Complex Networks

IXXI Summer School 2009

Guillaume Chelius guillaume.chelius@inria.fr
INRIA / ENS Lyon / IXXI

Some references

- CCNR: Center for Complex Network Research (Barabasi Lab)
  http://barabasilab.com
- Evolution & structure of the Internet, **R. Pastor-Satoras, A. Vespignagni**
- **A.-L. Barabási, R. Albert**
  Emergence of scaling in random networks
  Science 286, 509–512 (1999) [PDF]
- **R. Albert, A.-L. Barabási**
  Statistical mechanics of complex networks
  Reviews of Modern Physics 74, 47-97 (2002) [PDF]
- Dynamical processes in complex networks, **A. Barrat, M. Barthelemy, A. Vespignani**, Cambridge University Press (2008)

Contributors

- Alain Barrat
  Alain.Barrat@cpt.univ-mrs.fr
  http://www.cpt.univ-mrs.fr/~barrat/

- Eric Fleury
  Eric.Fleury@inria.fr
  http://perso.ens-lyin.fr/eric.fleury

- Matthieu Latapy
  Matthieu.Latapy@lip6.fr
  http://www-rp.lip6.fr/~latapy/

Contents

What is a network ?

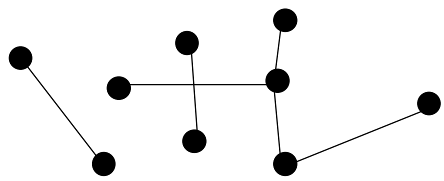
points linked by edges

individuals	**social relations**
computers	**wires**
web pages	**hyperlinks**
airports	**flights**
proteins	**reactions**
....	**....**

What is a network ?

points linked by edges

➔ modeled by a graph G=(V,E)



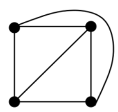
Auto-organisation

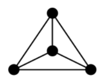
Dynamic evolution

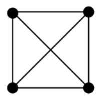
What is a network ?

A graph captures a structure: *n* nodes / vertices, *m* edges
A graph is not a drawing; one graph, several drawings !

E.g.: K4 = a, b, c, d ; a-b, a-c, a-d, b-c, b-d, c-d







What is a network ?

Many type of networks / graphs :

- Directed vs undirected networks
- Weighted vs unweighted networks
- Digraphs
- …

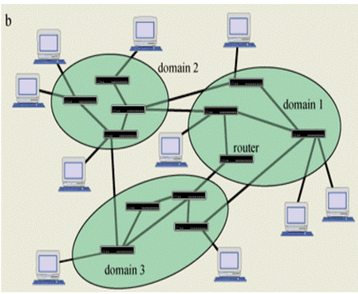
What is a complex network ?

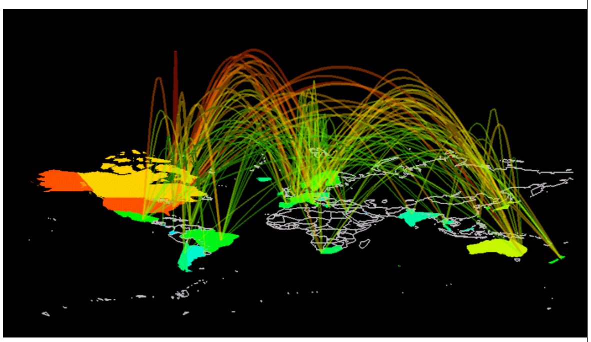
A complex network is a network build from observations / measurements of interactions occuring in the « real world » :

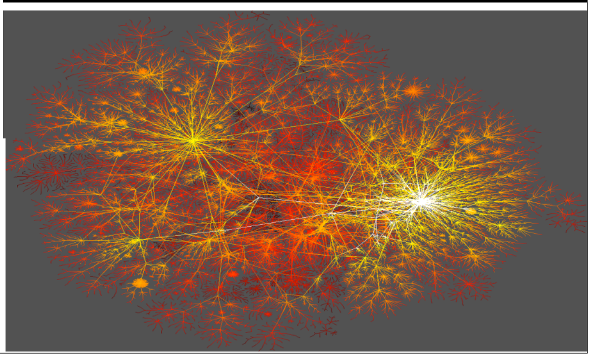
- Biology
- Sociology
- Transportation
- Chemistry
- Epidemiology
- …

As opposed to absract graphs, mathematical objects.

