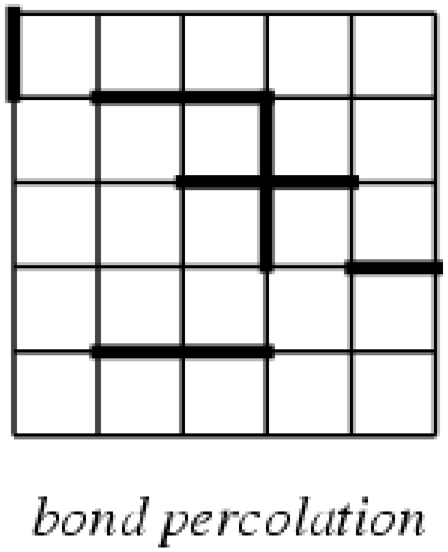
Lecture 9: Processes on Networks I

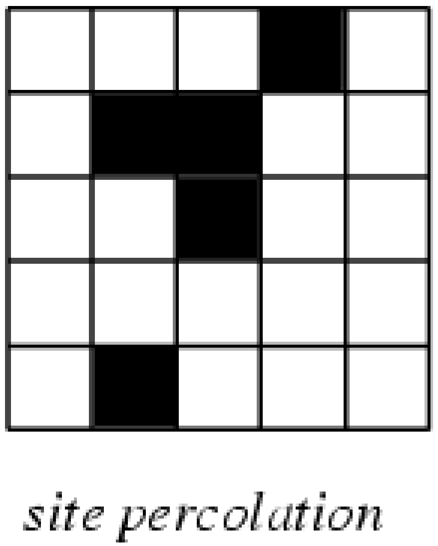
Network Resilience and Robustness

Network Resilience and Percolation in networks

- Q: If a given fraction of nodes or edges are removed…
  - how large are the connected components?
  - what is the average distance between nodes in the components

- Related to percolation
  - We say the network percolates when a giant component forms.





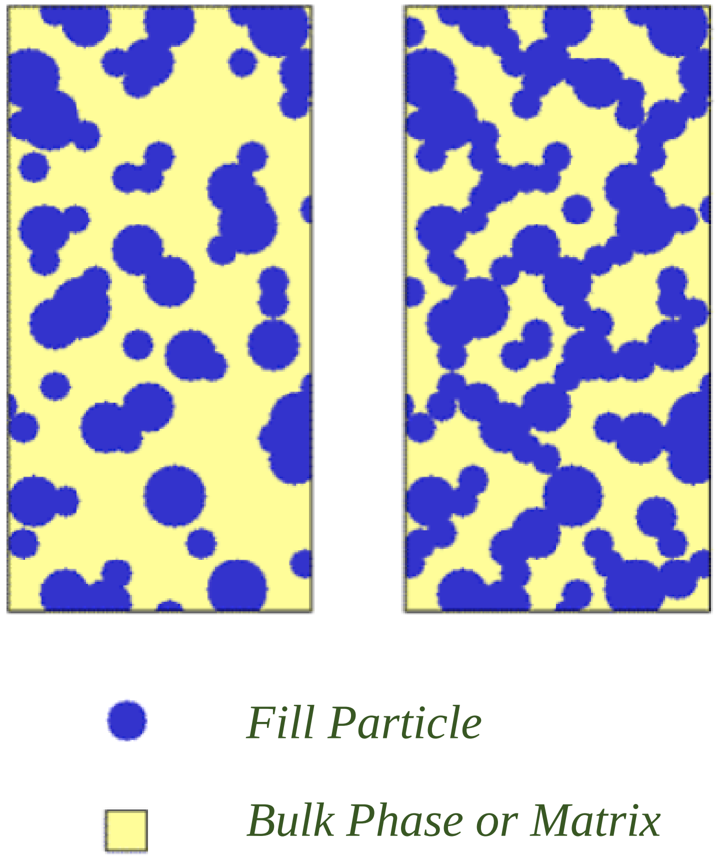
Source: http://mathworld.wolfram.com/BondPercolation.html

Percolation in networks

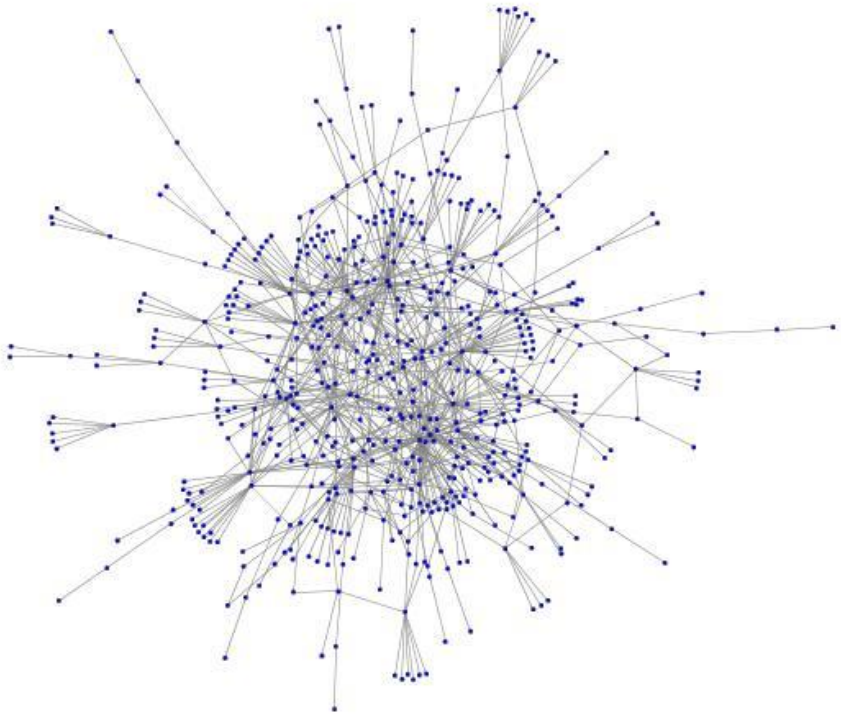
On a rainy day…

- Raindrops falling on the ground: initially, wet regions are isolated and we can find a dry path
- After some point, wet regions become connected and we can find a wet path
- There is a critical density where sudden change happens!

Below the Percolation Threshold *Above the Percolation Threshold*



Percolation in networks



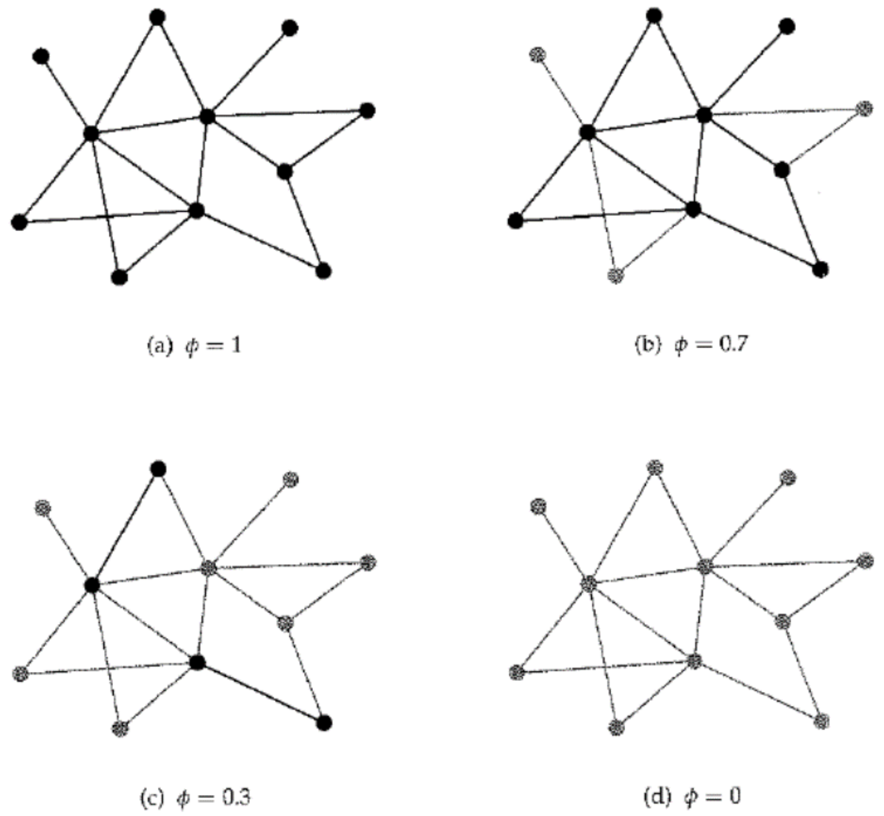
- Percolation can be extended to networks of arbitrary topology
- We say the network percolates when a giant component forms

Percolation in networks

Random Removal

- With probability p, keep node
- With probability $1-p$, remove node and its incident links
- If we remove nodes uniformly at random with probability p, will the remaining network still consist of a large connected cluster (aka the giant component)?
- If so, then we say that the network is ***resilient*** (or robust) to random removal of nodes

Percolation in networks



A depiction of the site percolation process on a small network for various values of the occupation probability ϕ. Gray denotes vertices that have been removed, along with their associated edges, and black denotes those that are still present. The networks in panels (a) and (b) are above the percolation threshold while those in panels (c) and (d) are below it.

When ϕ (occupation probability) is large the vertices tend to be connected together, forming a ***giant component*** (**in percolation, we usually refer components as ***clusters***) that fills most of the network (although there may be small components also). As ϕ decreases there comes a point where the giant component breaks apart and we are left only with small components.

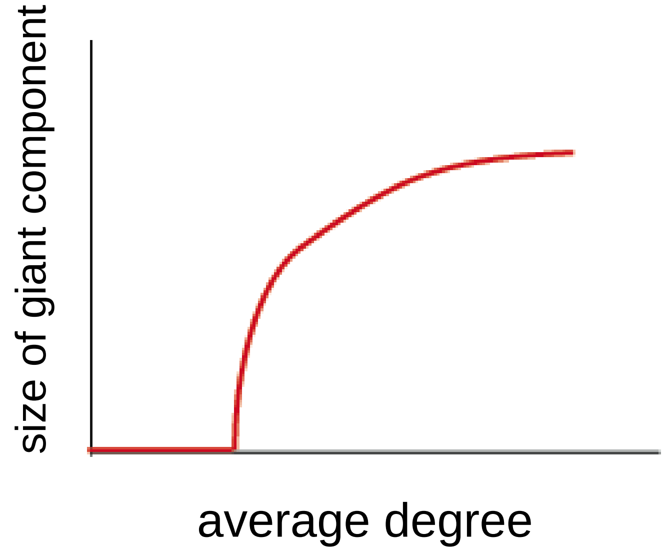
The formation or dissolution of a giant component is called a ***percolation transition***. When the network contains a giant component we say that it ***percolates*** and the point at which the percolation transition occurs is called the ***percolation threshold***.

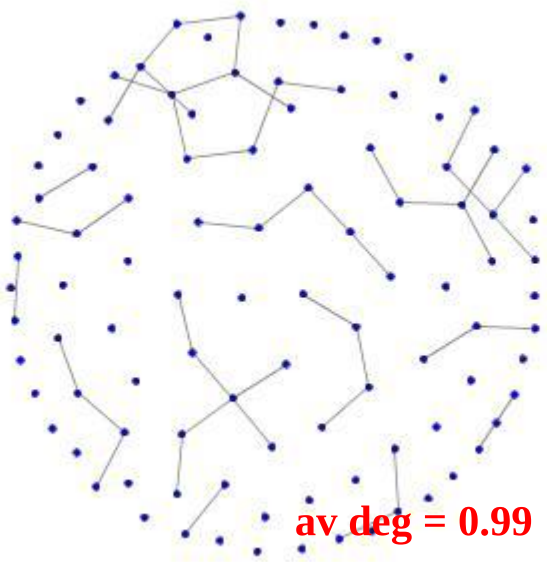
Percolation in networks

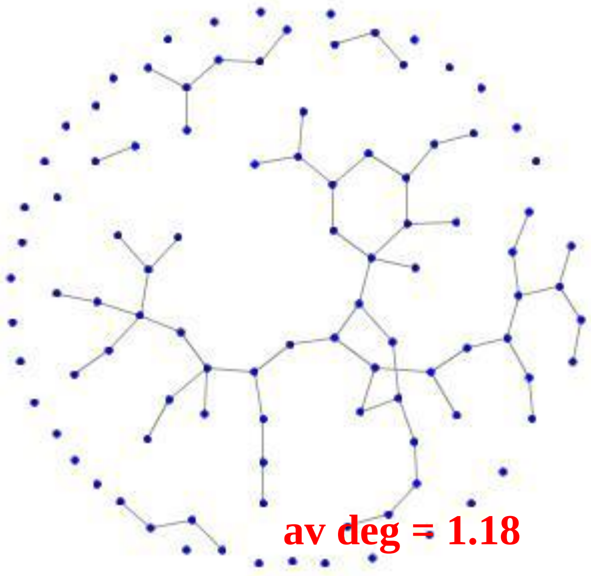
Percolation threshold in Erdos-Renyi Graphs

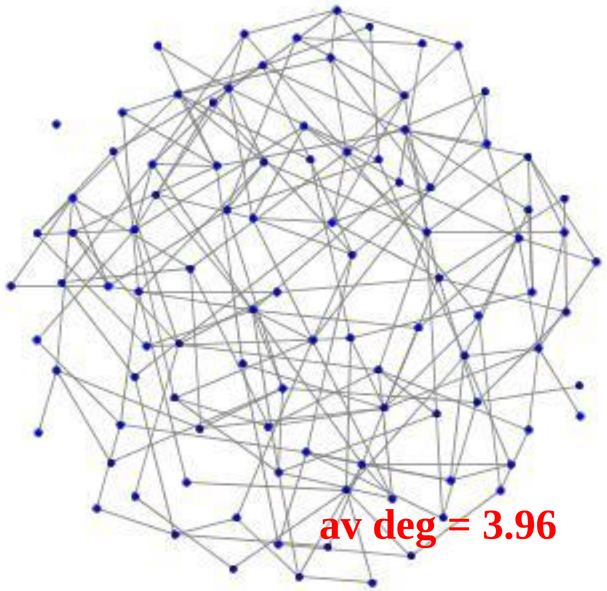
Percolation threshold: how many edges need to be added before the giant component appears?

As the average degree increases to z = 1, a giant component suddenly appears



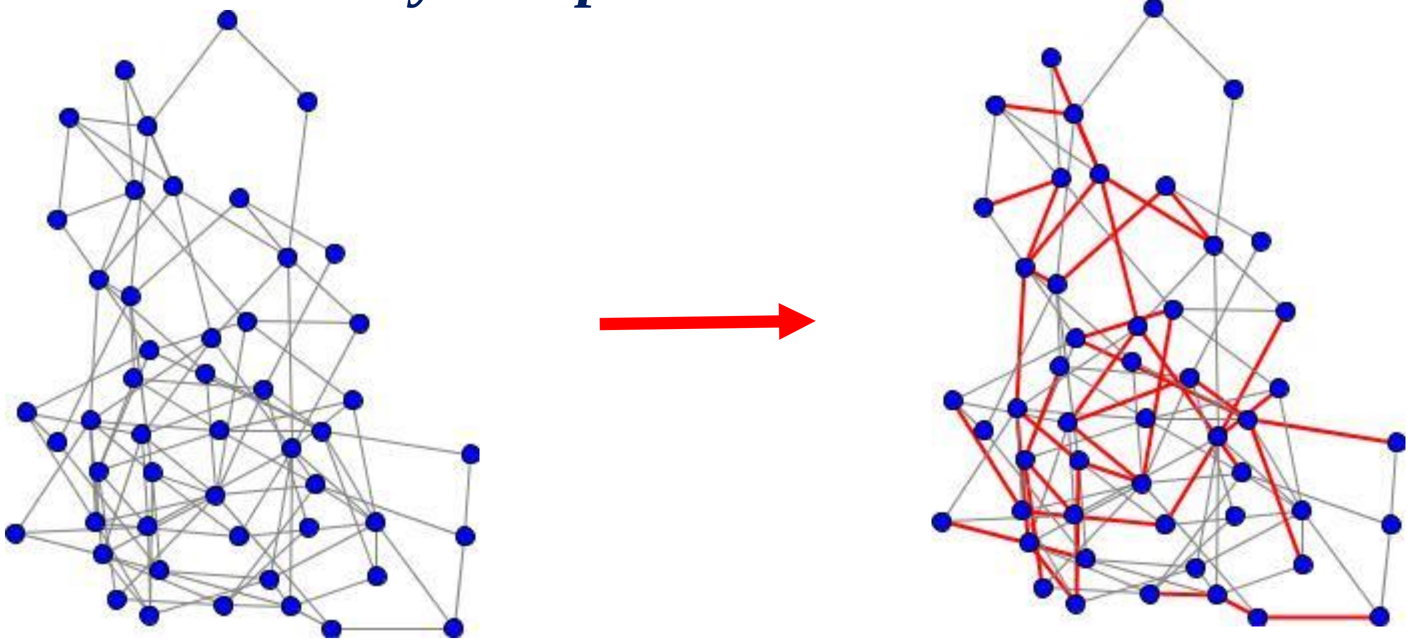






Percolation in networks

Edge Percolation in Erdos-Renyi Graphs

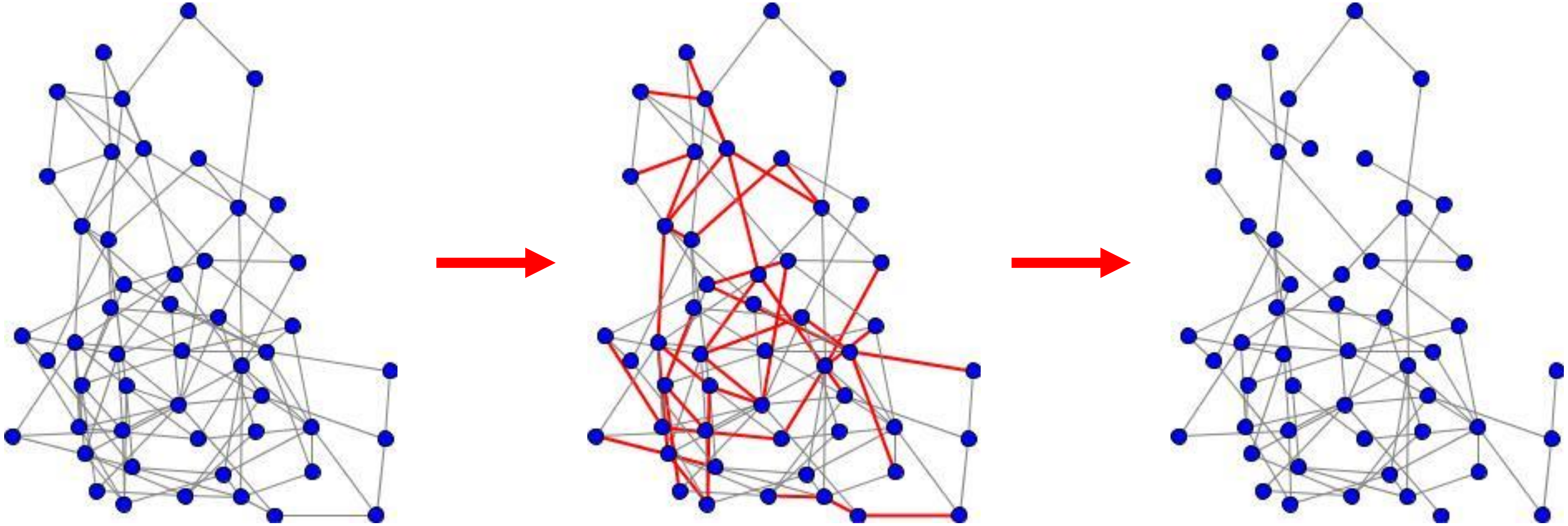


QUESTION:

In this network each node has average degree 4.64, if you removed 25% of the edges, by how much would you reduce the giant component?

