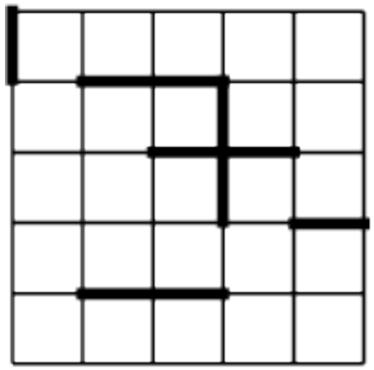
Network resilience

Outline

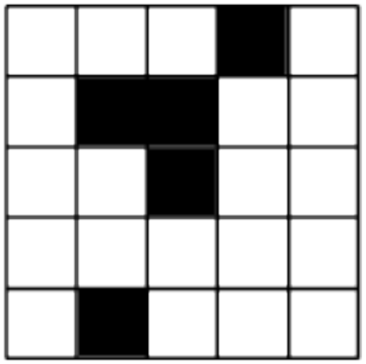
- network resilience
  - effects of node and edge removal
  - example: power grid
  - example: biological networks

Network resilience

- 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 (previously studied on lattices):



bond percolation

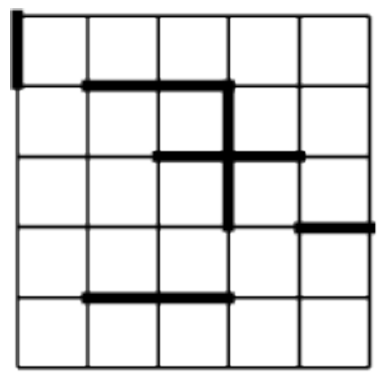


site percolation

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

Bond percolation in Networks

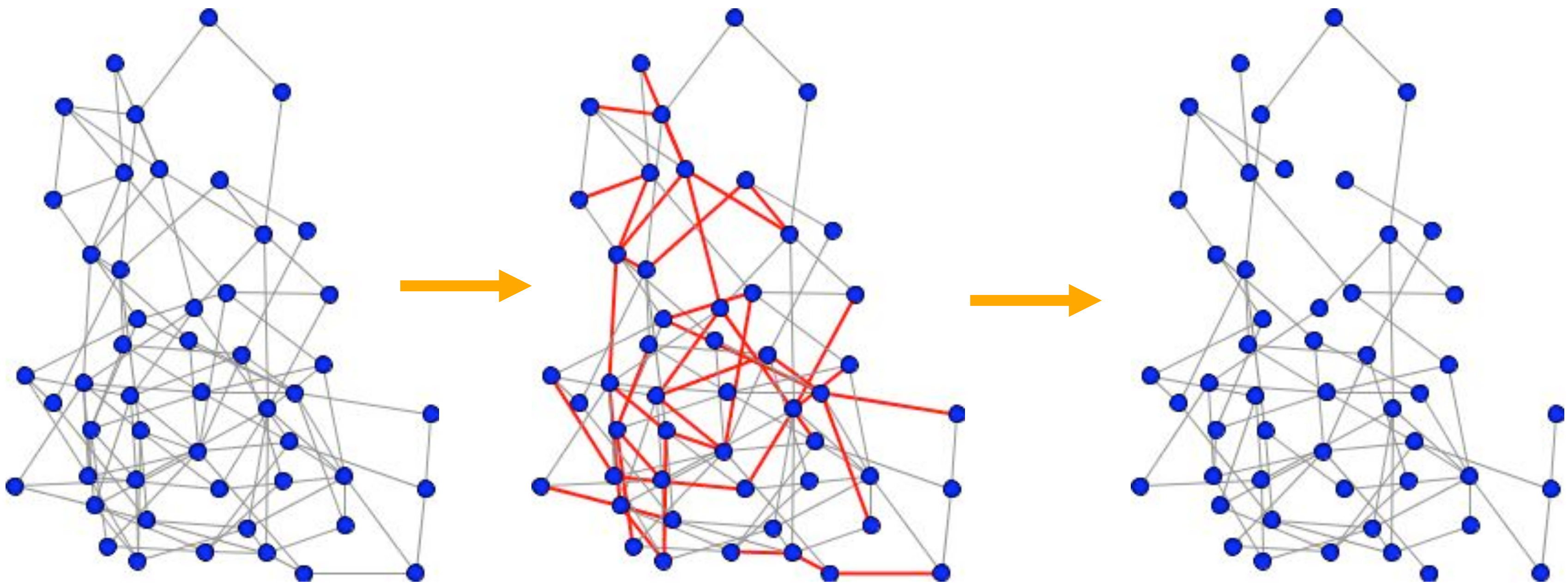
- Edge removal
  - bond percolation: each edge is removed with probability (1-p)
    - corresponds to random failure of links
  - targeted attack: causing the most damage to the network with the removal of the fewest edges
    - strategies: remove edges that are most likely to break apart the network or lengthen the average shortest path
    - e.g. usually edges with high betweenness



bond percolation

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

Edge percolation



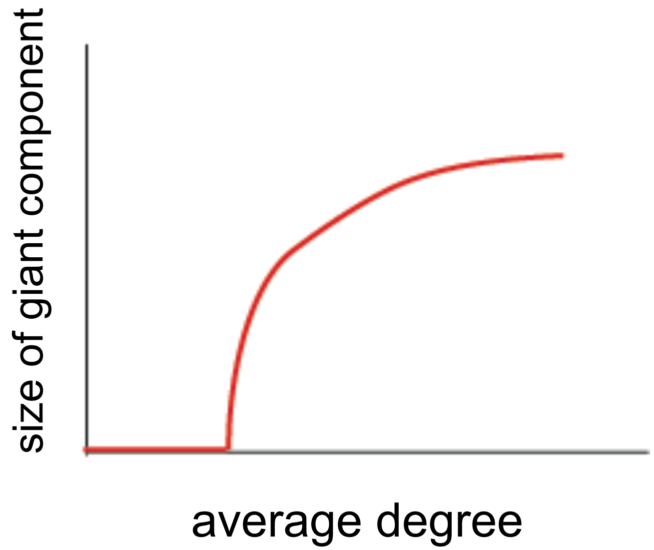
How many edges would you have to remove to break up an Erdos Renyi random graph? e.g. each node has an average degree of 4.6

50 nodes, 116 edges, average degree 4.64

after 25 % edge removal

76 edges, average degree 3.04 – still well above percolation threshold

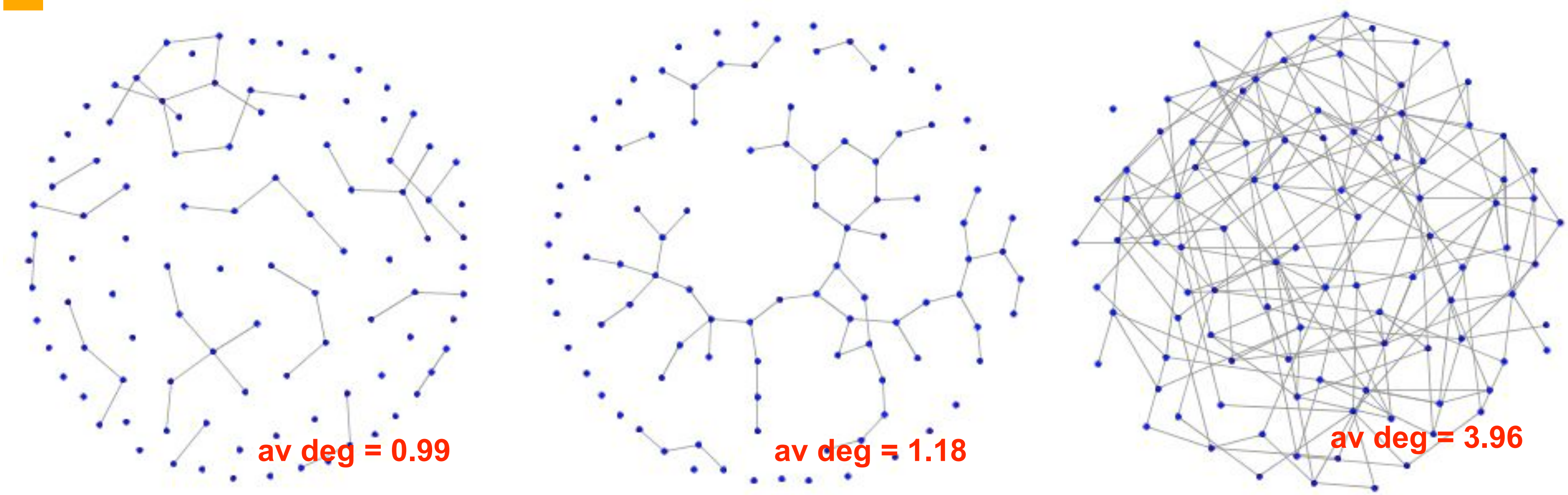
Percolation threshold in Erdos-Renyi Graphs



