Complex Networks: Ubiquity, importance and implications

A.Vespignani

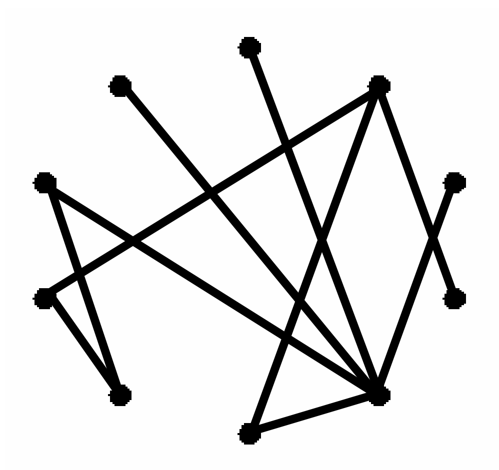
Indiana University School of informatics

the biocomplexity institute
Indiana University

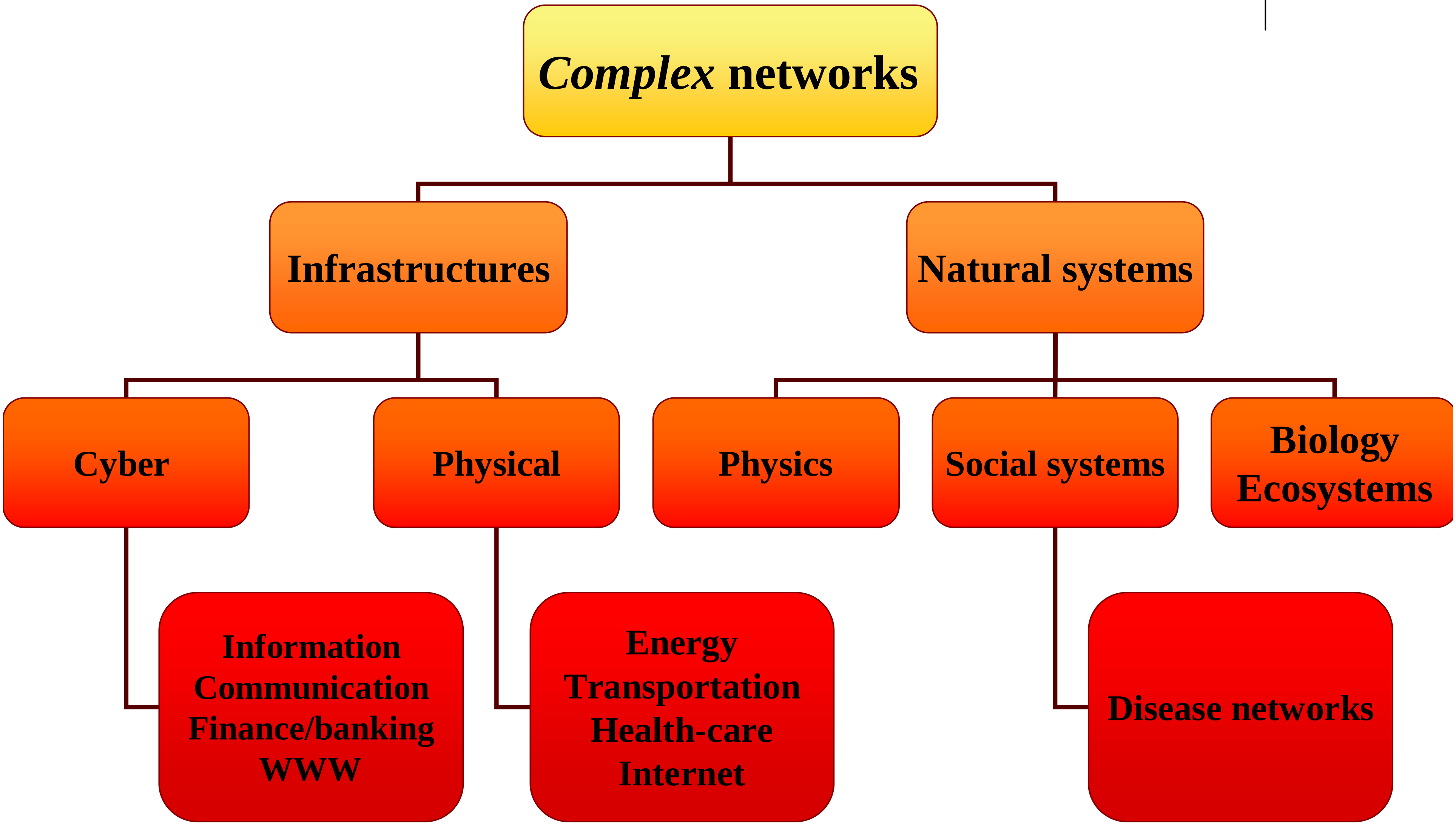
A network is a system that allows its abstract/mathematical representation as a graph

Vertices (nodes) = elements of the system

Edges (links) = interactions/relations among the elements of the system

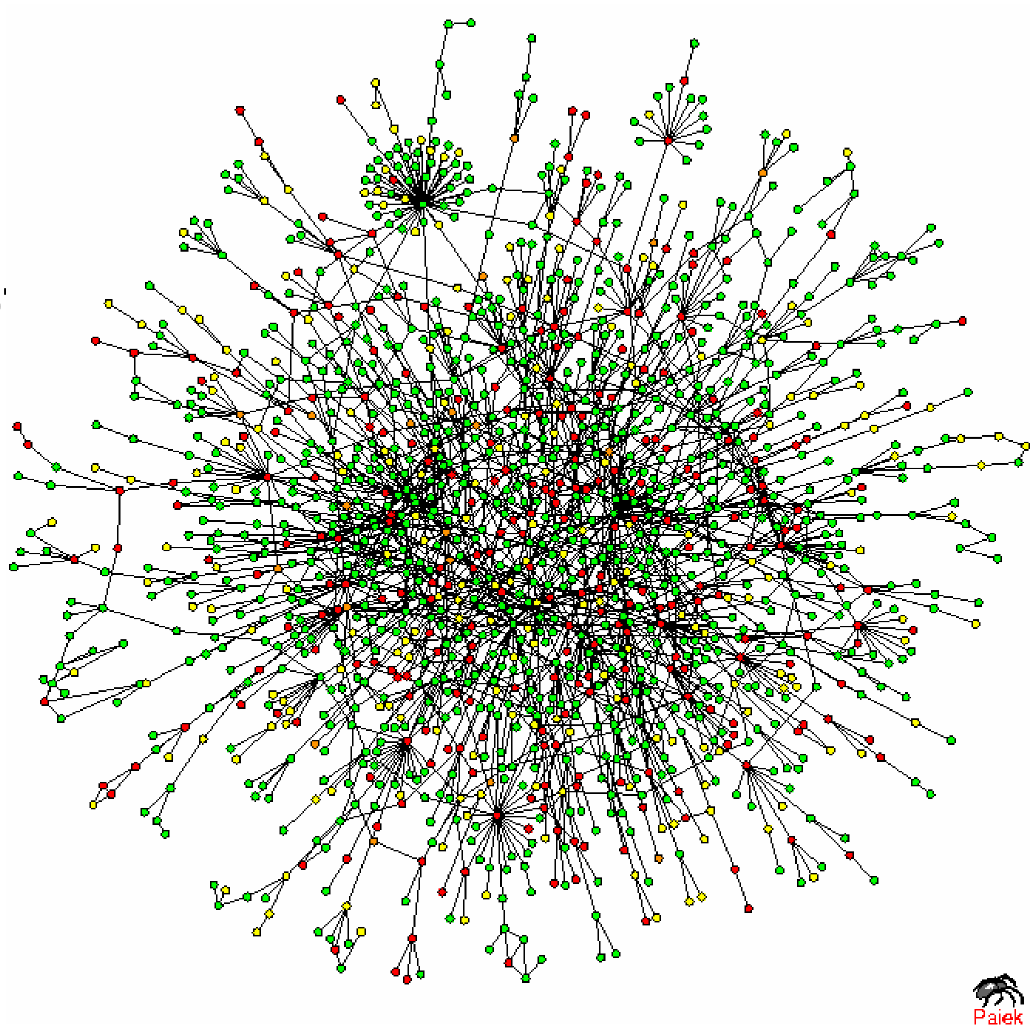


Many *complex* systems (from the molecular level to the scale of large communications infrastructures) can be regarded as a collection of inhomogeneously and generically interacting units.