Percolation in networks

Edge Percolation in Erdos-Renyi Graphs



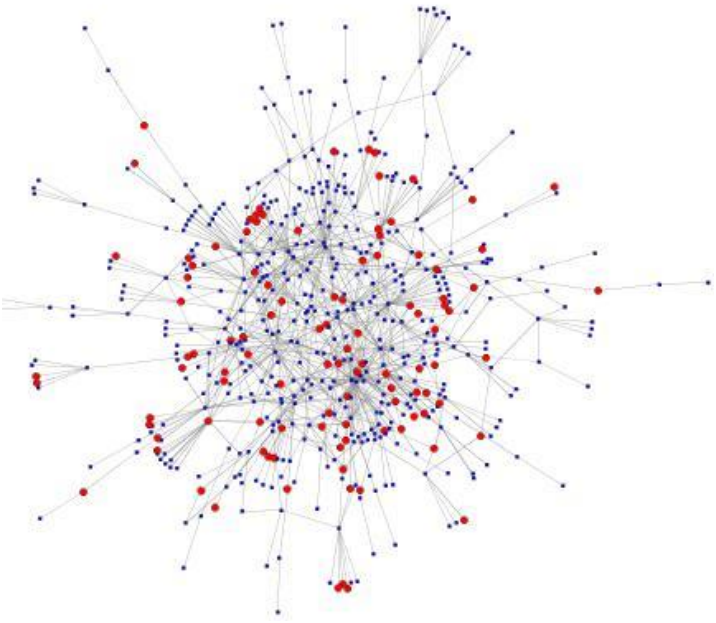
How many edges would you have to remove to break up an Erdos Renyi random graph?
e.g. each node has an average degree of 4.64

- 50 nodes, 116 edges, average degree 4.64
- After 25% edge removal → 76 edges, average degree 3.04
- Still well above percolation threshold

Percolation in networks

Scale-free networks are resilient with respect to random attack

- Example: Gnutella network
  - 20% of nodes removed



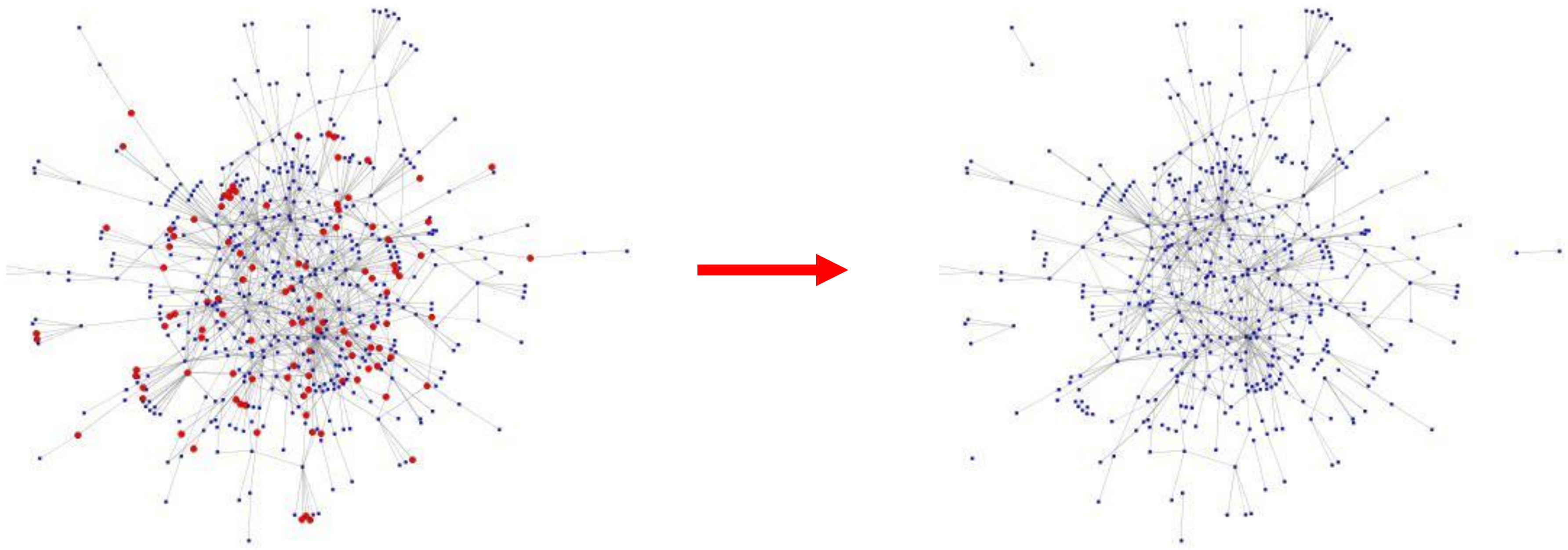
574 nodes in giant component

427 nodes in giant component

Percolation in networks

Targeted attacks are effective against Scale-free networks

- Example: Gnutella network (a large peer-to-peer network)
  - 22 most connected nodes removed (2.8% of the nodes)



574 nodes in giant component

301 nodes in giant component

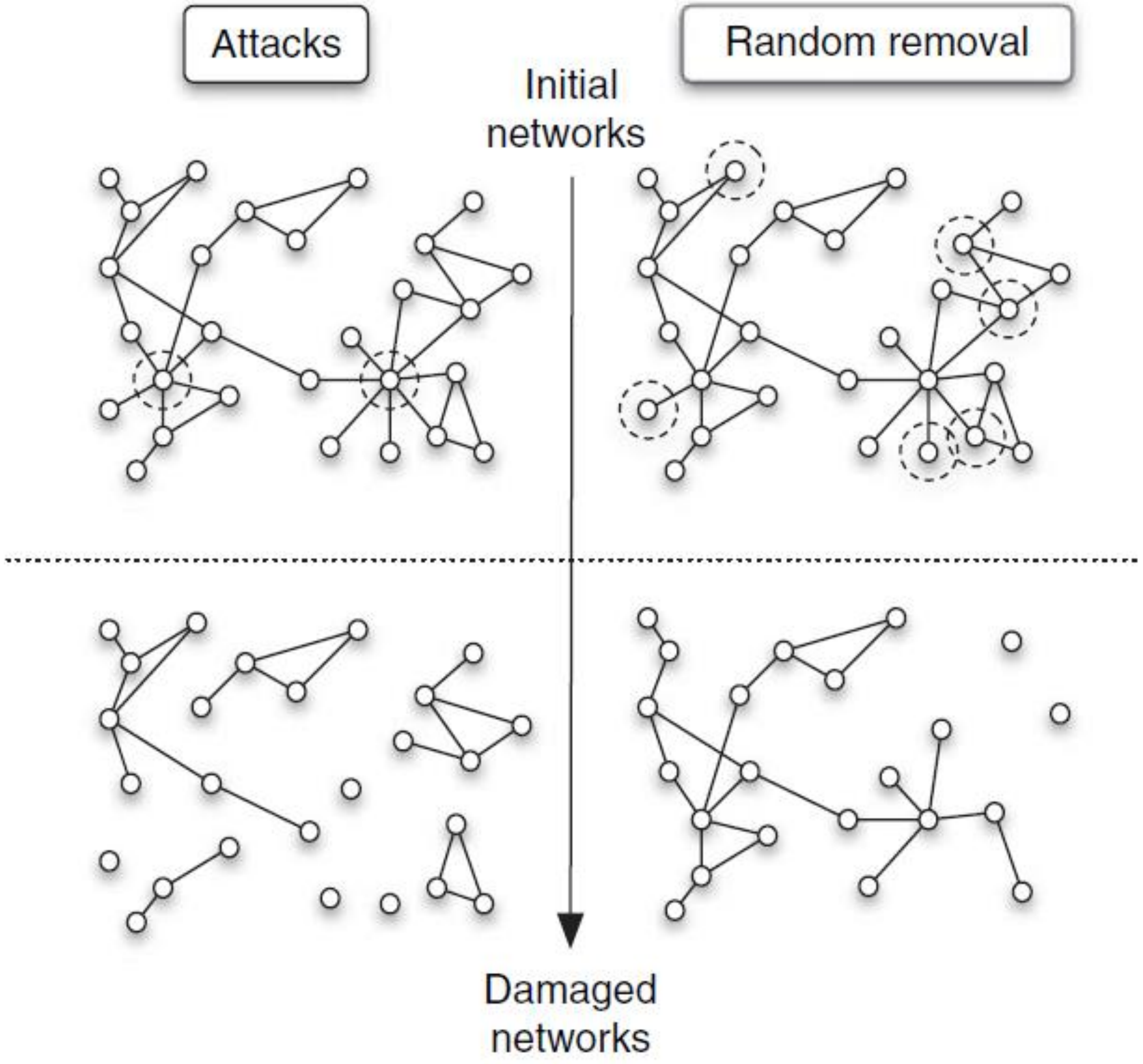
Percolation in networks

Question:

- Why is removing high-degree nodes more effective?
  - *it removes more edges*
  - *it removes more nodes*
  - *it targets the periphery of the network*

Network Resilience

random failures vs. attacks

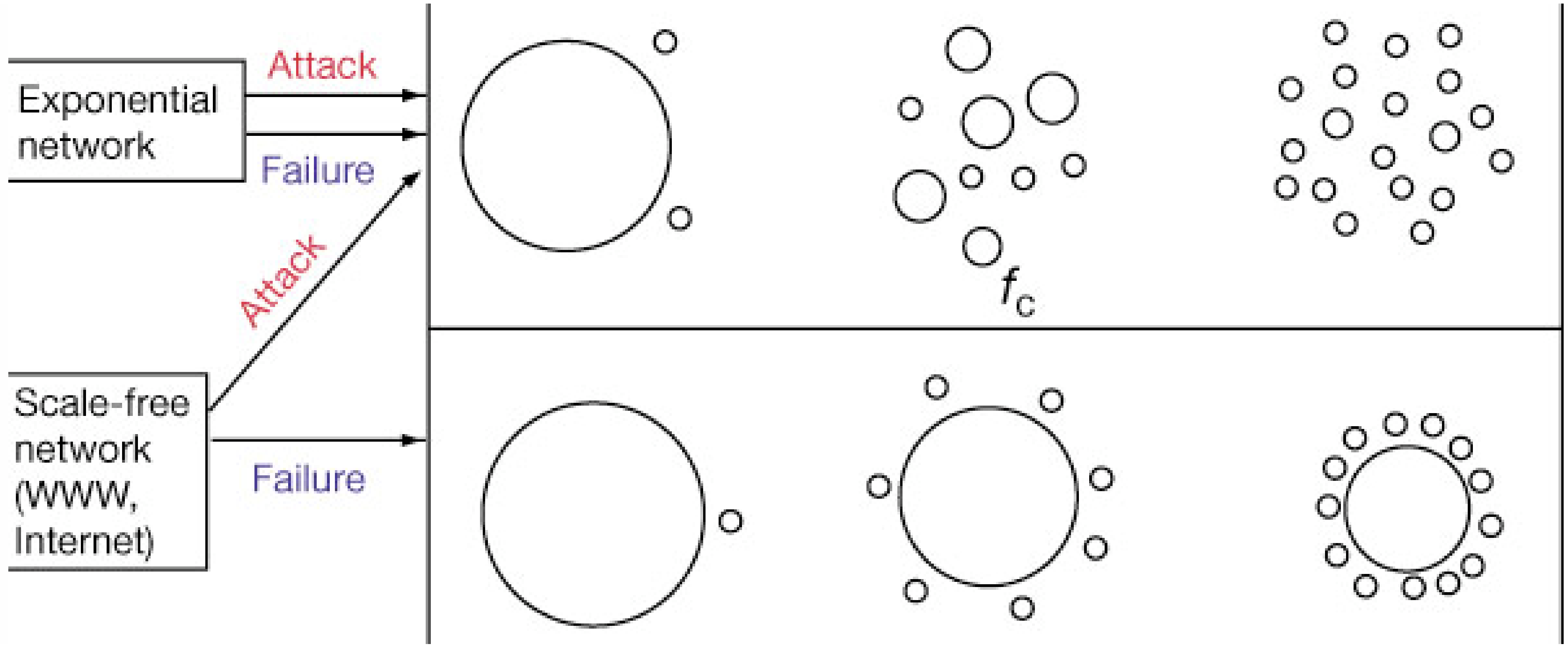


Schematic comparison of random and targeted removal. On the left, we show the effects of targeted attacks on the two nodes with largest degrees. The resulting network is made of small disconnected components. In the case of random removal of six nodes (right column), the damage is much less significant as there is still a path between most of the nodes of the initial network.

Barrat et al., *Dynamical Processes on Complex Networks*, 2008.

Network Resilience

random failures vs. attacks

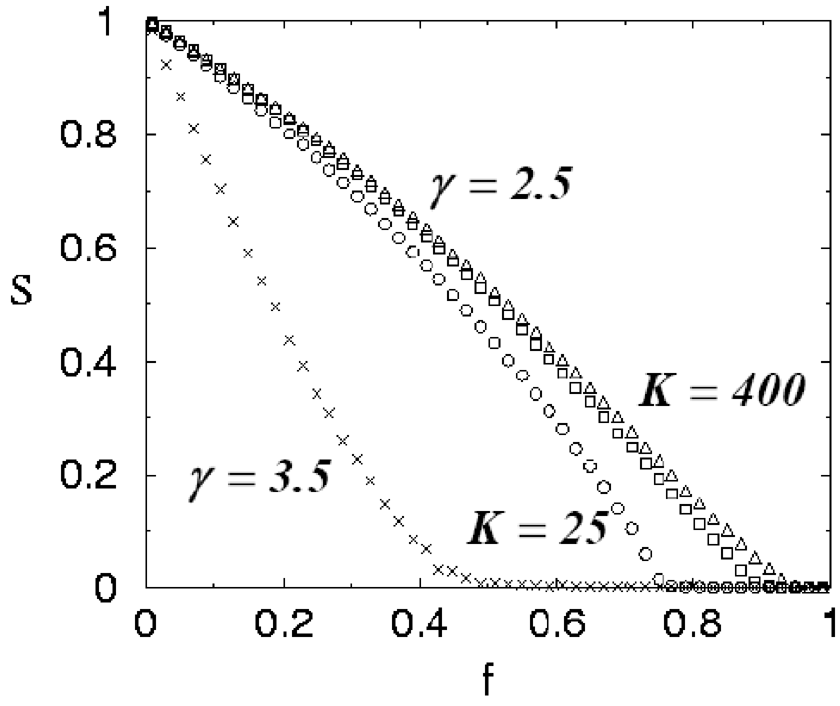


Source: Error and attack tolerance of complex networks. Réka Albert, Hawoong Jeong and Albert-László Barabási.

Network Resilience

Percolation Threshold scale-free networks

What proportion of the nodes must be removed in order for the size (S) of the giant component to drop to ~0?



K: max degree

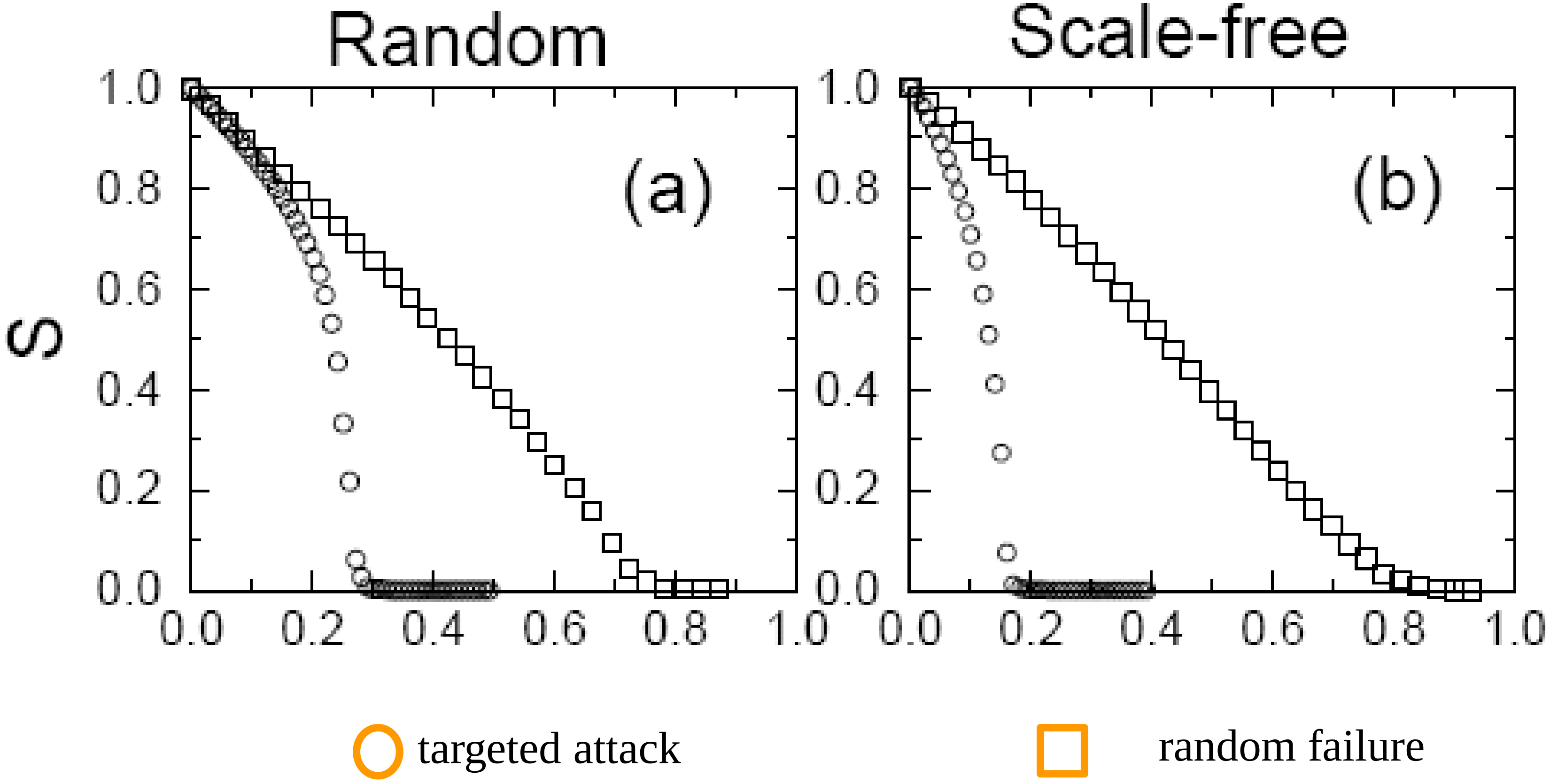
- For scale free graphs there is always a giant component
  - ➢ the network always percolates