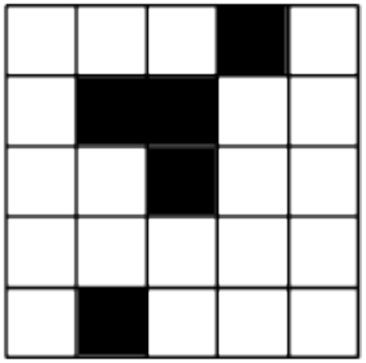
Percolation threshold: the point at which the giant component emerges

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

Edge removal is the opposite process –as the average degree drops below 1 the network becomes disconnected



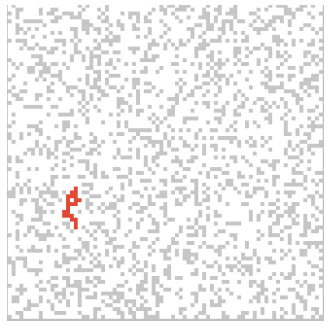
Site percolation on lattices



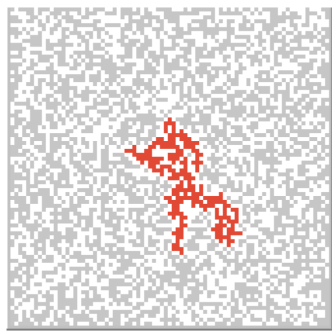
site percolation

Interactive demonstration:
http://projects.si.umich.edu/netlearn/NetLogo4/LatticePercolation.html

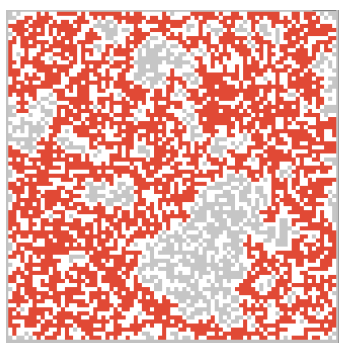
Fill each square with probability p



- **low p:** small isolated islands



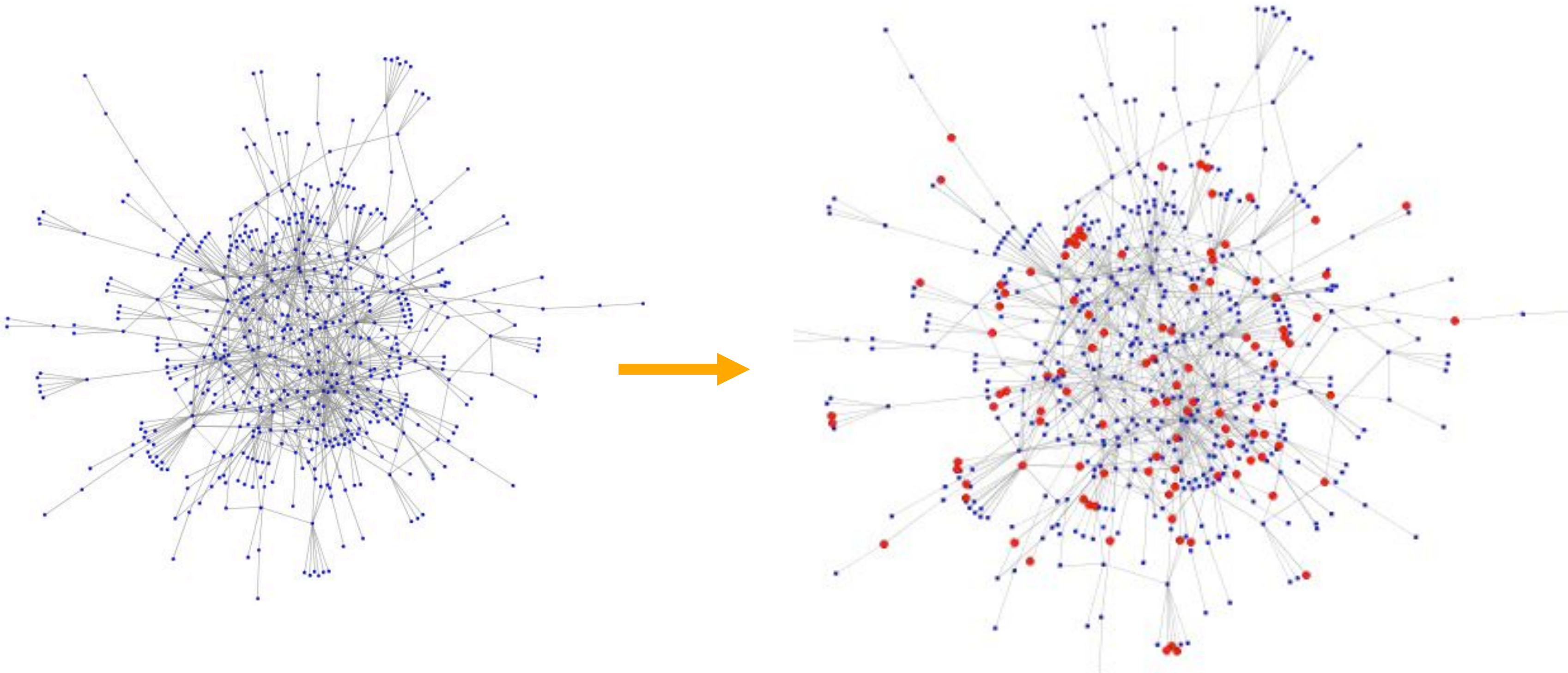
- **p critical**: giant component forms, occupying finite fraction of infinite lattice.
  Size of other components is power law distributed



- **p above critical**: giant component rapidly spreads to span the lattice.
  Size of other components is O(1).