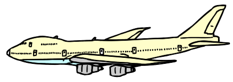
Internet

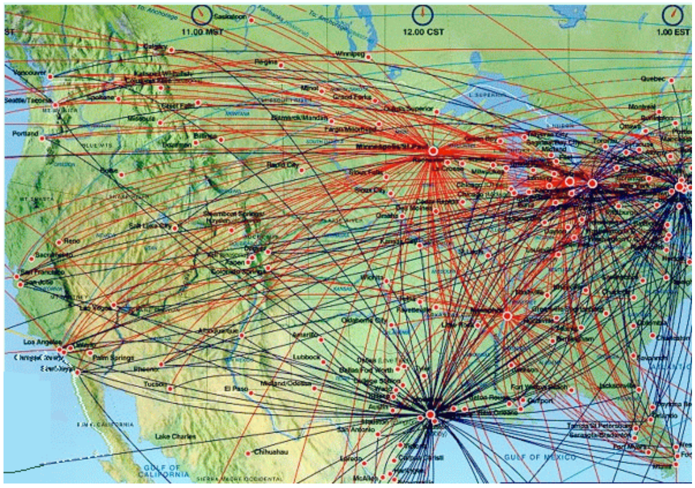




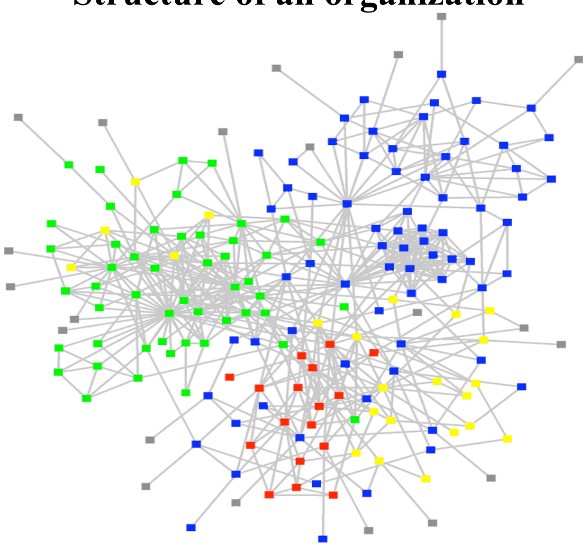


Airplane route network





Structure of an organization



Red, blue, or green: departments

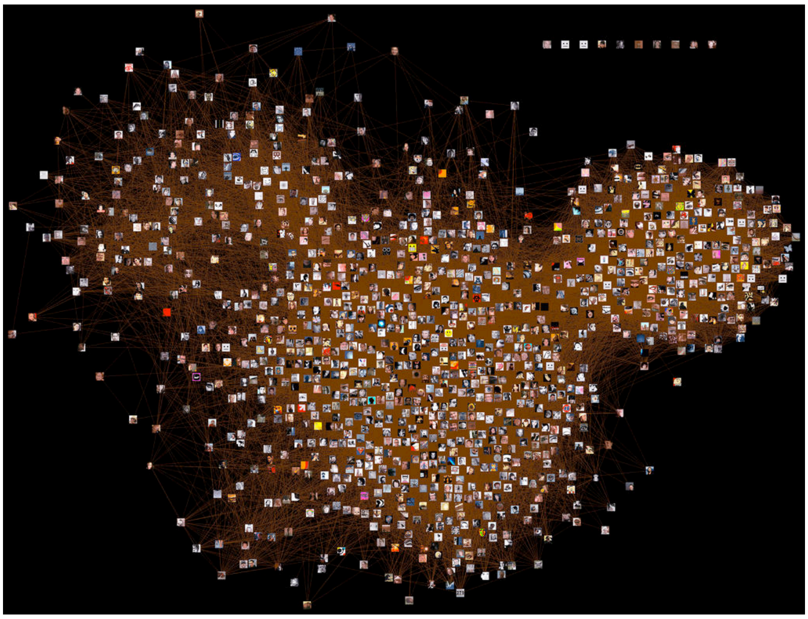
Yellow: consultants

Grey: external experts

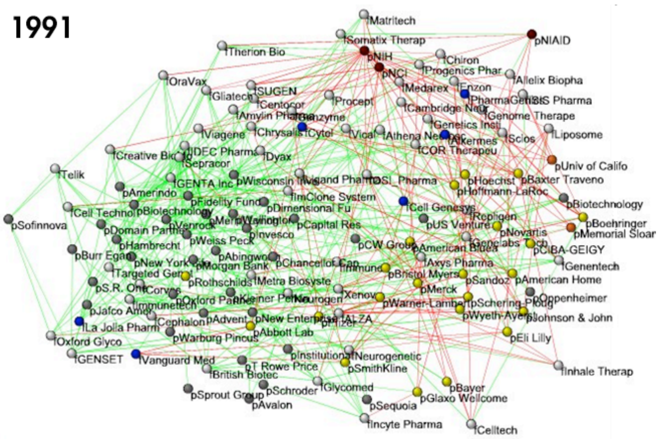
www.orgnet.com

- http://www.blogopole.fr/
- http://www.flickr.com/photos/gustavog/

Blog / WWW



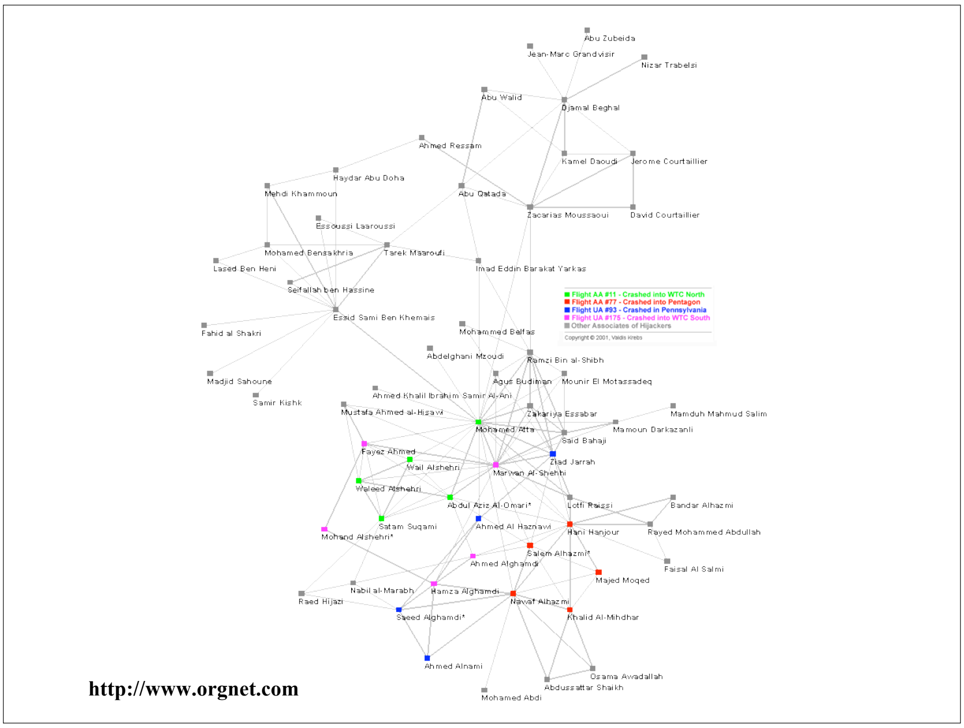
Business ties in US biotech-industry

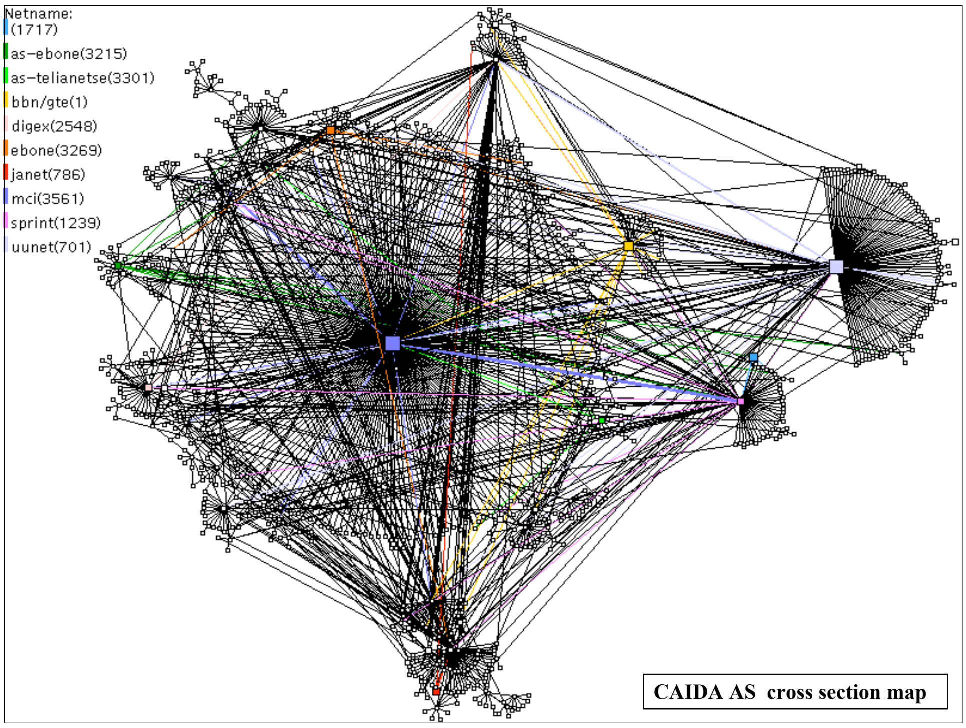


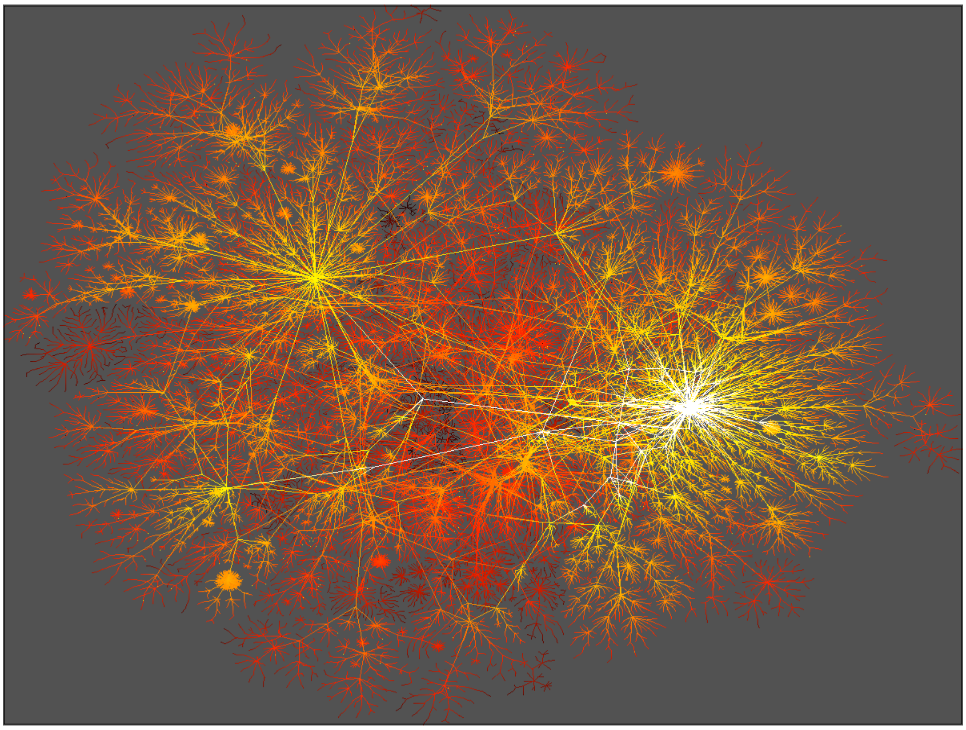
Nodes: companies
investment
pharma
research labs
public
biotechnology

