Documents are here:

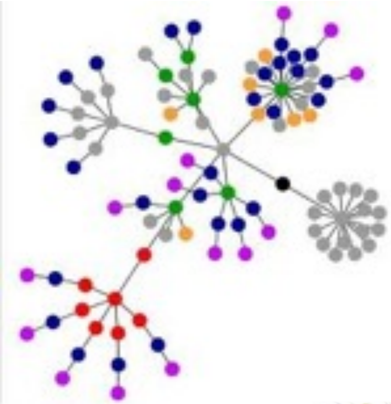
GRAPH THEORY [9]

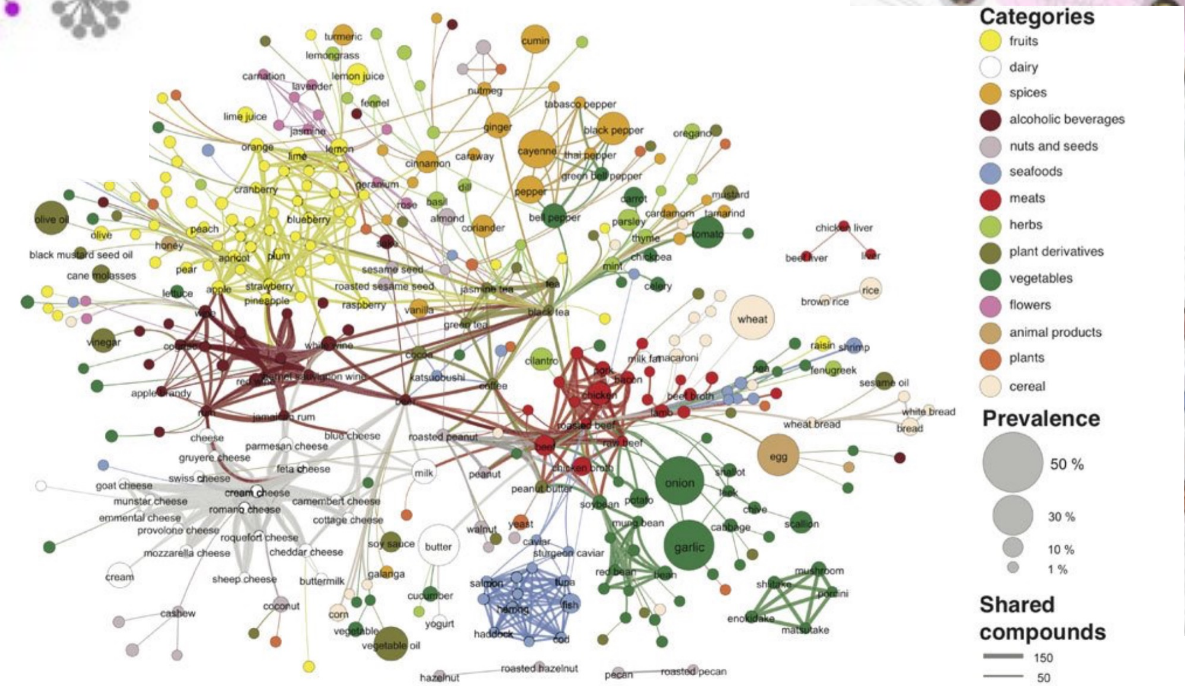
Complex Networks Analysis

https://www-l2ti.univ-paris13.fr/~viennet/ens/2024-USTH-Graphs

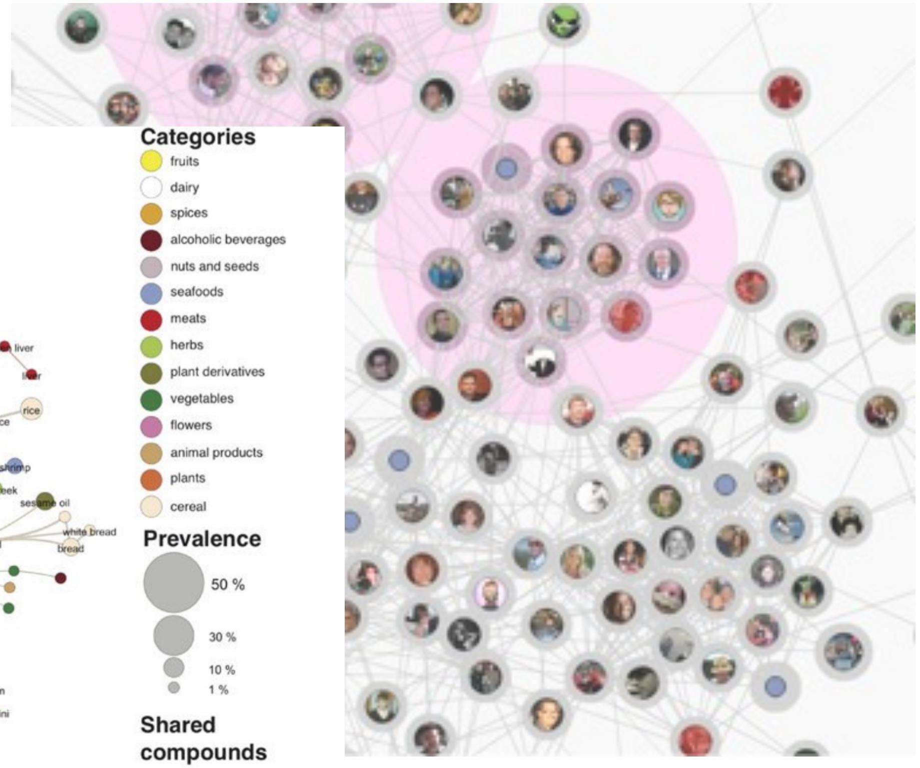
Emmanuel Viennet
emmanuel.viennet@univ-paris13.fr

Complex Network





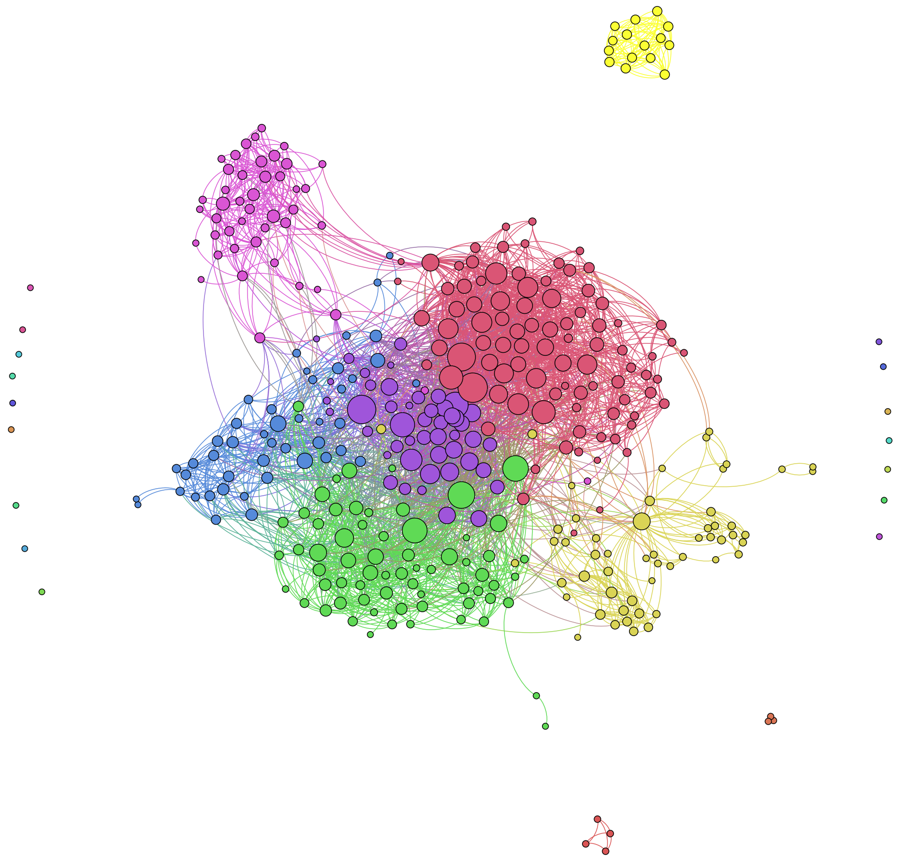
A flavor network that captures the flavor compounds shared by culinary ingredients. Each node denotes an ingredient, the node color indicates food category, and node size reflects the ingredient prevalence in recipes. Two ingredients are connected if they share a significant number of flavor compounds, link thickness representing the number of shared compounds between the two ingredients. (Barabasi et al 2012)



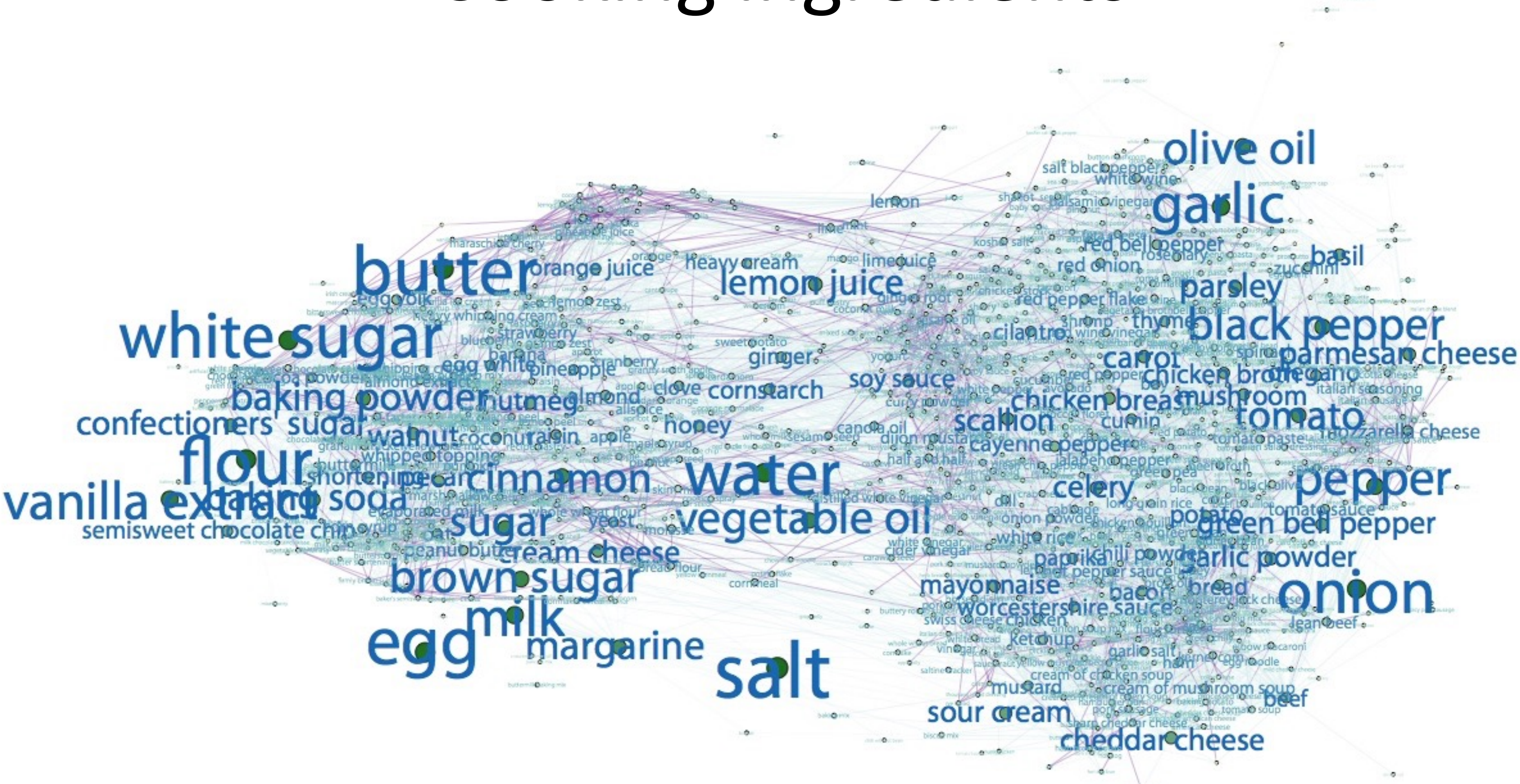
A lot of real world phenomena can be modeled as *complex networks*