Source: site percolation, http://mathworld.wolfram.com/BondPercolation.html

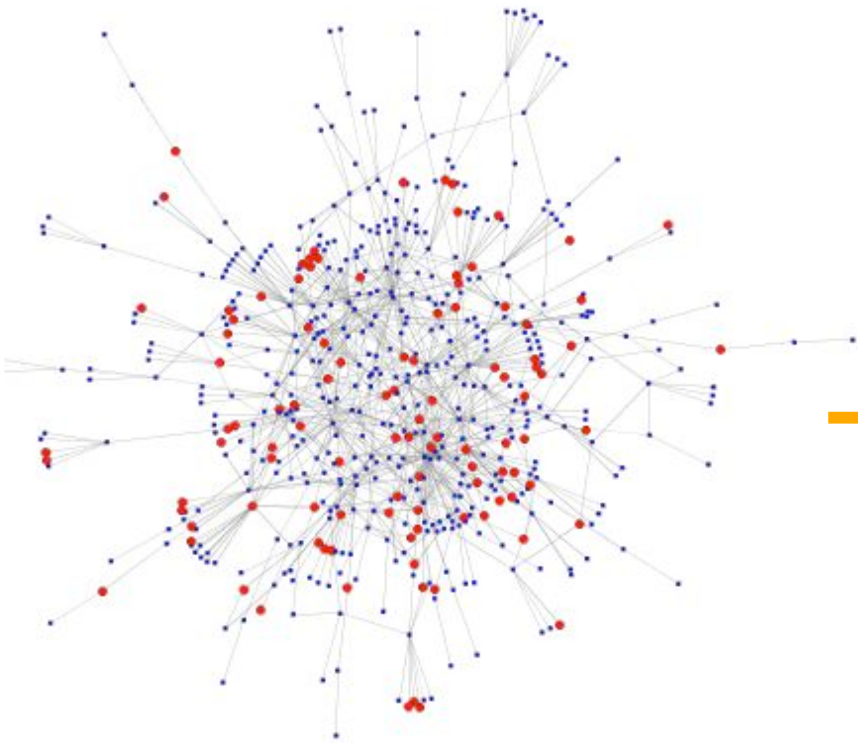
Percolation on Complex Networks



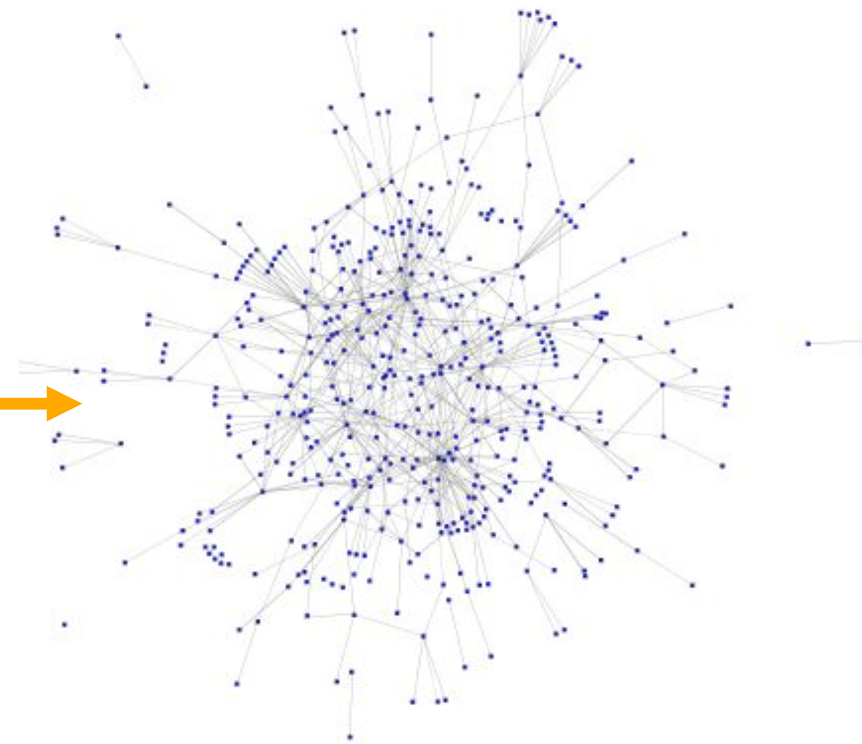
- Percolation can be extended to networks of arbitrary topology.
- We say the network percolates when a giant component forms.

Scale-free networks are resilient with respect to random attack

- Example: gnutella network, 20% of nodes removed



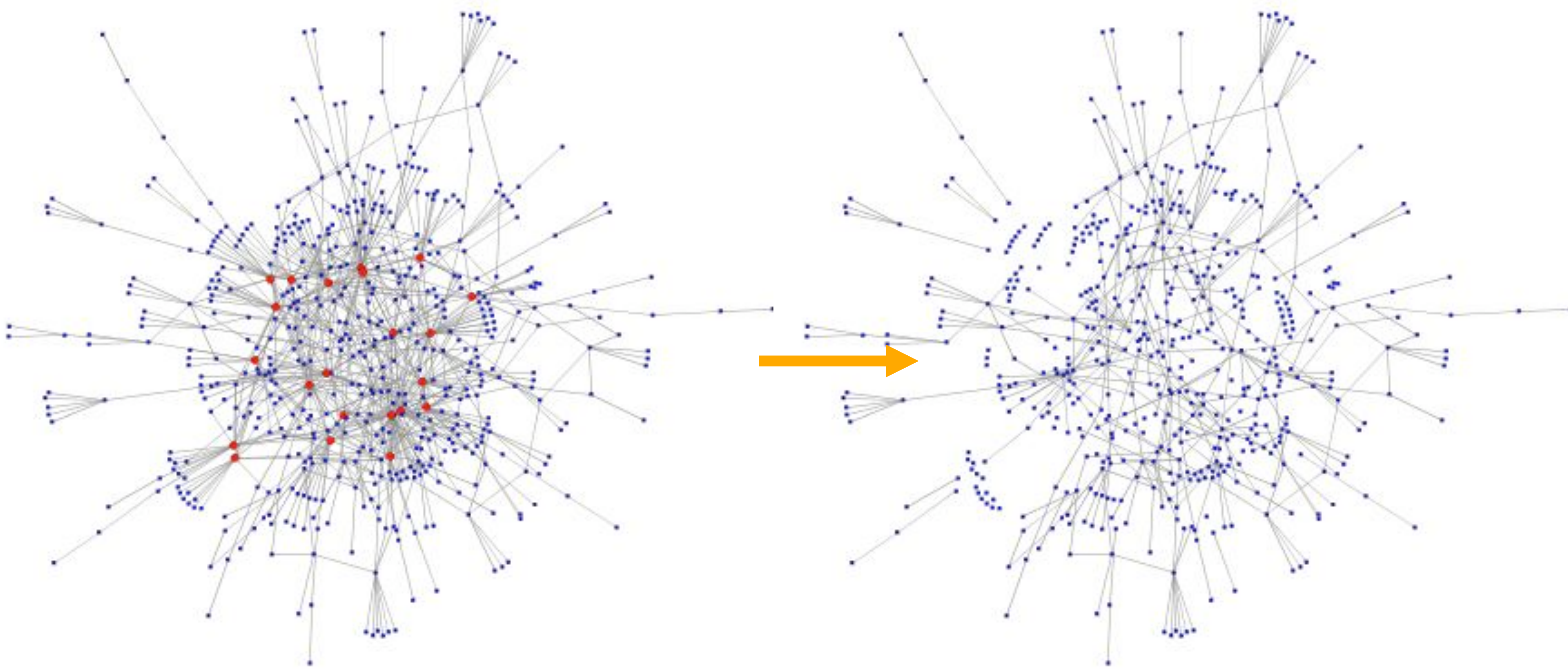
574 nodes in giant component



427 nodes in giant component

Targeted attacks are affective against scale-free networks

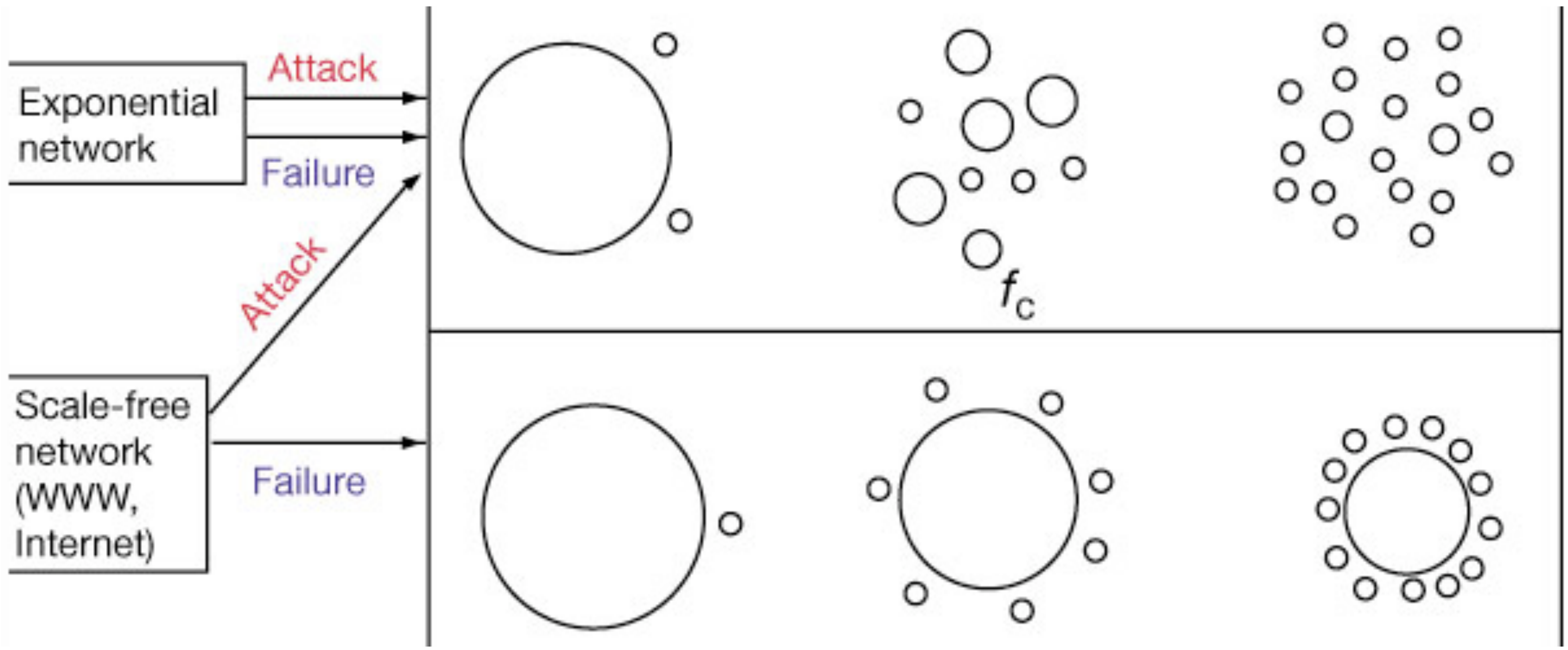
- Example: same gnutella network, 22 most connected nodes removed (2.8% of the nodes)