Protein Interaction network (PIN)

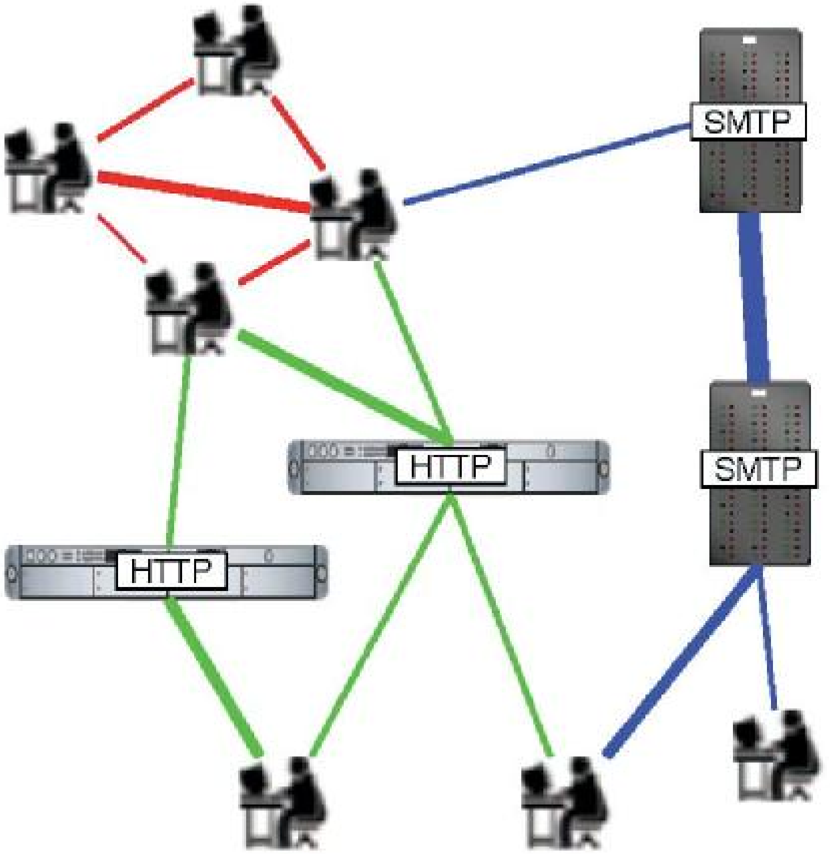
- **A protein interaction network is the set of binary interactions among the proteins o a given proteome**



- **Nodes**: proteins
- **Links**: physical interactions (binding)

Social networks

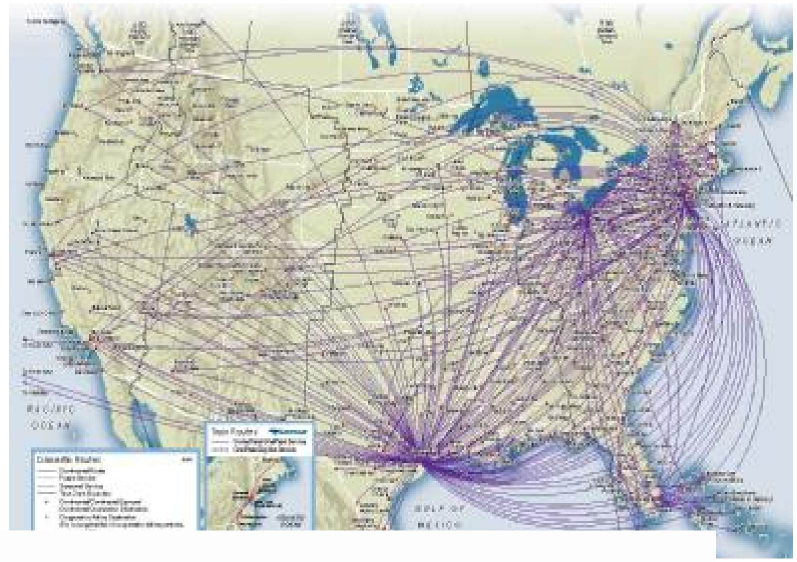
- **Nodes**
  - Individuals
- **Edges**
  - Relationship
  - Communications
  - Interactions