Facebook graph (2010)

Social Network (facebook, zoom)

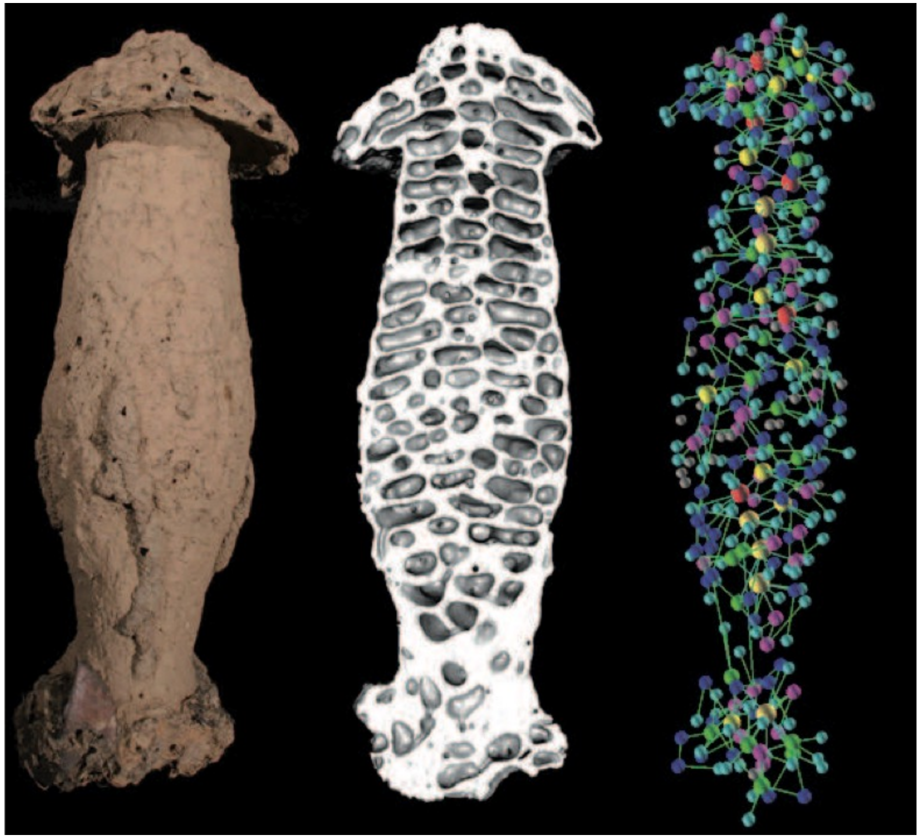


Cooking ingredients

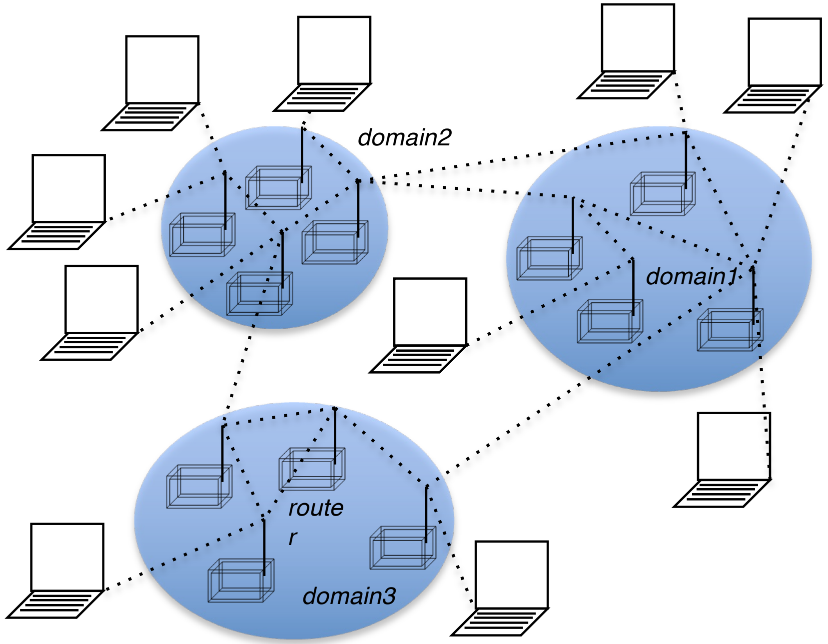


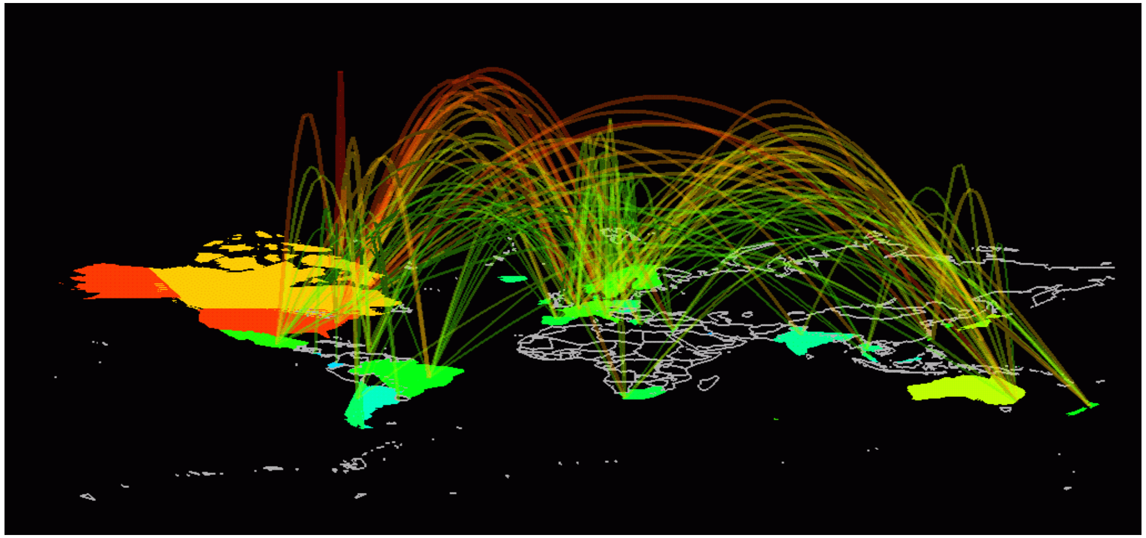
Termite mounds

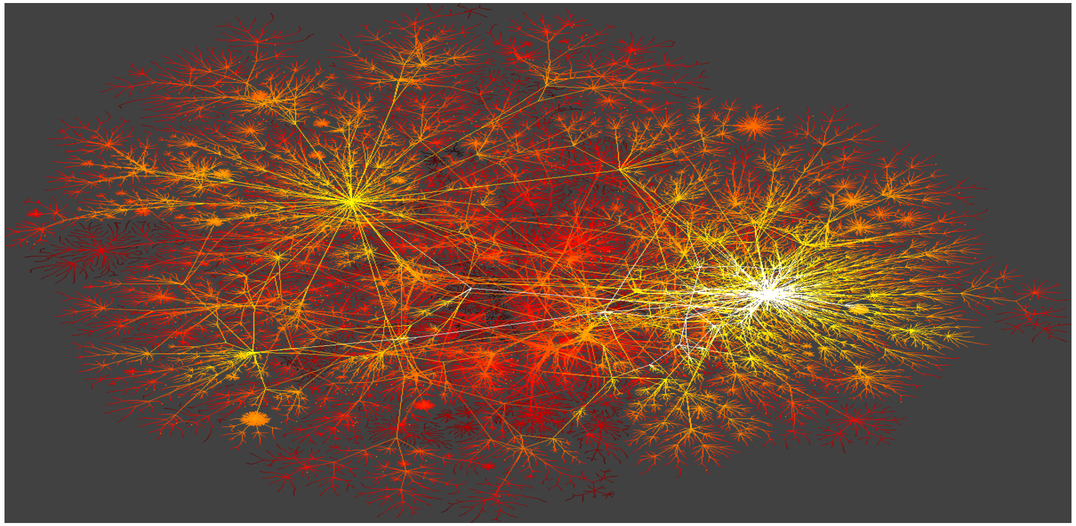
The galleries of a termite mound form a complex graph