Links: collaborations
financial
R&D

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

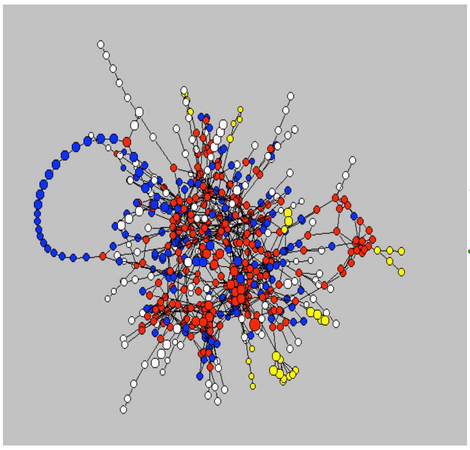

Abu Zubeida
Jean-Marc Grandvisir
Nizar Trabelsi
Abu Walid
Djamal Beghal
Ahmed Ressam
Kamel Daoudi
Jerome Courtaillier
Haydar Abu Doha
Mehdi Khammoun
Abu Qatada
Zacarias Moussaoui
David Courtaillier
Essoussi Laaroussi
Mohamed Bensakhria
Tarek Maaroufi
Lased Ben Heni
Imad Eddin Barakat Yarkas
Seifallah ben Hassine
Essid Sami Ben Khemais
Fahid al Shakri
Mohammed Belfas
Abdelghani Mzoudi
Ramzi Bin al-Shibh
Madjid Sahoune
Agus Budiman
Mounir El Motassadeq
Samir Kishk
Ahmed Khalil Ibrahim Samir Al-Ani
Mustafa Ahmed al-Hisawi
Zakariya Essabar
Mamduh Mahmud Salim
Mohamed Atta
Mamoun Darkazanli
Said Bahaji
Fayez Ahmed
Ziad Jarrah
Wail Alshehri
Marwan Al-Shehhi
Waleed Alshehri
Abdul Aziz Al-Omari*
Lotfi Raissi
Bandar Alhazmi
Satam Suqami
Ahmed Al Haznawi
Hani Hanjour
Rayed Mohammed Abdullah
Mohand Alshehri*
Salem Alhazmi*
Ahmed Alghamdi
Faisal Al Salmi
Majed Moqed
Nabil al-Marabh
Hamza Alghamdi
Raed Hijazi
Nawaf Alhazmi
Saeed Alghamdi*
Khalid Al-Mihdhar
Ahmed Alnami
Osama Awadallah
Abdussattar Shaikh
Mohamed Abdi
Flight AA #11 - Crashed into WTC North
Flight AA #77 - Crashed into Pentagon
Flight UA #93 - Crashed in Pennsylvania
Flight UA #175 - Crashed into WTC South
Other Associates of Hijackers
Copyright © 2001, Valdis Krebs
http://www.orgnet.com

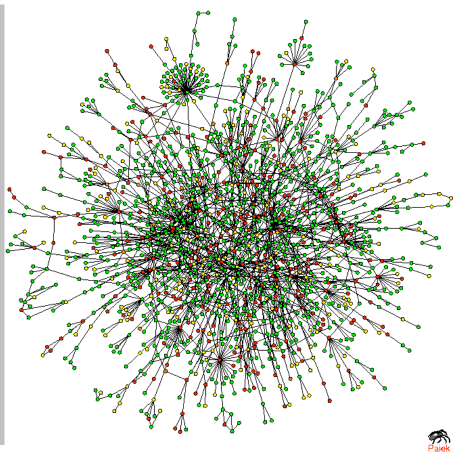

Netname:
(1717)
as-ebone(3215)
as-telianetse(3301)
bbn/gte(1)
digex(2548)
ebone(3269)
janet(786)
mci(3561)
sprint(1239)
uunet(701)
CAIDA AS cross section map



Metabolic Network

Protein Interactions





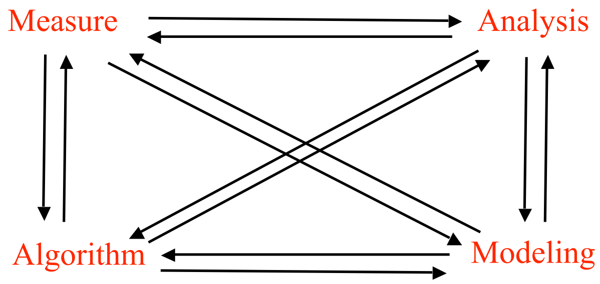
Complex networks

➔ Most of the complex networks offer some common non-trivial properties…

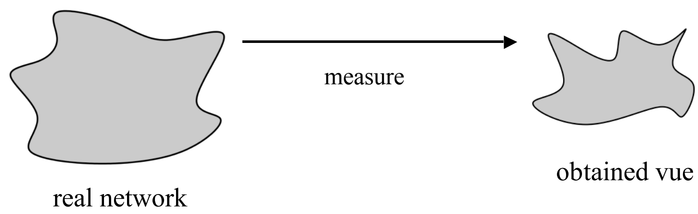
➔ … and some common issues…

- Measure
- Analysis
- Modeling
- Algorithm

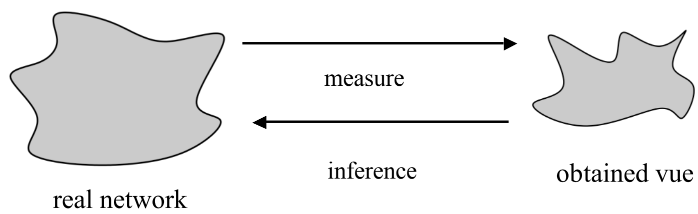
Complex networking : issues



Measure



Measure



What can we say about the real object given the measure ?

- Impact on the observed properties
- Impact of the properties on the view
- Mesures targeting some properties

Analysis

Describe

Extract some useful information

Analysis

Describe

Extract some useful information

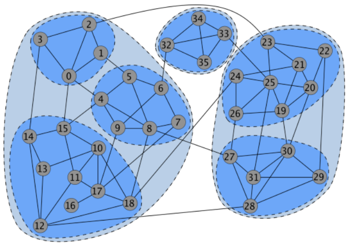
statistics

density

degree

local density

correlations



Modeling

Generation of realistic graphs

(i.e. given the observed properties)

motivations: formal approaches, simulation, algorithms, meaning.

Modeling

Generation of realistic graphs

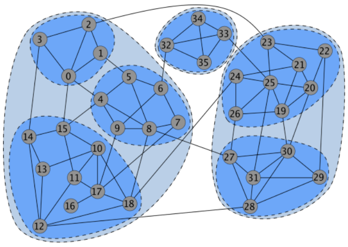
(i.e. given the observed properties)

motivations: formal approaches, simulation, algorithms, meaning.

state of the art:

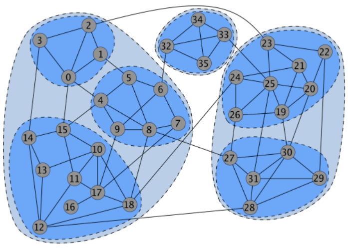
- Size, density, distances: easy
- Degree: consensus, models
- Clustering: no consensus

Algorithms



- Specific problems
- Revisited problems
- Space restrictions

Algorithms



- Specific problems
- Revisited problems
- Space restrictions

- Take profit of observed properties
- Approximated results often enough
- Experimental, empiric algorithms

➔ New way of doing algorithms

Global view

Measure —— Analysis —— Modeling —— Algorithm

Global view

Measure —— Analysis —— Modeling —— Algorithm

- Measure → data → analysis → properties → model

Global view

Measure —— Analysis —— Modeling —— Algorithm

- Measure → data → analysis → properties → model
- Model →formalism → metrology, algorithm

Global view

Measure —— Analysis —— Modeling —— Algorithm

- Measure → data → analysis → properties → model
- Model →formalism → metrology, algorithm
- Analysis → calculuous → algorithm → computation → analysis

Global view

Measure —— Analysis —— Modeling —— Algorithm

- Measure → data → analysis → properties → model
- Model →formalism → metrology, algorithm
- Analysis → calculuous → algorithm → computation → analysis
- Metrology → robust properties → analysis
- …

Contents

I. Complex Networks, an introduction
II. <u>Network properties</u>
III. Network classes
IV. Network models
V. Applications
VI. Case studies
    I. Contact monitoring
    II. Radar for the Internet

Graph modeling

- Graph G = (V, E)
  → What about dynamic network ?