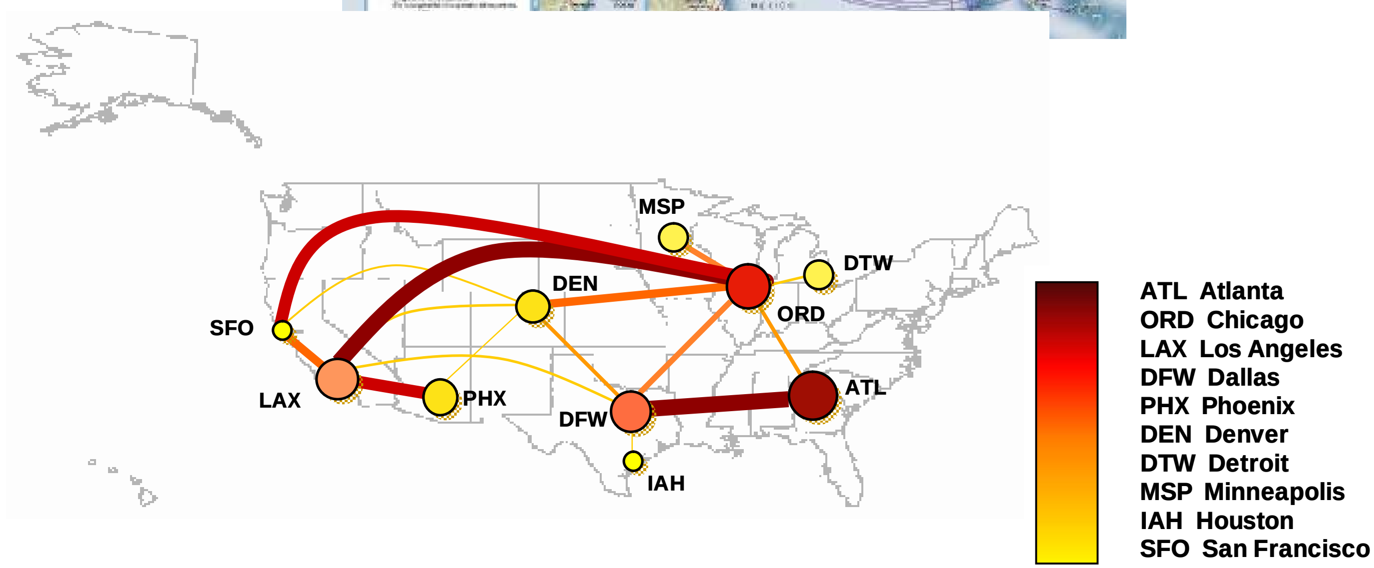


Social network analysis

Airport network

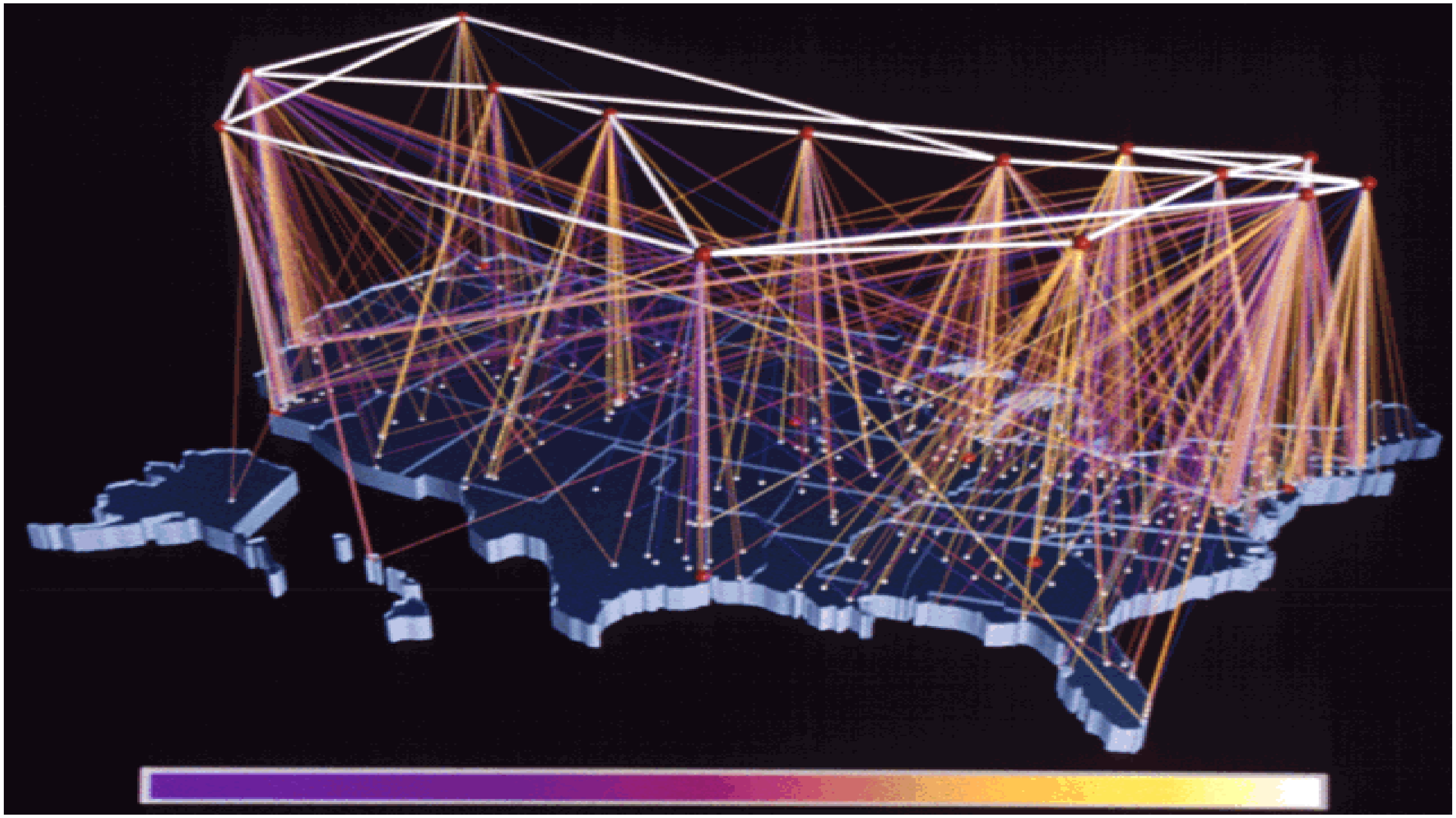
- Each edge is characterized by weight *wij* defined as the number of passengers in the year





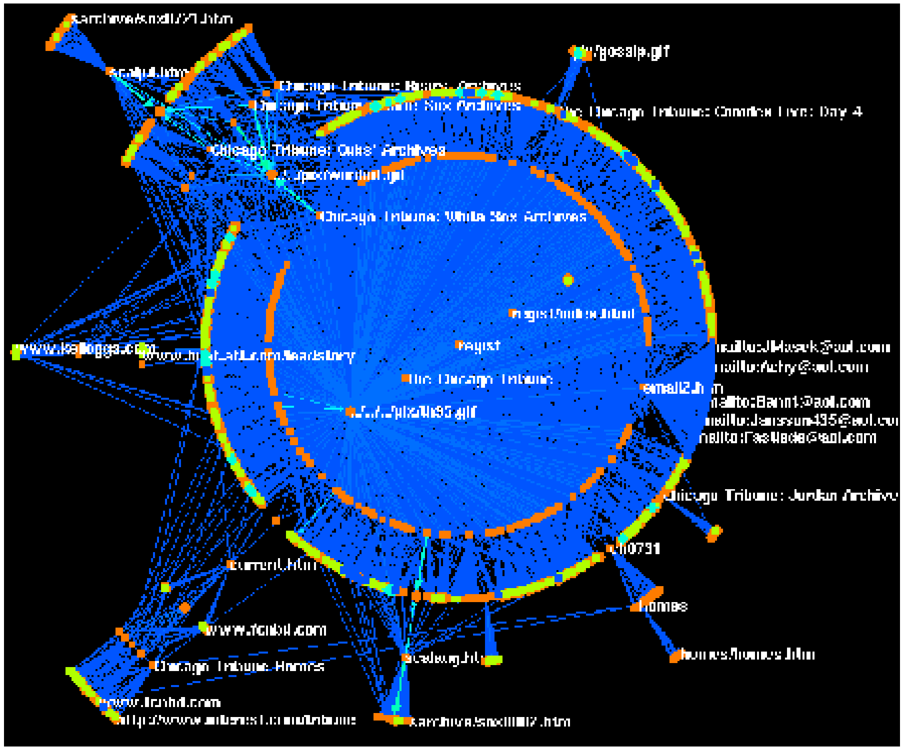
Physical Internet

- Computers (routers)
- Satellites
- Modems (??)
- Phone cables
- Optic fibers
- Wireless



The World-Wide Web

- Virtual network to find and share information
  - Web pages
  - Hyperlinks



What’s new ??

- Euler circa 1736 (Koningsberg problem)
- Moreno ’34 (sociogram) – social
- Erdos ’60-70 (random graph theory) – math
- ……….

Comp… networks

Size does matter !!!

- Starting in ’96 large scale internet and web measurements
- High throughput experiments in Biology
- Electronic databases and indexing

Networks sizes…

- Biology
  - Genome(s), regulatory networks, metabolic networks protein interaction networks (10^3- **10^4** nodes)
- Social network
  - Co-authorship, citations, patents, grants, e-mails, P2P, instant messaging………(10^3 – **10^7** nodes)
- Physical Internet
  - ISP level (10^4 nodes), Router level (10^5-10^6), Host level (10^7 nodes)
- WWW
  - Web pages (Url address 10^8 -**10^9** nodes)

What's new….

- Data size shifts (10^2 -> 10^8 elements)

  (Complexity??)

- Different domains (biology, info-structures, infrastructures, social, scientometrics)

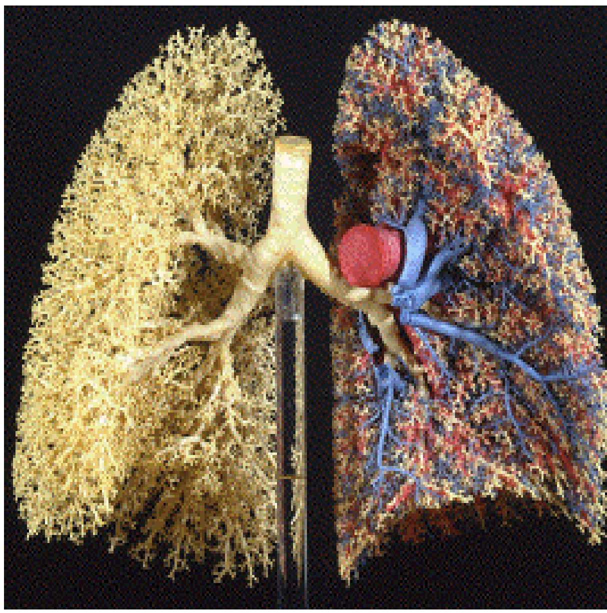
  (universality ??)

- Large scale longitudinal studies (time series)

  (dynamical modeling??)

Complexity = very complicated

Complications at all scales (compatible with the finite world)



Complex Features

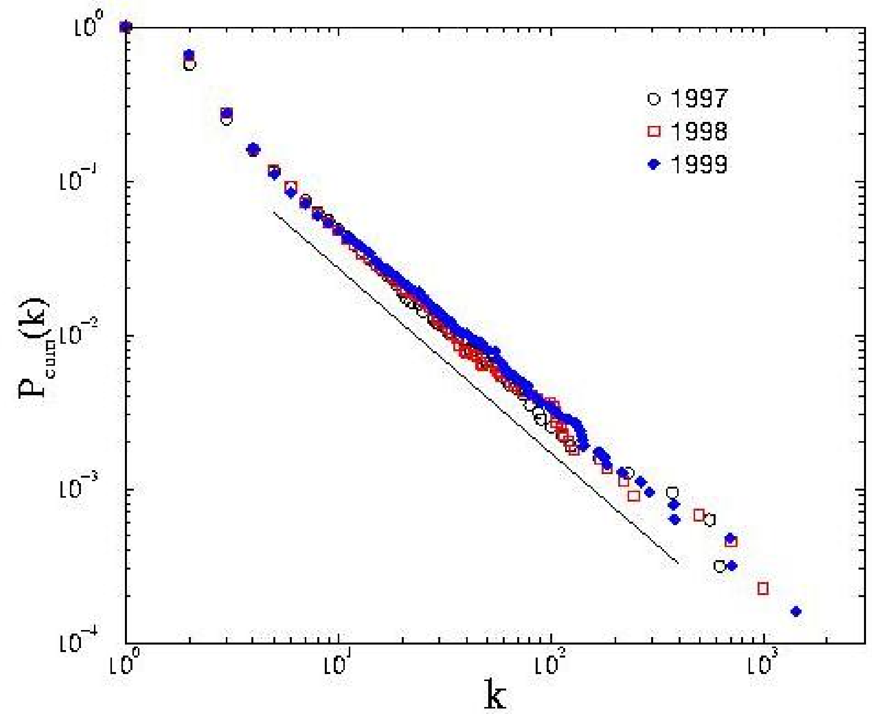
(symptoms)

Heavy tailed statistical distribution (degree etc.)