574 nodes in giant component

301 nodes in giant component

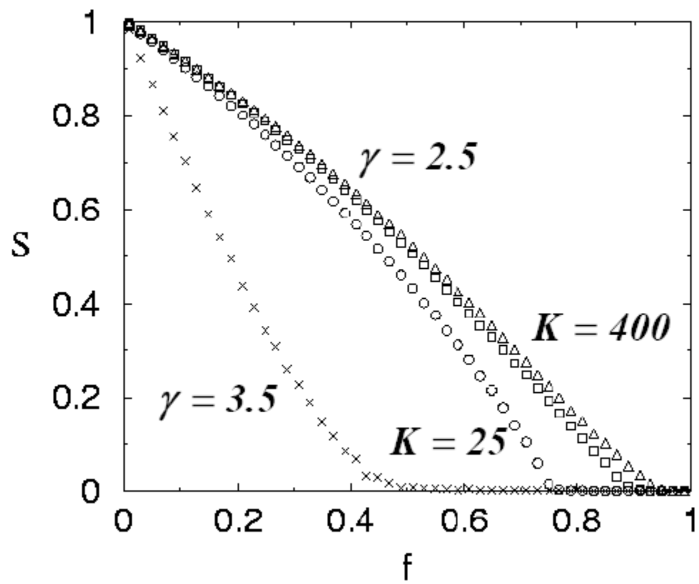
random failures vs. attacks



Source: Error and attack tolerance of complex networks. Réka Albert, Hawoong Jeong and Albert-László Barabási. Nature 406, 378-382(27 July 2000); http://www.nature.com/nature/journal/v406/n6794/abs/406378A0.html

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?

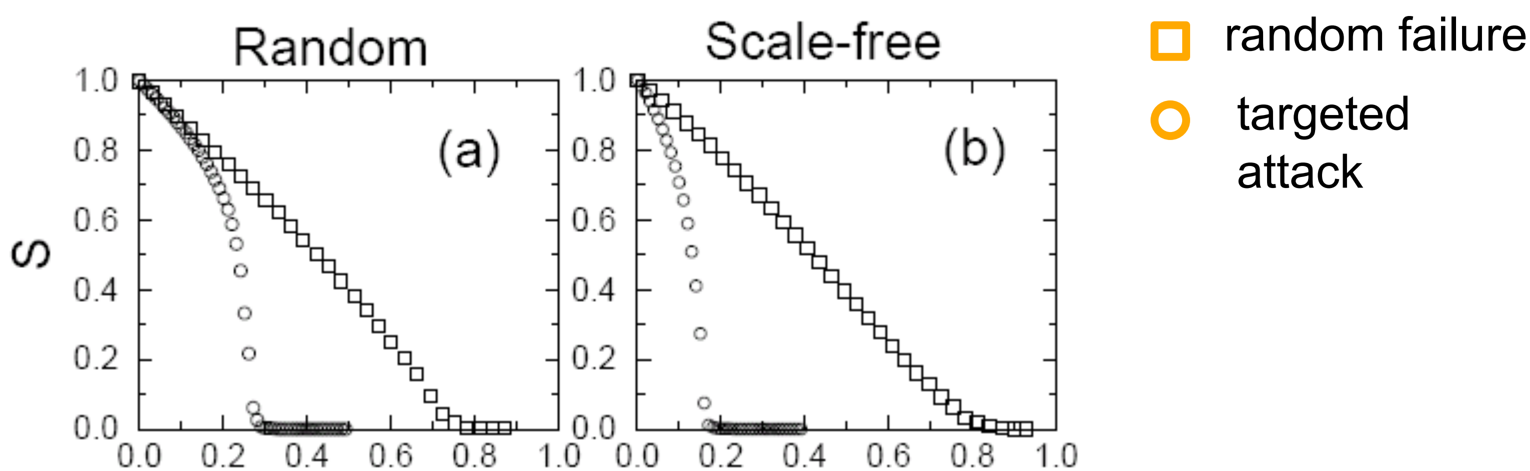


- For scale free graphs there is always a giant component (the network always percolates)

Source: Cohen et al., Phys. Rev. Lett. 85, 4626 (2000)