Internet Computer Network

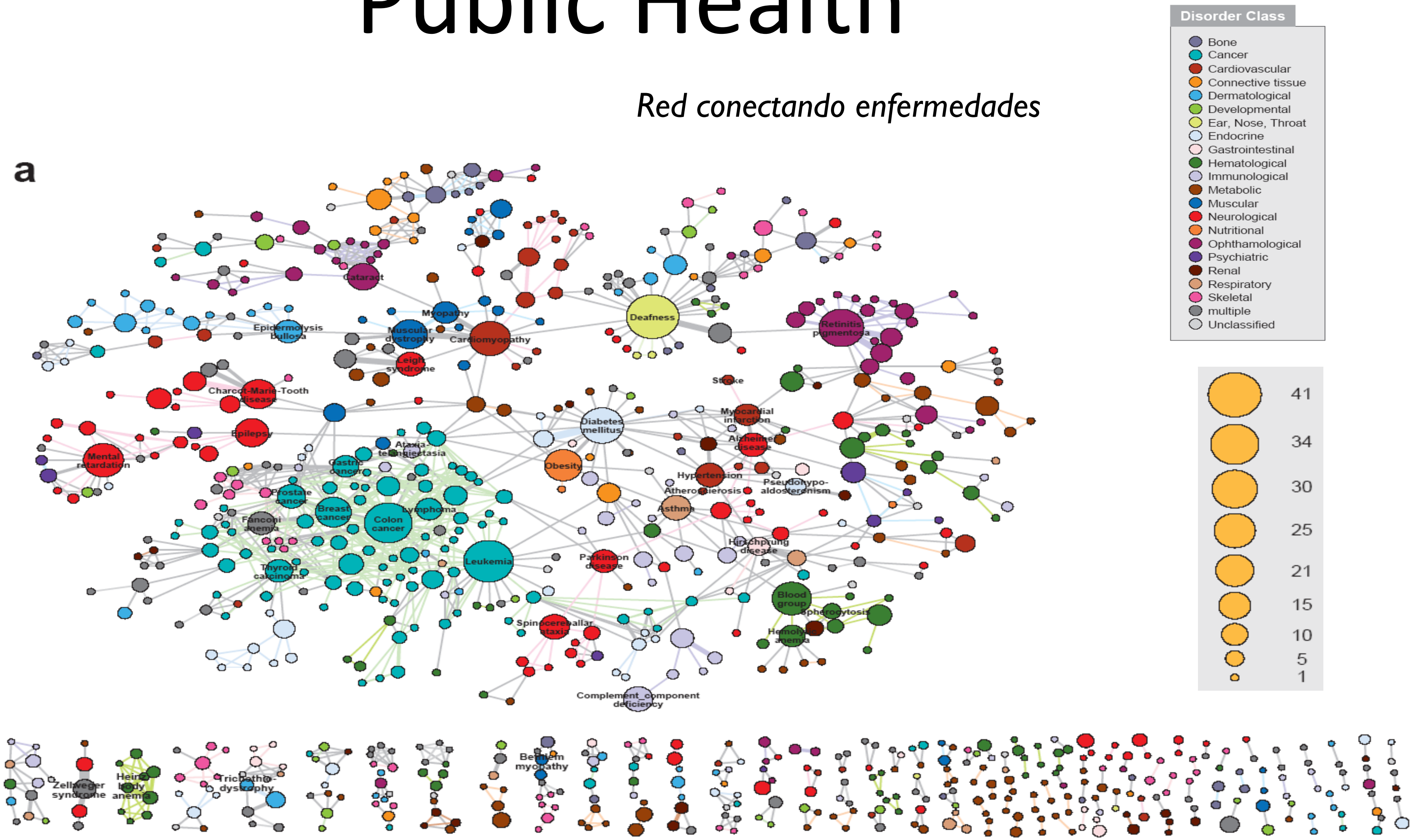




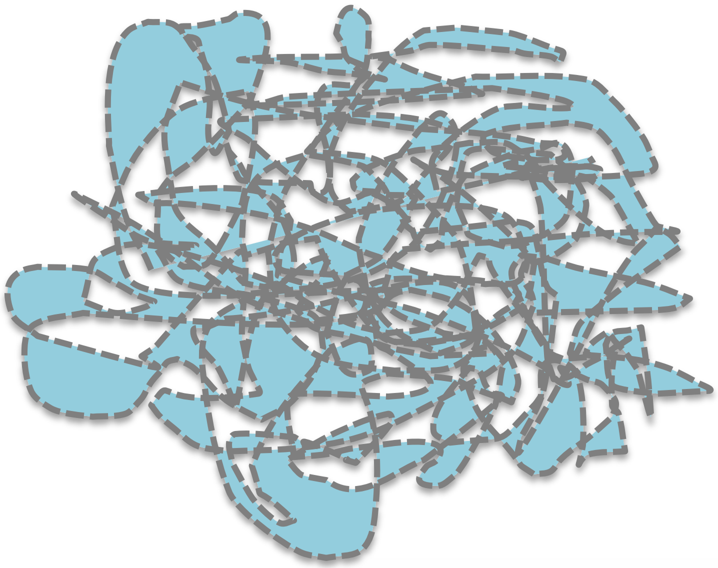


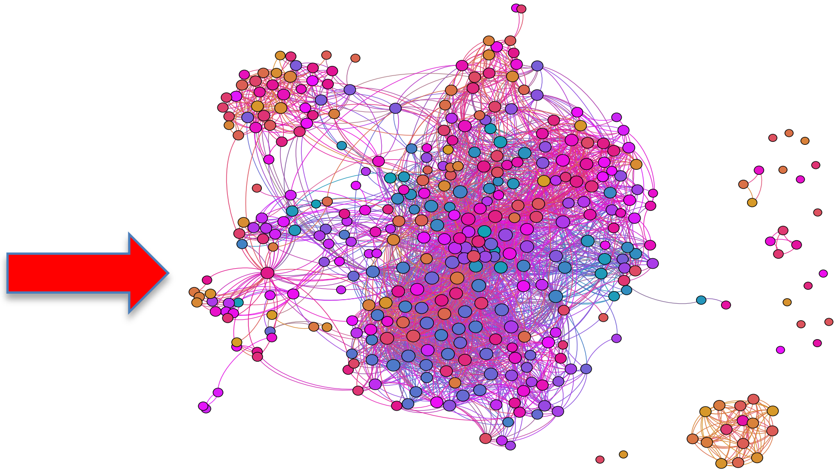
Public Health

Red conectando enfermedades



Graph to model complex systems





Credit: Lada Adamic

History of Complex Networks

- Graph Theory: 1735, Euler

- Social Networks (sociology): 1930… (Moreno)
- Communication Networks, Internet 1960…
- Ecological Networks : 1979

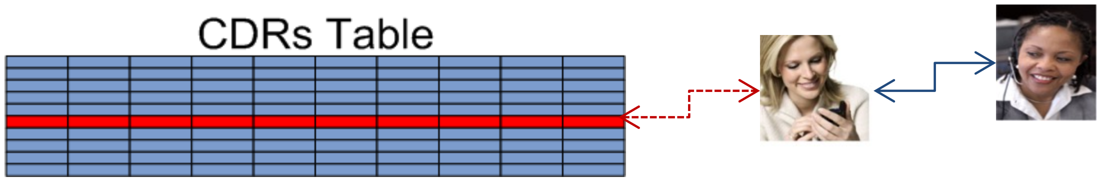
- Web: 1990s (Barabasi, *scale free graphs*…)

- Social Web (Web2.0): 2000s
  - Data mining, processing data from huge graphs

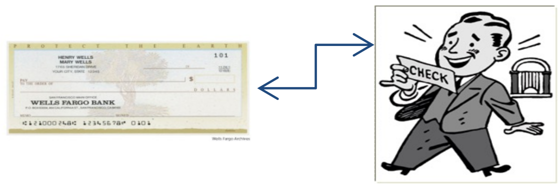
Applications
of
Social Network Analysis

Networked data in the industry

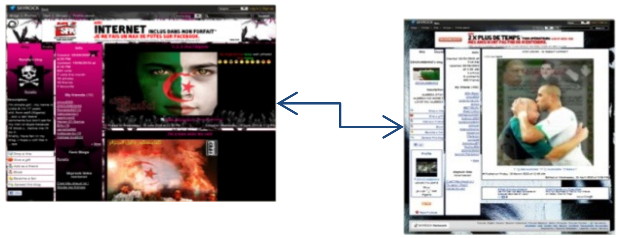
- Telecoms
  - call data



- Banks
  - Tranfer (checks, money transfer,...)
  - Credit card transactions



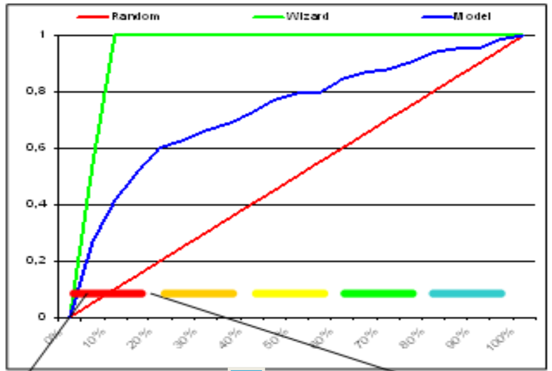
- Social apps, blogs
  - friends, followers
  - Posts and comments...

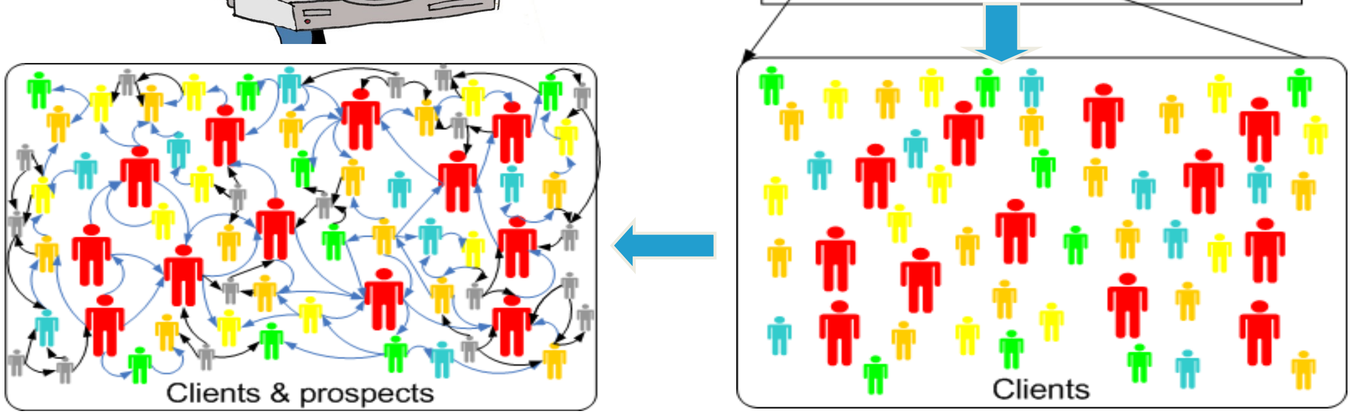


- Distribution
  - Customers buying the products

Example 1: application to marketing

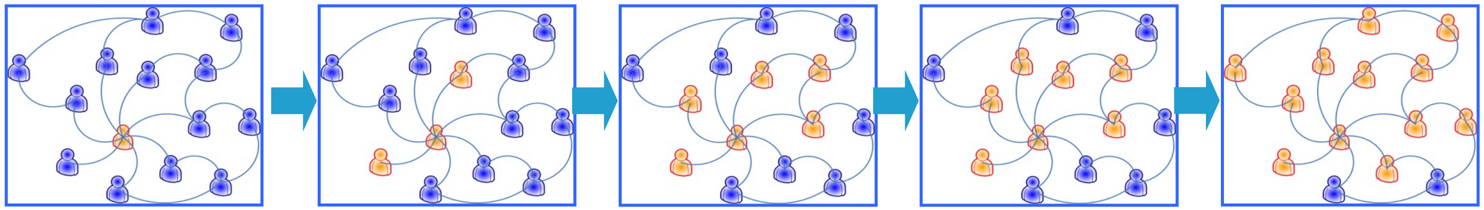
- Direct marketing actions to some customers
  - target is defined by a model