Source: Cohen et al., Resilience of the Internet to Random Breakdowns

Network Resilience

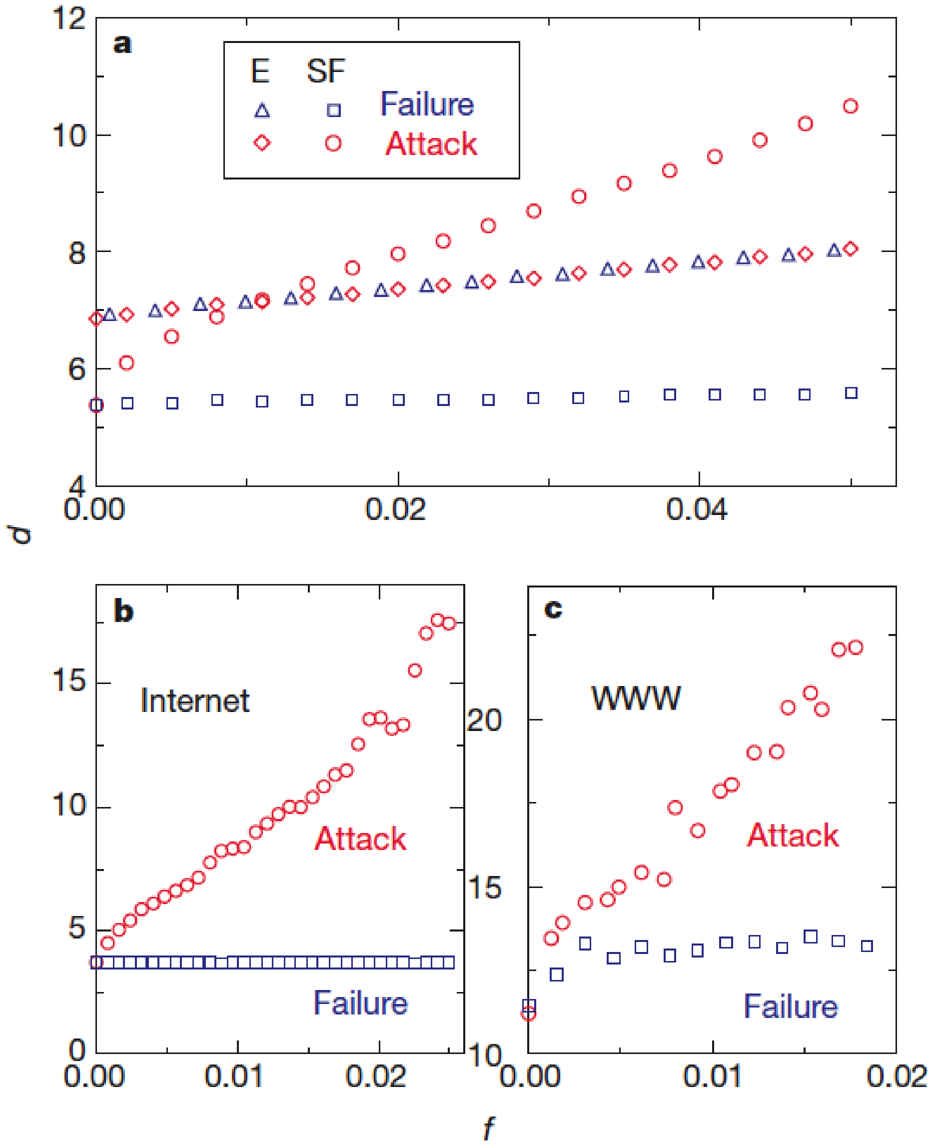
Network resilience to targeted attacks

- Scale-free graphs are resilient to random attacks, but sensitive to targeted attacks.
- For random networks there is smaller difference between the two



Source: Error and attack tolerance of complex networks. Réka Albert, Hawoong Jeong and Albert-László Barabási

Network Resilience

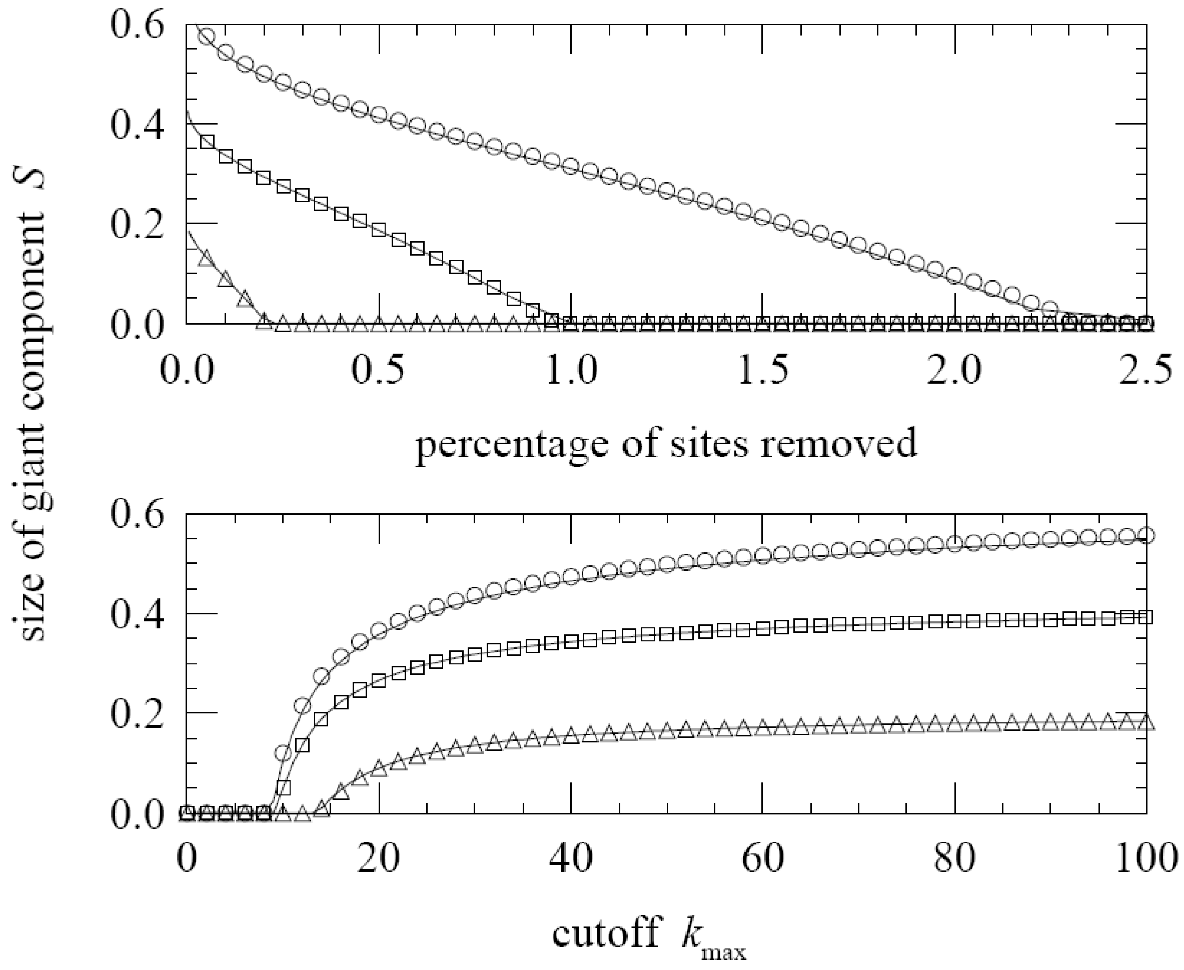


Changes in the diameter d of the network as a function of the fraction f of the removed nodes. **a**) Comparison between the random (E) and scale-free (SF) network models, each containing *N*=10000 nodes and 20,000 links (<k>= 4). The blue symbols correspond to the diameter of the random (triangles) and the scale-free (squares) networks when a fraction f of the nodes are removed randomly (error tolerance). Red symbols show the response of the random (diamonds) and the scale-free (circles) networks to attacks, when the most connected nodes are removed. The *f* dependence of the diameter for different system sizes (*N*=1000; 5,000; 20,000), apart from a logarithmic size correction, overlap with those shown in **a)**, indicating that the results are independent of the size of the system. The diameter of the unperturbed (*f* = 0) scale-free network is smaller than that of the random network, indicating that scale-free networks use the links available to them more efficiently, generating a more interconnected web. **b**) The changes in the diameter of the Internet under random failures (squares) or attacks (circles). The topological map of the Internet, containing 6,209 nodes and 12,200 links (<k>=3.4), collected by the National Laboratory for Applied Network Research (http://moat.nlanr.net/ Routing/rawdata/). **c**) Error (squares) and attack (circles) survivability of the World-Wide Web, measured on a sample containing 325,729 nodes and 1,498,353 links3, such that <k>=4.59.

Source: Error and attack tolerance of complex networks. Réka Albert, Hawoong Jeong and Albert-László Barabási

Network Resilience

Skewness of power-law networks and effects on targeted attack



The degree distribution of this model takes the form

$$p_k = \begin{cases} 0 & \text{for } k = 0, \\ Ck^{-\tau} e^{-k/\kappa} & \text{for } k \geq 1. \end{cases}$$

The network has a total of 10^7 nodes in the simulation.

Source: D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Network robustness and fragility: Percolation on random graphs*, Phys. Rev. Lett., 85 (2000), pp. 5468–5471.

Network Resilience

NETWORK	RANDOM FAILURES (REAL NETWORK)	RANDOM FAILURES (RANDOMIZED NETWORK)	ATTACK (REAL NETWORK
Internet	0.92	0.84	0.16
WWW	0.88	0.85	0.12
Power Grid	0.61	0.63	0.20
Mobile-Phone Call	0.78	0.68	0.20
Email	0.92	0.69	0.04
Science Collaboration	0.92	0.88	0.27
Actor Network	0.98	0.99	0.55
Citation Network	0.96	0.95	0.76
E. Coli Metabolism	0.96	0.90	0.49
Yeast Protein Interactions	0.88	0.66	0.06

This table shows the estimated critical threshold f_c for random node failures (second column) and attacks (fourth column) for ten reference networks. The third column (randomized network) offers f_c for a network whose N and L coincides with the original network, but whose nodes are connected randomly to each other. For most networks f_c for random failures exceeds f_c^{ER} for the corresponding randomized network, indicating that these networks display enhanced robustness. Three networks lack this property: the power grid, a consequence of the fact that its degree distribution is exponential, and the actor and the citation networks, which have a very high $\langle k \rangle$, diminishing the role of the high $\langle k^2 \rangle$.

Network Resilience and Cascade Failure

Cascade Failure:

In a network, the activity of a node depends on the activity of its neighboring nodes. Therefore, the failure of a node can at times induce the failure of the nodes that connect to it. In the case when the failure of a node (or a few nodes) induces a series of failure of nodes, this becomes a cascade effect. (Similar to the Domino effect)

- A *cascading failure* is a process in a system of interconnected parts in which the failure of one or few parts can trigger the failure of other parts and so on.
- Each node has a load and a capacity that says how much load it can tolerate.
- When a node is removed from the network its load is redistributed to the remaining nodes.
- If the load of a node exceeds its capacity, then the node fails

Network Resilience and Cascade Failure

Cascade Failure:

Cascading failure effects are observed in systems of rather different nature. Their size distribution is well approximated with the power law, meaning that most cascades are too small to be noticed. However, a few are so huge that they have global impact.

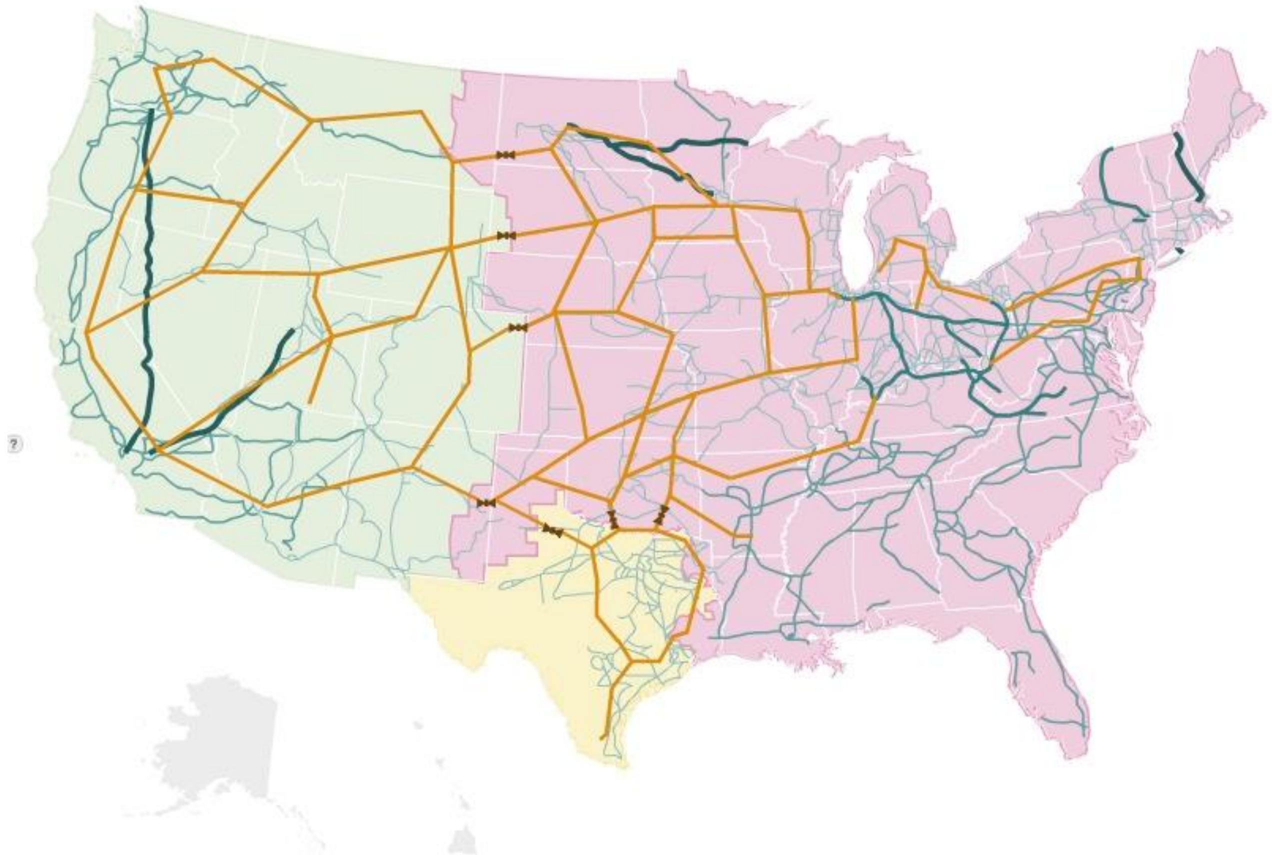
SOURCE	EXPONENT	CASCADE
Power grid (North America)	2.0	Power
Power grid (Sweden)	1.6	Energy
Power grid (Norway)	1.7	Power
Power grid (New Zealand)	1.6	Energy
Power grid (China)	1.8	Energy
Twitter Cascades	1.75	Retweets

Barabasi, *Network Science*, 2016

Network Resilience and Cascade Failure

Example: Power Grid

- Electric power flows simultaneously through multiple paths in the network.

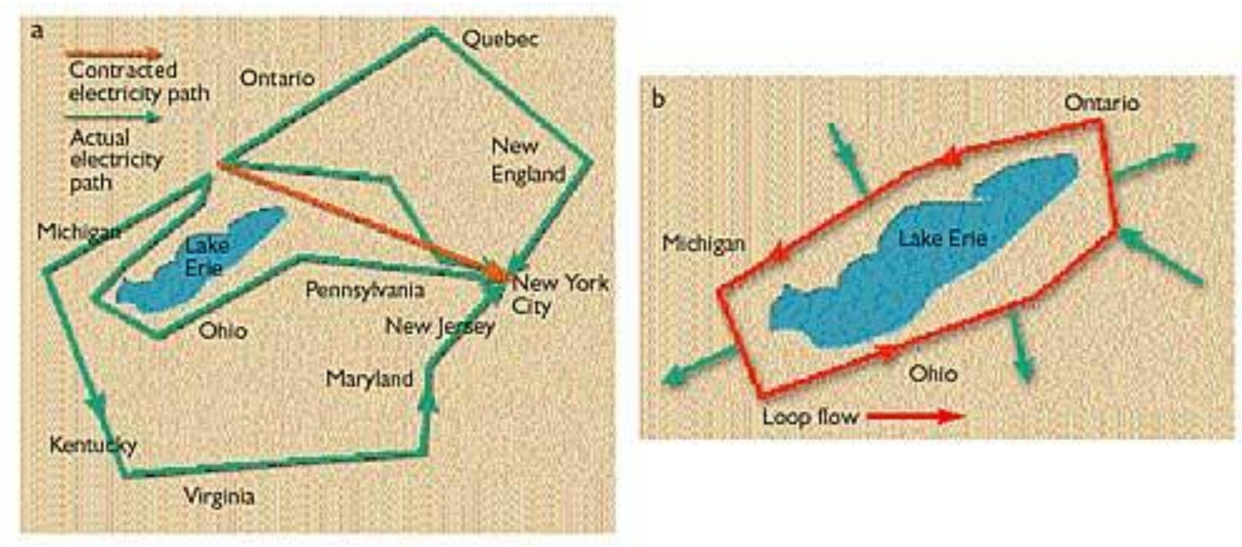


Network Resilience and Cascade Failure

Example: Power Grid

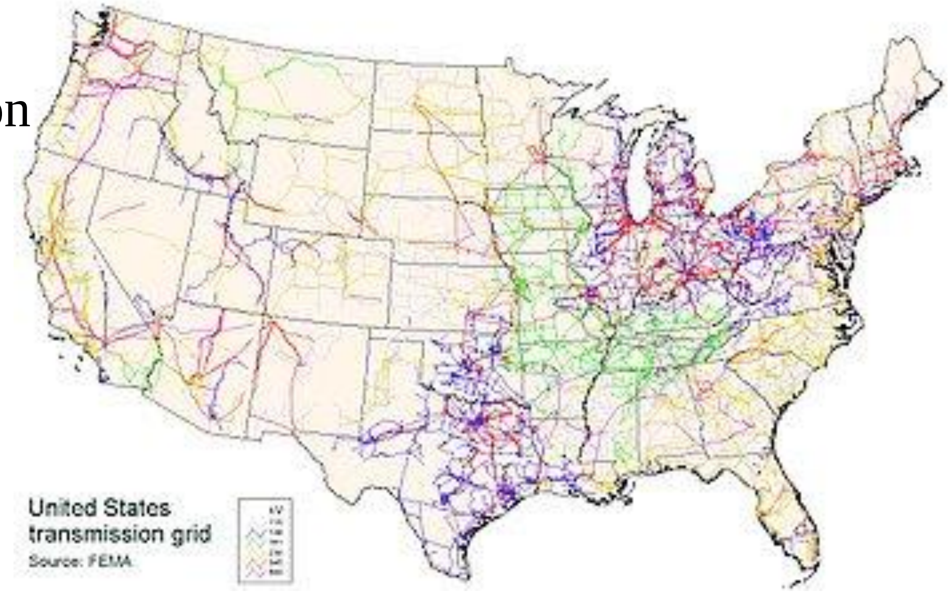
- Electric power does not travel just by the shortest route from source to sink, but also by parallel flow paths through other parts of the system.

- Where the network jogs around large geographical obstacles, such as the Rocky Mountains in the West or the Great Lakes in the East, loop flows around the obstacle are set up that can drive as much as 1 *GW* of power in a circle, taking up transmission line capacity without delivering power to consumers.



