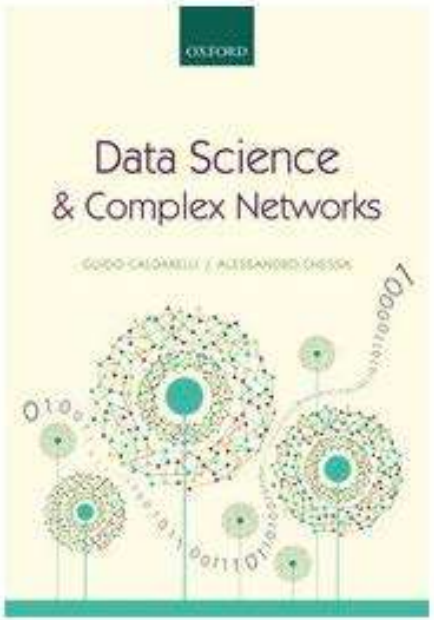
¿Quieres recibirlo el viernes 16 feb.? Cómpralo antes de 3 hrs y 33 mins y elige **Envío 1 día** al completar tu pedido. Ver detalles

Nota: Otros vendedores ofrecen este producto a un precio inferior, potencialmente sin las ventajas de Amazon Prime.

Networks constitute the backbone of complex systems, from the human brain to computer

Caldarelli & Chessa

- Books
  - foundations: any of the books shown before
  - applications: http://book.complexnetworks.net/



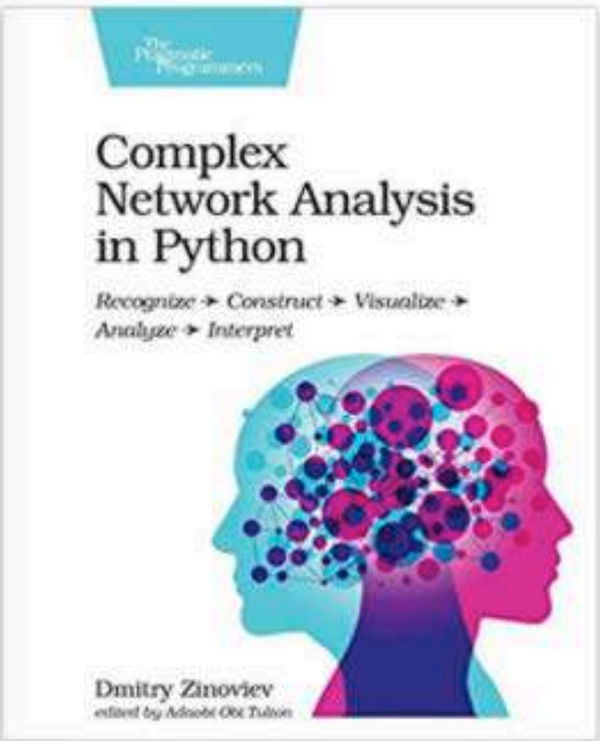
Data Science and Complex Networks

Real Case Studies with Python

Guido Caldarelli and **Alessandro Chessa**

- Clearly presents the theoretical concepts
- Exposition is based on Data
- Every concept is shown with code (Python)
- Dedicated companion website for download of code, data, and platform to test personal progress

New new new (2018)



Complex Network Analysis in Python (Inglés) Tapa blanda – 29 ene 2018
de Dmitry Zinoviev (Autor)
Sé el primero en opinar sobre este producto

Ver los formatos y ediciones

Tapa blanda
EUR 34,60

3 Nuevo desde EUR 34,01

Recíbelo entre el 27 feb. - 5 mar. al elegir **Entrega estándar** durante la tramitación del pedido. Ver detalles

Nota: Este producto no disfruta de las ventajas de Amazon Prime (más información). Ofertas con envío gratuito de Amazon Prime disponibles en otras opciones de compra.

PDF book available:
https://pragprog.com/book/dzcnapy/complex-network-analysis-in-python
Source code: https://pragprog.com/titles/dzcnapy/source_code

Science AAAS

Home News Journals Topics Careers

Science Science Advances Science Immunology Science Robotics Science Signaling Science Translational Medici

PERSPECTIVE | NETWORK ANALYSIS

Network analytics in the age of big data

SHARE

Nataša Pržulj, Noël Malod-Dognin

Department of Computer Science, University College London, London, UK.

Email: natasa(at)cs.ucl.ac.uk

+ See all authors and affiliations

Science 08 Jul 2016:
Vol. 353, Issue 6295, pp. 123-124
DOI: 10.1126/science.aah3449

ELSEVIER

Physics Reports

Volume 635, 27 May 2016, Pages 1–44

Combining complex networks and data mining: Why and how

Combining complex networks and data mining: Why and how

M. Zanin[a, b], D. Papo[c], P.A. Sousa[b], E. Menasalvas[c], A. Nicchi[d], E. Kubik[e], S. Boccaletti[f]

Show more

http://dx.doi.org/10.1016/j.physrep.2016.04.005

Get rights and content

“Take home message”

- Big Data <=> Big Networks
- Importance of networks
  - Methodologically
  - Application
- Privacy & anonimicity
- We are the data