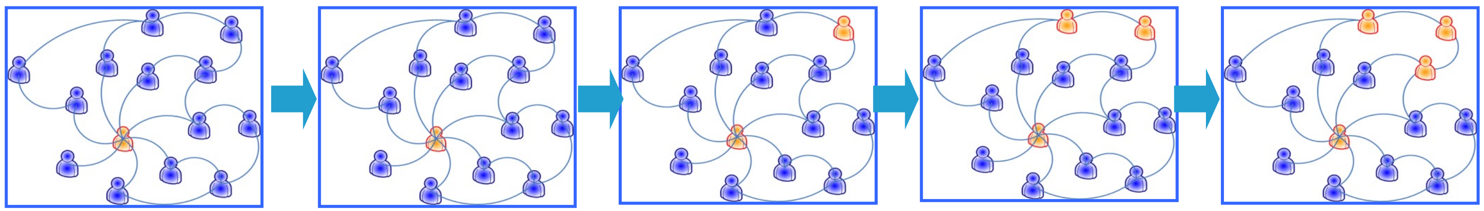


Example 1: application to marketing

- When customers interact, a behavior can become **viral:** a client can influence her friend



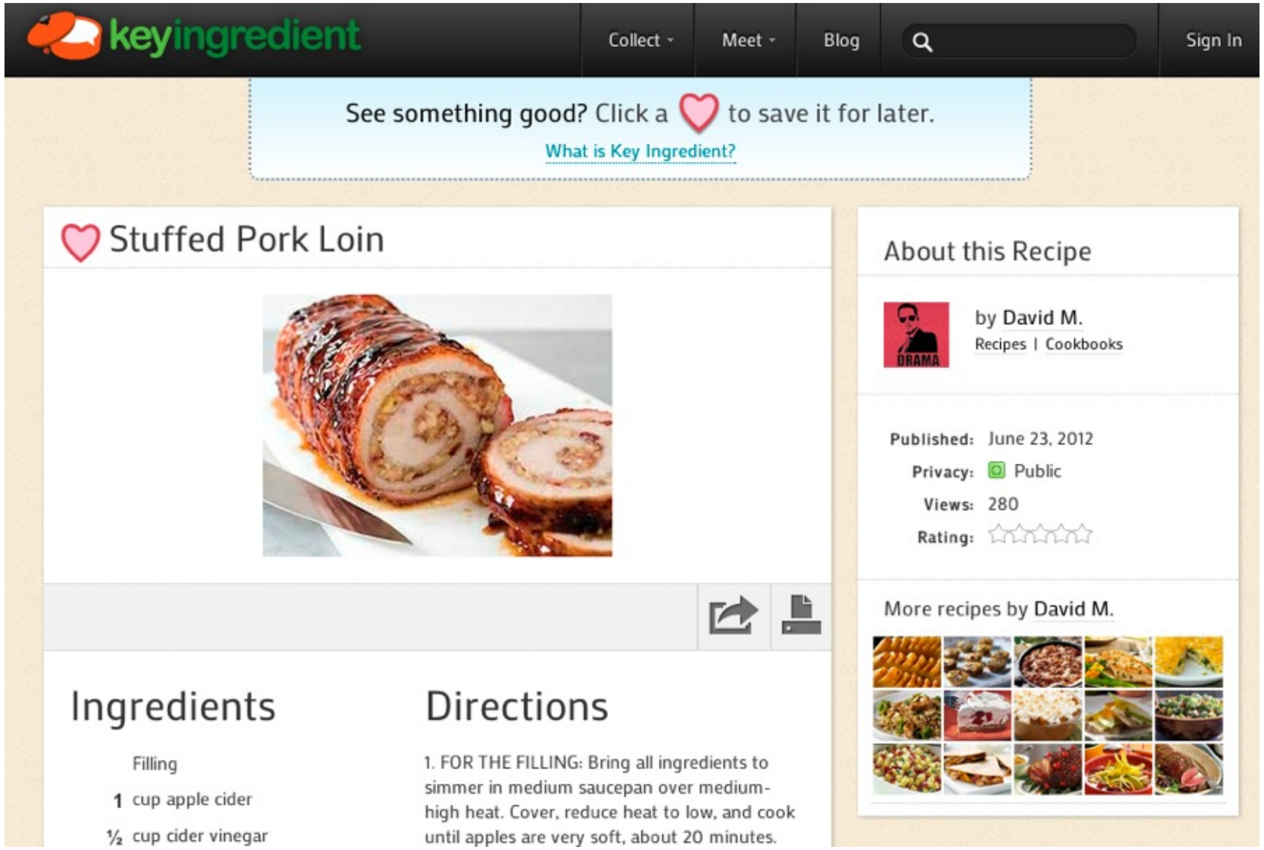
… but sometimes not



A successful viral campaign requires

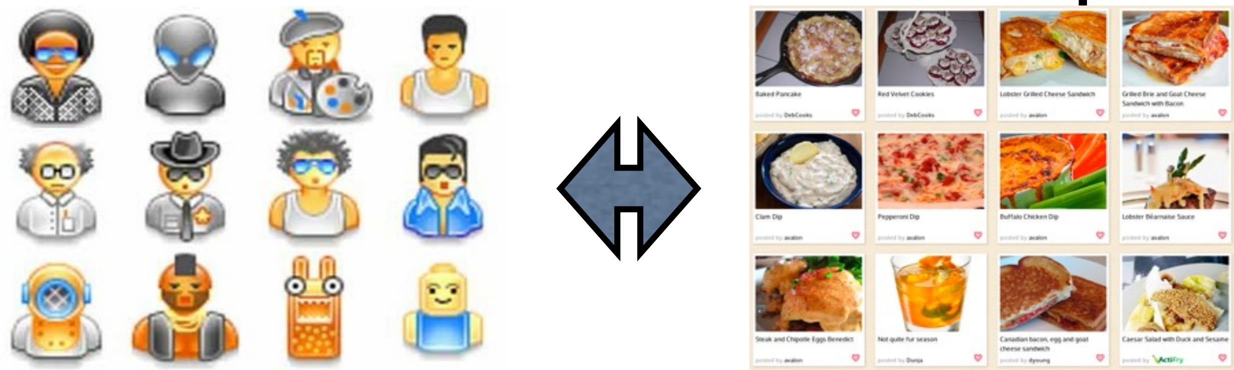
- a good understanding of the **roles** of the nodes
- creating a correct **propagation model**

Example 2: social web platform eg food recipes



Example 2: social web platform

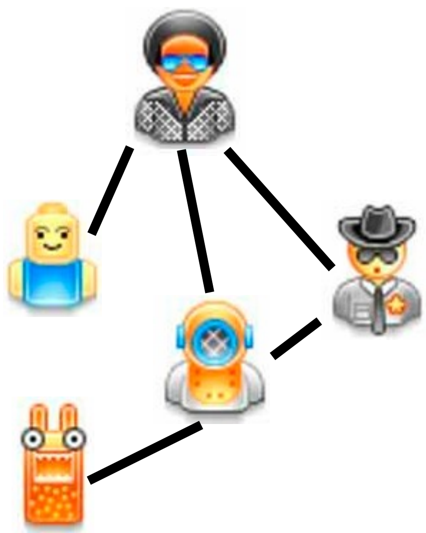
Users Recipes

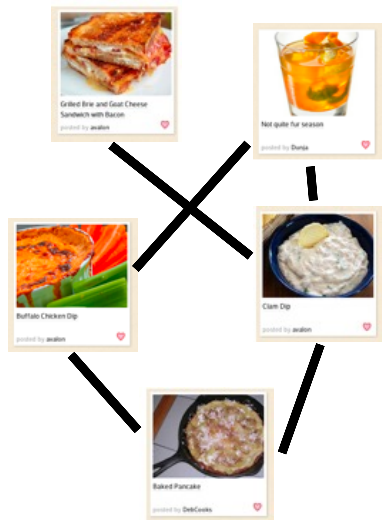


- Blogs (associated to the users or to the recipes)
- Users’ ratings
- *Tags*
- Comments on recipes

Example 2: social web platform

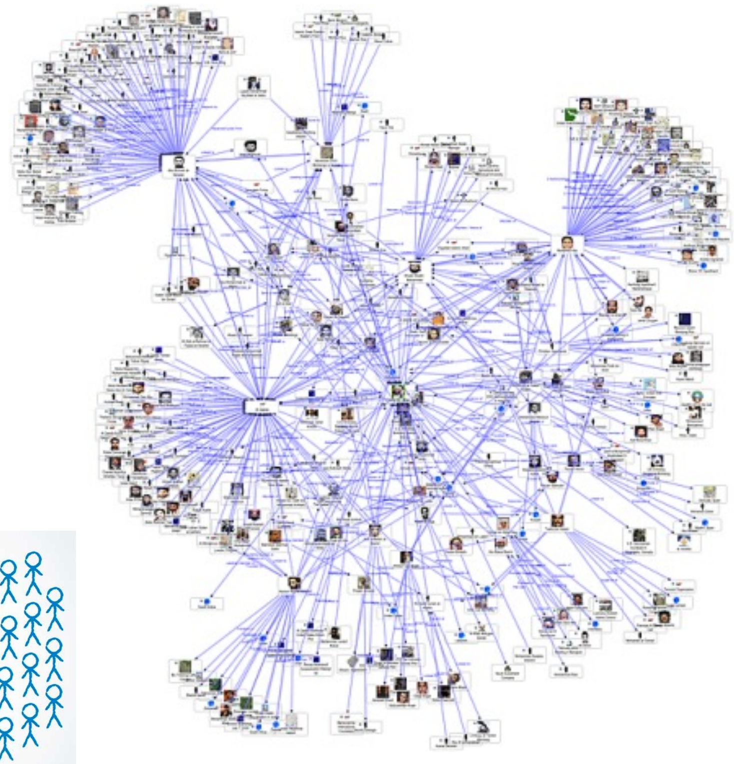
- One can build several graphs



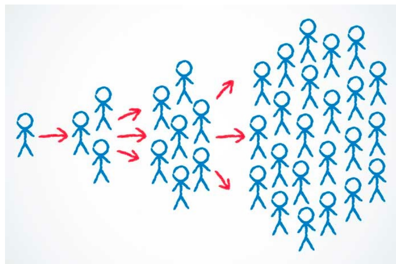


Unipartites or Bipartites graphs

Some important application of Social Network Analysis



- Analyze users behavior, understand customers
  - Link analysis
    - Security (finance, intelligence)
    - fraud detection (banks, telcos)
  - Community analysis
    - clustering/segmentation
    - community management
    - detect hot groups/topics, emergence, predict evolution
- Use the network
  - Viral marketing
    - identify influencers
    - build diffusion models