Network Resilience and Cascade Failure

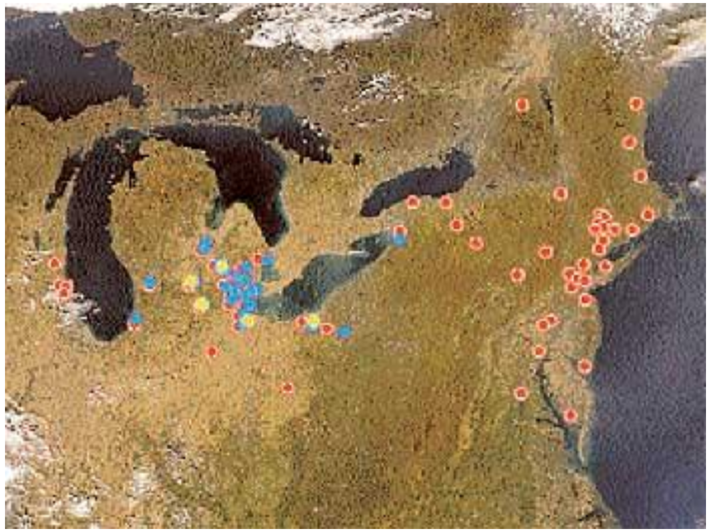
Power Grid and Cascade Failure:



- Vast system of electricity generation, transmission & distribution is essentially a single network
- Power flows through all paths from source to sink (flow calculations are important for other networks, even social ones)
- All AC lines within an interconnect must be in sync
- If frequency varies too much (as line approaches capacity), a circuit breaker takes the generator out of the system
- Larger flows are sent to neighboring parts of the grid – triggering a cascading failure

Network Resilience and Cascade Failure

Cascade Failure:

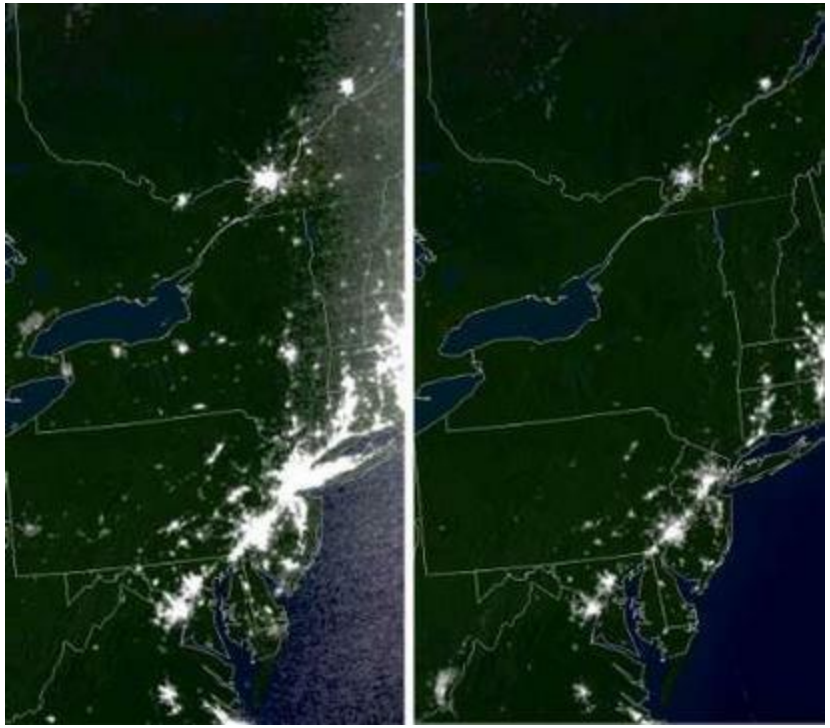


- **1:58 p.m.** The Eastlake, Ohio, First Energy generating plant shuts down (maintenance problems).
- **3:06 p.m.** A First Energy 345-kV transmission line fails south of Cleveland, Ohio.
- **3:17 p.m**. Voltage dips temporarily on the Ohio portion of the grid. Controllers take no action, but power shifted by the first failure onto another power line causes it to sag into a tree at 3:32 p.m., bringing it offline as well. While Mid West ISO and First Energy controllers try to understand the failures, they fail to inform system controllers in nearby states.
- **3:41 and 3:46 p.m.** Two breakers connecting First Energy's grid with American Electric Power are tripped.
- **4:05 p.m.** A sustained power surge on some Ohio lines signals more trouble building.
- **4:09:02 p.m.** Voltage sags deeply as Ohio draws 2 GW of power from Michigan.
- **4:10:34 p.m.** Many transmission lines trip out, first in Michigan and then in Ohio, blocking the eastward flow of power. Generators go down, creating a huge power deficit. In seconds, power surges out of the East, tripping East coast generators to protect them.

Source: Eric J. Lerner, "What's wrong with the electric grid?"

Network Resilience and Cascade Failure

Actual satellite images of the effect of the blackout



20 hours prior to blackout | 7 hours after blackout

Source: NOAA, U.S. Government

Network Resilience and Cascade Failure

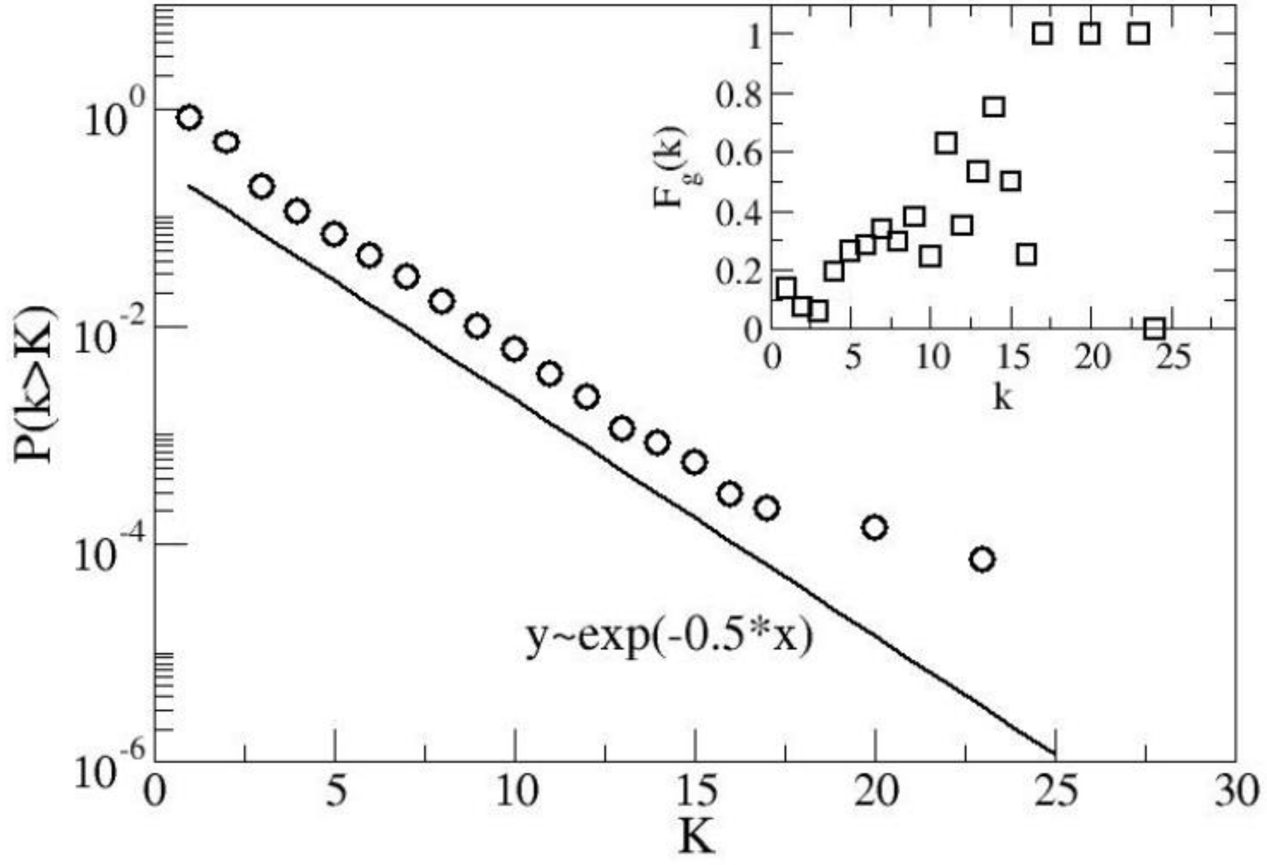
Case study: North American power grid

- Nodes: generators, transmission substations, distribution substations
- Edges: high-voltage transmission lines

- 14,099 substations:
  - N_G 1,633 generators,
  - N_D 2,179 distribution substations
  - N_T the rest transmission substations
- 19,657 edges

Modeling cascading failures in the North American power grid, R. Kinney, P. Crucitti, R. Albert, and V. Latora, Eur. Phys. B, 2005

Degree distribution is exponential



$$P(k > K) \approx \exp(-0.5K)$$

Source: Albert et al., ‘Structural vulnerability of the North American power grid

Efficiency of a path

- efficiency e [0,1]
  - 0 if no electricity flows between two endpoints,
  - 1 if the transmission lines are working perfectly
- harmonic composition for a path

$$e_{path} = \left[\sum_{edges} \frac{1}{e_{edge}} \right]^{-1}$$

- ➢ path A, 2 edges, each with e=0.5, e_{path} = 1/4
- ➢ path B, 3 edges, each with e=0.5 e_{path} = 1/6
- ➢ path C, 2 edges, one with e=0 the other with e=1, e_{path} = 0

➡ *Simplifying assumption*: electricity flows along most efficient path

Efficiency of the network

- Efficiency of the network:
  - average over the most efficient paths from each generator to each distribution station

$$E = \frac{1}{N_G N_D} \sum_{i \in G_G} \sum_{j \in G_D} \epsilon_{ij}$$

ε_{ij} is the efficiency of the most efficient path between *i* and *j*

Capacity and node failure

- Assume capacity (C_i) of node i is proportional to initial load L_i

$$C_i = \alpha L_i(0) \quad i = 1, 2..N$$

  - L represents the weighted betweenness of a node
  - α is the tolerance parameter

  - Each neighbor of a node is impacted as follows

$$e_{ij}(t+1) = \begin{cases} e_{ij}(0)/\frac{L_i(t)}{C_i} & \text{if } L_i(t) > C_i \\ e_{ij}(0) & \text{if } L_i(t) \leq C_i \end{cases}$$

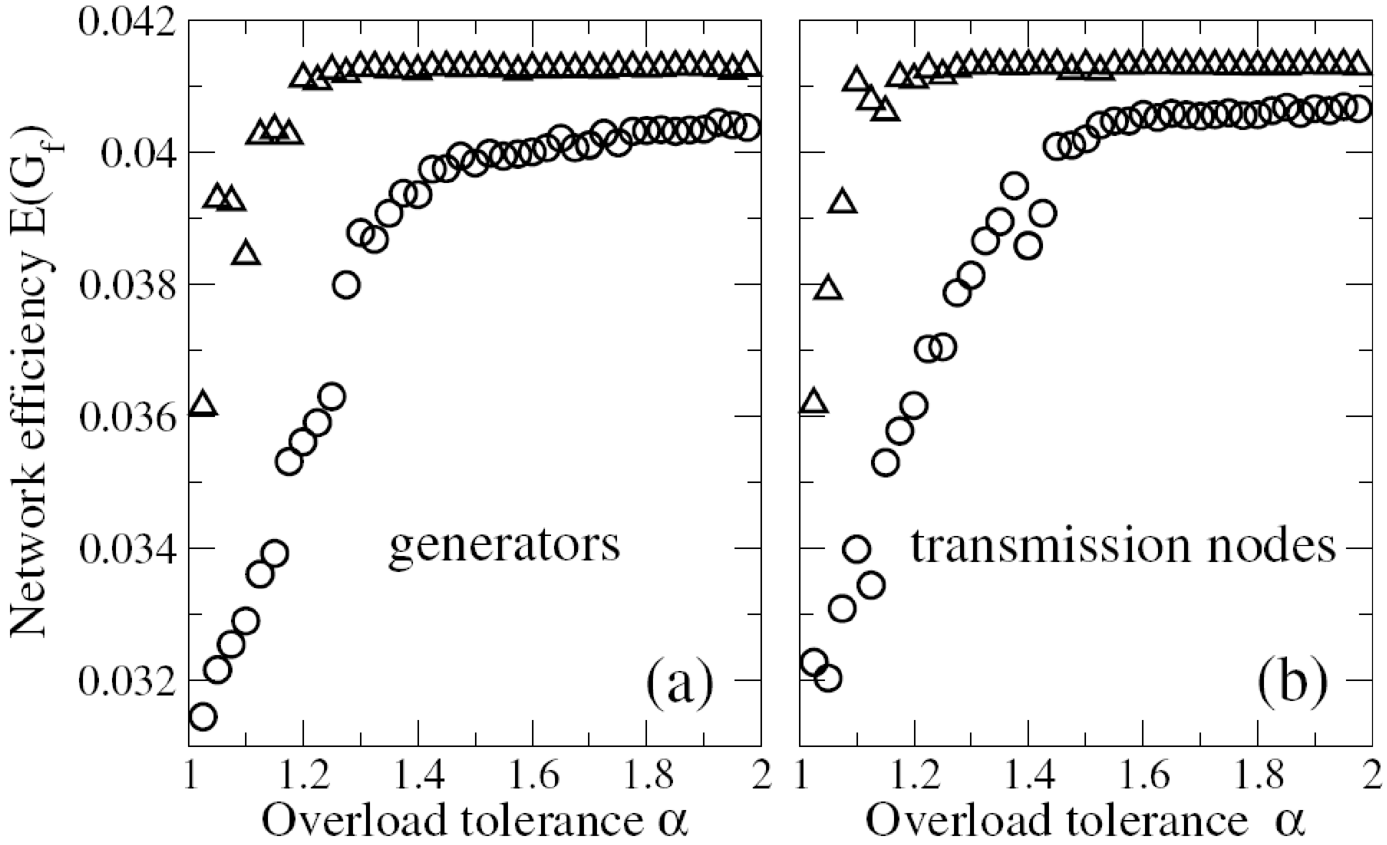
load exceeds capacity

- Load is distributed to other nodes/edges

- The greater a (reserve capacity), the less susceptible the network to cascading failures due to node failure

Power grid structural resilience

- efficiency is impacted the most if the node removed is the one with the highest load



○ highest load generator/transmission station removed

Source: Modeling cascading failures in the North American power grid; R. Kinney, P. Crucitti, R. Albert, and V. Latora

Network Resilience and Robustness

Summary

- Resilience depends on topology
- Also depends on what happens when a node fails
  - e.g. in power grid load is redistributed
  - in protein interaction networks other proteins may start being produced or cease to do so
- In biological networks, more central nodes without which cannot function

Synchronization in Networks

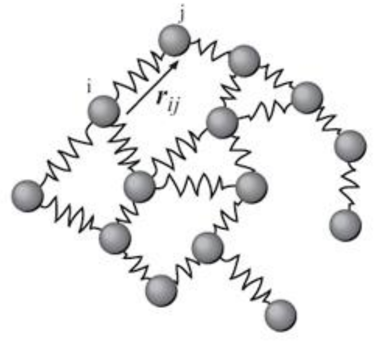
Synchronization in Networks

Many systems can be modeled as a collection of oscillators coupled to each other via an interaction matrix, e.g., earthquakes, ecosystems, neurons, animal and insect behavior, etc. These coupled oscillators may sometimes display synchronized behavior, i.e. follow a common dynamical evolution. Famous examples include crickets that chirp in unison, or flashing fireflies.

Synchronization of firefly flashing

Linearly coupled identical oscillators

For a linearly coupled identical oscillator system in which the units are interconnected through an interaction matrix, we can describe it by the following set of coupled equations,



$$\frac{dx_i}{dt} = F(x_i) + \sigma \sum_{j=1}^{N} C_{ij} H(x_j) \ , \qquad i = 1, 2, \dots, N$$

where H is a fixed output function, σ denotes the interaction strength, and C_{ij} is the coupling matrix. One can further assume that the coupling between two units depends only on the difference between their outputs, and the above equation becomes

$$\frac{dx_i}{dt} = F(x_i) + \sigma \sum_{j \in V(i)} \left[H(x_i) - H(x_j)\right]$$

V(i) is the set of neighbors of *i*, and C_{ij} becomes the Laplacian matrix L_{ij}.

Linearly coupled identical oscillators

Denote by s(t) the evolution of the uncoupled oscillators, i.e., σ equals to zero, all of the oscillators will then be in the same synchronized state. The stability of this synchronized state can be studied by using the master stability function approach that considers a small perturbation $x_i = s + \xi_i$ of the system, with ξ_i close to the synchronized state. The synchronized state is then stable if and only if the dynamical evolution drives the system back to a synchronized state by a steady decrease of the perturbation ξ_i, otherwise the synchronized state is unstable.

Expand F and H as $F(x_i) = F(s) + \xi_i F'(s)$; $H(x_i) = H(s) + \xi_i H'(s)$, the evolution equation becomes

$$\frac{d\xi_i}{dt} = F'(s)\,\xi_i + \sigma \sum_j \left[L_{ij} H'(s)\right] \xi_j$$

Linearly coupled identical oscillators

Remember the eigenvalues ($\lambda_i, i = 1, \ldots, N$) of the Laplacian are real and non-negative, and one of them is 0. We therefore have $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N \equiv \lambda_{max}$.

The equation can be decoupled into a set of equations involving each of the eigenvectors ζ_i, which are a linear combination of ξ_j

$$\frac{d\zeta_i}{dt} = [F'(s) \ + \sigma\lambda_i H'(s)]\zeta_i$$

At short times, one can assume that *s* almost does not change, and these decoupled equations can be solved, with the solutions

$$\zeta_i = \zeta_i^0 \exp\{[F'(s) + \sigma\lambda_i H'(s)]t\}$$

where ζ_i^0 is the initial imposed perturbation. These equations show that the perturbation will either increase or decrease exponentially, depending on the signs of the exponential, i.e. $\Lambda_i = [F'(s) + \sigma\lambda_i H'(s)]$

Linearly coupled identical oscillators

The eigenvalue $\lambda_1 = 0$ gives $\Lambda_1 = F'(s)$, related to the evolution of each single unit, which is either chaotic (if $F'(s) > 0$) or periodic.

For the synchronized state to be stable, all the other components of the perturbation ζ_i have to decrease, which means that all the Λ_i, *i* = 2, . . . , N, need to be negative.

The master stability function is therefore defined as

$$\Lambda(s) = max_s\ (F'(s) + \alpha H'(s))$$

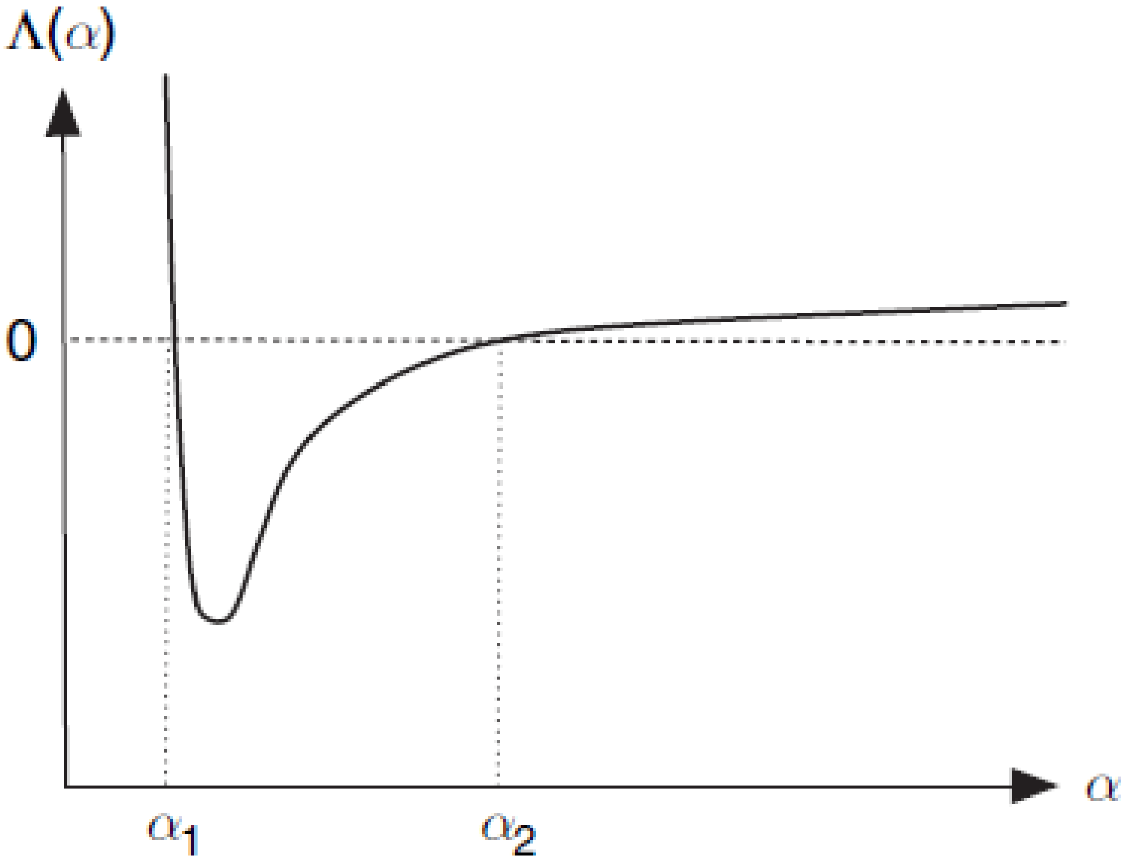
where the max is taken over the trajectory defined by $\frac{ds}{dt} = F(s(t))$, and the stability condition of the synchronized state means that all $\sigma\lambda_i$, for i = 2, . . . , N, are located in the negative region of the master stability function, i.e. $\Lambda(\sigma\lambda_i) \leq 0$.