- Size of the network
  – $N = |V|$, number of nodes / vertices
  – $M = |E|$, number of edges

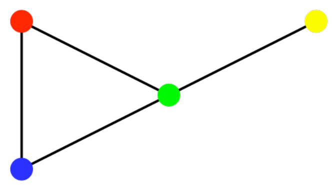
- Weighted, directed graphs
  – Social relationship graphs
  – Traffic networks

Basic properties

- Path
- Connectivity of the network
  – GCC, GWCC, GSCC…
- Shortest path → distance
- Distance → eccentricity, radius, diameter, *etc*.
- Node degree

Graph density

- Number of edges vs number of potential edges



N = 4, M = 4, delta = 0.66…

Centrality measures

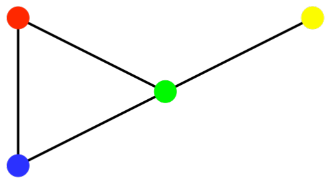
➔ measures the centrality of a node i in the network

Various definitions

May be based on degree, distance or betweenness.

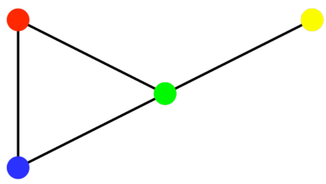
Degree centrality

- Classifies nodes according to the number of neighbors
- Based on inbound and outbound degree
- Refined: the Bonacich index
  - Considers the degree of the neighbors
  - Computed through iterations.



Proximity centrality

Centrality based on the distance between a node and its peers



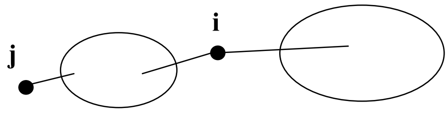
Betweenness centrality

Betweenness centrality of a node i:

for each pair (l,m) in the graph, there are

- s^{lm} shortest paths between l and m
- s_i^{lm} shortest paths going through i

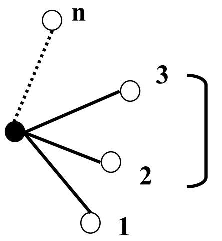
b_i is the sum of s_i^{lm} / s^{lm} over all pairs (l,m)



bi is large

bj is small

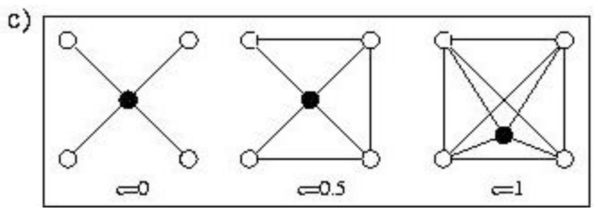
Clustering coefficient



Higher probability to be connected

$$C = \frac{\text{\# of links between 1,2,…n neighbors}}{n(n-1)/2}$$

Clustering: My friends will know each other with high probability! (typical example: social networks)



Satistical characterization

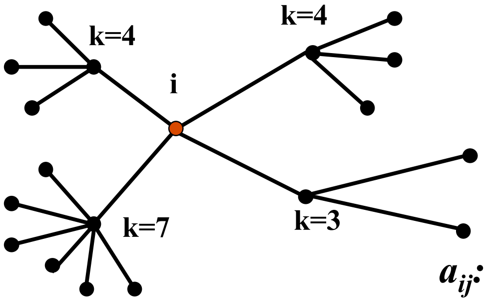
- Degree distribution
  - Homogeneous vs heterogeneous networks
  - *n*-th moment of the degree distribution

- Betweenness distribution
  - Average betweenness linked to the average distance in the network
  - Notion of hubs

Topological correlations: assortativity

$$k_{nn,i} = \frac{1}{k_i}\Sigma_j\ a_{ij}k_j$$

$$k_{nn}(k) = \frac{\Sigma_i\ \delta(k_i - k)k_{nn,i}}{\Sigma_i\ \delta(k_i - k)}$$



ki=4

knn,i=(3+4+4+7)/4=4.5

a_{ij}: Adjacency matrix

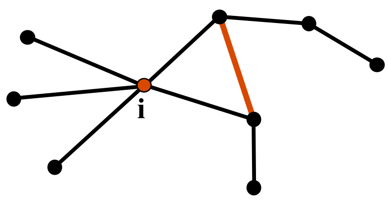
Assortativity

- Assortative behaviour: growing $k_{nn}(k)$
- Disassortative behaviour: decreasing $k_{nn}(k)$

Topological correlations: clustering

$$c_i = \frac{1}{k_i(k_i-1)} \Sigma_{j\neq k}\, a_{ij} a_{ik} a_{jk}$$

$$C(k) = \frac{\Sigma_i\, \delta(k_i - k) c_i}{\Sigma_i\, \delta(k_i - k)} \qquad C = < c_i >$$

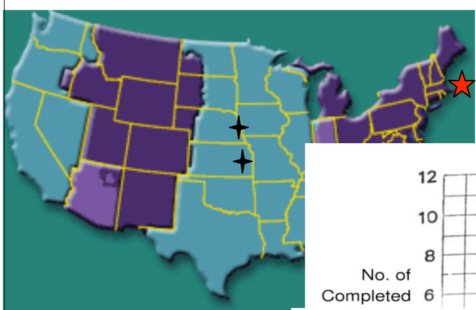


ki=5

ci=0.1

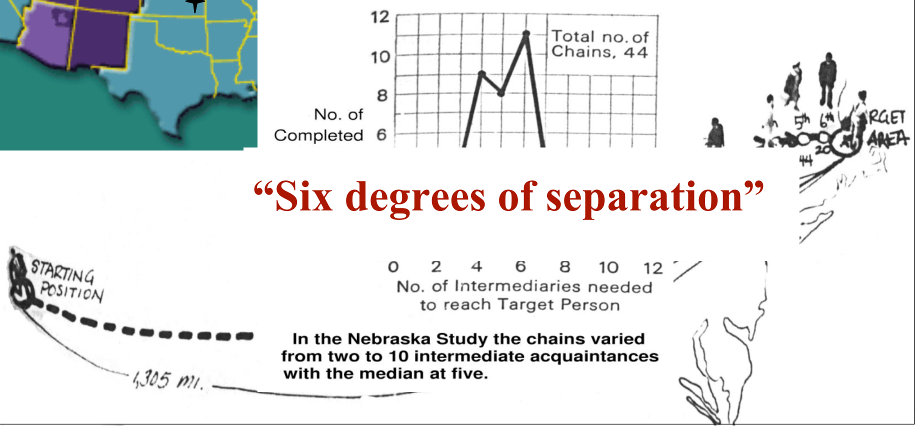
Contents

Social networks: Milgram's experiment



Milgram, *Psych Today* **2**, 60 (1967)

Dodds et al., *Science* **301**, 827 (2003)

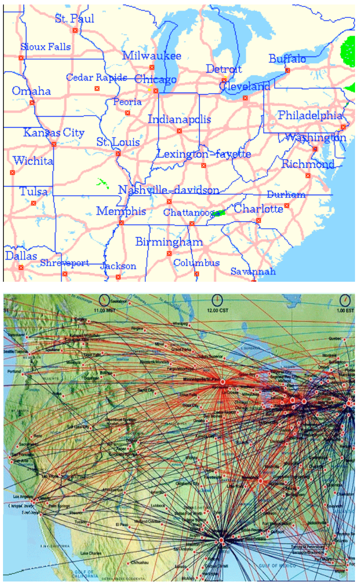


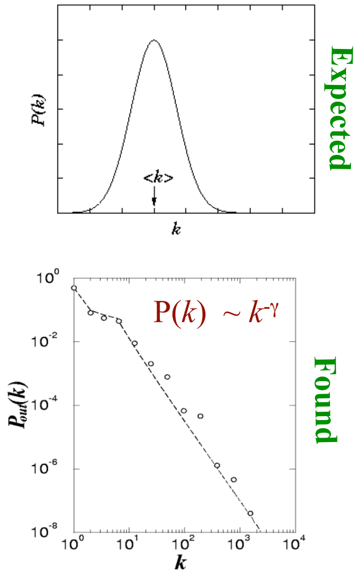
"Six degrees of separation"