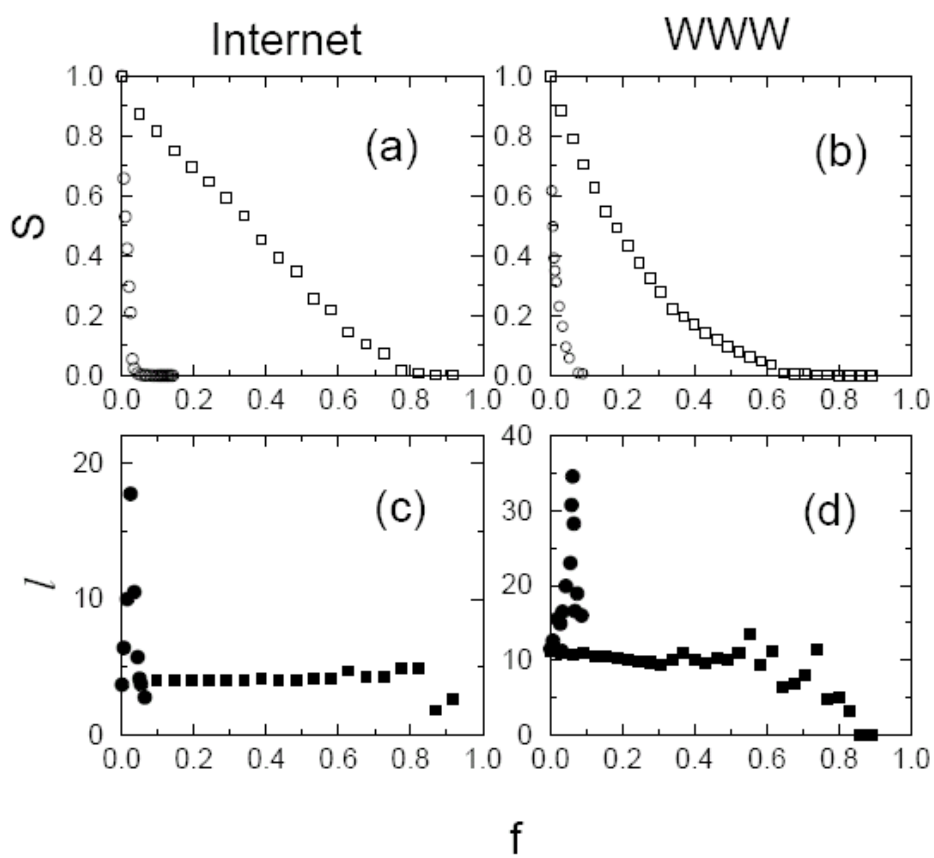
Network resilience to targeted attacks

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



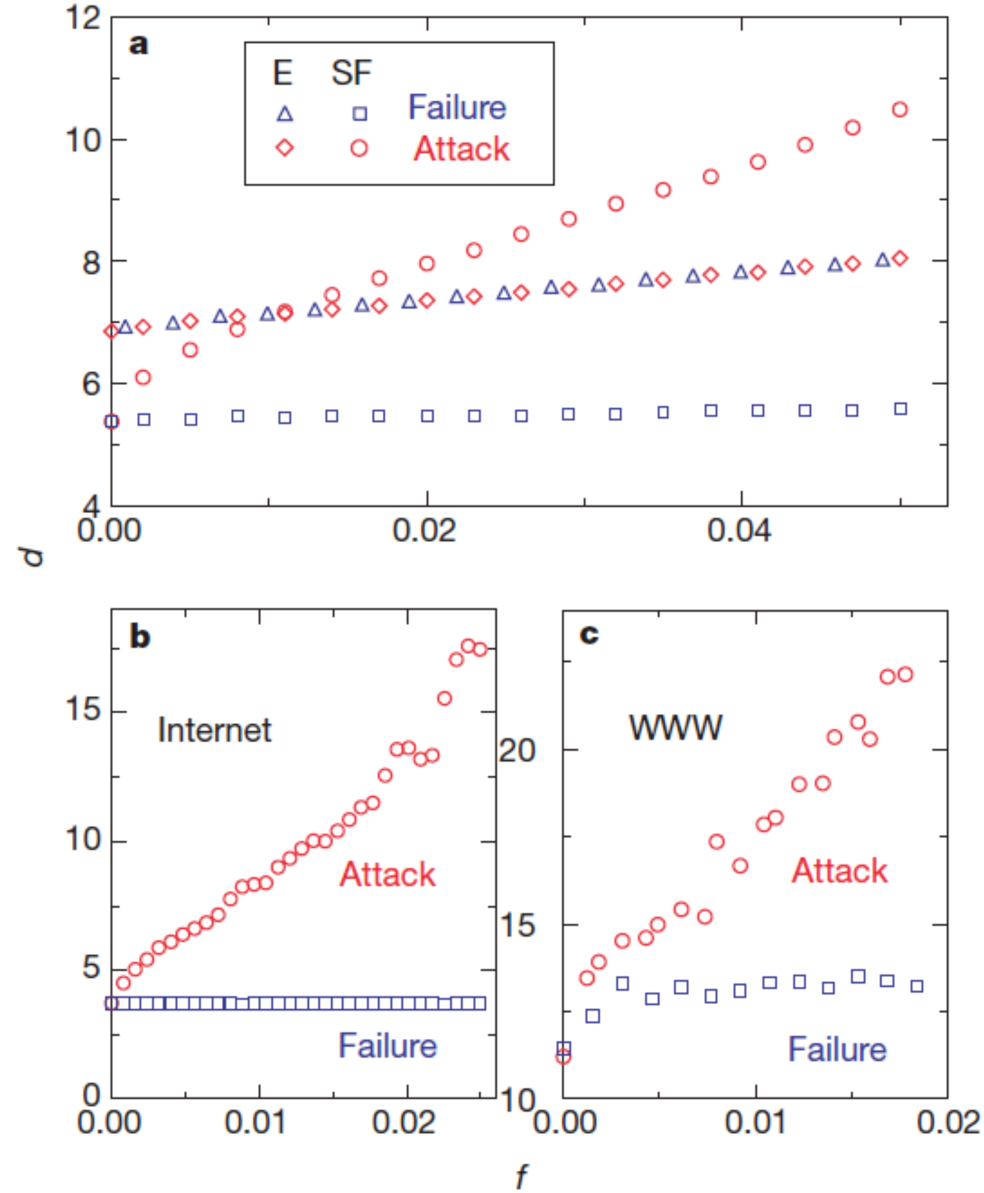
Source: Error and attack tolerance of complex networks. Réka Albert, Hawoong Jeong and Albert-László Barabási. Nature 406, 378-382(27 July 2000); http://www.nature.com/nature/journal/v406/n6794/abs/406378A0.html

Real networks



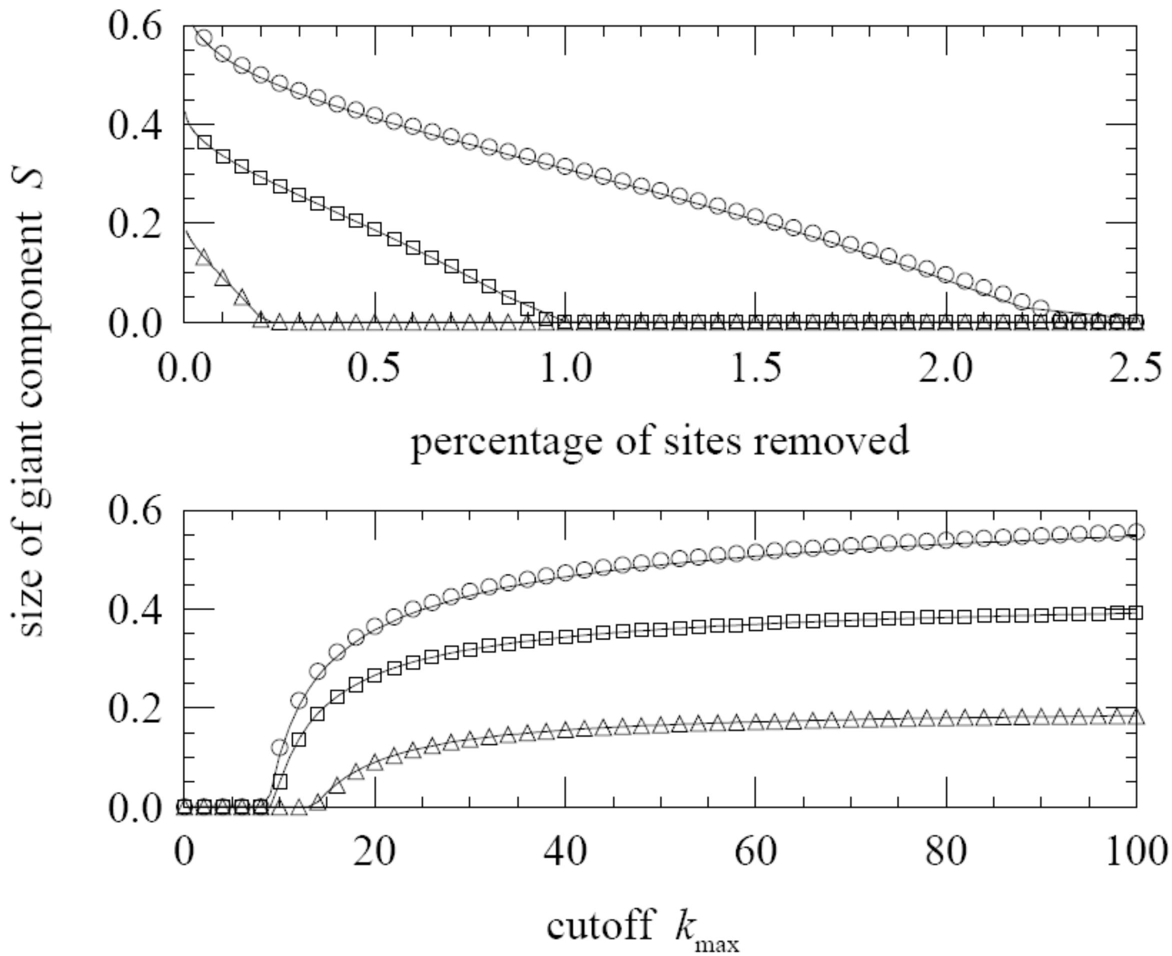
Source: Error and attack tolerance of complex networks. Réka Albert, Hawoong Jeong and Albert-László Barabási. Nature 406, 378-382(27 July 2000); http://www.nature.com/nature/journal/v406/n6794/abs/406378A0.html

- the first few % of nodes removed



Source: Error and attack tolerance of complex networks. Réka Albert, Hawoong Jeong and Albert-László Barabási. Nature 406, 378-382(27 July 2000); http://www.nature.com/nature/journal/v406/n6794/abs/406378A0.html