Linearly coupled identical oscillators

For $\alpha > 0$:

- If $\Lambda(\alpha)$ is always positive, the synchronized state is never stable.
- If $\Lambda(\alpha)$ decreases and becomes negative for all $\alpha > \alpha_c$, it is enough to have a large enough coupling strength ($\sigma > \alpha_c/\lambda_2$) to ensure synchronization.

Linearly coupled identical oscillators



Typical master stability function. The synchronized state is stable only if all the positive eigenvalues λ_i of the Laplacian are such that $\Lambda(\sigma\lambda_i) \leq 0$, which corresponds to $\sigma\lambda_i \in [\alpha_1, \alpha_2]$ for $i = 2, \ldots, N$.

Barahona, M. and Pecora, L. M. (2002), Synchronization in small-world systems, *Phys. Rev. Lett.* **89**, 054101.

The Kuramoto Model

In real-world situations it is frequent to find that the oscillators are not identical. The Kuramoto model simulates this kind of situations.

$$\frac{dx_i}{dt} = \omega_i + K \sum_{j\sim i} \sin\left(x_i - x_j\right) \ ,$$

where $\sum_{j\sim i}$ represents the sum over pairs of adjacent nodes.

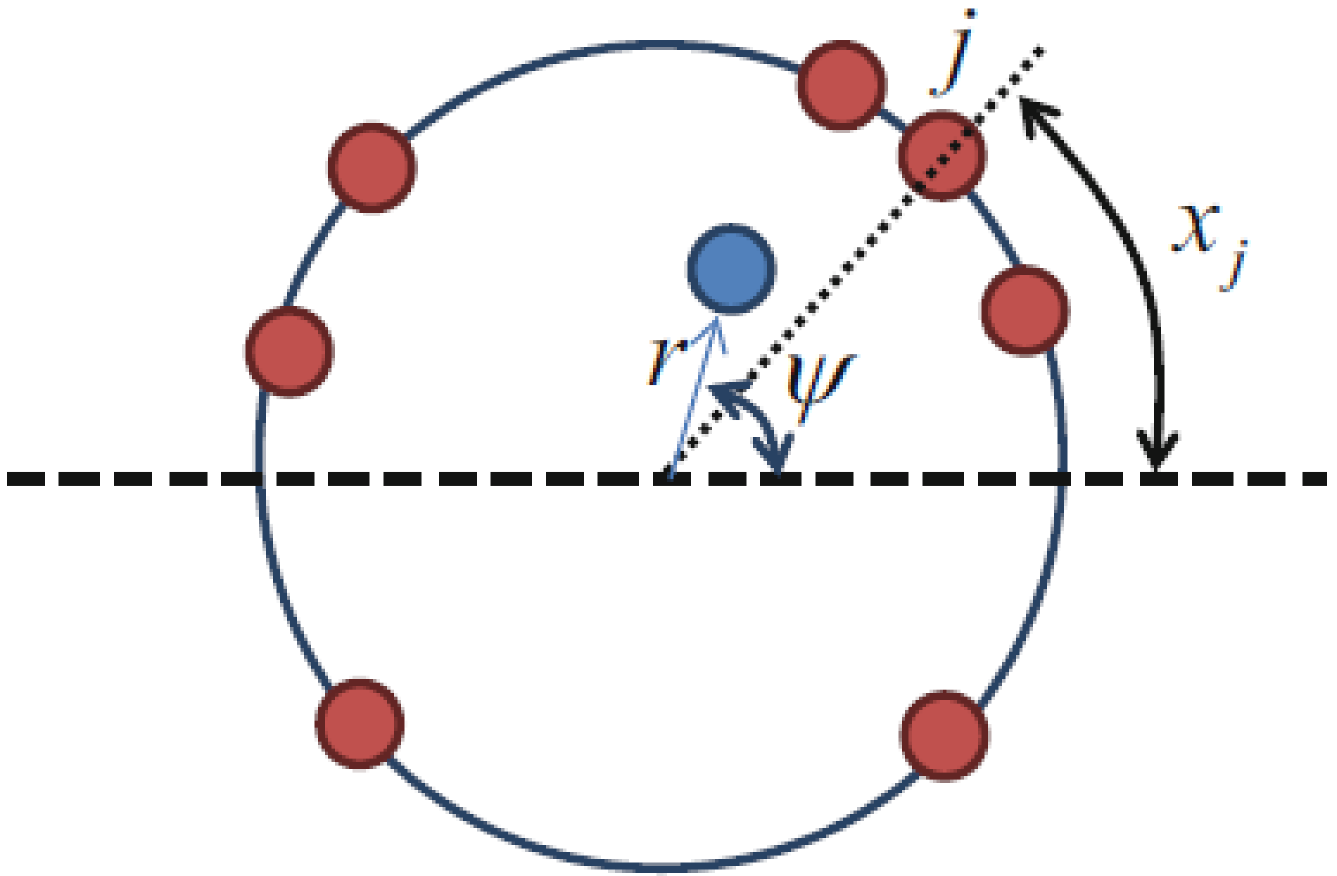
Macroscopic coherence of the system is characterized by the order parameter

$$r(t)e^{i\psi(t)} = \frac{1}{N} \sum_{j=1}^{N} e^{ix_j(t)}$$

or,

$$r = \left| \frac{1}{N} \sum_{j=1}^{N} e^{ix_j(t)} \right|$$

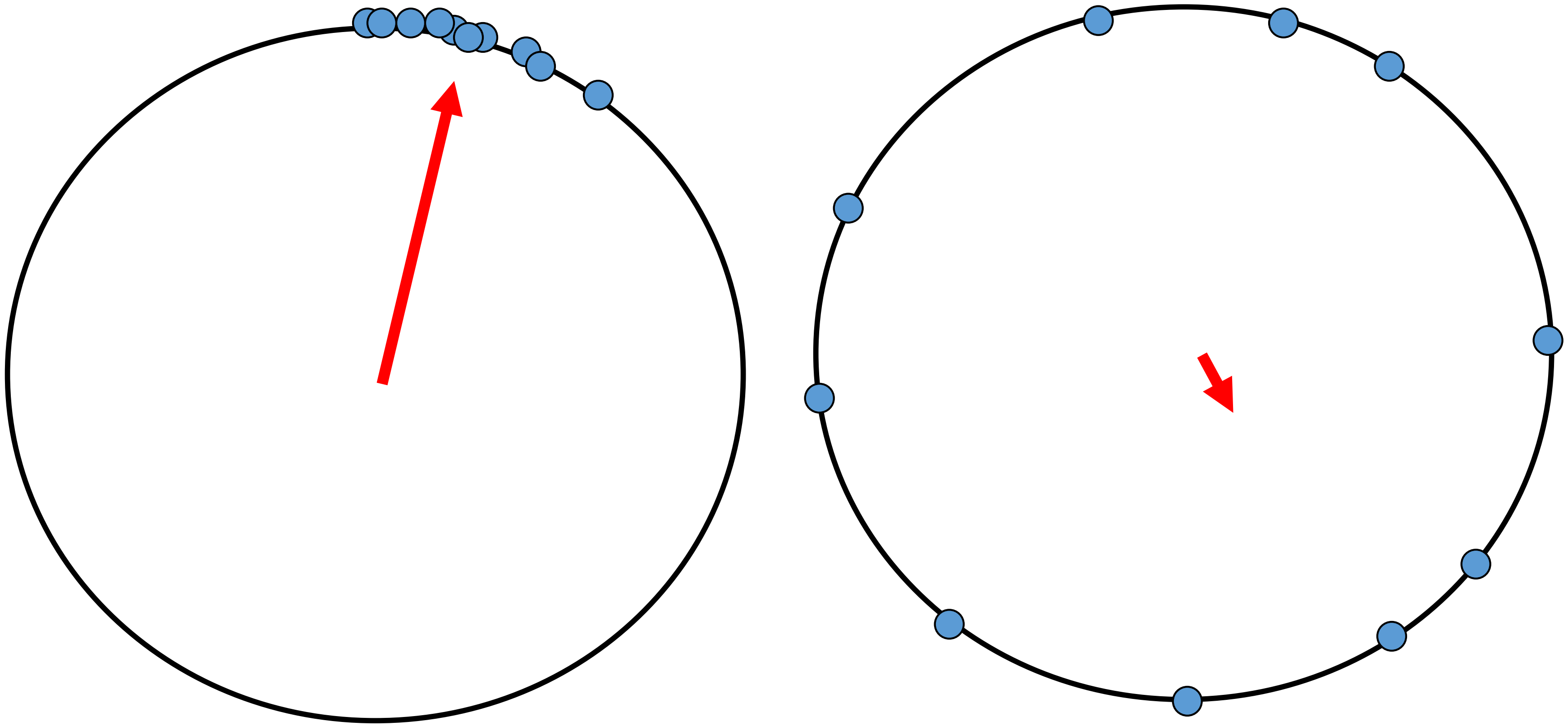
The Kuramoto Model



Representation of the Kuramoto model where the phases correspond to points moving around a unit circle in the complex plane. The *blue circle* represents the center of mass of these points, which is defined by the order parameter

The Kuramoto Model

Order Parameter Measures Coherence



$r \approx 1$ $r \approx 0$

The Kuramoto Model

Assume a complete network of oscillators, i.e., each pair of oscillators is connected, coupled with the same strength $K = K^0/N$, with K^0 finite. Multiplying both sides of the order parameter by $e^{ix_i(t)}$and taking imaginary parts to obtain

$$r\sin(\psi - x_i) = \frac{1}{N}\sum_{j=1}^{N}\sin(x_j - x_i)$$

So that

$$\frac{dx_i}{dt} = \omega_i + K^0 r\sin(x_i - \psi)$$

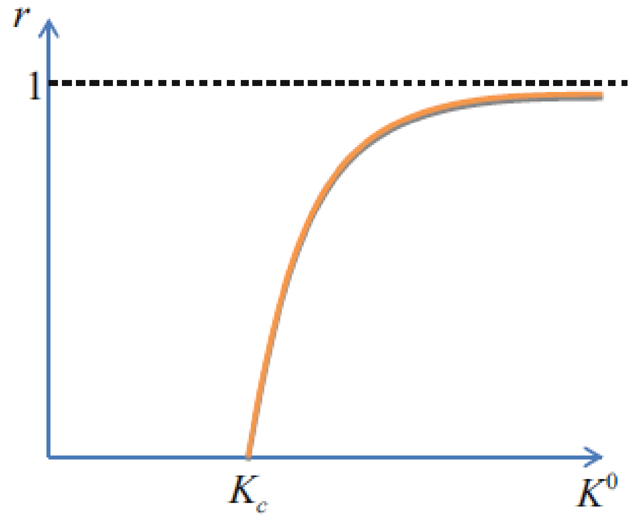
in which the interaction term is given by a coupling with the mean phase ψ and the intensity is proportional to the coherence *r*.

There is a transition to synchrony at a critical value of the coupling constant K^0.

When $N, t \to \infty$, we have

$$r = \begin{cases} 0 & K^0 < K_c \\ (K^0 - K_c)^{\beta} & K^0 \geq K_c \end{cases}$$

and β = 1/2



Mean-field results of the Kuramoto model

Kuramoto Model on a Network

The network is introduced by means of the adjacency matrix ***A***

$$\boxed{\frac{dx_i}{dt} = \omega_i + K\sum_{m=1}^{N} A_{mn}\sin(x_i - x_j)}$$

m is not connected to *n* $\Longleftrightarrow$ $A_{nm} = 0$.

The nonzero elements of ***A*** can have any positive or negative value and correspond to the interaction strength at each link.

Barrat et al., *Dynamical Processes on Complex Networks*, 2008.

The Kuramoto Model

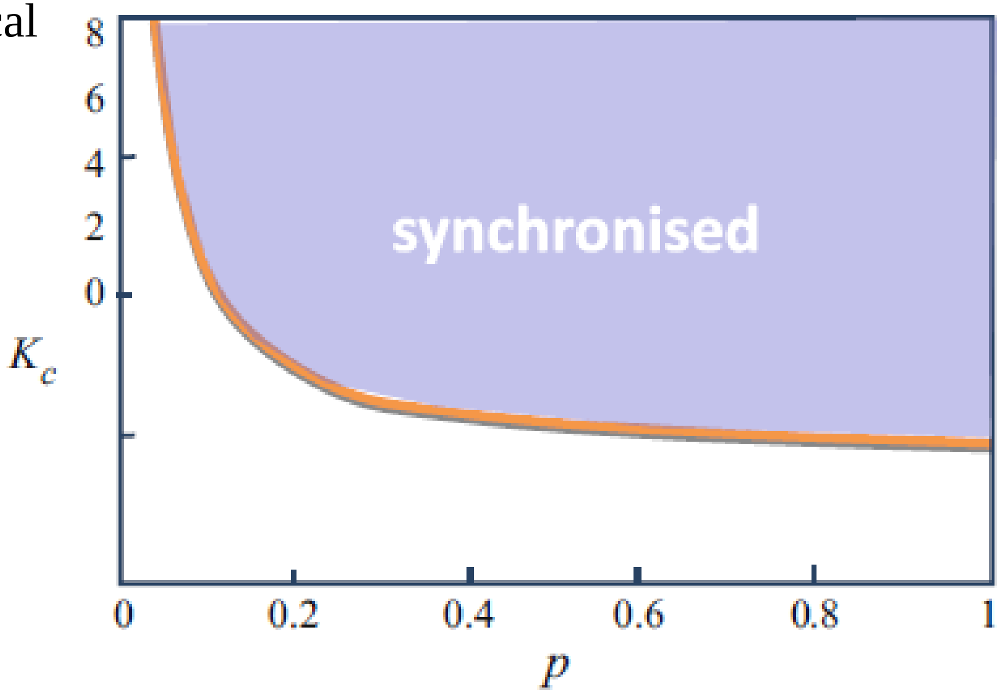
In small-world networks it has been empirically found that the order parameter is scaled as

$$r(N,K) = N^{-\beta/v} F\left[(K-K_c)N^{1/v}\right]$$

where F is a scaling function and describes the divergence of the typical correlation size $(K-K_c)^{-v}$.

In the case of oscillators with varying degree, the critical coupling parameter scales as

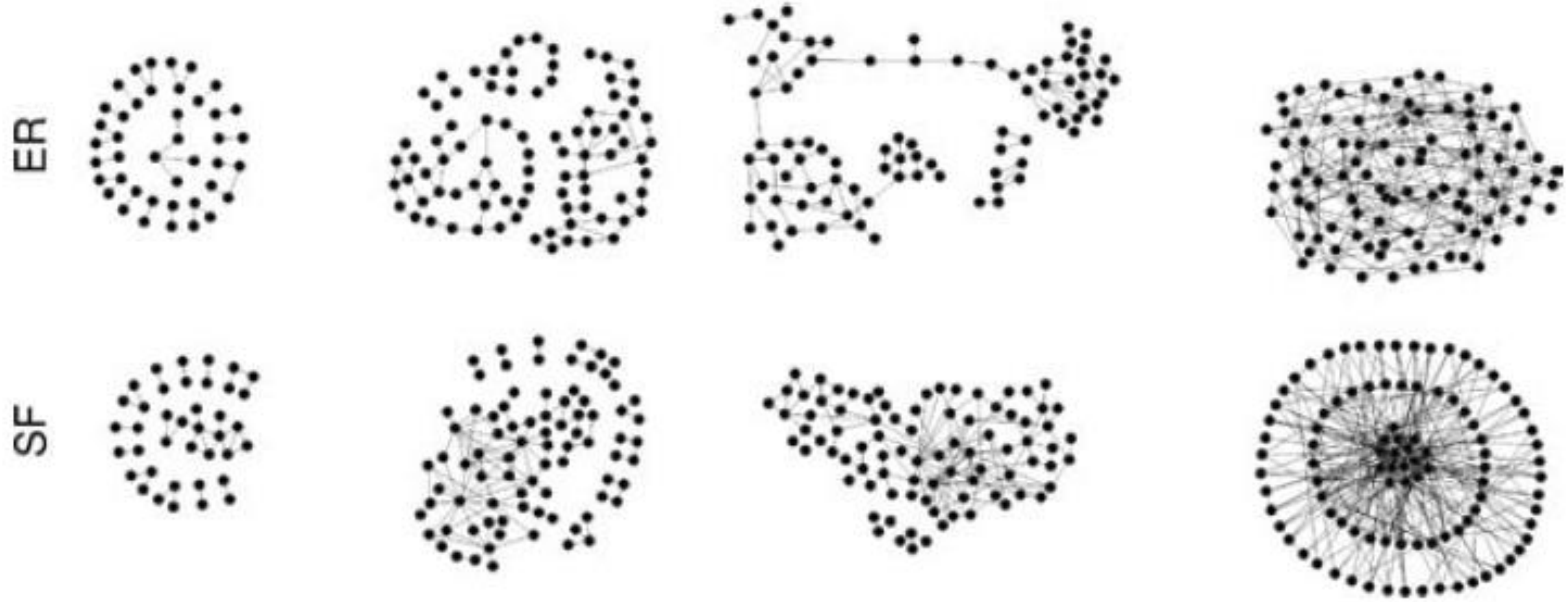
$$K_c = c\,\frac{\langle k\rangle}{\langle k^2\rangle}$$



Relation of the critical coupling parameter K_c with the rewiring probability p in the Watts-Strogatz model of small-world networks

Synchronization in Networks

The network topology can affect the synchronization behavior in the network.



Synchronized clusters for two different topologies: homogeneous and scale-free. Small networks of 100 nodes are displayed for the sake of visualization. From left to right, the coupling intensity is increasing. Depending on the topology, different routes to synchronization are observed. In the case of Erdos–Renyi networks (ER), a percolation-like mechanism occurs, while for scale-free networks (SF), hubs are aggregating more and more synchronized nodes.

Models of coupled heterogeneous oscillators

All-to-all Network | Sine coupled phase oscillators

Kuramoto model

All-to-all Network | More general dynamics

Ott *et al* '02; Pikovsky *et al '96*
Baek *et al* '04; Topaj *et al* '02

More general network | Sine coupled phase oscillators

Ichinomiya, Phys. Rev. E '04; '05
Restrepo *et al.*, Phys. Rev. E '05; Chaos 06

More general Network | More general dynamics

Restrepo et al., PRL'06; Physica D'06

Some Remarks

- For a large class of networks, there is a transition to synchrony at a critical coupling constant determined by the maximum eigenvalue of the adjacency matrix.

- A larger maximum eigenvalue of the adjacency matrix favors a lower threshold for synchronization.

- Heterogeneity in the degree distribution, randomness in the couplings, and positive degree correlations favors synchronization.