P(k) **= probability that a node has *k* links**

$$P(k) \sim k^{-g}$$

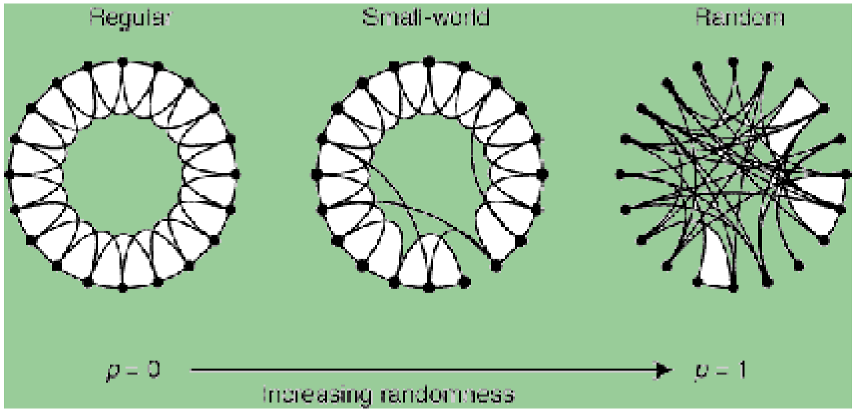
- $<k>$= const
- $<k^2>$ ® ¥

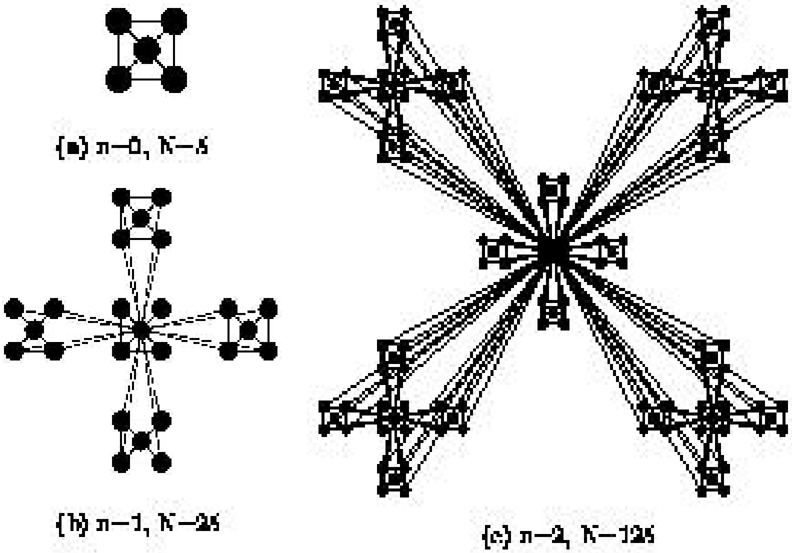


Scale-free properties ⟷ **Diverging fluctuations**

(more) Complex features

- Small-world + clustering
- Assortativity/correlations
- Community structure
- Motifs
- ....

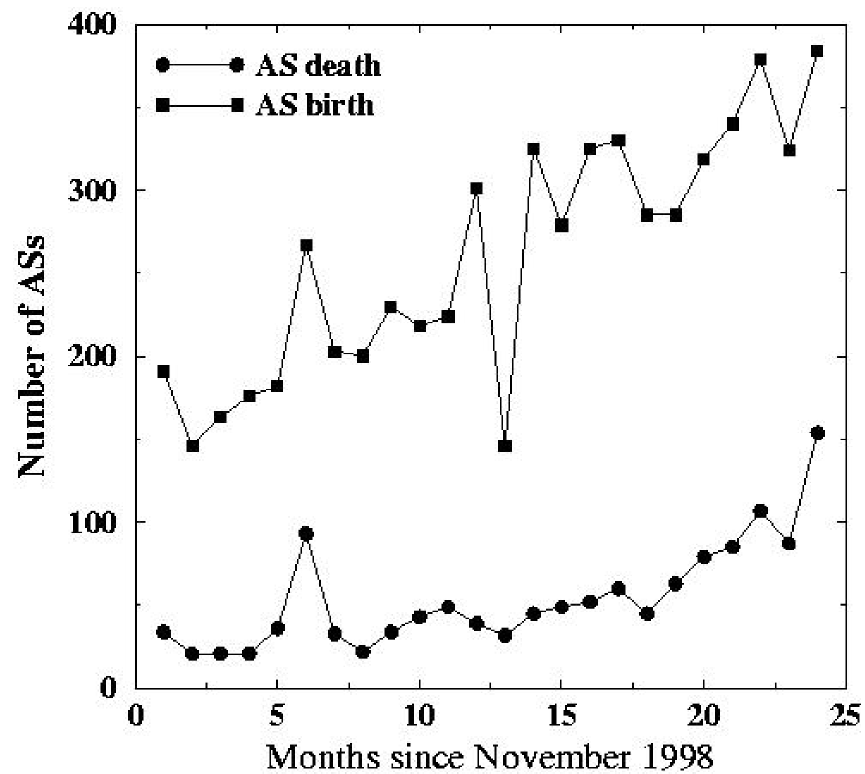




The Internet growth

- 1997
  - **3112 AS**
- 2000
  - **9107 AS**

- **<u>In 1999:</u>**
  - **3410 new AS**
  - **1713 lost AS**



Self-organizing and evolving system

Main ingredient for complex systems

- Many interacting units
- Dynamical evolution
- Self-organization

Supervising entity **Project/blueprint**

↓

Non-trivial architecture
Unexpected emergent properties
Cooperative phenomena

Complexity features

Shift of focus:

static construction → **dynamical evolution**

- To each network realization ***X*** corresponds probability *P(X)*

$$P(X) = F(q_i, z_i)$$

- θ_i = set of model parameters
- Z_i = network statistical observables

Parameters θ_i to be estimated from the real data

(vast traditions in the social literature/ large amount of techniques)

Shift of focus:

static construction → **dynamical evolution**

- To each network realization $\boldsymbol{X}$ corresponds probability $P(X)$

$$\partial_t P(x,t) = \sum_{y \neq x} \left[P(y,t) w_{(y \to x)} - P(x,t) w_{(x \to y)} \right]$$

The focus shifts on $w_{(x \to y)}$

i.e. the dynamical rules governing the evolution from one configuration to the other (transition rates/probability)

The rich-get-richer mechanism

(Barabasi& Albert 1999)

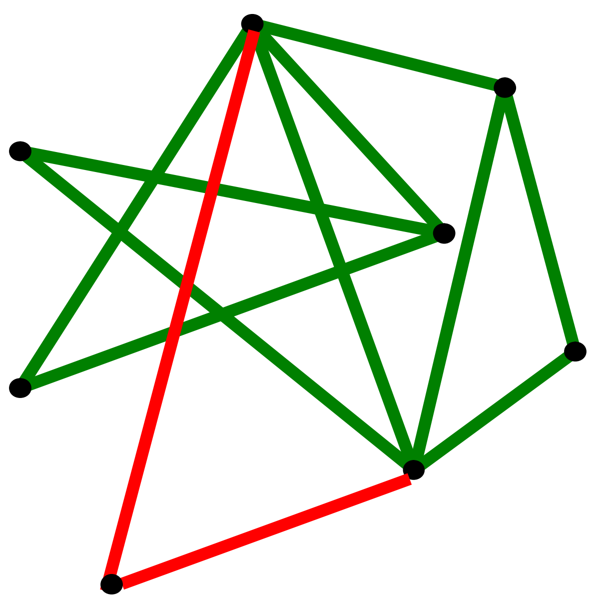
GROWTH :