Skewness of power-law networks and effects and targeted attack

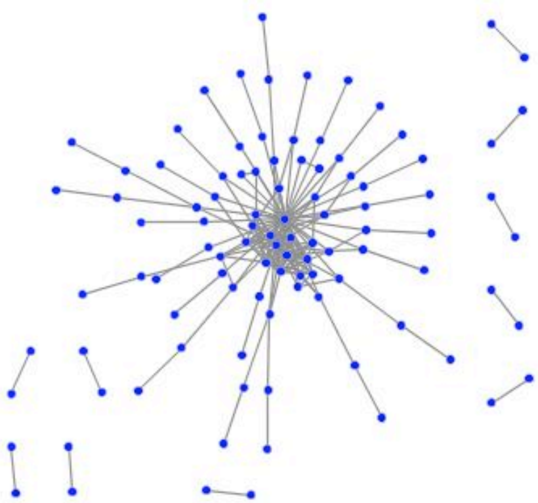


% of nodes removed, from highest to lowest degree

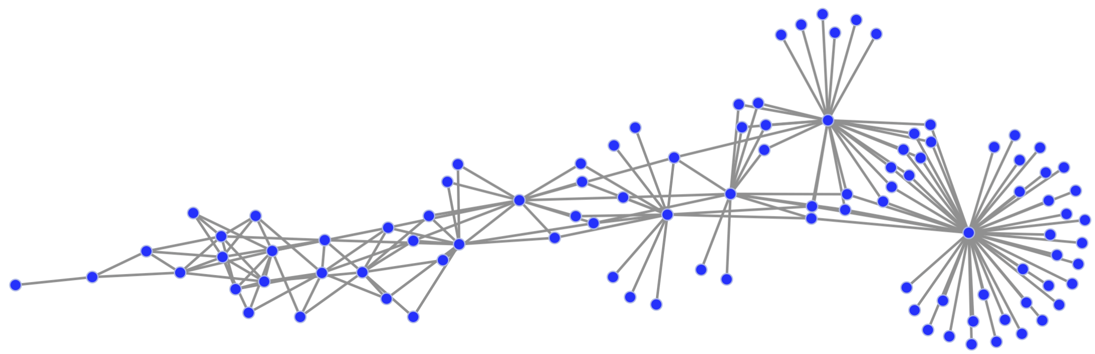
Source: D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, *Network robustness and fragility: Percolation on random graphs,* Phys. Rev. Lett., 85 (2000), pp. 5468–5471.

degree assortativity and resilience

will a network with positive or negative degree assortativity be more resilient to attack?



assortative



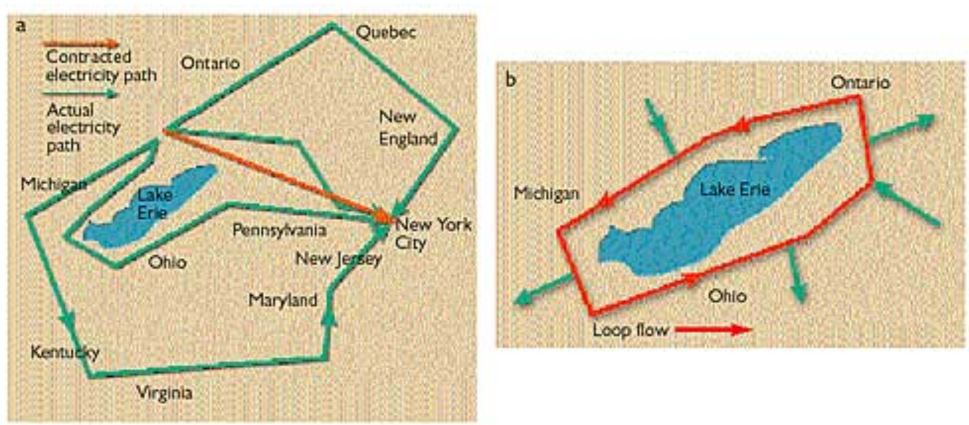
disassortative

Is it really that simple?

- Internet?
- Terrorist networks?

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.



Source: Eric J. Lerner, http://www.aip.org/tip/INPHFA/vol-9/iss-5/p8.html

Cascading failures

- 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

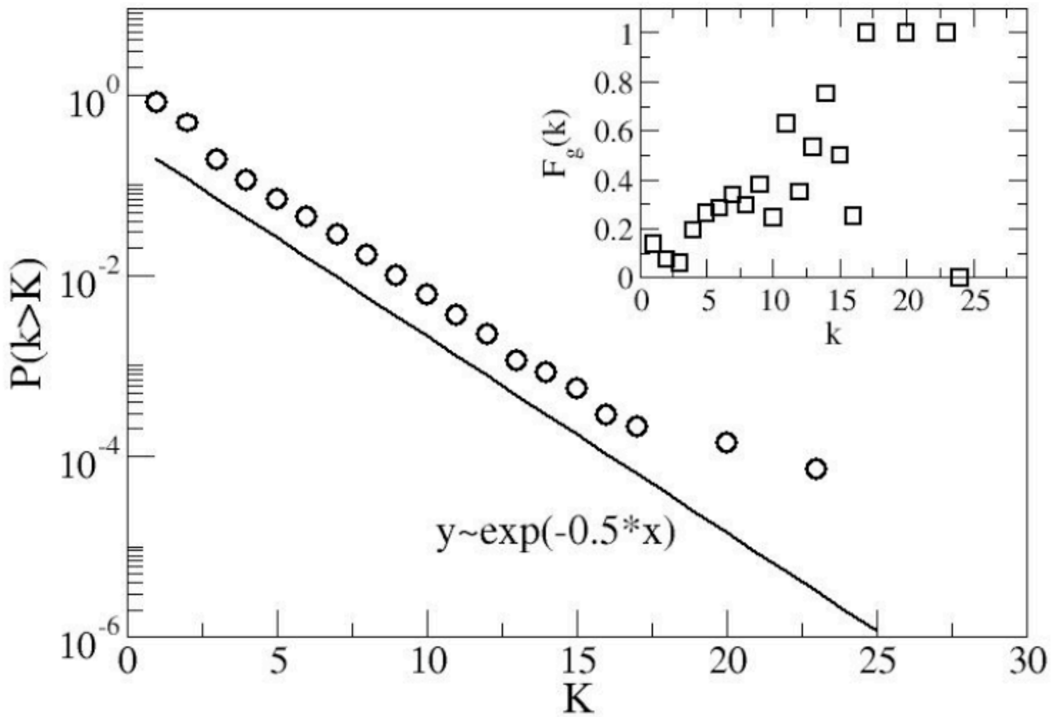
Case study: North American power grid

Modeling cascading failures in the North American power grid

R. Kinney, P. Crucitti, R. Albert, and V. Latora, Eur. Phys. B, 2005

- Nodes: generators, transmission substations, distribution substations
- Edges: high-voltage transmission lines
- 14099 substations:
  - N_G 1633 generators,
  - N_D 2179 distribution substations
  - N_T the rest transmission substations
- 19,657 edges

Degree distribution is exponential



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

Source: Albert et al., 'Structural vulnerability of the North American power grid, Phys. Rev. E 69, 025103 (2004)

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
  - path B, 3 edges, each with e=0.5
  - path C, 2 edges, one with e=0 the other with e=1
- 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