Exp. vs. Scale-Free

Exponential Network

Scale-free Network



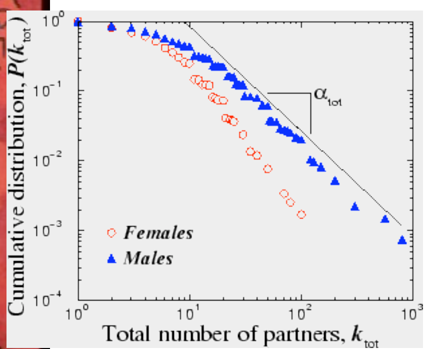


R. Albert, H. Jeong, A-L Barabasi, *Nature*, **401** 130 (1999).

Swedish sex-web

Nodes: people (Females; Males)

Links: sexual relationships



4781 Swedes; 18-74;
59% response rate.

Liljeros et al. Nature 2001

Scale-free Networks in Real World

- Metabolic networks：$\gamma \fallingdotseq 2.24$（depending on species）
  - Node ⇔ Chemical compound
  - Edge ⇔ Chemical reaction (almost equivalently, enzyme)
- Protein interaction networks：$\gamma \fallingdotseq 2.2$
- WWW：$\gamma \fallingdotseq 2.1$
  - Node ⇔ Web page
  - Edge ⇔ Link between web pages
- Movie stars：$\gamma \fallingdotseq 2.3$
  - Node ⇔ Actor/Actress
  - Edge ⇔ Act in the same movie

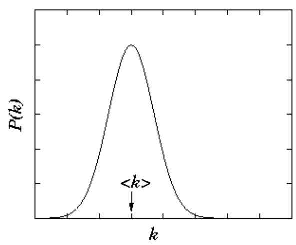
Main Features of complex networks

•Many interacting units
•Self-organization
•Small-world
•Scale-free heterogeneity
•Dynamical evolution

Standard graph theory

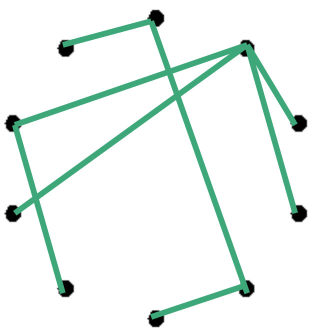
Random graphs

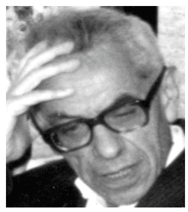
•Static
•Ad-hoc topology



Contents

Erdös-Rényi model (1960)



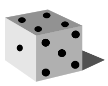


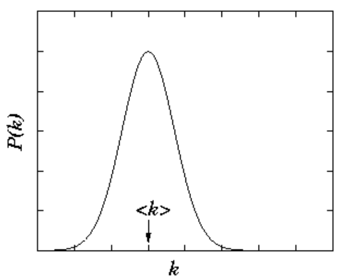
Connect with probability p

Pál Erdös (1913-1996)

p=1/6
N=10 $\langle k \rangle \sim 1.5$

Poisson distribution





short distances (log N)

BUT...

Asymptotic behavior

Lattice

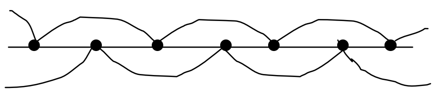
$$L(N) = N^{1/d}$$

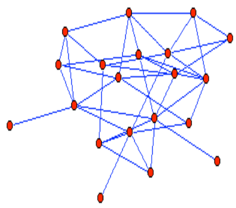
$$C(N) \approx const.$$

Random graph

$$L(N) = \log N$$

$$C(N) \approx N^{-1}$$





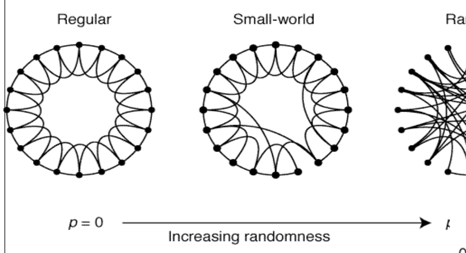
Work of Watt and Strogatz

- Asks why we see the small world pattern and what implications it has for the dynamical properties of social networks.
- Their contribution is to show that the globally significant changes can result from locally insignificant network change.

Assumptions of WS Model

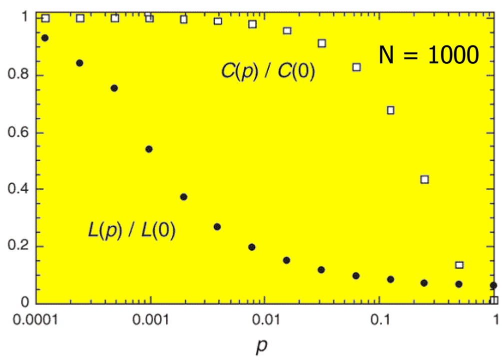
- Undirected Edges
- Un-weighted Edges:- no weight associated with edges
- Sparse:- Number of Edges << total number of edges
- Connected:- All vertices are connected. (no isolated clusters)

In-between: Small-world networks



N nodes forms a regular lattice.
With probability p,
each edge is rewired randomly

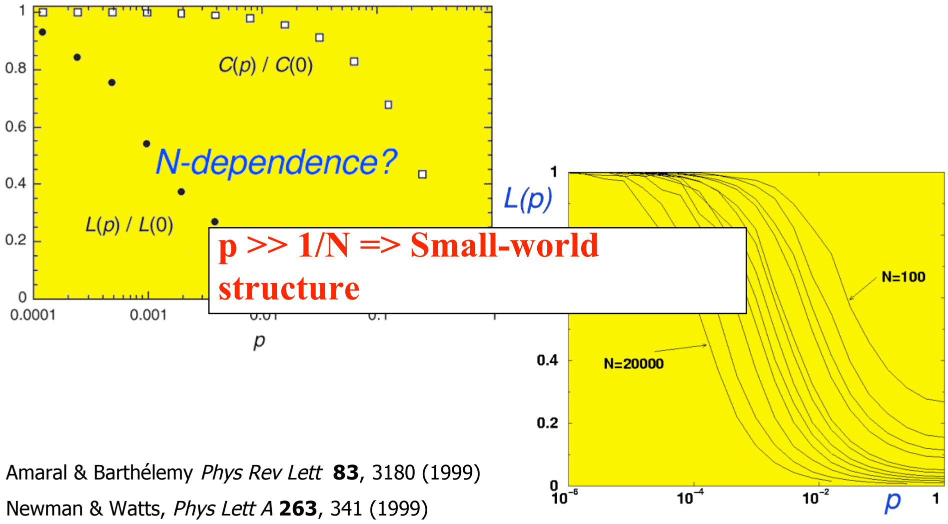
=>Shortcuts



- Large clustering coeff.
- Short typical path

Watts & Strogatz,
Nature **393**, 440 (1998)

Size-dependence



p >> 1/N => Small-world structure

Amaral & Barthélemy *Phys Rev Lett* **83**, 3180 (1999)

Newman & Watts, *Phys Lett A* **263**, 341 (1999)

Barrat & Weigt, *Eur Phys J B* **13**, 547 (2000)

Small World Graphs

Examples !! **Studied by Watts-Strogatz**

- Kevin Bacon Graph (KBG)
- Power Grid (Western US)
- C. elegans Worm

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Is that all we need ?

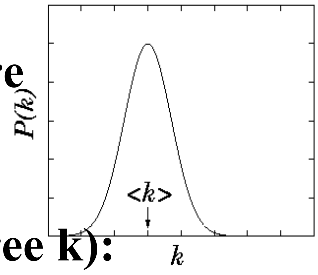
NO, because...