- Predictive modeling
  - Churn prediction, x-sell/up-sell
  - Recommender systems

Challenges of Social Networks Analysis

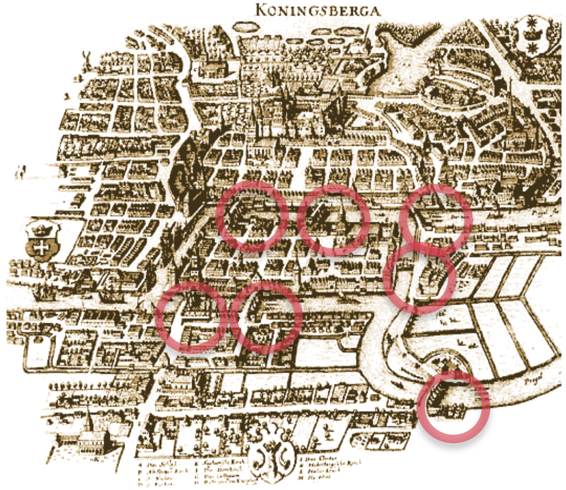
- Big Data: very large amount/rate of transactions
  - Network links => impossible to model on random subsamples
- Data is moving fast => scalability of the models
  - Example in telecoms (CDRs):

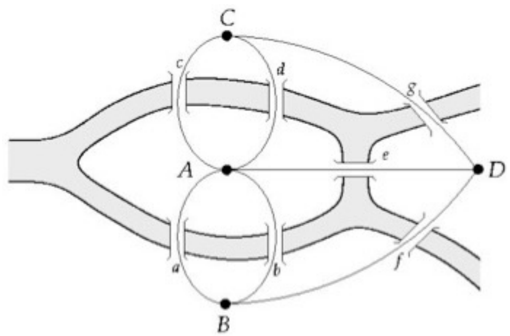
	Rows (Millions)	Nodes (Millions)
One day	150	20
One week	1 100	45
One month	4 360	70
Three months	13 080	90

Graph and Networks in Science & Technology

- 1735 Graph Theory (Euler)
- 1930 Social networks (Moreno)
- 1950 Random networks (Erdős-Rényi)
- 1960-70 Some applications to telecom networks, biology, ecology
- 1990 Web, scale free (Barabasi-Albert)
- 2000 Social Web (2.0), data mining, big graphs
- Present: networks are everywhere, lot of industrial application

7 bridges of Königsberg





(most) real graphs are sparse

In most cases, the average degree of a node does not depend on the size of the graph.

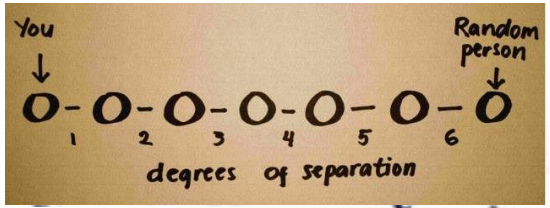
The adjacency matrix is thus **sparse** (most elements are zeroes)

	#Nodes	#Links		Average degree
WWW (ND Sample):	N=325 729;	$L=1.4\ 10^6$	$L_{max}=10^{12}$	$\langle k \rangle=4.51$
Protein (*S. Cerevisiae*):	N= 1 870;	L=4 470	$L_{max}=10^7$	$\langle k \rangle=2.39$
Coauthorship (Math):	N= 70 975;	$L=2\ 10^5$	$L_{max}=3\ 10^{10}$	$\langle k \rangle=3.9$
Movie Actors:	N=212 250;	$L=6\ 10^6$	$L_{max}=1.8\ 10^{13}$	$\langle k \rangle=28.78$

(Source: Albert, Barabasi, RMP2002)

Some properties of complex networks

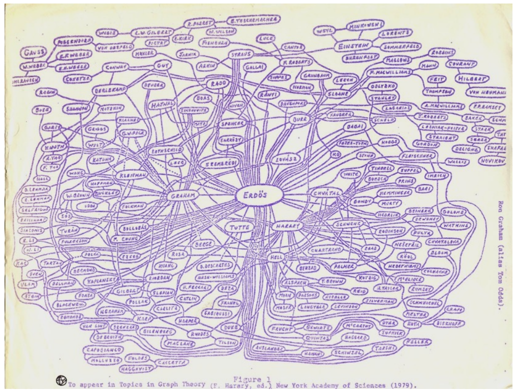
Small worlds



Six degrees of separation

https://en.wikipedia.org/wiki/Six_degrees_of_separation

http://www.ams.org/mathscinet/collaborationDistance.html



E-mail Study Corroborates Six Degrees of Separation

By Dan Cho

ces are, you donʼt personally know any
ralian policemen, Estonian archival
ctors or Norwegian army
inarians. But you could probably get in
ı with one of these distant individuals
ıgh a friend, or a friend of a friend, or a
d of your friendʼs friend. The notion
every person on the planet is separated

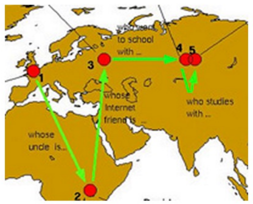


Image: COURTESY OF DUNCAN J. WATTS

Degree distribution

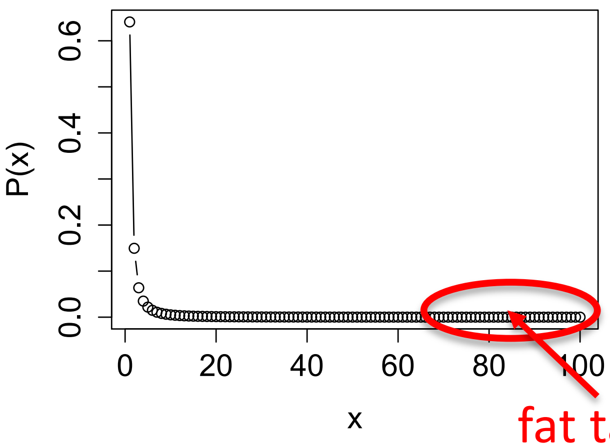
In complex networks, **most nodes have a low degree**, but some have a very high degree.

The degree distribution follows a power law:

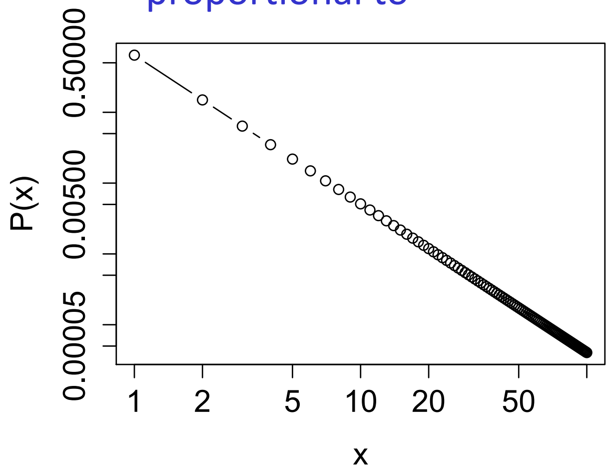
$$P(\deg = x) \propto x^{-\alpha}$$

= nb of nodes with degree *x*

proportional to



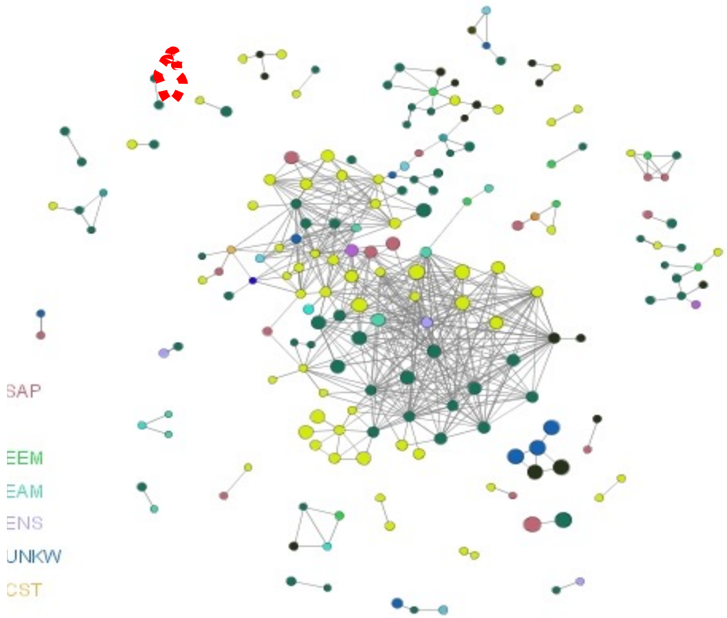
fat tail

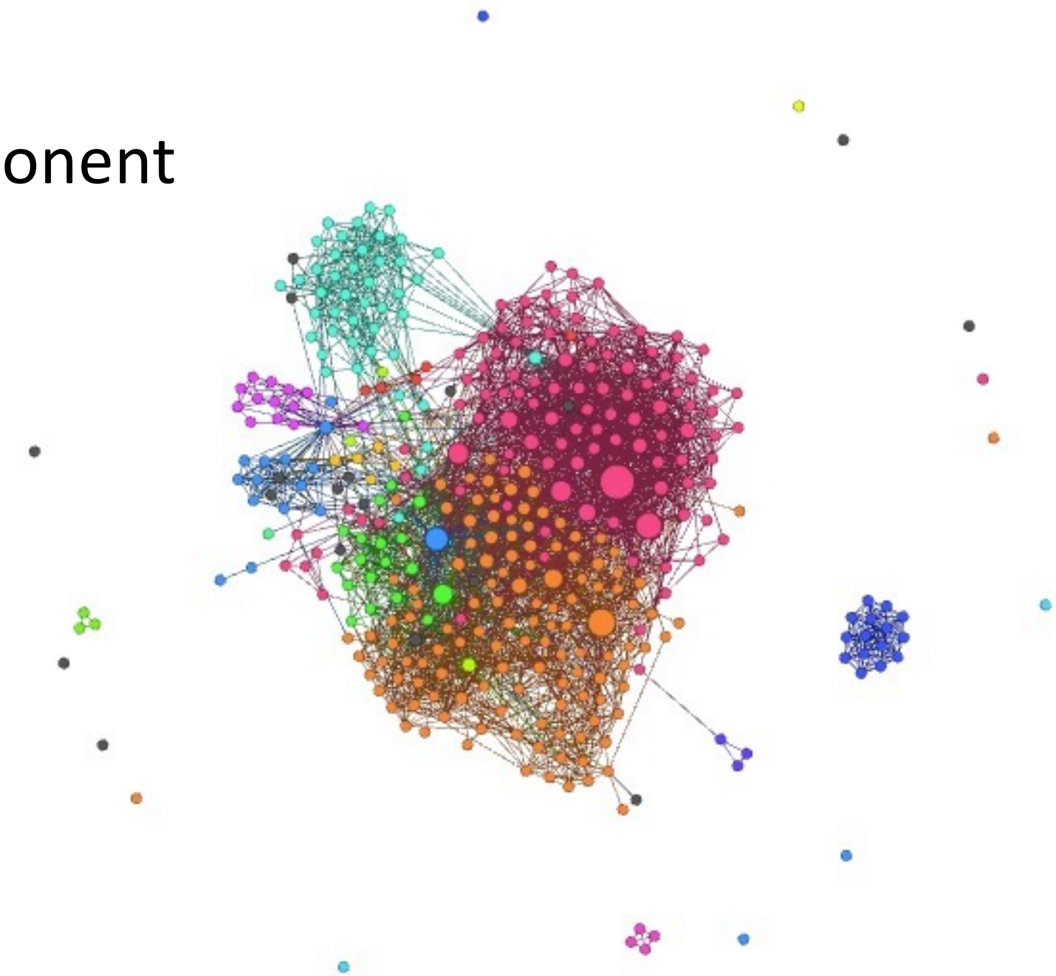


Linear in log/log scale

Connected components

Giant connected component





For more details, see: https://fr.coursera.org/lecture/algorithms-on-graphs/strongly-connected-components-OIOTT

Clustering coefficient

What are the odds that two of my friends know each other?

- *Clustering coefficient* of a node ("triangles")
  - measures how close its neighbors are to being a clique: i.e. the proportion of links between nodes in its 1st circle to the number of links that *could* possibly exist

If N_i is the neighborhood (or 1st circle of node n_i)

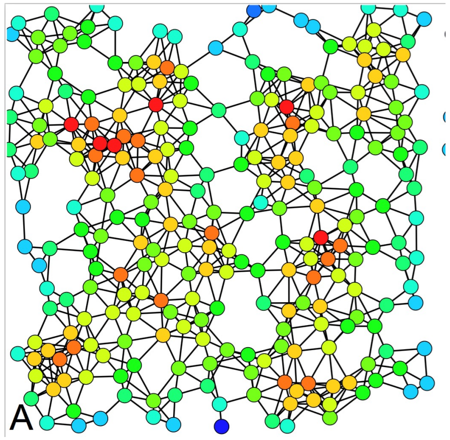
$$C_i = \left|\left\{e_{jk} \in E : n_j, n_k \in N_i\right\}\right| / Deg(n_i) \times \left[Deg(n_i) - 1\right]$$ (in a directed graph)

$C_A = 1$ $C_A = 1/3$ $C_A = 0$

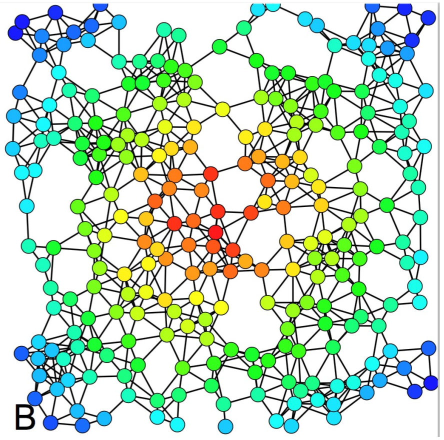
Centrality

Node importance in a network can be measured by

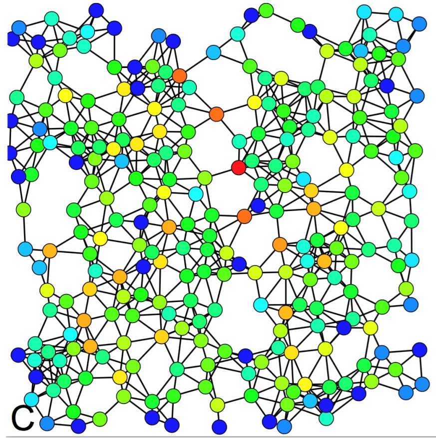
- *its degree (number of neighbors)*
- *proximity to other nodes (average)*
- *betweenness (number of shortest path passing through this node)*



Degree centrality



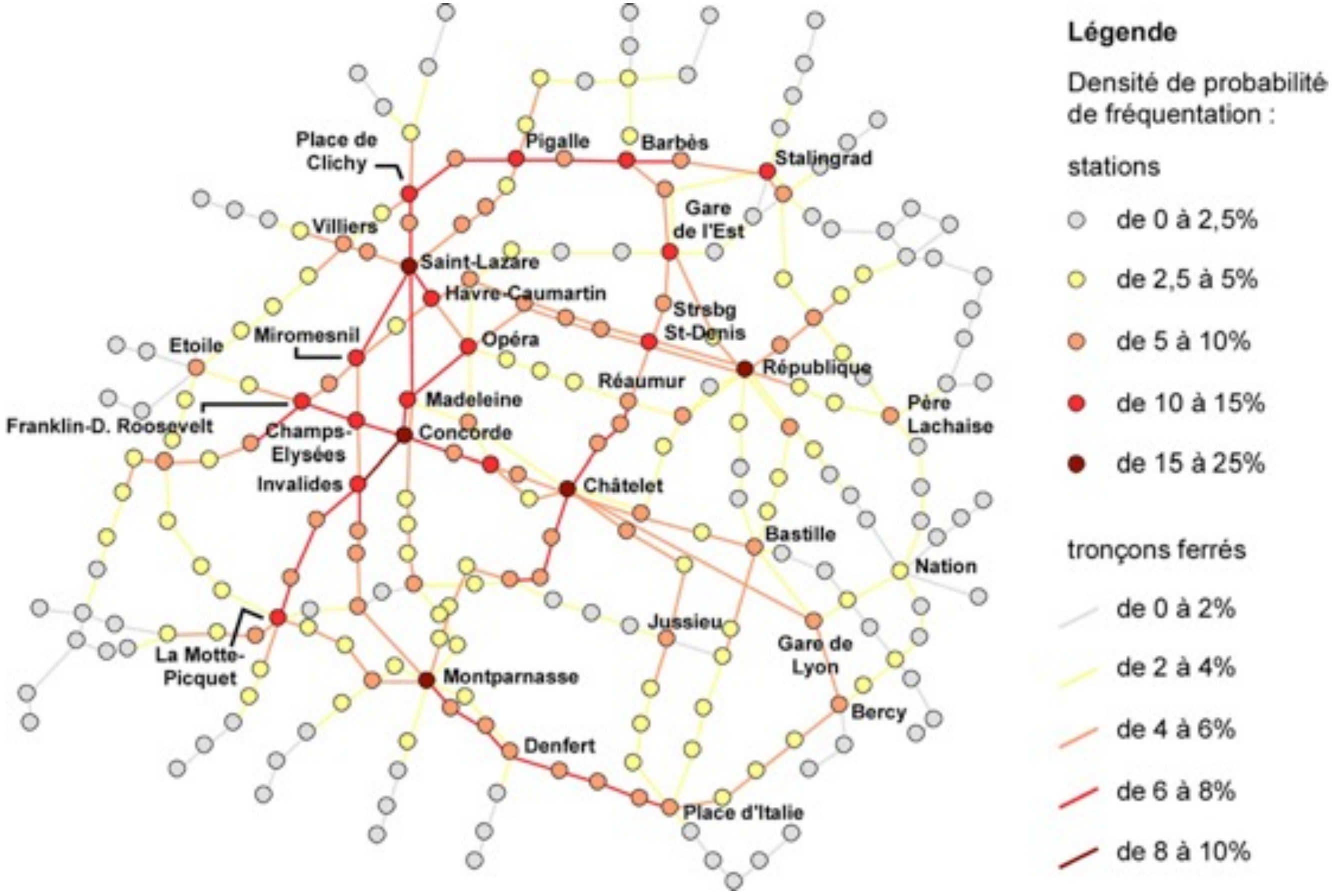
Proximity centrality



Betweenness centrality

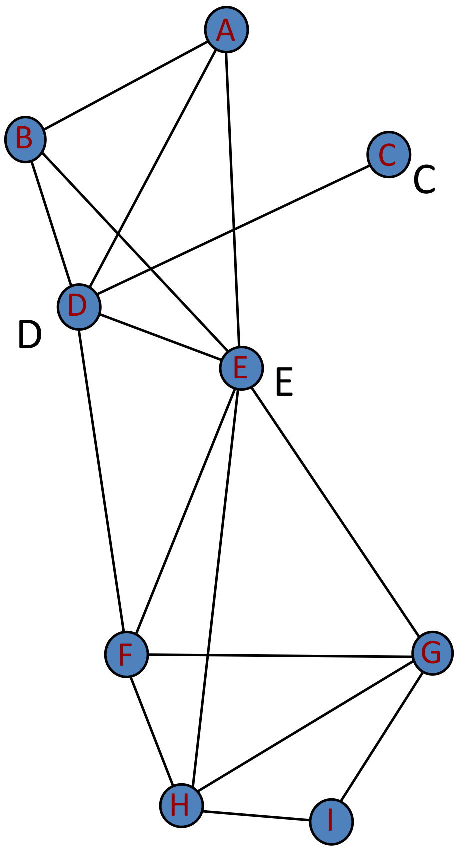
… among other, like spectral centrality, see https://en.wikipedia.org/wiki/Centrality

Example: betweenness centrality of stations in Paris subway



Note: I've lost the source for this figure, but you could start from https://github.com/totetmatt/gexf/tree/master/metro/Paris

Social roles

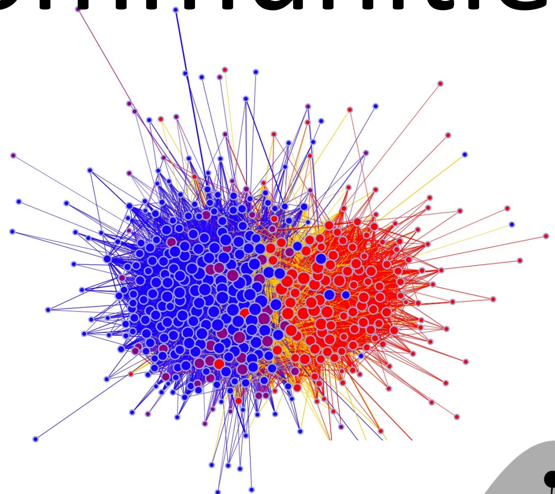


- **Structural roles**
  - central/peripheral
  - connectors
- **Roles based on actions**
  - Author, reader, commentator

See *Roles in social networks: methodologies and research issues*, M. Forestier et al. 2012.