Network Search

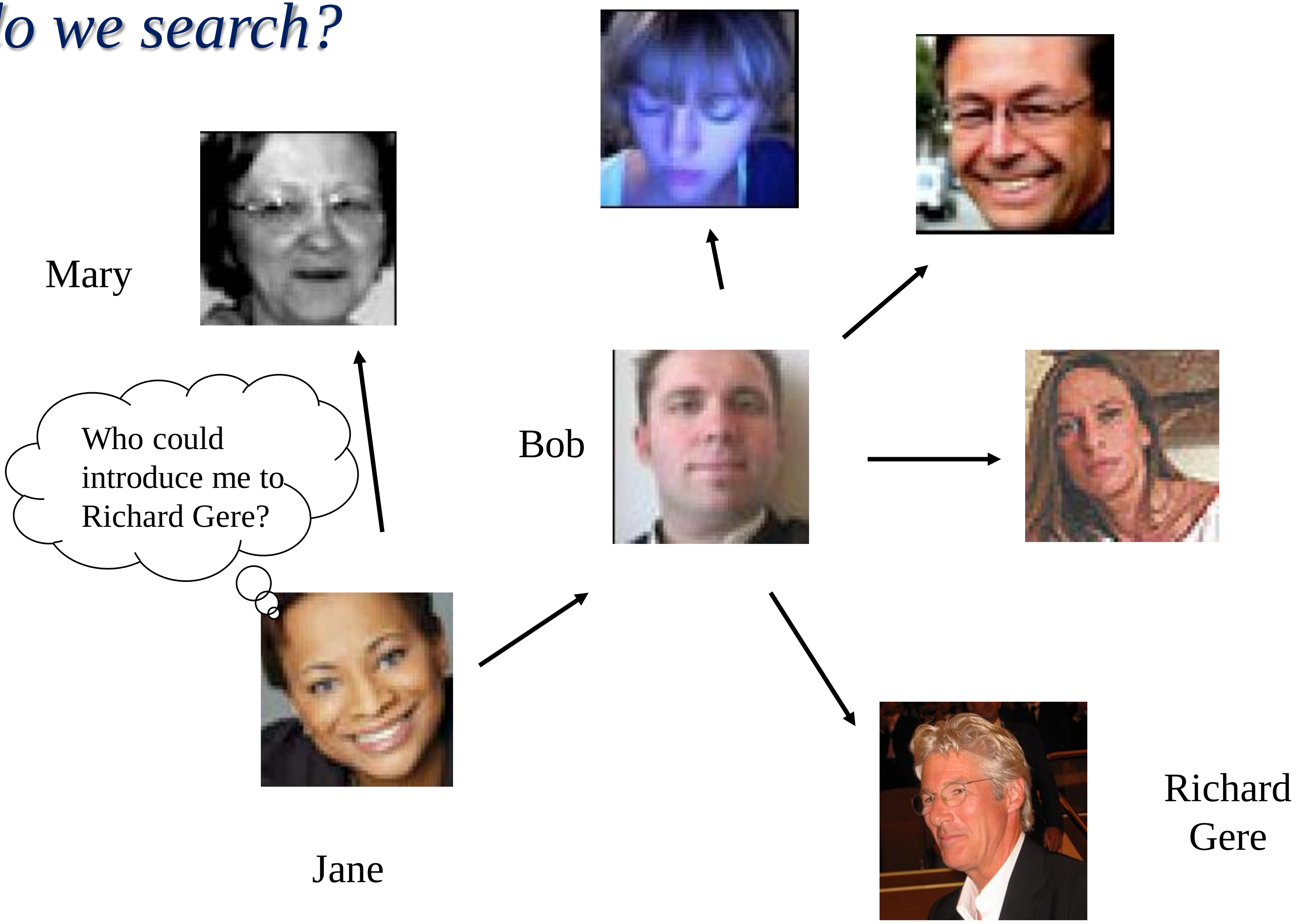
Network Search

The task of searching can correspond to various situations:

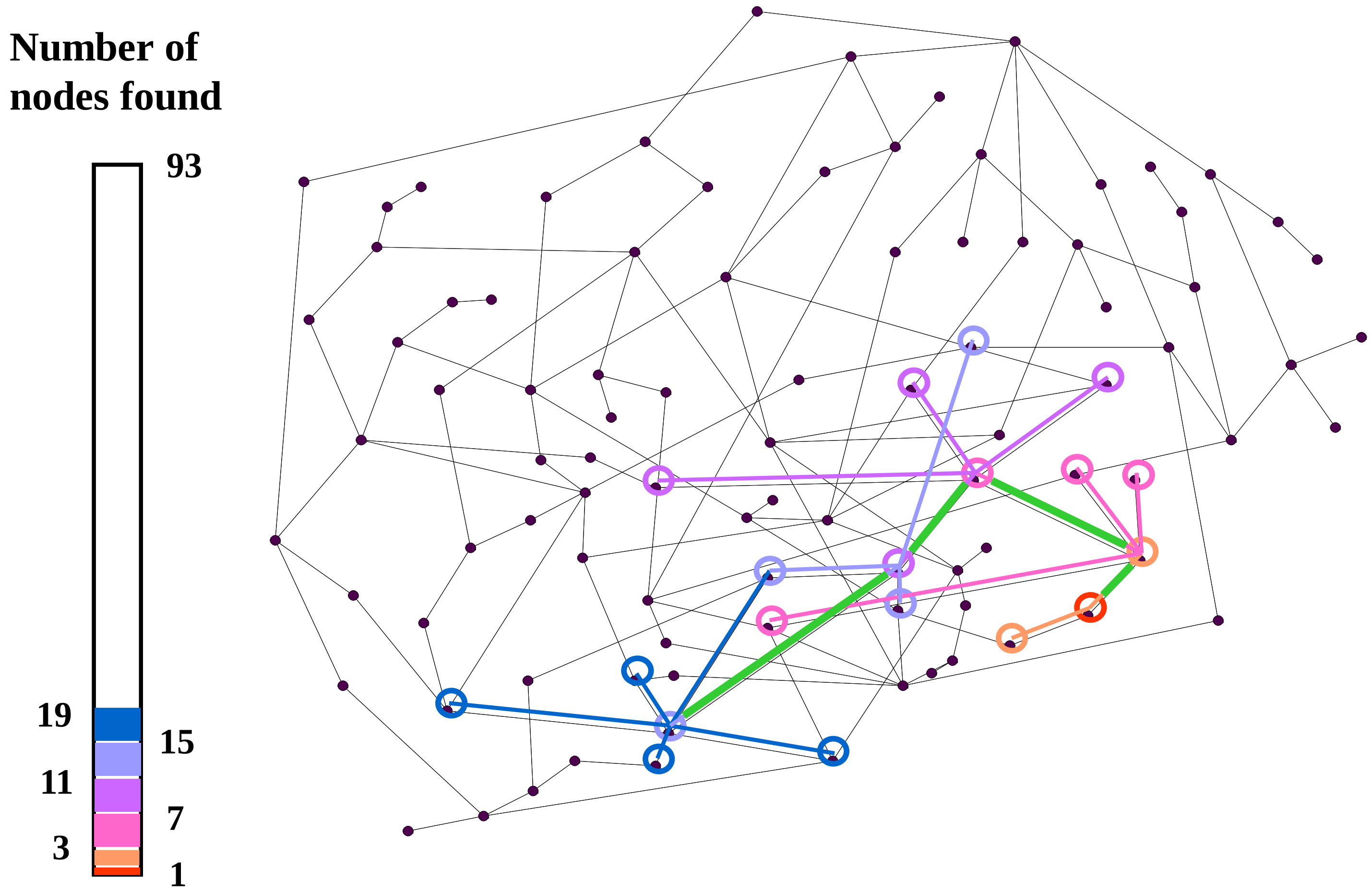
- Finding a particular node (possibly given some information on its geographical position, as in Milgram's experiment) in order to transmit a message, or

- Finding particular information without knowing a priori in which node it is stored (the case in Peer-to-Peer networks).

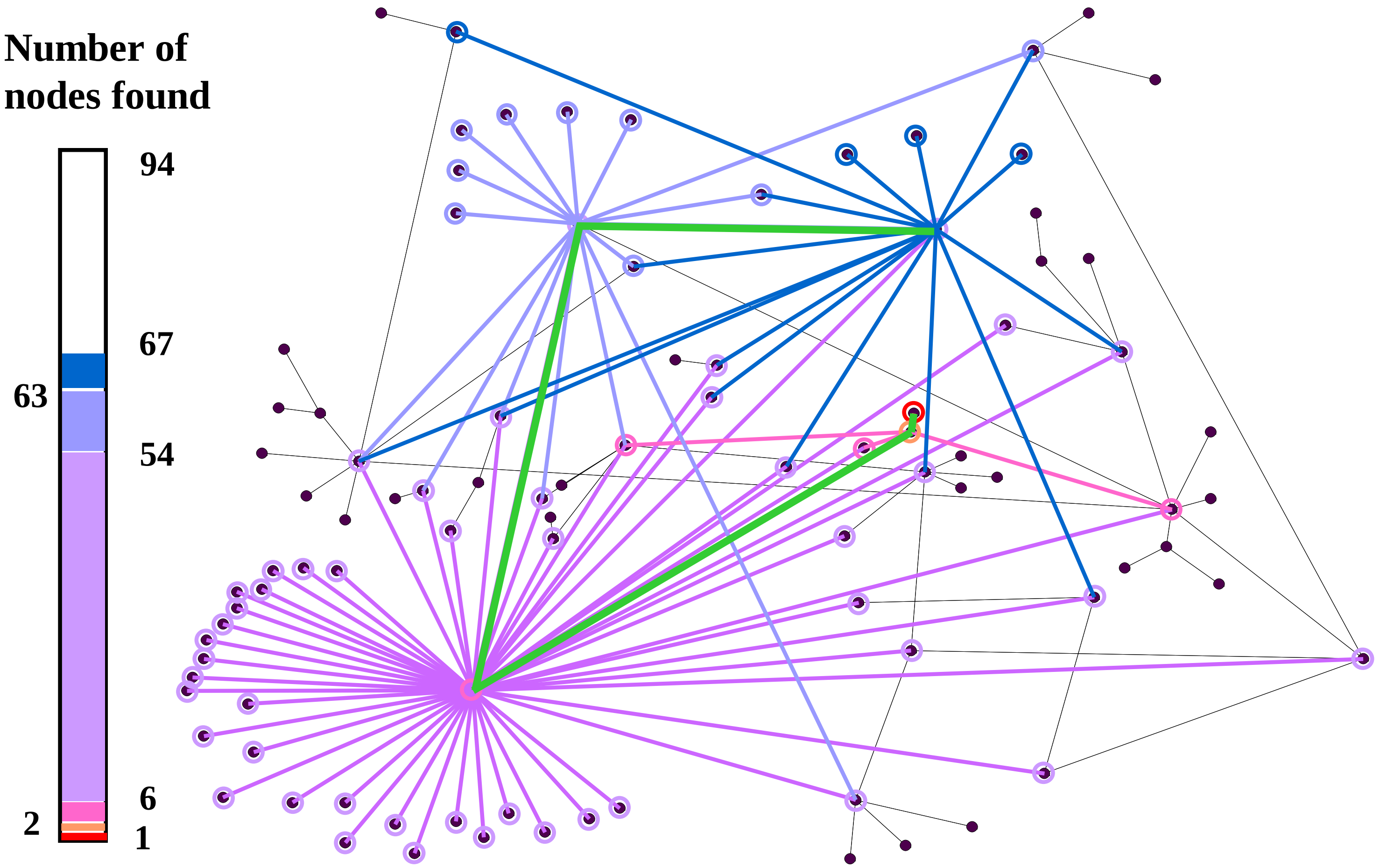
How do we search?



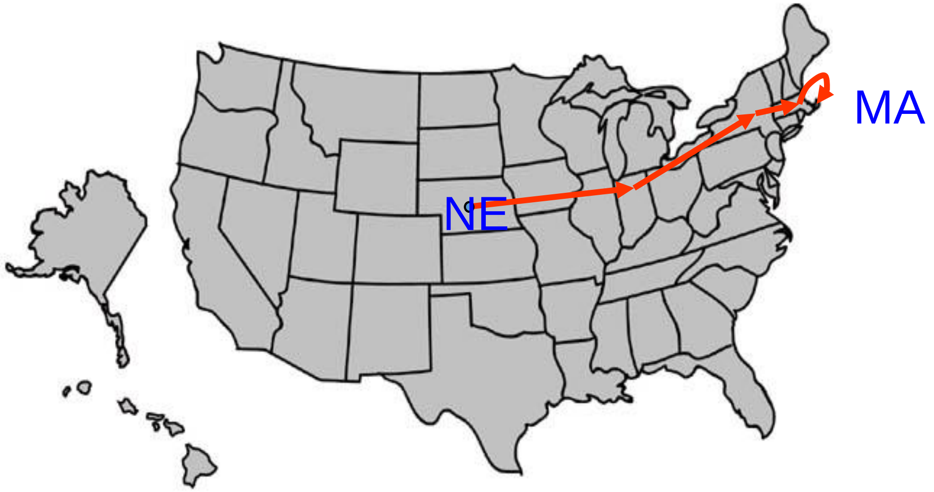
Poisson graph



Power-law graph



Recall the Small world experiments



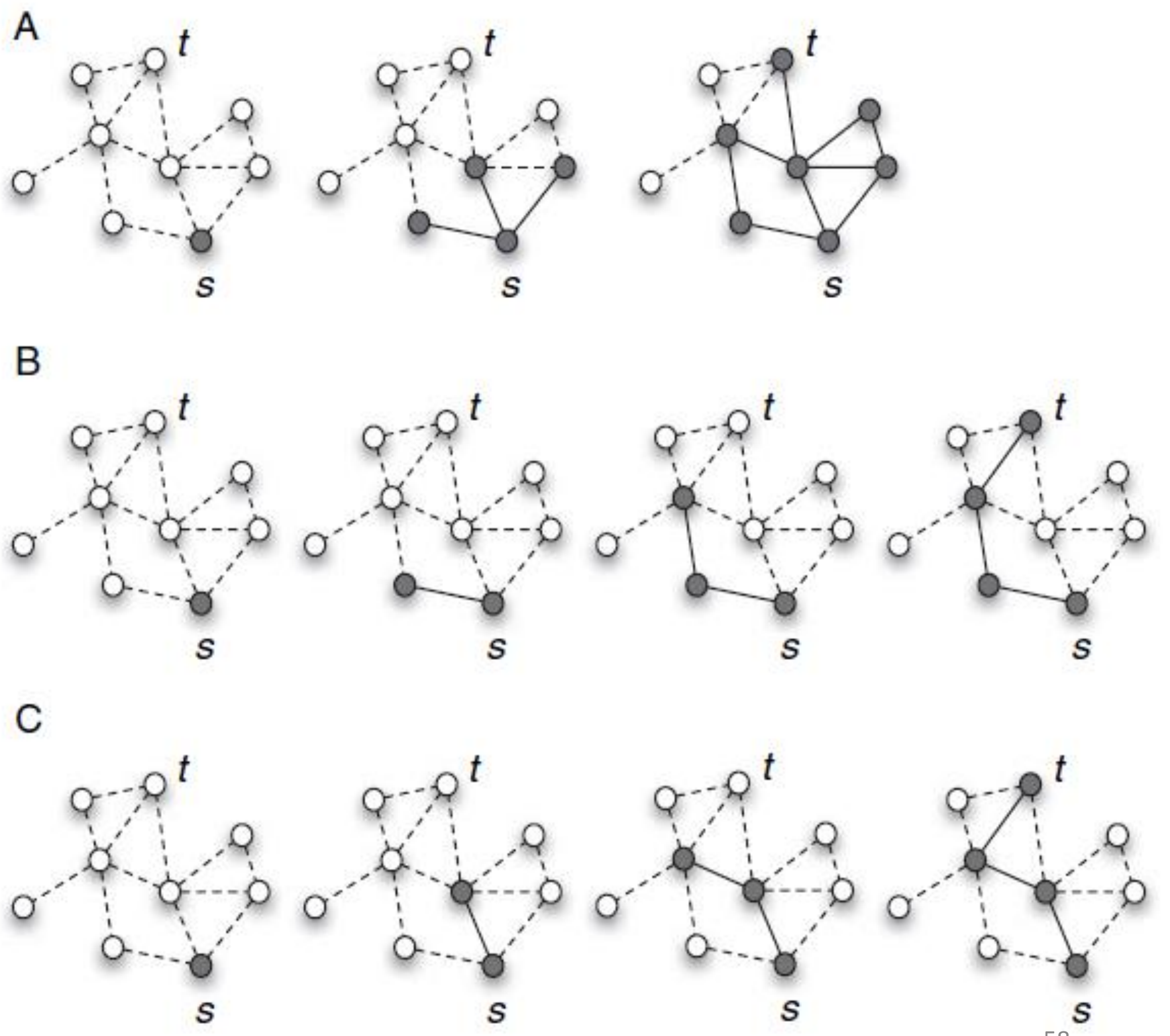
Milgram (1960's), Dodds, Muhamad, Watts (2003)

Given a target individual and a particular property, pass the message to a person you correspond with who is "closest" to the target.

Short chain lengths – six degrees of separation

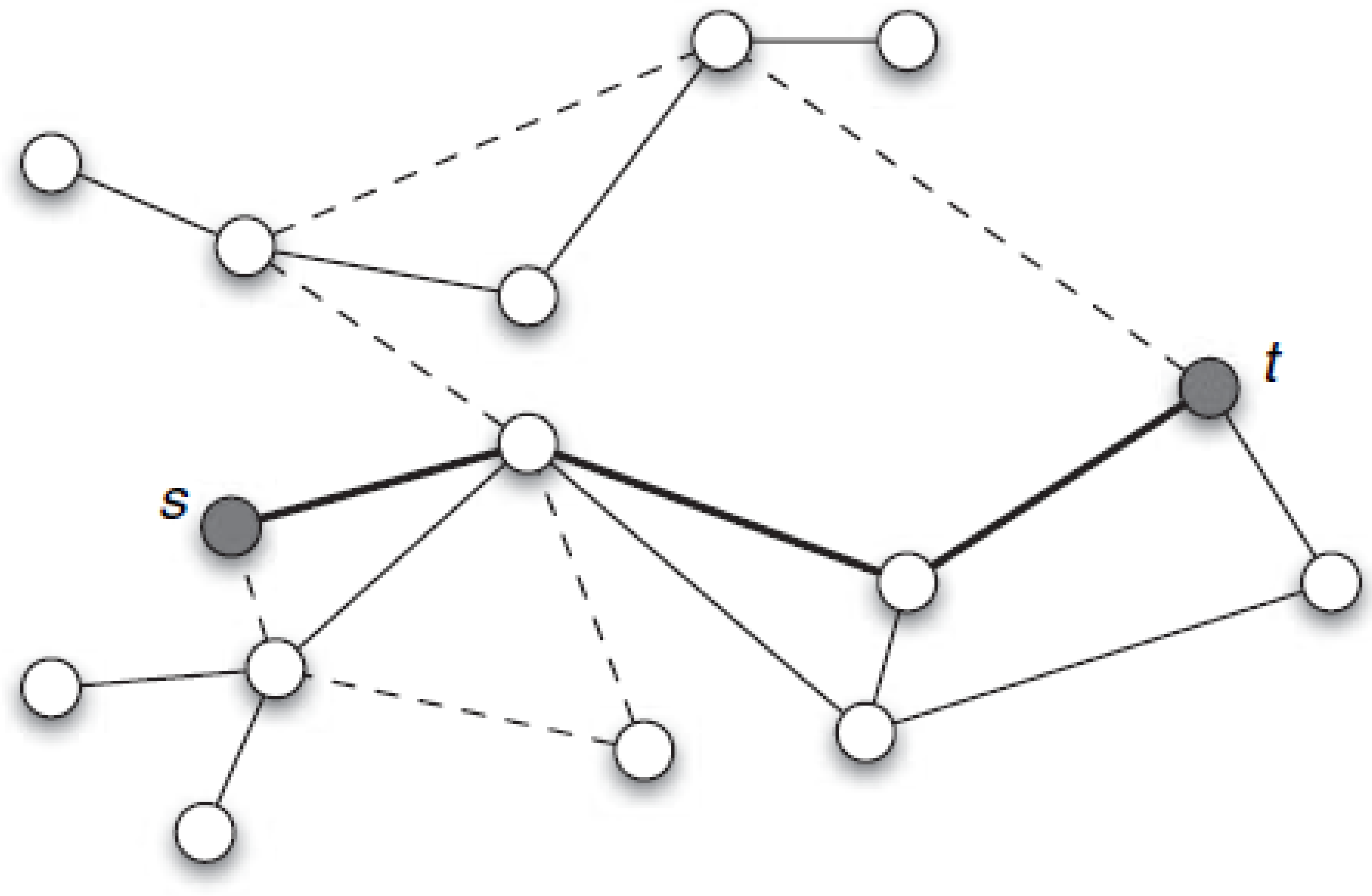
Typical strategy – if far from target choose someone geographically closer, if close to target geographically, choose someone professionally closer

Example of searching strategies



Schematic comparison of various searching strategies to find the target vertex *t*, starting from the source *s*. A, Broadcast search; B, Random walk; C, Degree-biased strategy. The broadcast search finds the shortest path, at the expense of high traffic

Example of searching strategies



Difference between the shortest path between nodes *s* and *t* (continuous thick line), of length 3, and an instance of a random walk (dashed line), here of length 6.