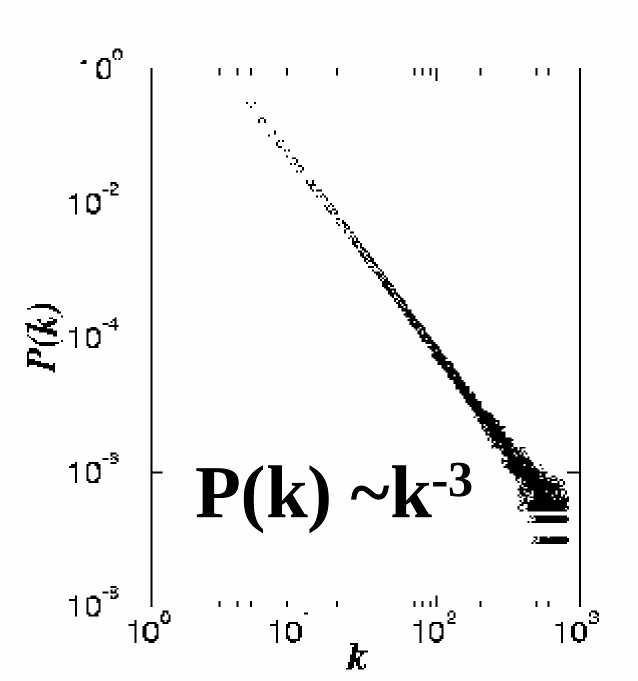
At every timestep we add a new node with m edges (connected to the nodes already present in the system).

PREFERENTIAL ATTACHMENT :

The probability Π that a new node will be connected to node i depends on the connectivity k_i of that node

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

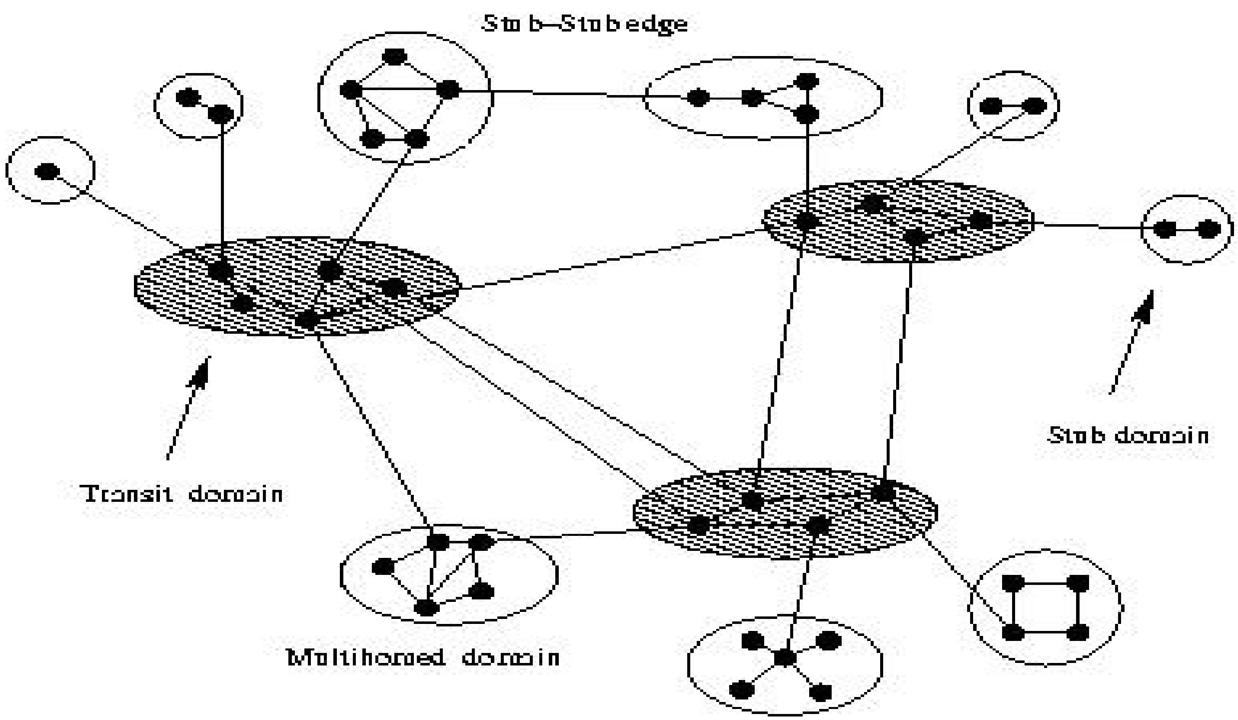




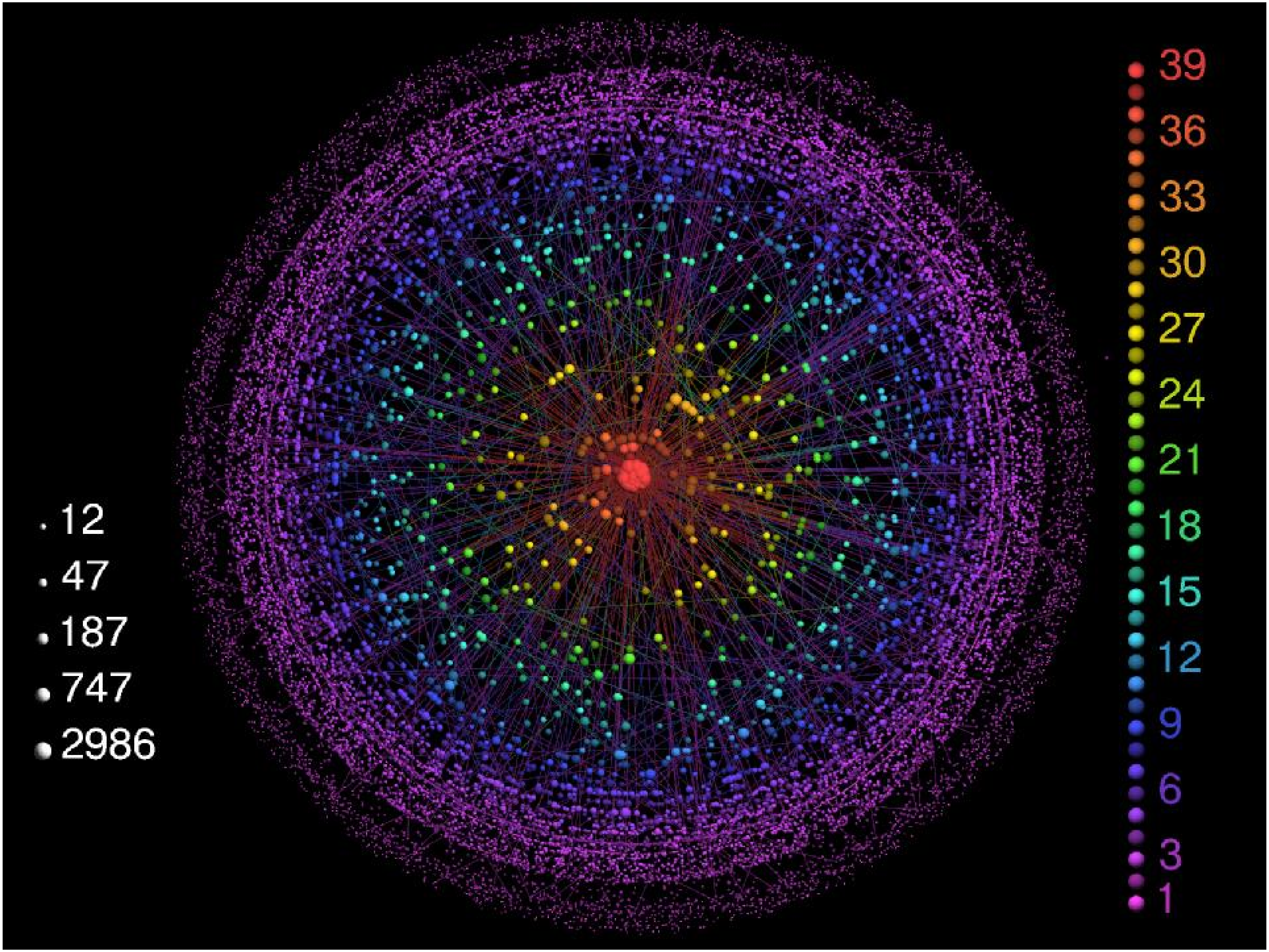
Dynamical approach (pro's)