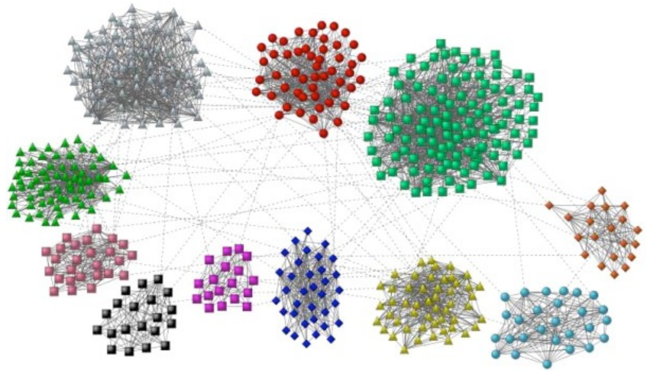
Communities
in
complex networks

Communities in complex networks

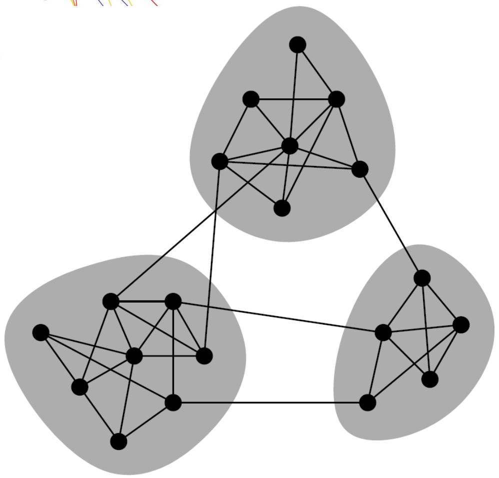


Political blogs

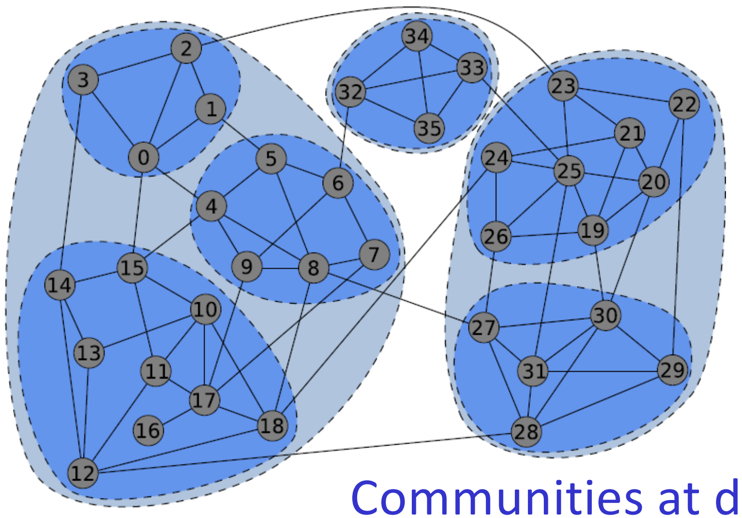
Group nodes in clusters



A. Lancichinetti, S. Fortunato (2009)



Link density higher inside communities



Communities at different scales

Credit: P. Pons, 2007

Many definitions for “communities”

→ many partitioning algorithms

A *community* is a set of nodes such that

- Nodes are similar (considering attributes)
  - Persons, Web pages …
- or Highly connected
  - Quasi-clique …
- or Local link density > C^{st} * Global density
  - Cliques, triangles…
- More links inside than outside

Modularity of a graph partition

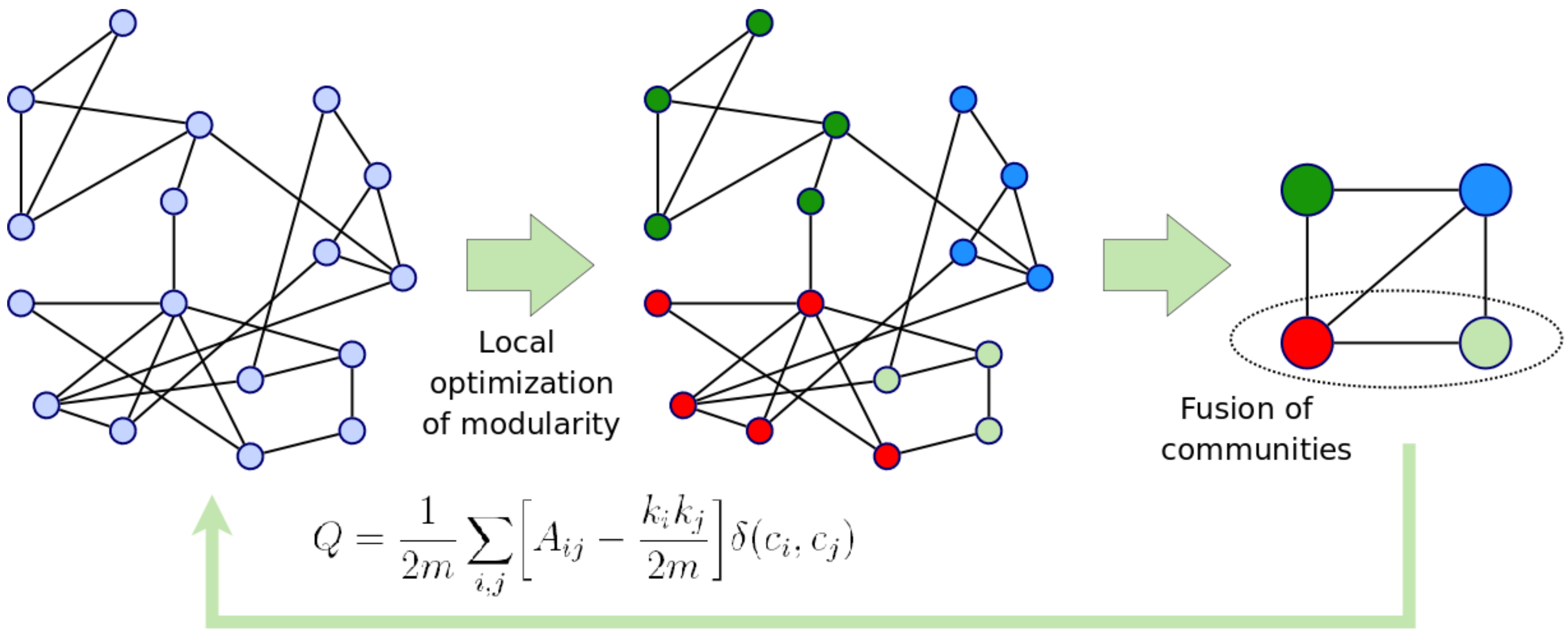
- Assume random networks do not have communities
- Consider a network with N nodes and L links
- Partition it into n_c communities, each with N_c nodes connected by L_c links, where $c=1,\ldots, n_c$
- For **each community**, measure the difference between the **network connections** (A_{ij}) and the **expected links** if the network were randomly wired

$$M = \sum_{c=1}^{n_c} M_c = \sum_{c=1}^{n_c} \frac{1}{2L} \sum_{i,j \in C_c} \left(A_{ij} - \frac{k_i \cdot k_j}{2L} \right) = \sum_{c=1}^{n_c} \left[\frac{L_C}{L} - \left(\frac{k_c}{2L} \right)^2 \right]$$

where k_c is the total degree in community c

Finding communities: Louvain algorithm

Local greedy algorithm



very fast (process millions of nodes in less than one minute)