How do people find short paths?

How to choose among hundreds of acquaintances?

Strategy:
Simple greedy algorithm - each participant chooses correspondent who is closest to target with respect to the given property

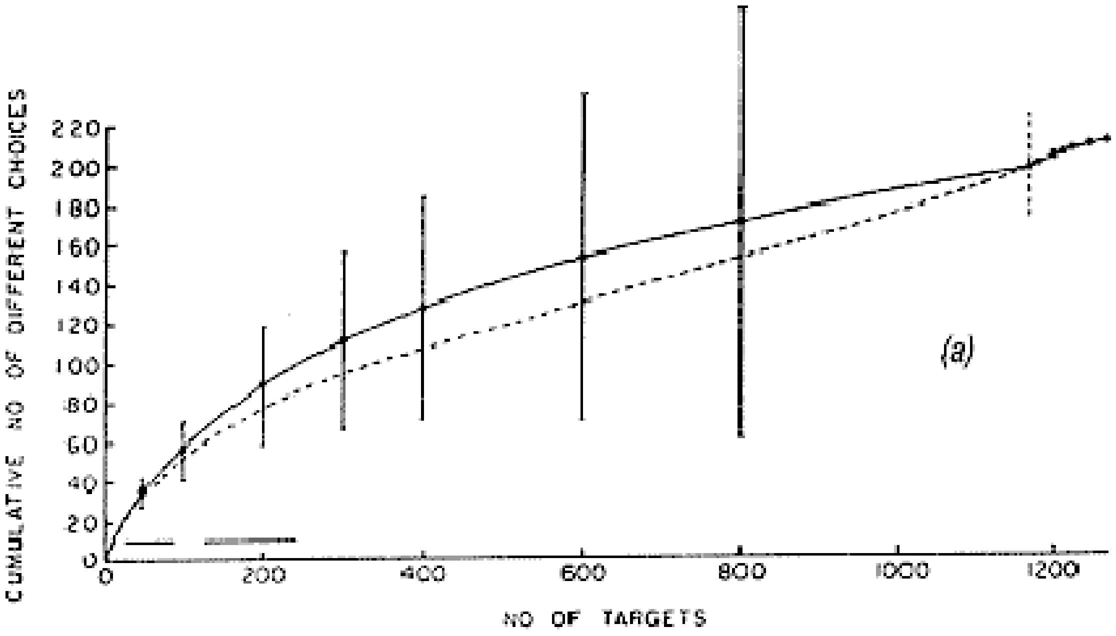
Models

high degree nodes
Adamic, Puniyani, Lukose, Huberman (2001), Newman(2003)

geography
Kleinberg (2000)

hierarchical groups
Watts, Dodds, Newman (2001), Kleinberg(2001)

Recall: The Reverse small world experiment



- Killworth & Bernard (1978)
- Given hypothetical targets (name, occupation, location, hobbies, religion…) participants choose an acquaintance for each target
- Acquaintance chosen based on
- (most often) Name, occupation and geographic location
- *Simple greedy algorithm*: most similar acquaintance

Source: 1978 Peter D. Killworth and H. Russell Bernard. "*The Reverse Small World Experiment Social Networks*" 1:159–92.

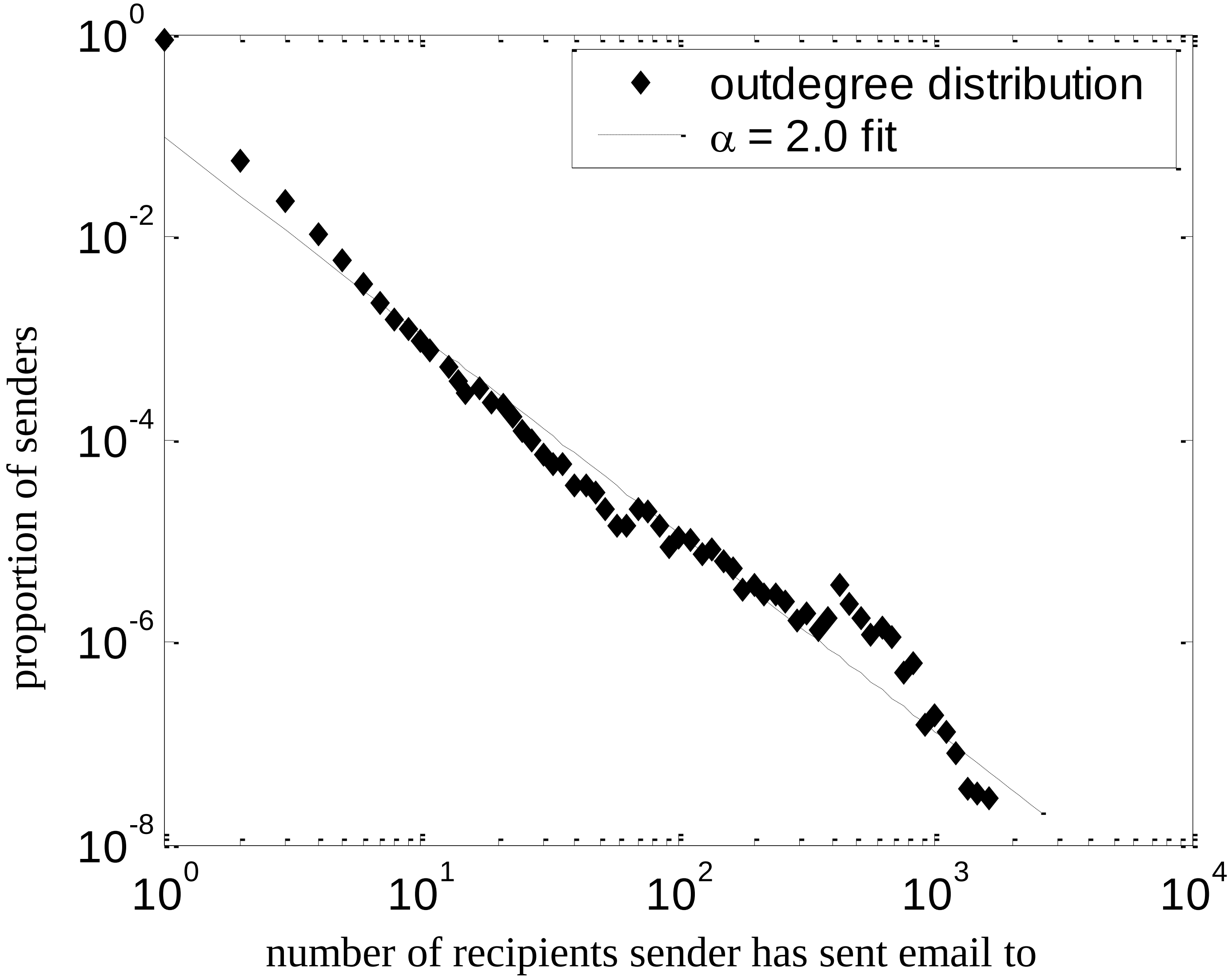
How many hops actually separate any two individuals in the world?

- People use only local information
  - not perfect in routing messages
- "*The accuracy of small world chains in social networks*"

  Peter D. Killworth, Chris McCarty, H. Russell Bernard & Mark House:
  - Analyzed 10,920 shortest path connections between 105 members of an interviewing bureau,
  - together with the equivalent conceptual, or 'small world' routes, which use individuals' selections of intermediaries
    - study the impact of accuracy within small world chains
  - The mean small world path length (3.23) is 40% longer than the mean of the actual shortest paths (2.30)
    - Model suggests that people make a less than optimal choice more than half the time

Strategy 1: High Degree Search

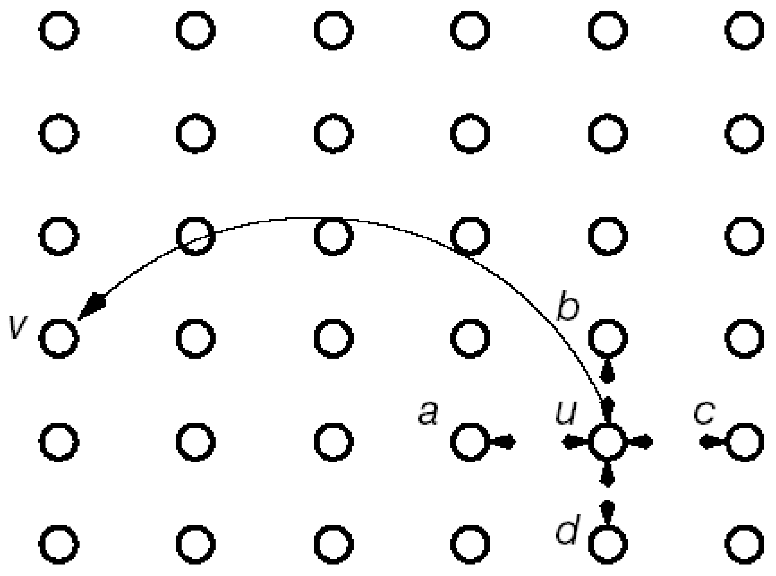
Power-law degree distribution of all senders of email passing through HP labs



Strategy 2: Geography

Recall: Kleinberg's Spatial search

Kleinberg, "*The Small World Phenomenon, An Algorithmic Perspective*"
Proc. 32nd ACM Symposium on Theory of Computing. (Nature 2000)



"*The geographic movement of the [message] from Nebraska to Massachusetts is striking. There is a progressive closing in on the target area as each new person is added to the chain*",
S.Milgram 'The small world problem', Psychology Today 1,61,1967

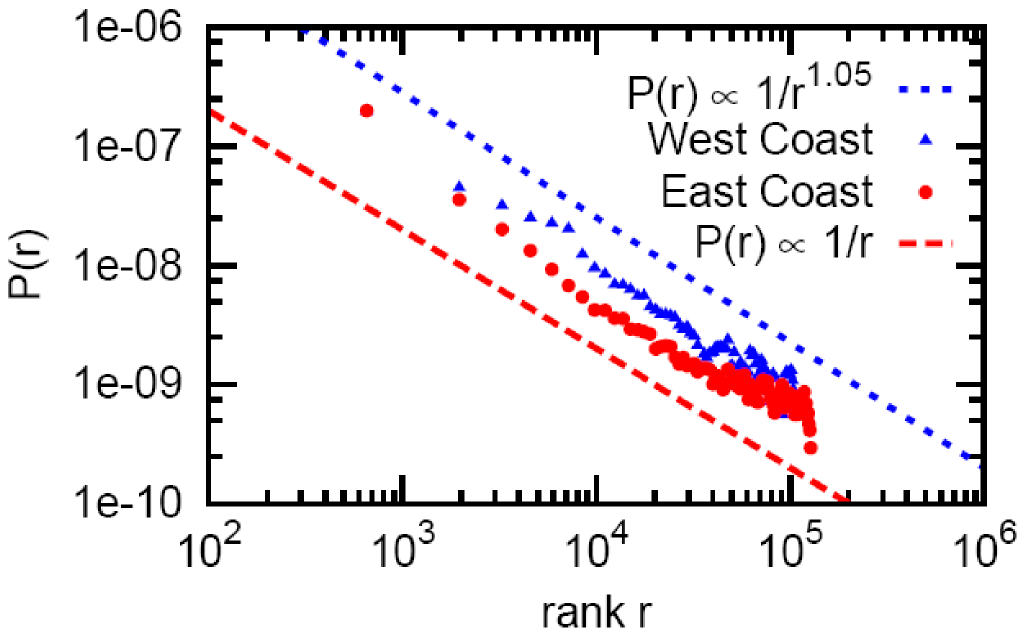
nodes are placed on a lattice and
connect to nearest neighbors

additional links placed with
p(link between u and v) = (distance(u,v))$^{-\alpha}$

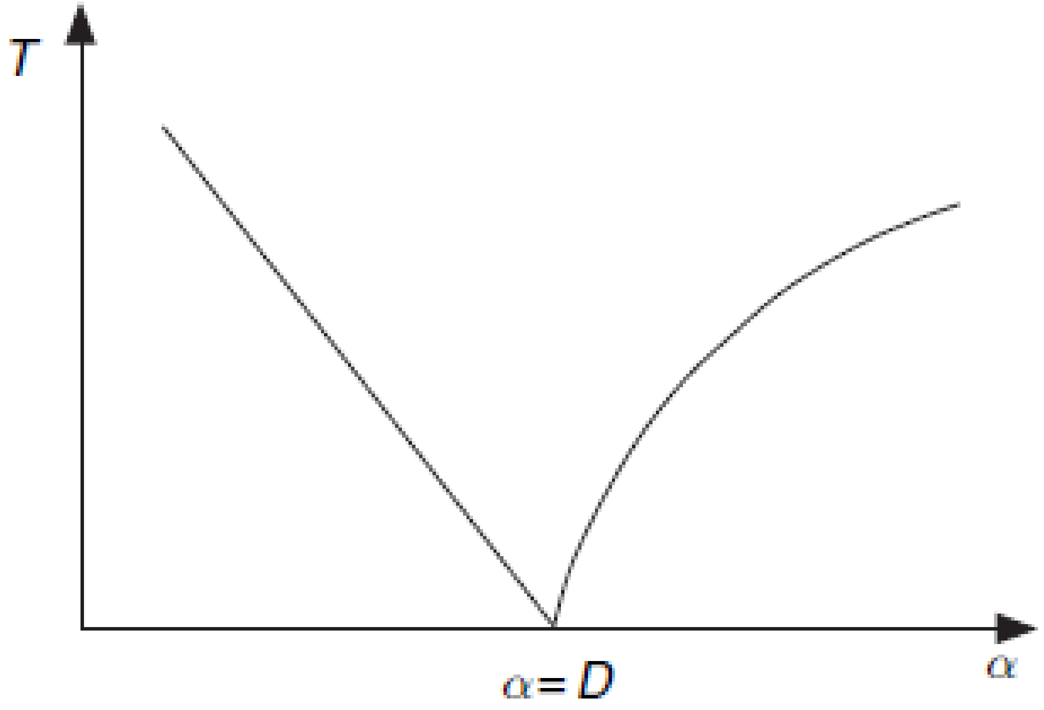
exponent that will determine navigability

Navigability in networks of variable geographical density

- Kleinberg assumed a uniformly populated 2D lattice
- But population is far from uniform
- population networks and rank-based friendship
  - probability of knowing a person depends not on absolute distance but on relative distance
    - i.e. how many people live closer $Pr[u \to v] \sim 1/rank_u(v)$