Random graphs,

Watts-Strogatz graphs are

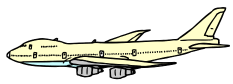
homogeneous graphs

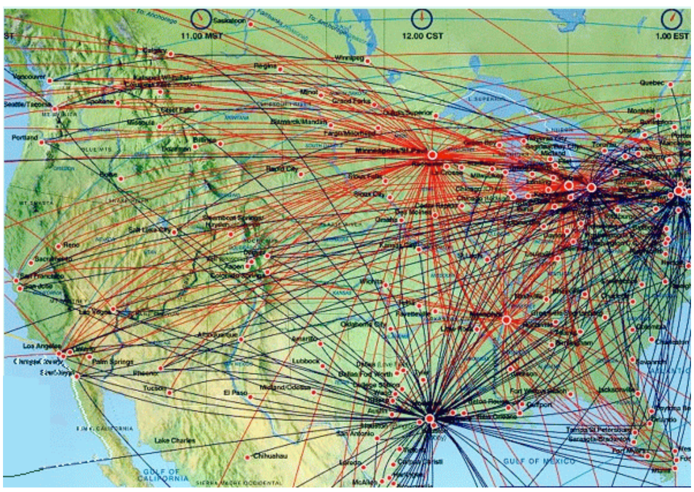
(small fluctuations of the degree k):

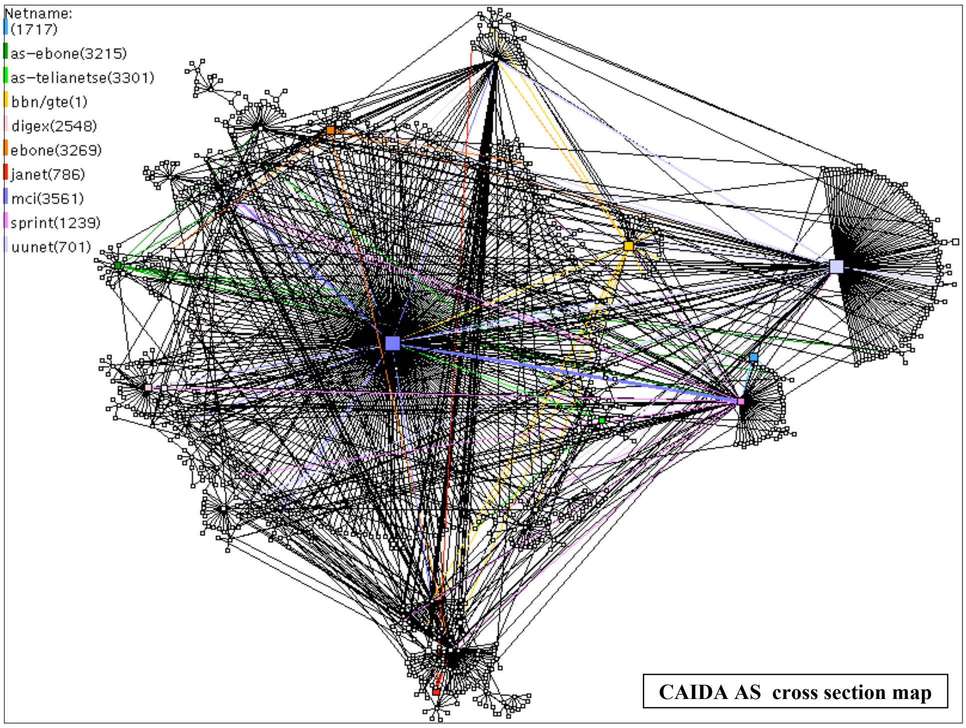


While.....

Airplane route network

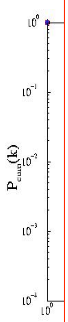






CAIDA AS cross section map

Topological characterization



•The Internet and the World-Wide-Web

•Protein networks

•Metabolic networks

•Social networks

•Food-webs and ecological networks

Are

Heterogeneous networks

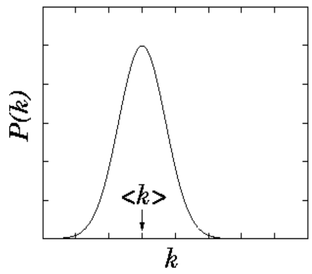
as *k* links

Scale-free properties

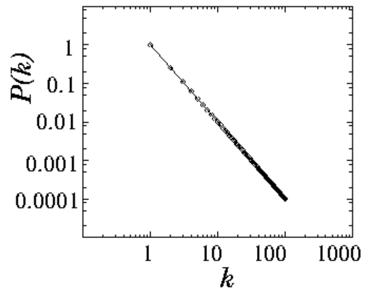
Diverging fluctuations

Exp. vs. Scale-Free

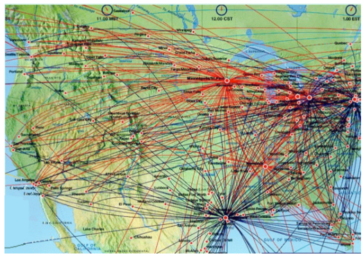
Poisson distribution



Power-law distribution



Exponential Network



Scale-free Network

Two important observations

(1) The number of nodes (N) is NOT fixed.

Networks continuously expand
by the addition of new nodes

Examples:
WWW : addition of new documents
Citation : publication of new papers

(2) The attachment is NOT uniform.

A node is linked with higher probability to a node that already has a large number of links.

Examples :
WWW : new documents link to well known sites (CNN, YAHOO, NewYork Times, etc)
Citation : well cited papers are more likely to be cited again

Origin of SF networks: Growth and preferential attachment

(1) Networks continuously expand by the addition of new nodes

WWW : addition of new documents

(2) New nodes prefer to link to highly connected nodes.

WWW : linking to well known sites

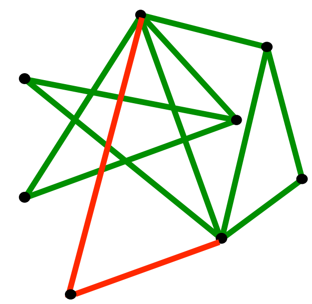
GROWTH:

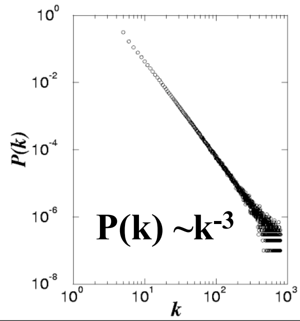
add a new node with m links

PREFERENTIAL ATTACHMENT:

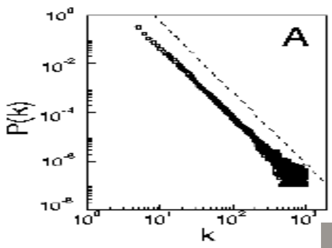
the probability that a node connects to a node with k links is proportional to k.

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



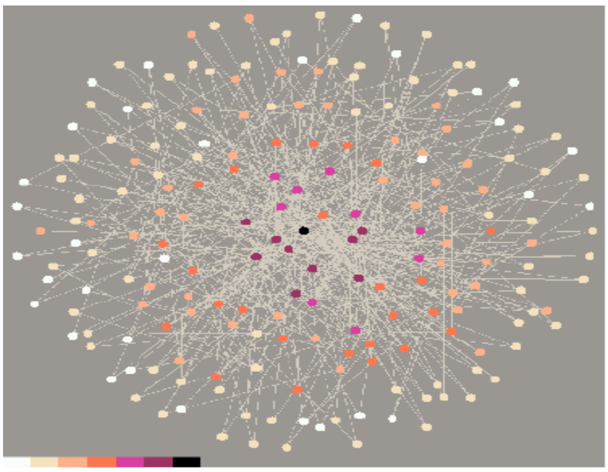


Barabási & Albert, *Science* **286,** 509 (1999)



Connectivity distribution

BA network



More models

•Generalized BA model (Redner et al. 2000)

Non-linear preferential attachment : $\Pi(k) \sim k^{\alpha}$

Initial attractiveness : $\Pi(k) \sim A+k^{\alpha}$ 0)

Rewiring

(Albert et al. 2000)

•Highly clustered

(Dorogovtsev et al. 2001)

(Eguiluz & Klemm 2002)

•Fitness Model

(Bianconi et al. 2001)

$$\Pi(k_i) \cong \frac{\eta_i k_i}{\sum_j \eta_j k_j}$$

•Multiplicative noise

(Huberman & Adamic 1999)

(....)

Tools for characterizing the various models

- Connectivity distribution P(k)
  =>Homogeneous vs. Scale-free
- Clustering
- Assortativity
- ...