- Very intuitive/technically easier
- Extremely suitable for large scale simulations/monte-carlo approaches
- In non-equilibrium cases is the only viable approaches.
- Asymptotic and universality

The Internet Hierarchy



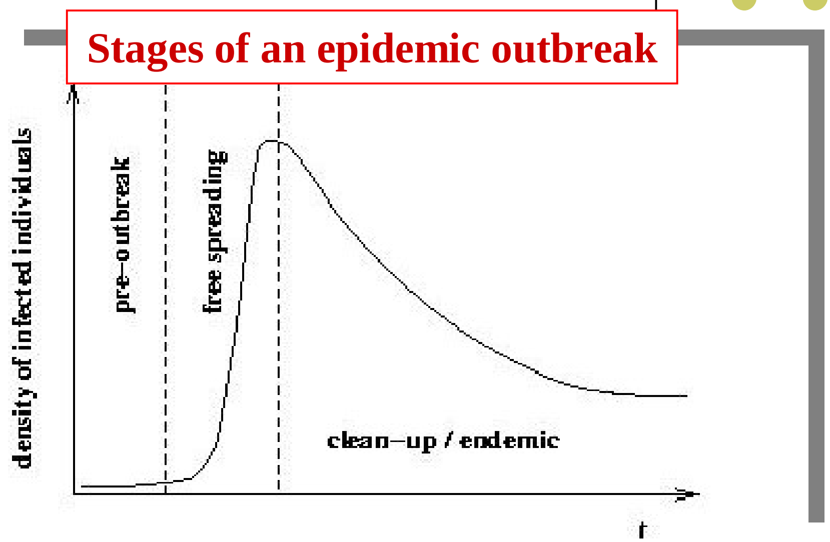
- **Stub AS** : has only one connection to another AS
- **Multi-homed AS**: two or more connections to other ASs but does not carry transit traffic
- **Transit AS**: Two or more connections to other ASs and carries transit traffic

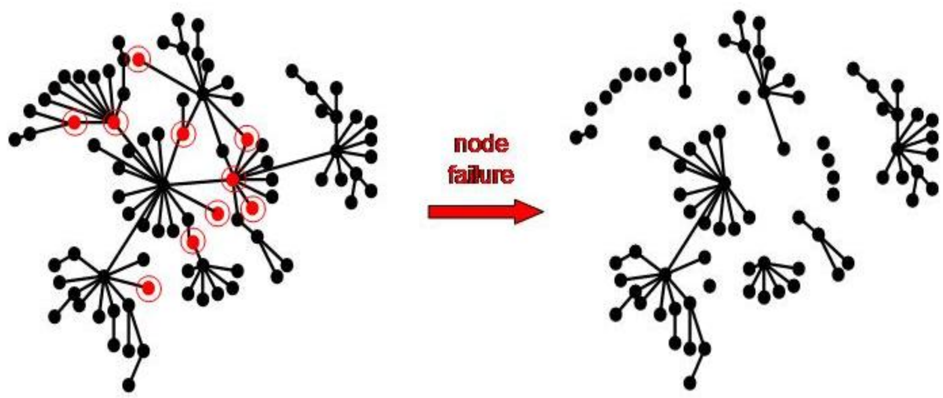


39
36
33
30
27
24
21
18
15
12
9
6
3
1
12
47
187
747
2986

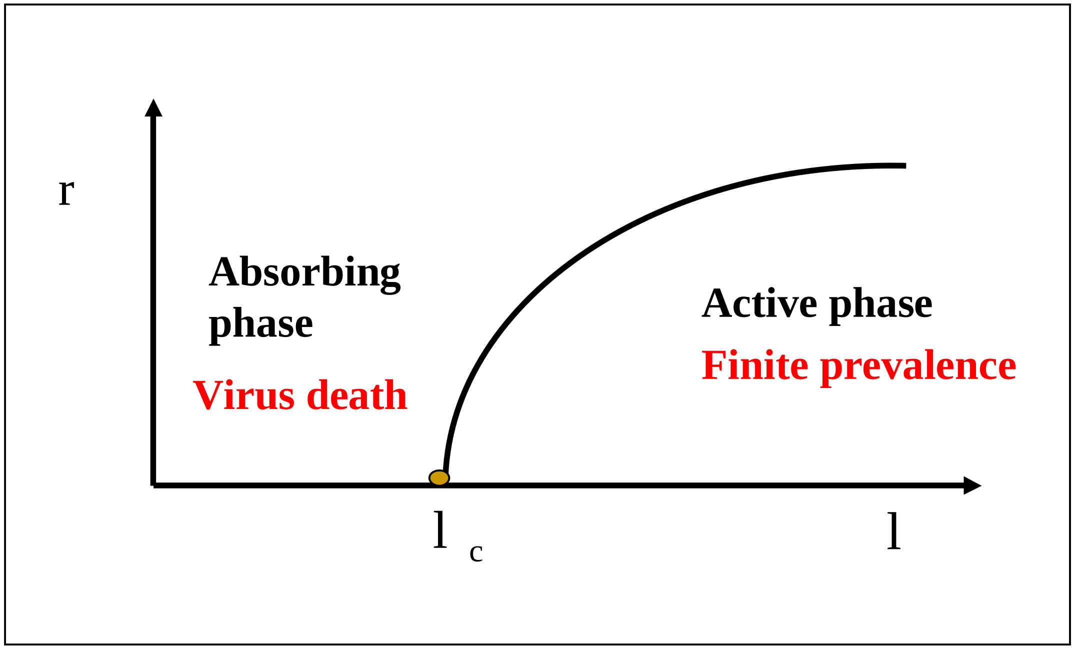
(more) Implications/applications

- Spreading, Diffusion and Finding
  - Epidemic modeling
  - Behavioral patterns
  - Competition/evolution
  - Search algorithm
  - Ranking algorithms
- Resilience and robustness
  - Avalanche
  - Congestions
  - Adaptive control





Epidemic threshold



ρ=I/N

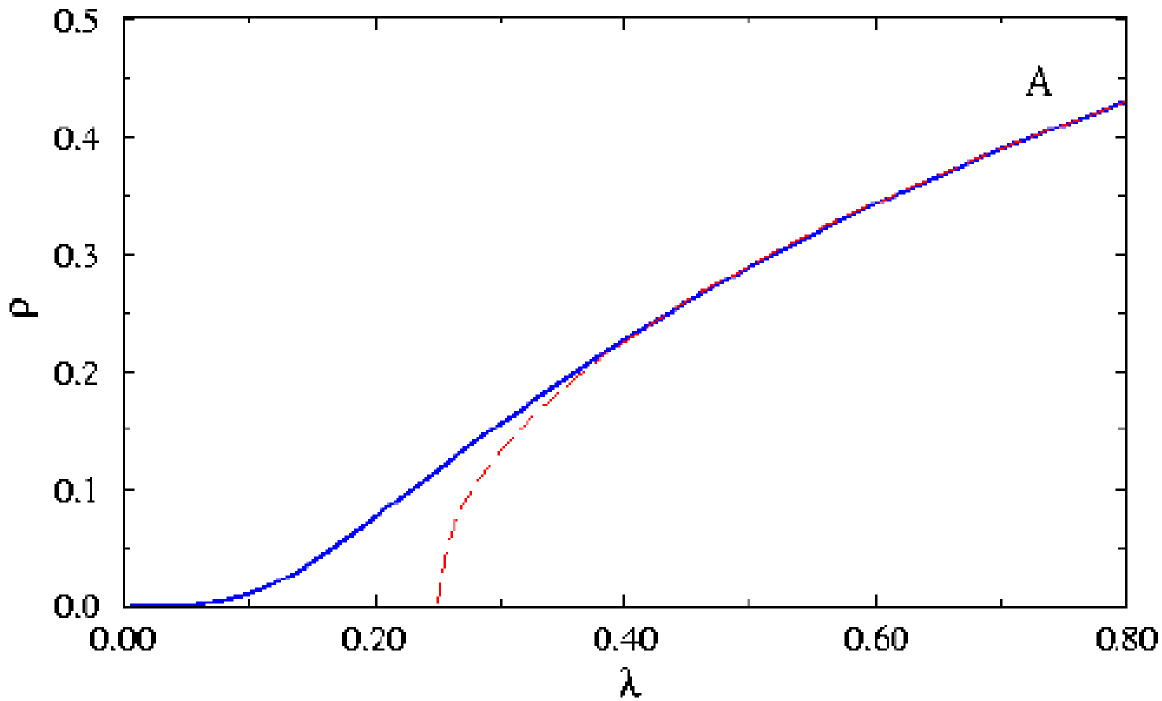
l **=spreading rate**

$\{\lambda=(\beta_0\ k)/\mu\}$

- The Epidemic threshold is related to the reproduction rate R
- The epidemic threshold is a general result (SIR,SIS, etc.)