Navigability in networks of variable geographical density

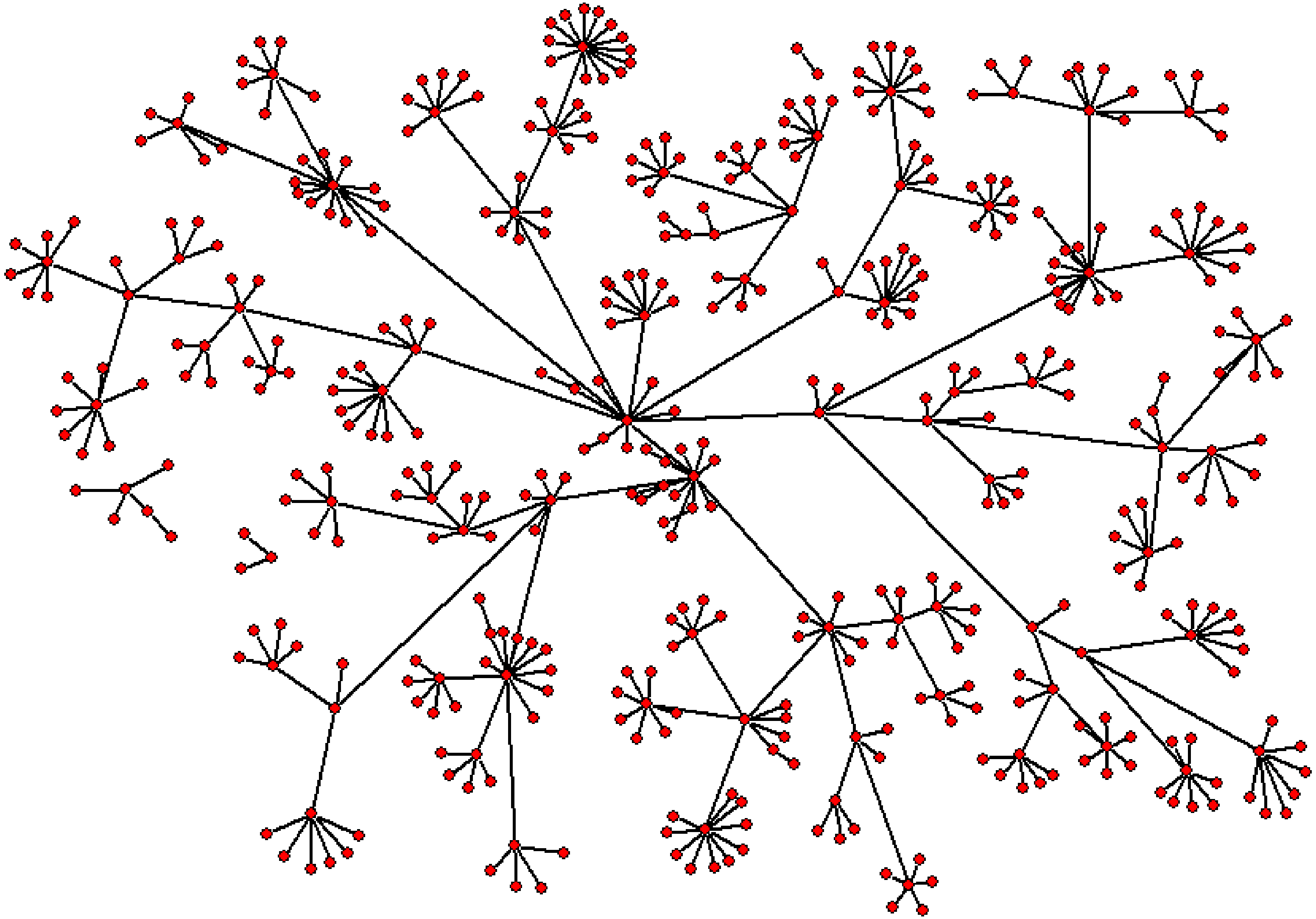


D is the spatial dimension

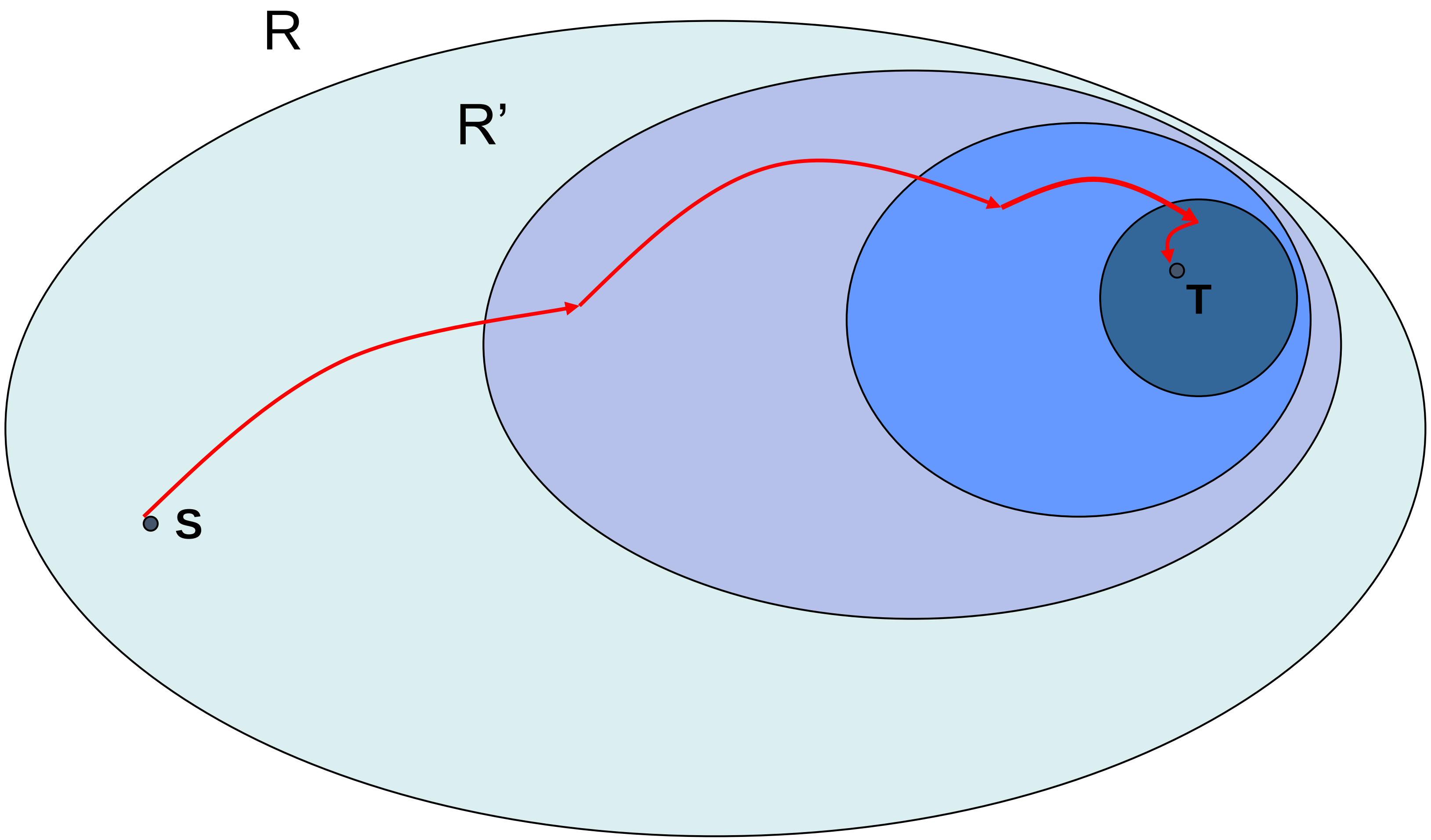
Schematic evolution of the delivery time as a function of α in Kleinberg's model of a small-world network. Figure adapted from Kleinberg (2000).

Strategy 3: Organizational Hierarchy

How about if we do not have geography? Does community structure help?

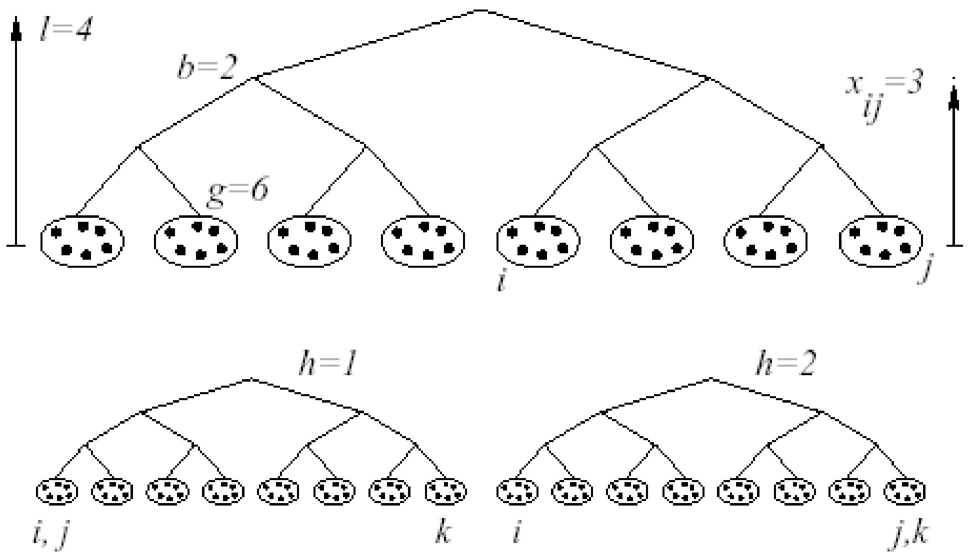


Why search is fast in hierarchical topologies



Example: Hierarchical models with multiple hierarchies

Individuals belong to hierarchically nested groups

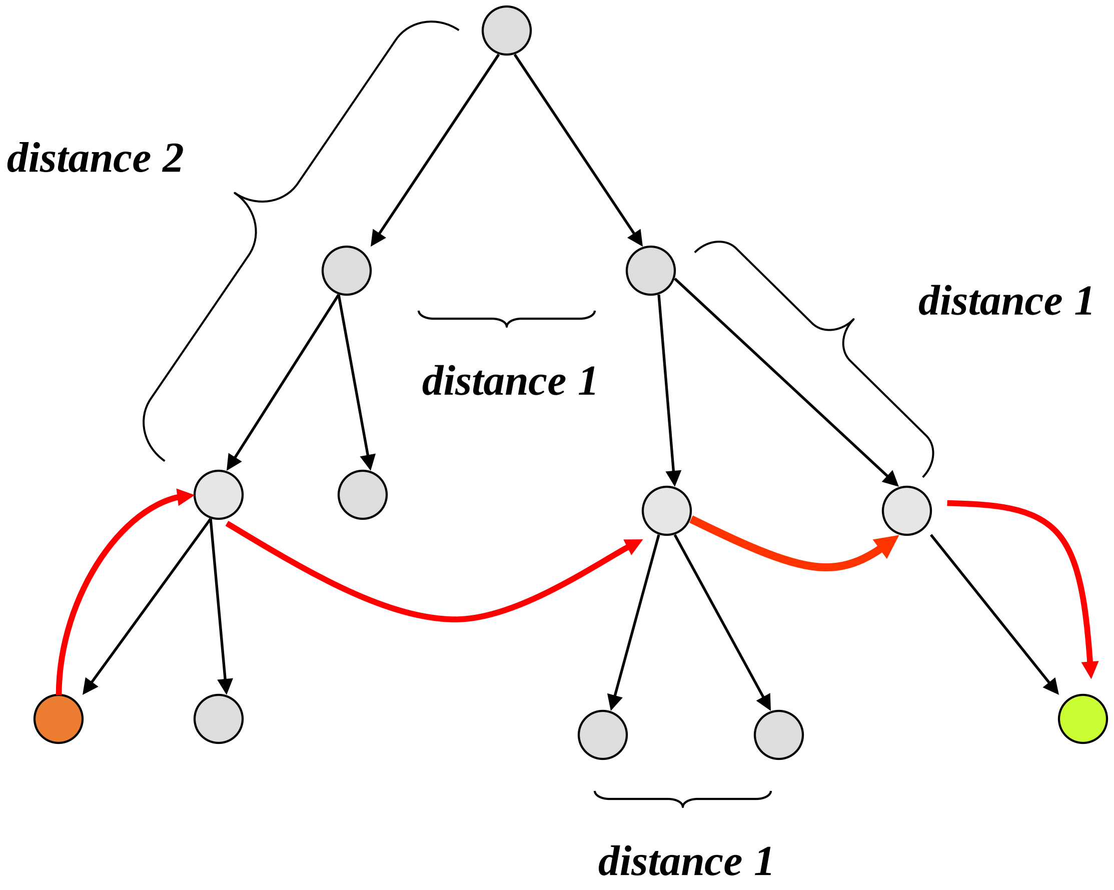


$$p_{ij} \sim e^{-\alpha x}$$

Multiple independent hierarchies h=1,2,..,H coexist corresponding to occupation, geography, hobbies, religion

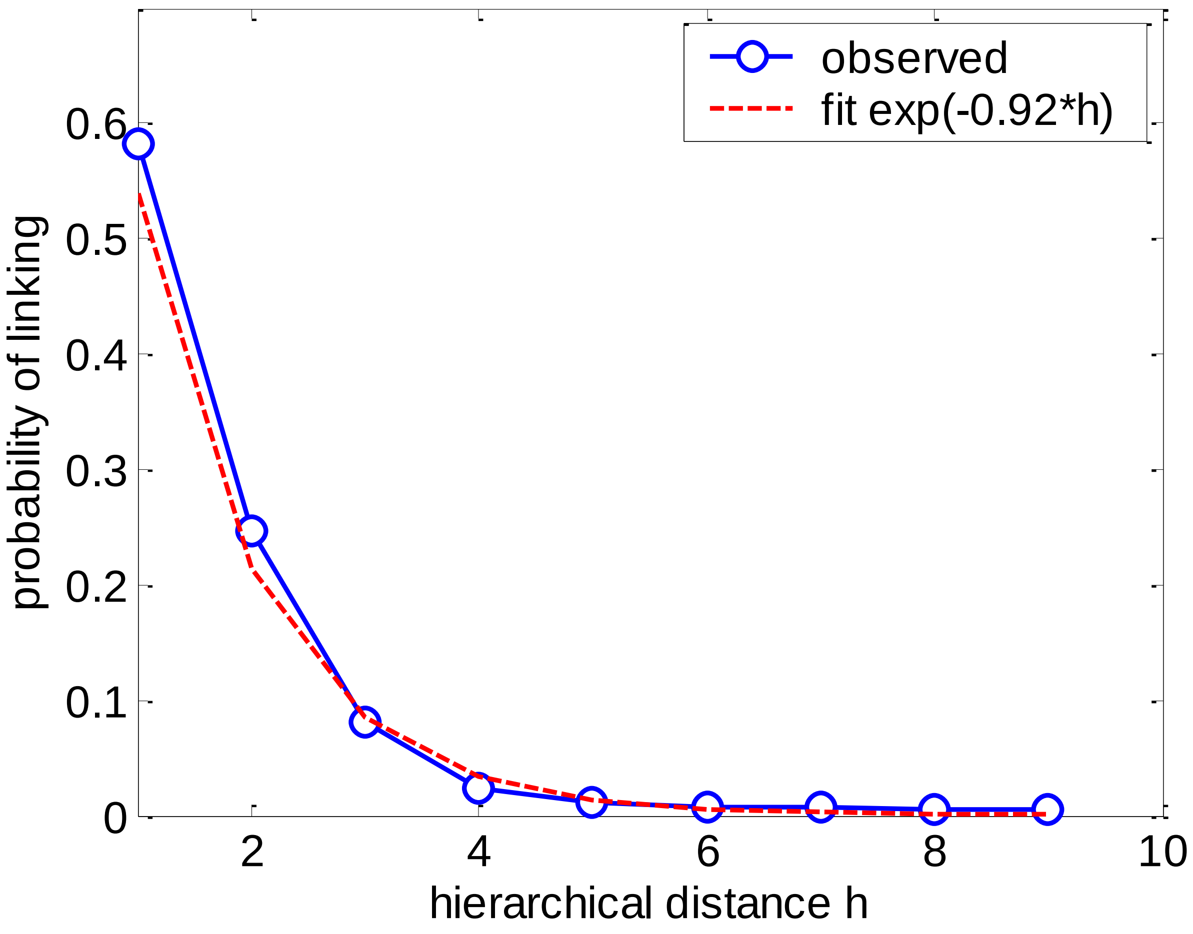
Source: Identity and Search in Social Networks: Duncan J. Watts, Peter Sheridan Dodds, and M. E. J. Newman

Example of search path



hierarchical distance = 5
search path distance = 4

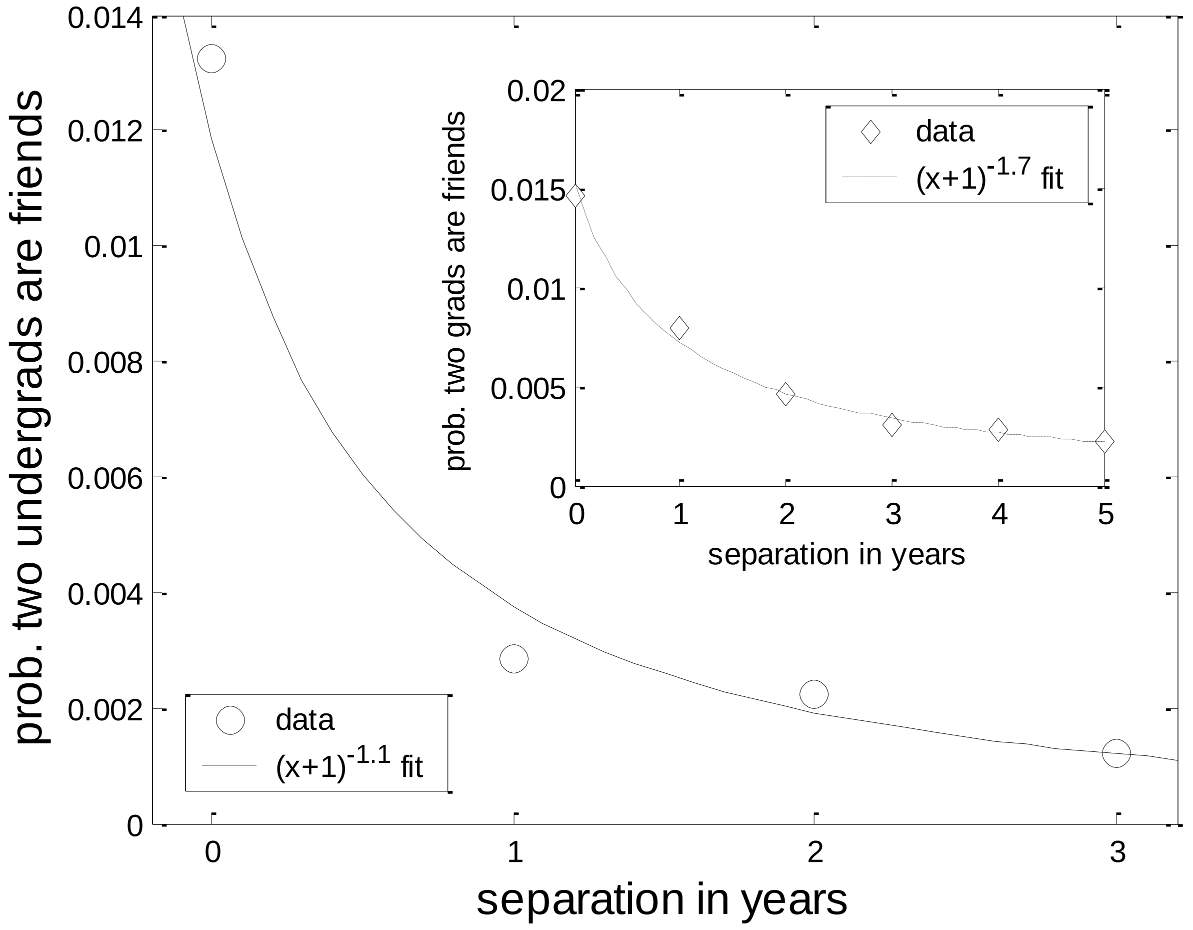
Probability of linking vs. distance in hierarchy



in the 'searchable' regime: $0 < \alpha < 2$ (Watts, Dodds, Newman 2001)

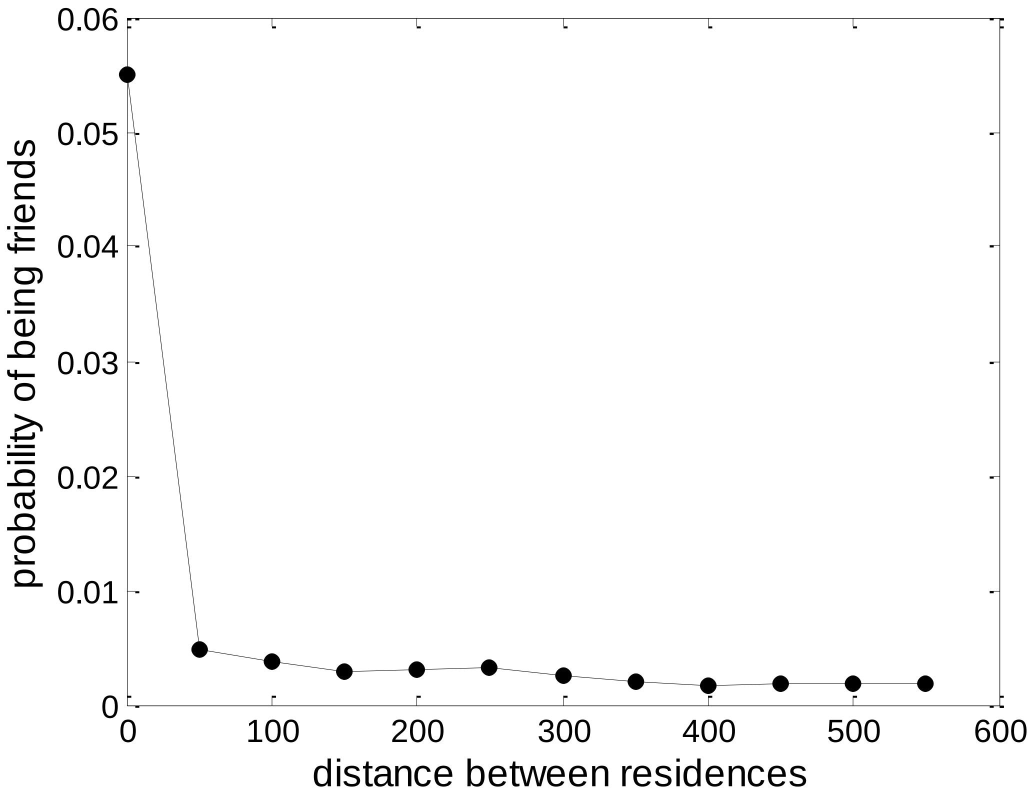
Problem: How to construct hierarchies?

Probability of linking by separation in years



source: Adamic and Adar, **[How to search a social network](#)**, Social Networks, 27(3), p.187-203, 2005.

Hierarchies not useful for other attributes: Geography



Other attributes: major, sports, freetime activities, movie preferences…

source: Adamic and Adar, **How to search a social network**, Social Networks, 27(3), p.187-203, 2005.

Comparison of the search time of the three strategies:

Search times using various strategies

Strategy	Median number of steps	Mean number of steps
High degree	16	43.2
Organizational hierarchy	4	5.0
Geography	6	11.7

The actual average shortest path is 3.1.

source: Adamic and Adar, **How to search a social network**, Social Networks, 27(3), p.187-203, 2005.

Summary

- Individuals associate on different levels into groups.
- Group structure facilitates decentralized search using social ties
- Hierarchy search faster than geographical search
- A fraction of ‘important’ individuals are easily findable
- Humans may be more resourceful in executing search tasks:
  - making use of weak ties
  - using more sophisticated strategies