=> Compare with real-world networks

Contents

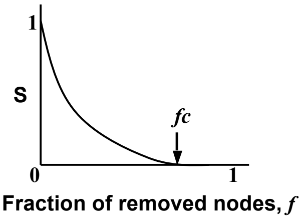
Consequences of the topological heterogeneity

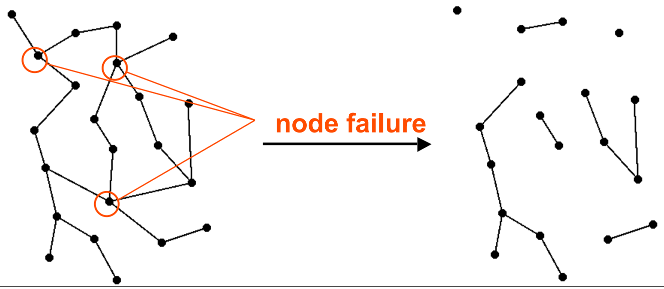
- **Robustness and vulnerability**
- **Propagation of epidemics**

Robustness

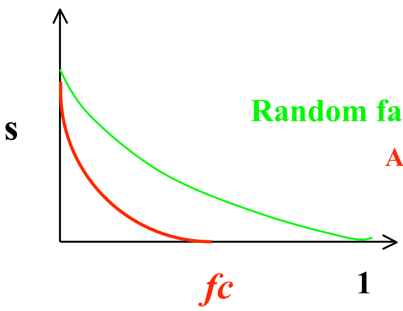
Complex systems maintain their basic functions even under errors and failures (cell → mutations; Internet → router breakdowns)

S: fraction of giant component





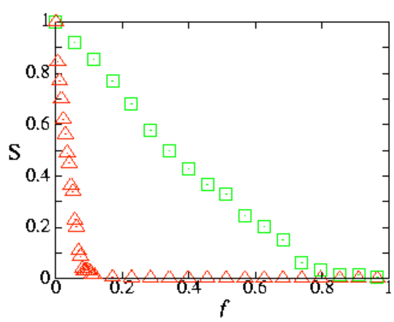
Case of Scale-free Networks



Random failure fc =1 (2 < g ≤ 3)

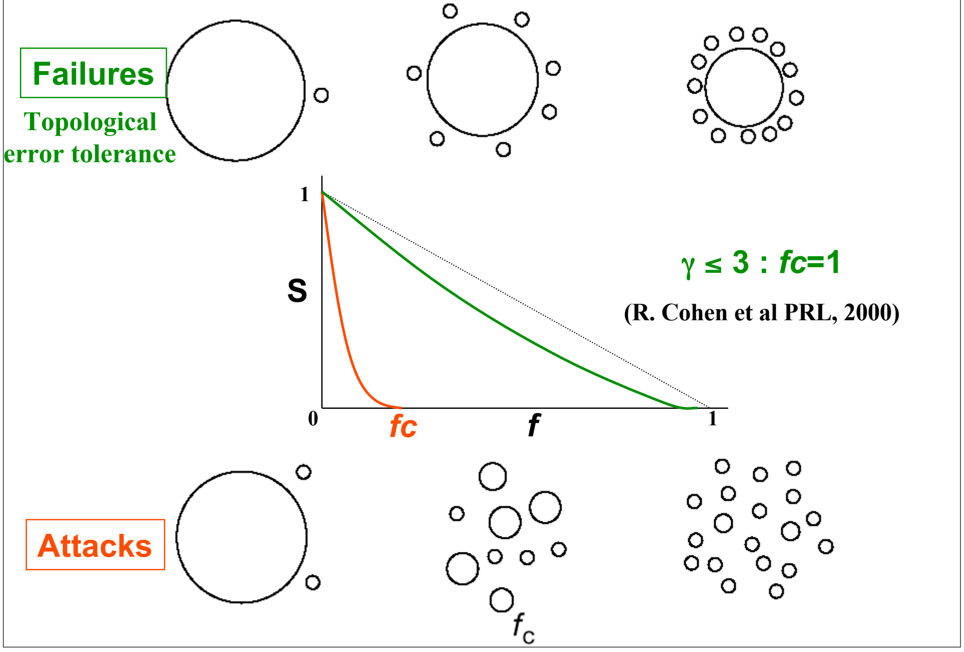
Attack =progressive failure of the most connected nodes fc <1

Internet maps



R. Albert, H. Jeong, A.L. Barabasi, Nature 406 378 (2000)

Failures vs. attacks



Other attack strategies

- Most connected nodes
- Nodes with largest betweenness
- Removal of links linked to nodes with large k
- Removal of links with largest betweenness
- Cascades
- …

Epidemic spreading on SF networks

Natural computer virus

- DNS-cache computer viruses
- Routing tables corruption

→ Internet topology

Data carried viruses

- ftp, file exchange, etc.

→ E-mail network topology

Computer worms

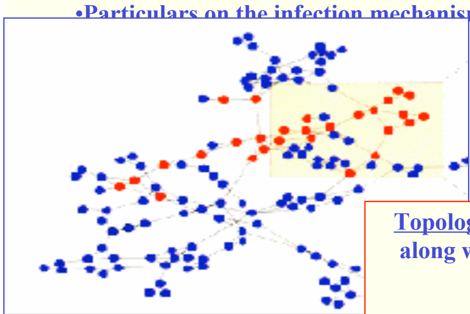
- e-mail diffusing
- self-replicating

Epidemiology → Air travel topology

Mathematical models of epidemics

Coarse grained description of individuals and their state

- Individuals exist only in few states:
- Healthy or Susceptible * Infected * Immune * Dead
- Particulars on the infection mechanism on each individual are neglected.



Topology of the system: the pattern of contacts along which infections spread in population is identified by a network

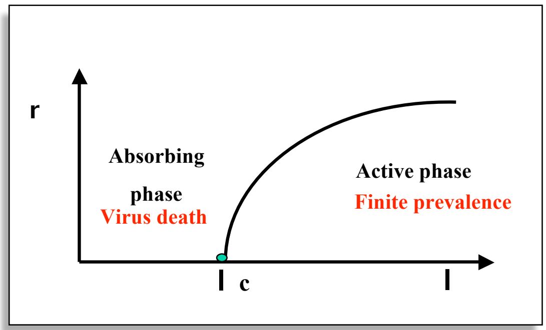
- Each node represents an individual
- Each link is a connection along which the virus can spread

SIS model:

•Each node is infected with rate n if connected to one or more infected nodes

•Infected nodes are recovered (cured) with rate d without loss of generality d =1 (sets the time scale)

•Definition of an effective spreading rate l =n/d



r =prevalence

•Non-equilibrium phase transition

•epidemic threshold = critical point

•prevalence r =order parameter

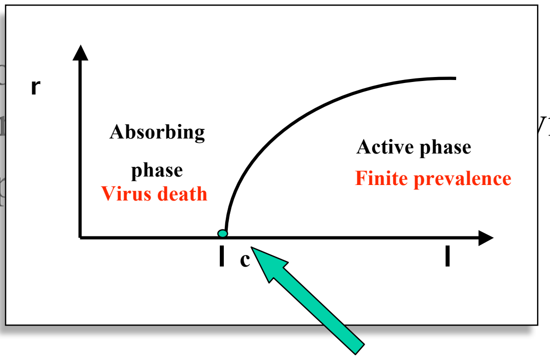
The epidemic threshold is a general result

The question of thresholds in epidemics is central

(in particular for immunization strategies)

What about computer viruses?

- Very lo… compar… …irus
- Small p…



Computer viruses ???

Long lifetime + low prevalence = computer viruses always tuned infinitesimally close to the epidemic threshold

???