Blondel et al., Fast unfolding of communities in large networks, 2008

Louvain algorithm: example

Belgian mobile phones

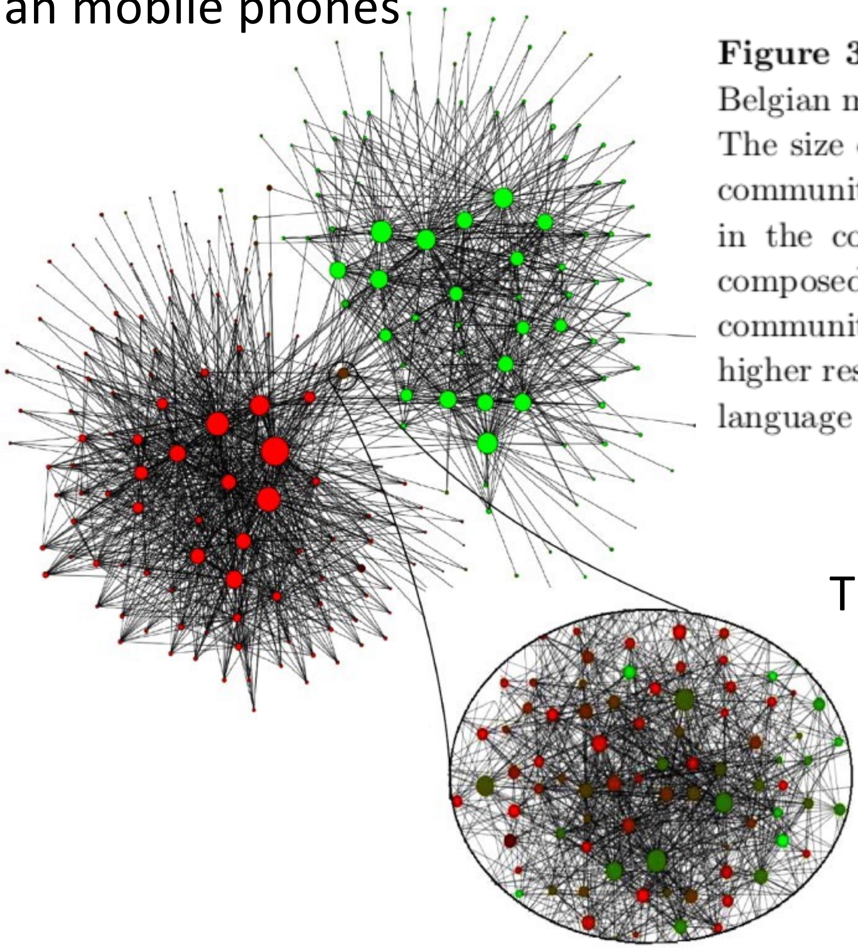


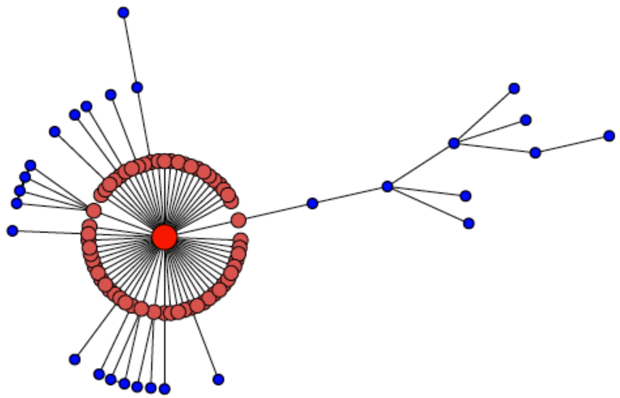
Figure 3. Graphical representation of the network of communities extracted from a Belgian mobile phone network. About 2M customers are represented on this network. The size of a node is proportional to the number of individuals in the corresponding community and its colour on a red-green scale represents the main language spoken in the community (red for French and green for Dutch). Only the communities composed of more than 100 customers have been plotted. Notice the intermediate community of mixed colours between the two main language clusters. A zoom at higher resolution reveals that it is made of several sub-communities with less apparent language separation.

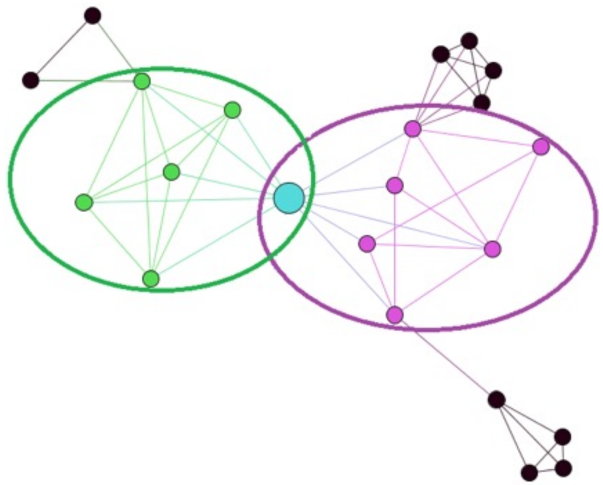
The algorithm provides a hierarchical segmentation

Blondel et al., Fast unfolding of communities in large networks, 2008

Local or ego-centric communities

Algorithms to extract a community of nodes in strong interaction with a given node (starting point)



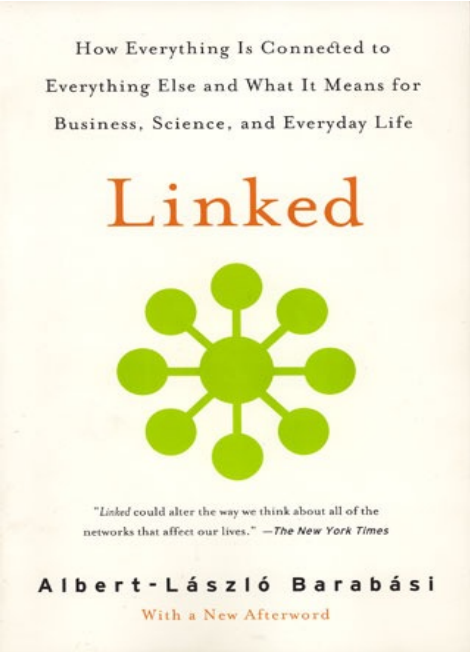


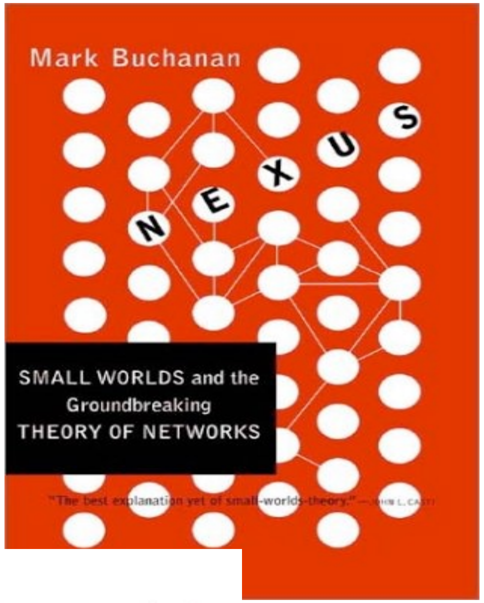
Blaise Ngonmang, Maurice Tchuente & Emmanuel Viennet 2012

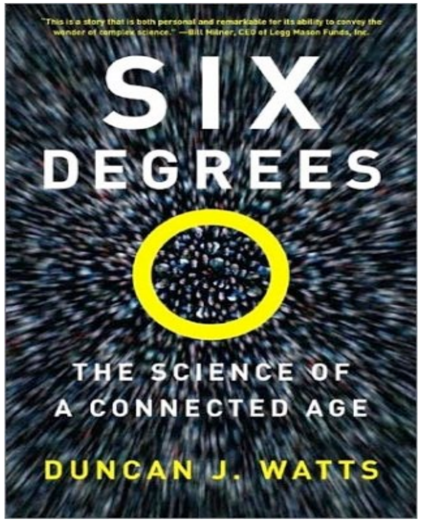
Conclusion

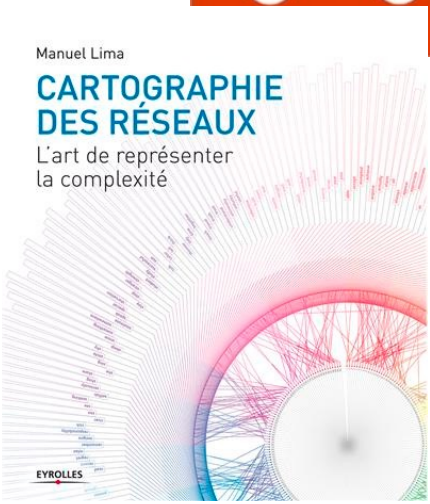
- We briefly introduced Complex Network concepts
- Powerful tools to model complex systems with links or transactions
- We defined some basic metrics: degree, centrality, clustering coefficient
- We have shown how to extract communities
- "Social variables" can help build better predictive models
- Other important topics:
  - Dynamic networks
  - Propagation models
  - Recommendation in social networks
  - Graph neural networks

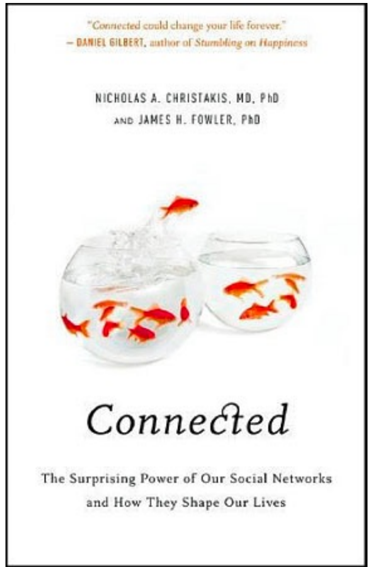
References: general books



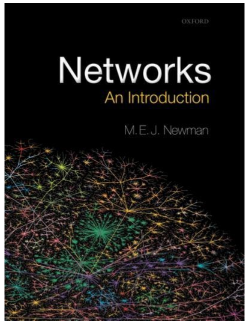




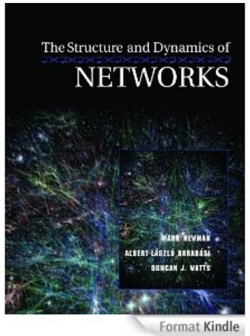




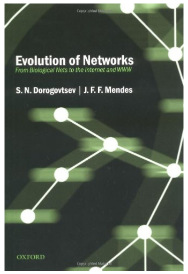
References: some specialized books



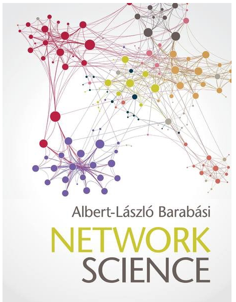
M. Newman. Networks, an introduction. (Oxford University Press, 2010).



Albert-László Barabási, Mark Newman, Duncan J. Watts
The Structure and Dynamics of Networks
Feb. 2013



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



A.-L. Barabasi: Network Science
(8/2016).
http://barabasi.com/networksciencebook

References: scientific papers

- **Community detection**
  - S. Fortunato, “Community detection in graphs,” *Physics Reports*, vol. 486, no. 3, 2010.
  - A. Lancichinetti and S. Fortunato, “Limits of modularity maximization in community detection,” *Physical Review E*, vol. 84, no. 6, 2011.
  - V. Blondel, J.-L. Guillaume et al. “Fast unfolding of communities in large networks” Journal of Statistical Mechanics: Theory and Experiment 2008.
  - B Ngonmang, M Tchuente, E Viennet, “Local community identification in social networks”, Parallel Processing Letters, 2008

Datasets

- Stanford Large Network Dataset Collection https://snap.stanford.edu/data
- Mark Newman's collection https://public.websites.umich.edu/~mejn/netdata
- The Colorado Index of Complex Networks} (ICON) https://icon.colorado.edu
- The KONECT Project http://konect.cc
- Interaction data from the Copenhagen Networks Study https://www.nature.com/articles/s41597-019-0325-x

Software

Gephi

Tulip

igraph library

NetworkX

Pajek

Cytoscape

NODEXL Network Graphs
The Social Media Research Foundation

UCINET