Scale-free graphs

The healthy phase does not exist



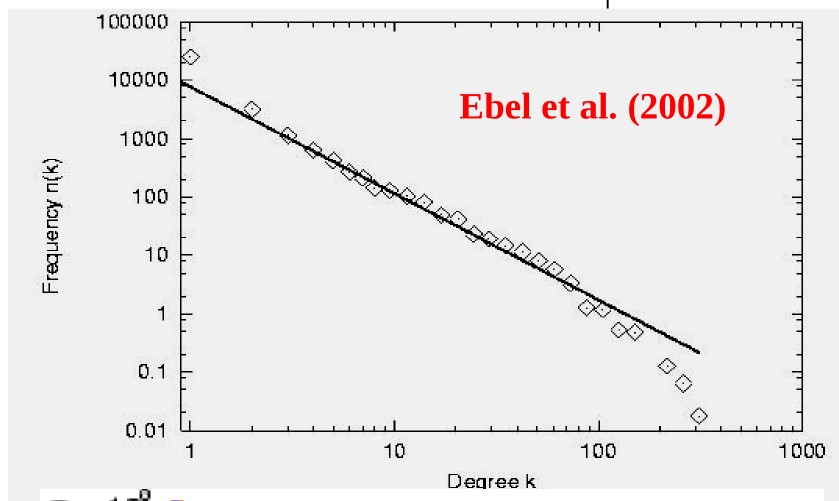
- **Absence of any epidemic threshold (critical point)**
- **Active state for any value of λ**
- **The infection pervades the system whatever spreading rate**
- **In infinite systems the infection is infinitely persistent (indefinite stationary state)**

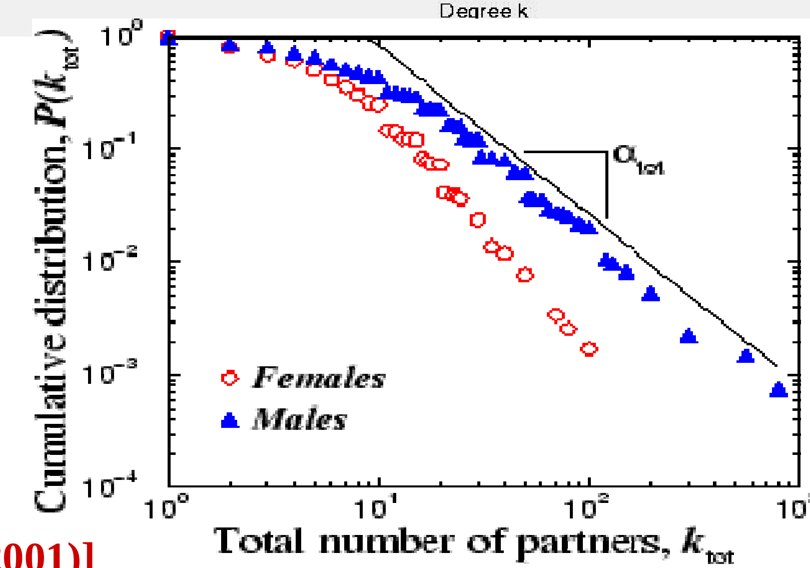
Pastor-Satorras & Vespignani, PRL 86, 3200(2001)

Relevance…..

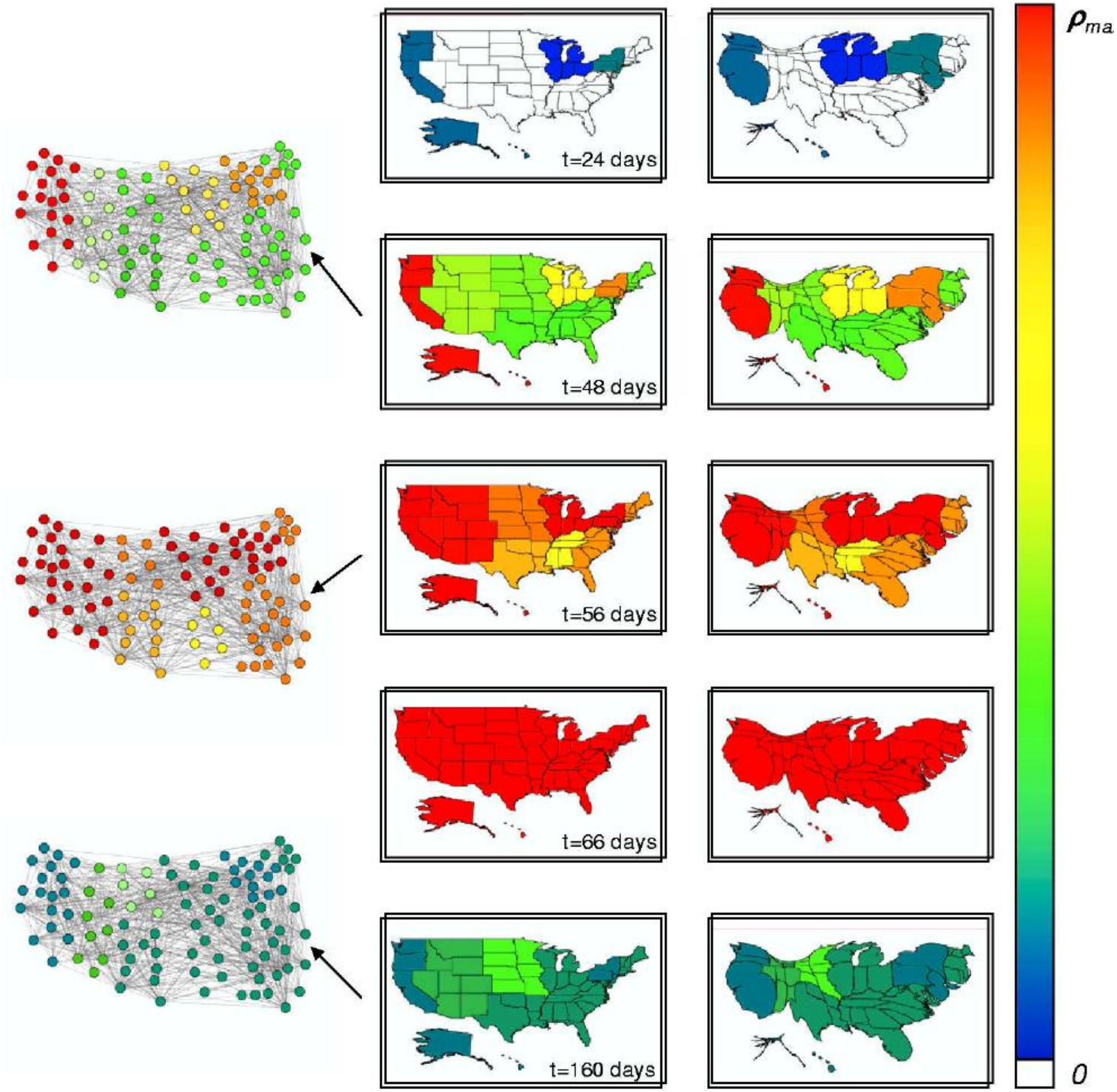
- Infostructures:
  - Computer viruses/worms
- Natural computer virus
  - DNS-cache computer viruses
  - Routing tables corruption
- Transportation networks
  - Airport network
  - Commuter networks
- The web of human sexual contacts

E-mail network



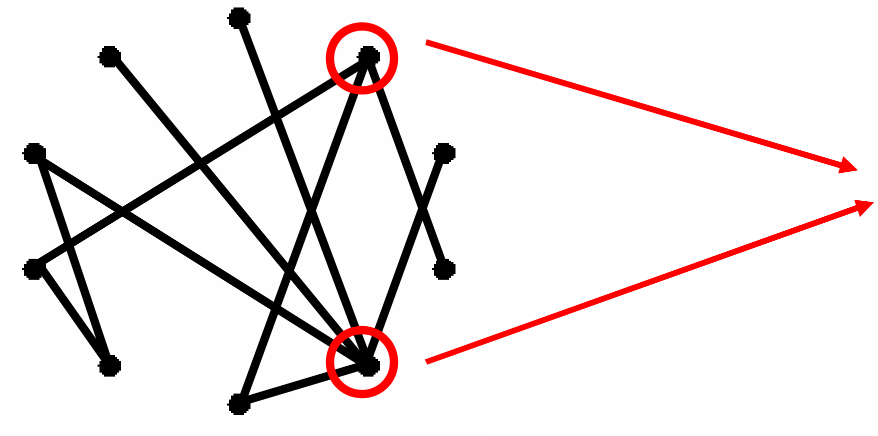


[Lilijeros et al., Nature (2001)]
[Schneeberger et al. STD (2004)]