SIS model on SF networks

SIS= Susceptible – Infected – Susceptible

Mean-Field usual approximation: all nodes are "equivalent" (same connectivity) => existence of an epidemic threshold 1/<k> for the order parameter ρ (density of infected nodes)

Scale-free structure => necessary to take into account the strong heterogeneity of connectivities => **ρ_k=density of infected nodes of connectivity k**

=>epidemic threshold $\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$

Epidemic threshold in scale-free networks

$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}$

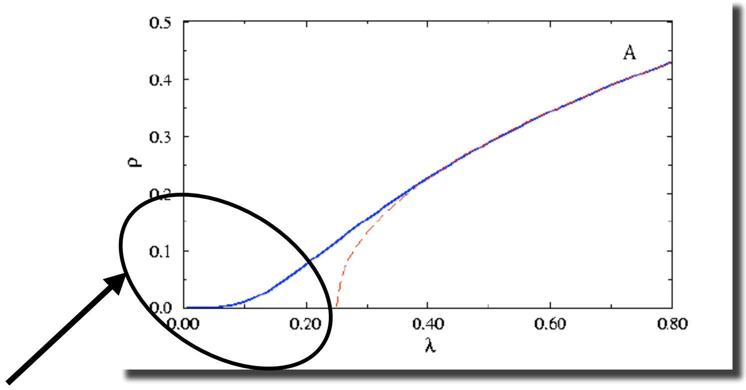
$\langle k^2 \rangle \rightarrow \infty$

$\lambda_c \rightarrow 0$

Order parameter behavior in an infinite system

$$\rho = 2e^{-1/m\lambda}$$

Rationalization of computer virus data



•Wide range of spreading rate with low prevalence (no tuning)

•Lack of healthy phase = standard immunization cannot

Results can be generalized to ge

scale-free connectivity distributions

•If $2 < g \leq 3$ we have absence of an epidemic threshold and no critical behavior.

•If $3 < g \leq 4$ an epidemic threshold appears, but it is approached with vanishing slope (no criticality).

•If $g > 4$ the usual MF behavior is recovered. SF networks are equal to random graph.

Main results for epidemics spreading on SF networks

•Absence of an epidemic/immunization threshold

•The network is prone to infections (endemic state always possible)

•Small prevalence for a wide range of spreading rates

•Progressive random immunization is totally ineffective

•Infinite propagation velocity

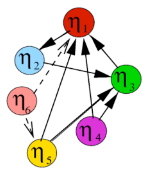
Very important consequences of the SF topology!

(NB: Consequences for immunization strategies)

Pastor-Satorras & Vespignani (2001, 2002), Boguna, Pastor-Satorras, Vespignani (2003),

First Mover Advantage

Network



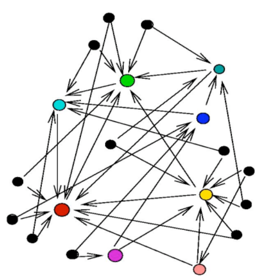
Scale Free Model: First mover advantage

Real Systems: nodes compete for links

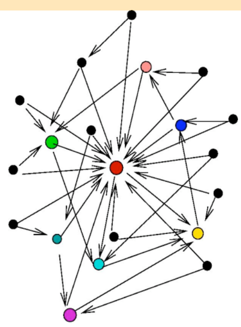
Fitness (h): $P(k) \sim hk$

$k(t) \sim t^{ha}$

Fit-gets-rich



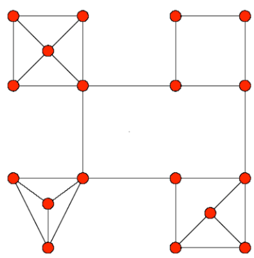
Bose-Einstein condensation

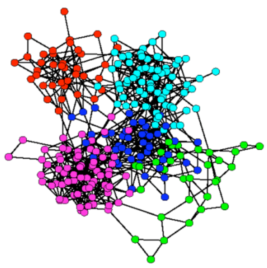


G. Bianconi and A.-L. Barabási, *Physical Review Letters* (2001)

Modularity

- Real networks are fragmented into group or modules
  - Society: Granovetter, M. S. (1973) ; Girvan, M., & Newman, M.E.J. (2001); Watts, D. J., Dodds, P. S., & Newman, M. E. J. (2002).
  - WWW: Flake, G. W., Lawrence, S., & Giles. C. L. (2000).
  - Biology: Hartwell, L.-H., Hopfield, J. J., Leibler, S., & Murray, A. W. (1999).
  - Internet: Vasquez, Pastor-Satorras, Vespignani(2001).
- Traditional view of modularity:

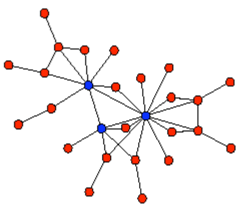




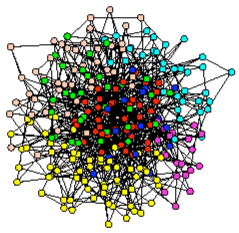
Ravasz, Somera, Mongru, Oltvai, A-L. B, ***Science*** **297,** 1551 (2002).

Modular vs. Scale-free Topology

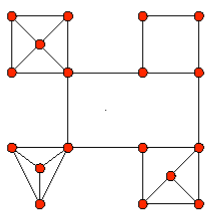
(a)



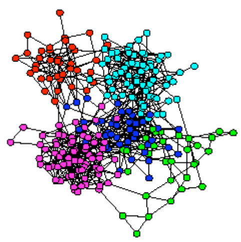
Scale-free



(b)



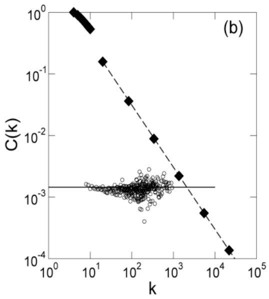
Modular

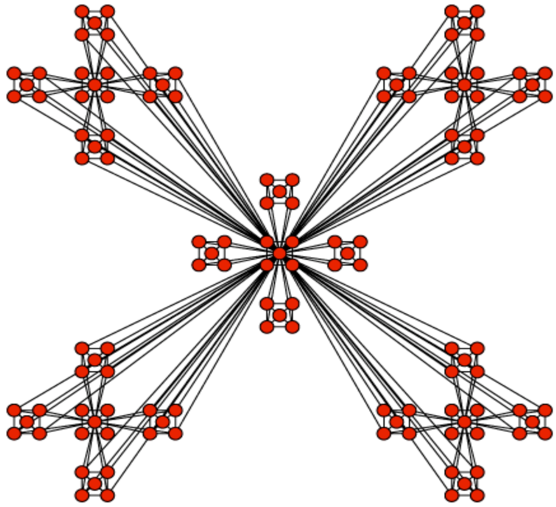


Hierarchical Networks

3. Clustering coefficient scales

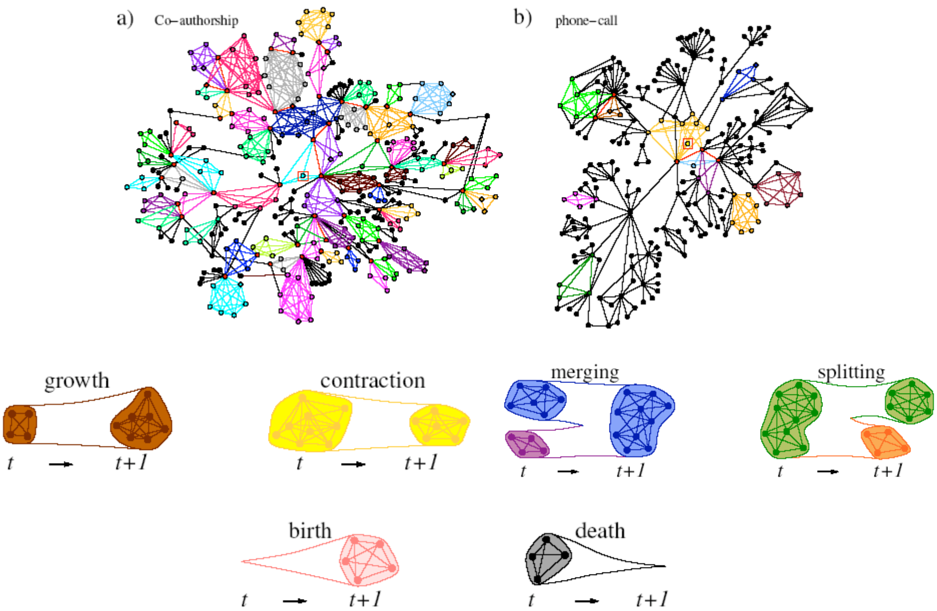
$$C(k) \sim k^{-1}$$





$$C(k) = \frac{\text{\# links between k neighbors}}{k(k-1)/2}$$

Finding Overlapping Communities in Networks



Palla, ALB, Vicsek