HK seed

Targeted immunization strategies



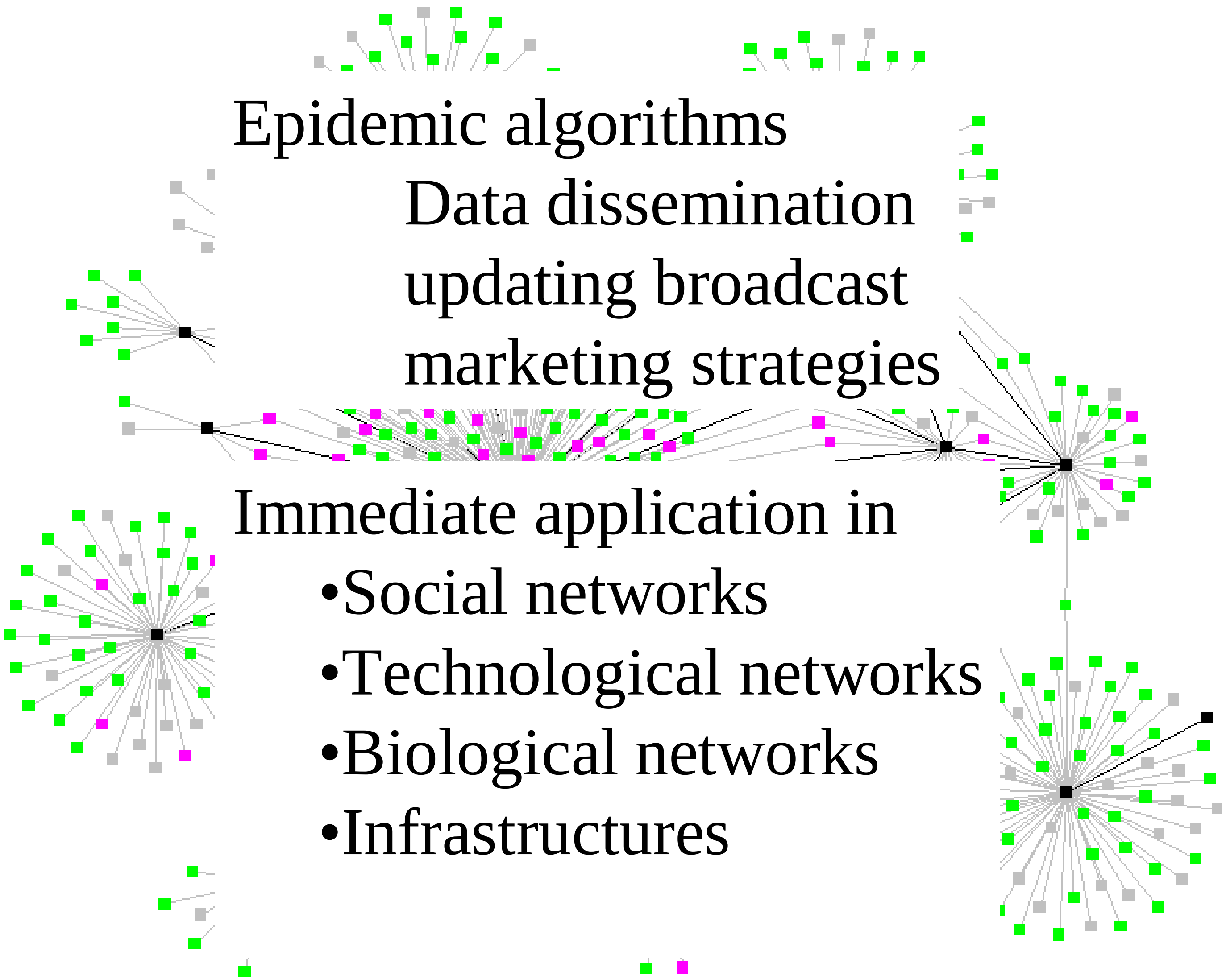
Progressive immunization of crucial nodes

Epidemic threshold is reintroduced

$$g_c = \exp(-2/m\lambda)$$

[Pastor Satorras &Vespignani, PRE 65, 036104 (2002)]

[Dezso & Barabasi cond-mat/0107420; Havlin et al. Preprint (2002)]



Epidemic algorithms

Data dissemination
updating broadcast
marketing strategies

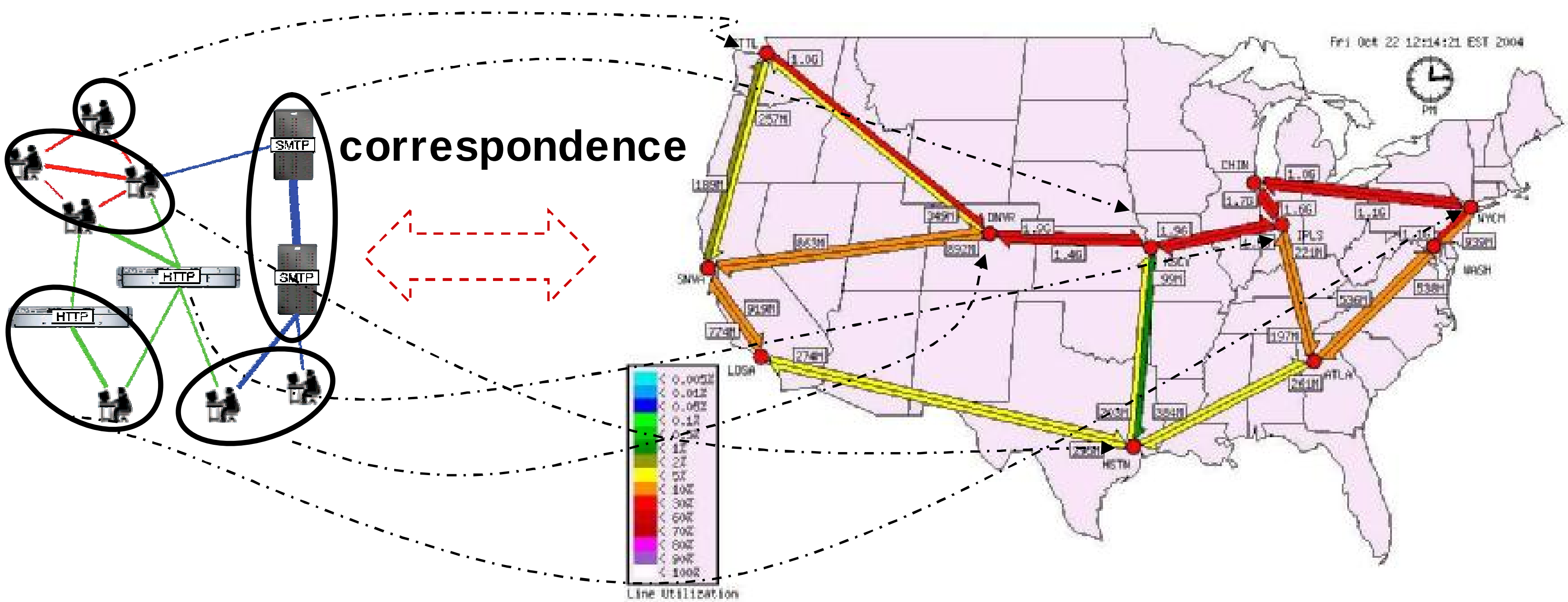
Immediate application in

- •Social networks
- •Technological networks
- •Biological networks
- •Infrastructures

black: opinion leaders
red: influenced
green: uninfluenced
grey: undecided

http://www.orgnet.com

The future/challenge……



Final remarks…..

- Complexity can be accounted for with compact mathematical characterizations (simplicity of complexity)
- Complex features may lead to a change of paradigm (rewarding exercise)
- Dealing with complexity is necessary to discriminate the effect/interplay of the "global and local"
- The study of complexity is needed to extract/identify the "conceptually relevant" from the "merely complicated"