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

- Impact of node removal
  - change in efficiency

$$D = \frac{E(G_0) - E(G_f)}{E(G_0)}$$

Capacity and node failure

- Assume capacity of each node is proportional to initial load

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

  - L represents the weighted betweenness of a node
  - Each

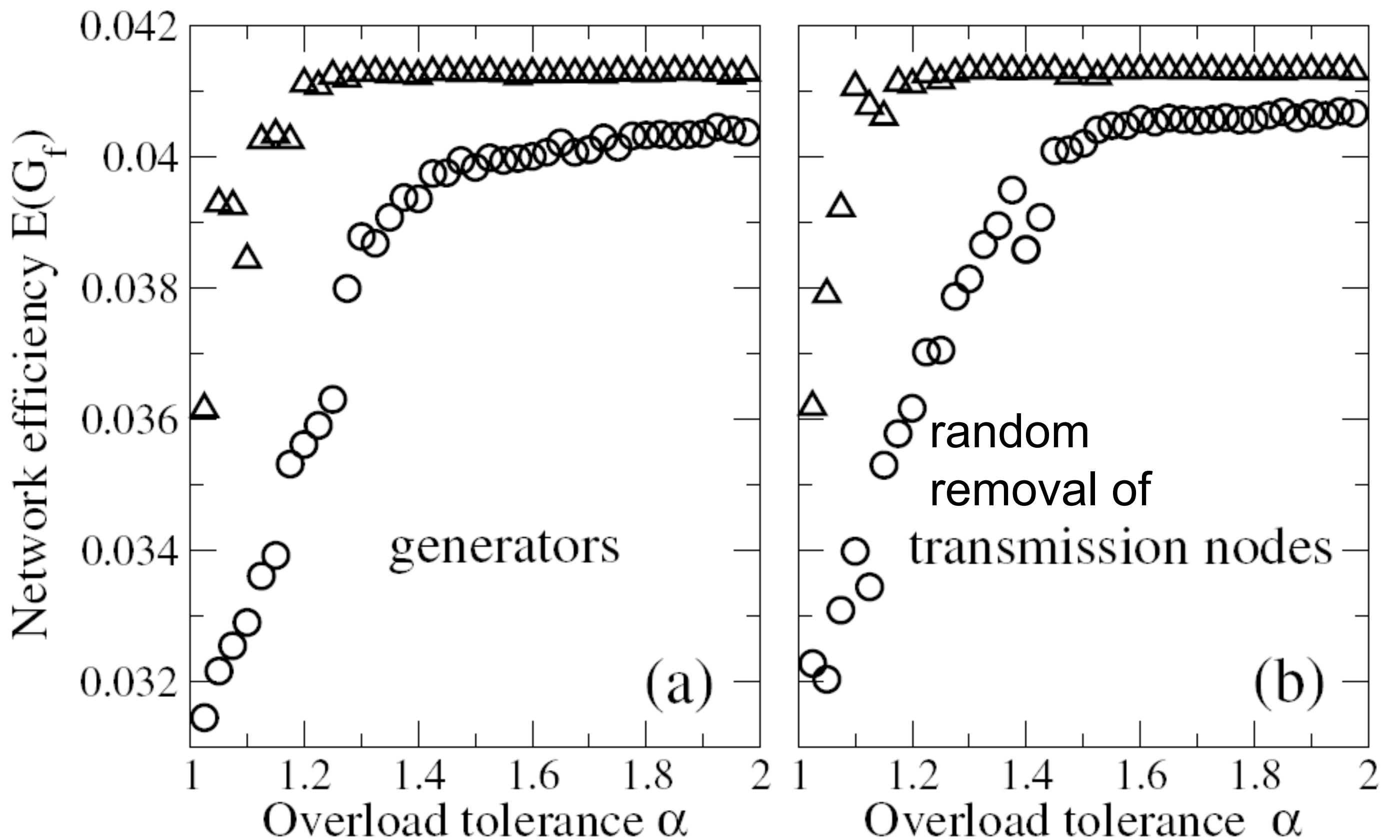
$$e_{ij}(t+1) = \begin{cases} e_{ij}(0)/\frac{L_i(t)}{C_i} \text{ if } L_i(t) > C_i \\ e_{ij}(0) \text{ if } L_i(t) \leq C_i \end{cases}$$

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, Eur. Phys. B, 2005

Biological networks

- In biological systems nodes and edges can represent different things
  - nodes
    - protein, gene, chemical (metabolic networks)
  - edges
    - mass transfer, regulation
- Can construct bipartite or tripartite networks:
  - e.g. genes and proteins

types of biological networks

genome

gene regulatory networks:

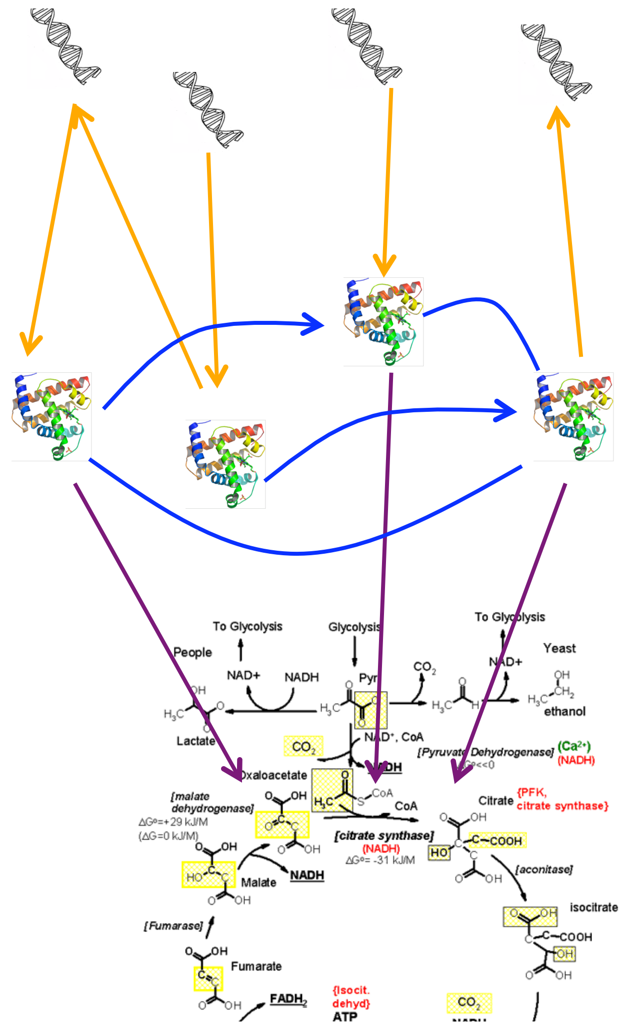
protein-gene interactions

proteome

protein-protein interaction networks

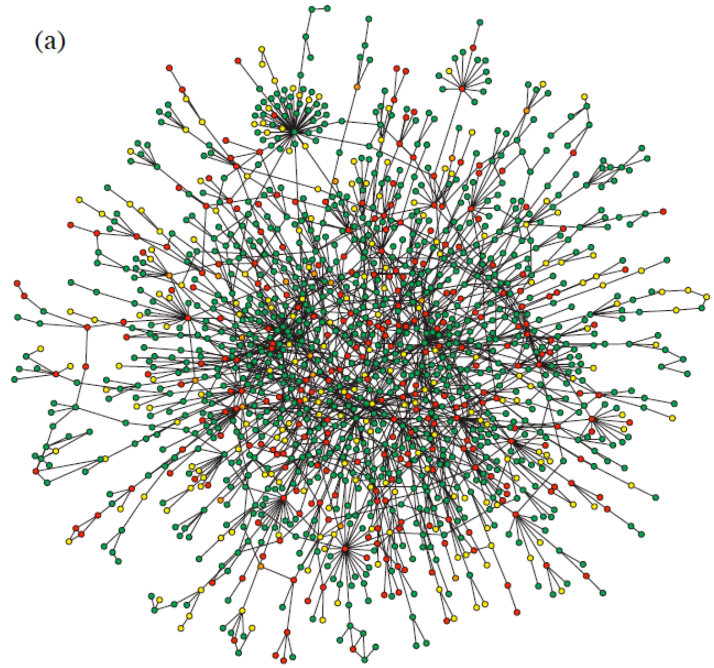
metabolism

bio-chemical reactions



protein-protein interaction networks

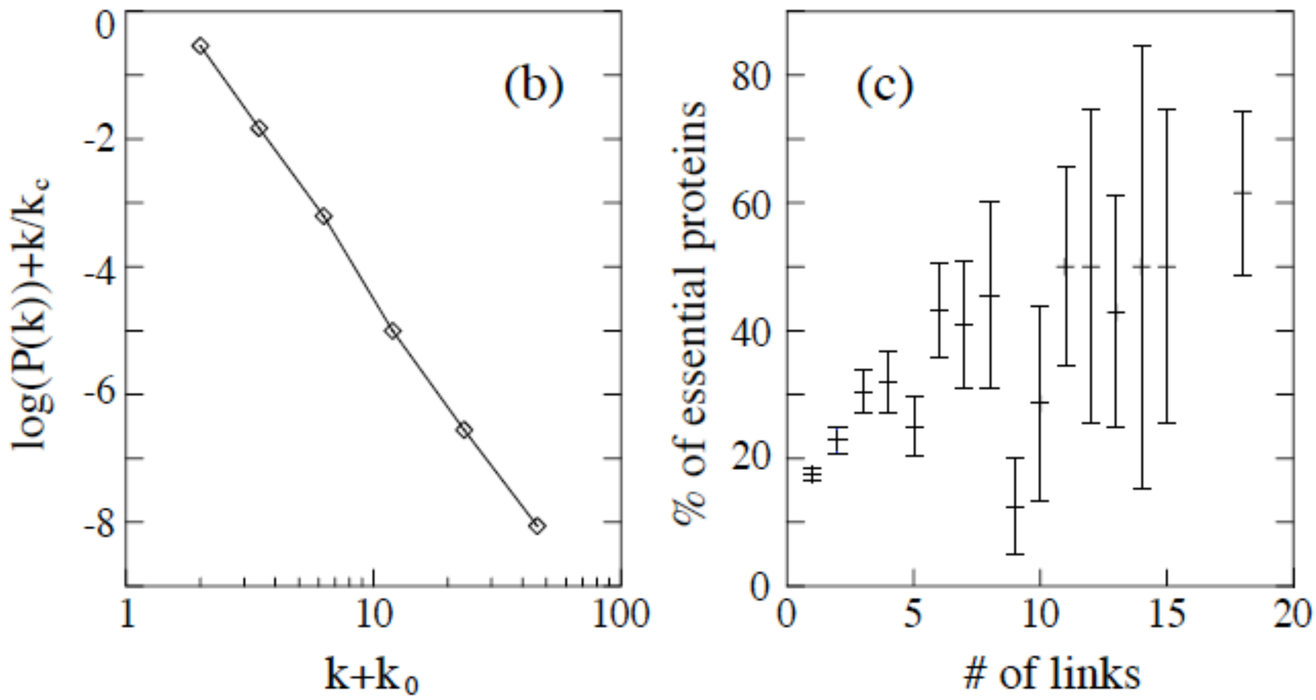
- Properties
  - giant component exists
  - longer path length than randomized
  - higher incidence of short loops than randomized



Source: Jeong et al, 'Lethality and centrality in protein networks', Nature 411, 41-42 (2001) | doi:10.1038/35075138

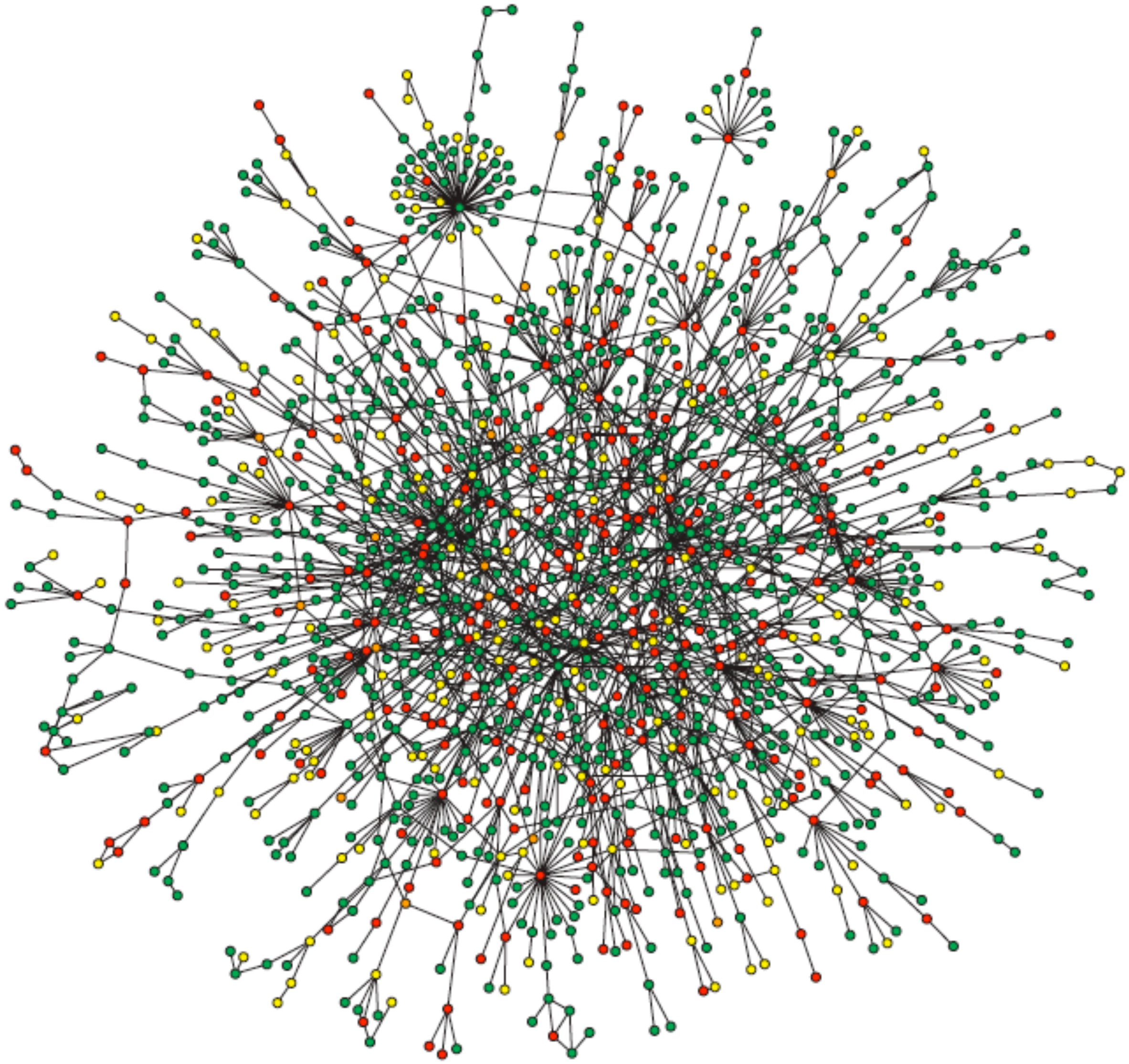
protein interaction networks

- Properties
  - power law distribution with an exponential cutoff
  - higher degree proteins are more likely to be essential



Source: Jeong et al, 'Lethality and centrality in protein networks', Nature 411, 41-42 (2001) | doi:10.1038/35075138

resilience of protein interaction networks



if removed:

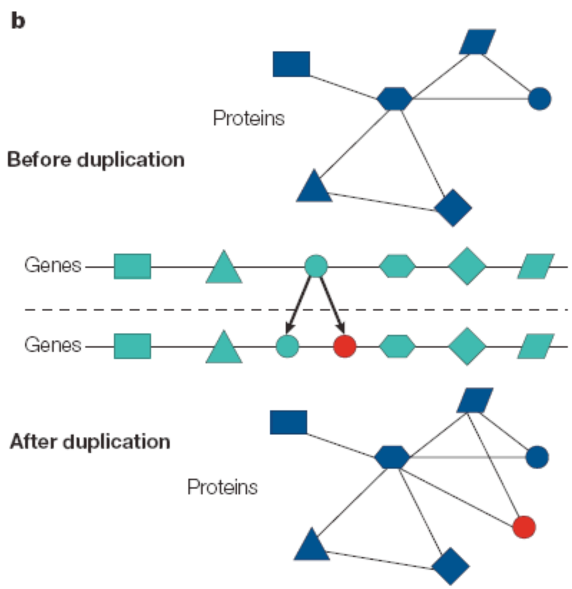
- lethal
- non-lethal
- slow growth
- unknown

Source: Jeong et al, 'Lethality and centrality in protein networks', Nature 411, 41-42 (2001) | doi:10.1038/35075138

Implications

- Robustness
  - resilient to random breakdowns
  - mutations in hubs can be deadly
- Evolution
  - most connected hubs conserved across organisms (important)
  - gene duplication hypothesis
    - new gene still has same output protein, but no selection pressure because the original gene is still present. So some interactions can be added or dropped
    - leads to scale free topology

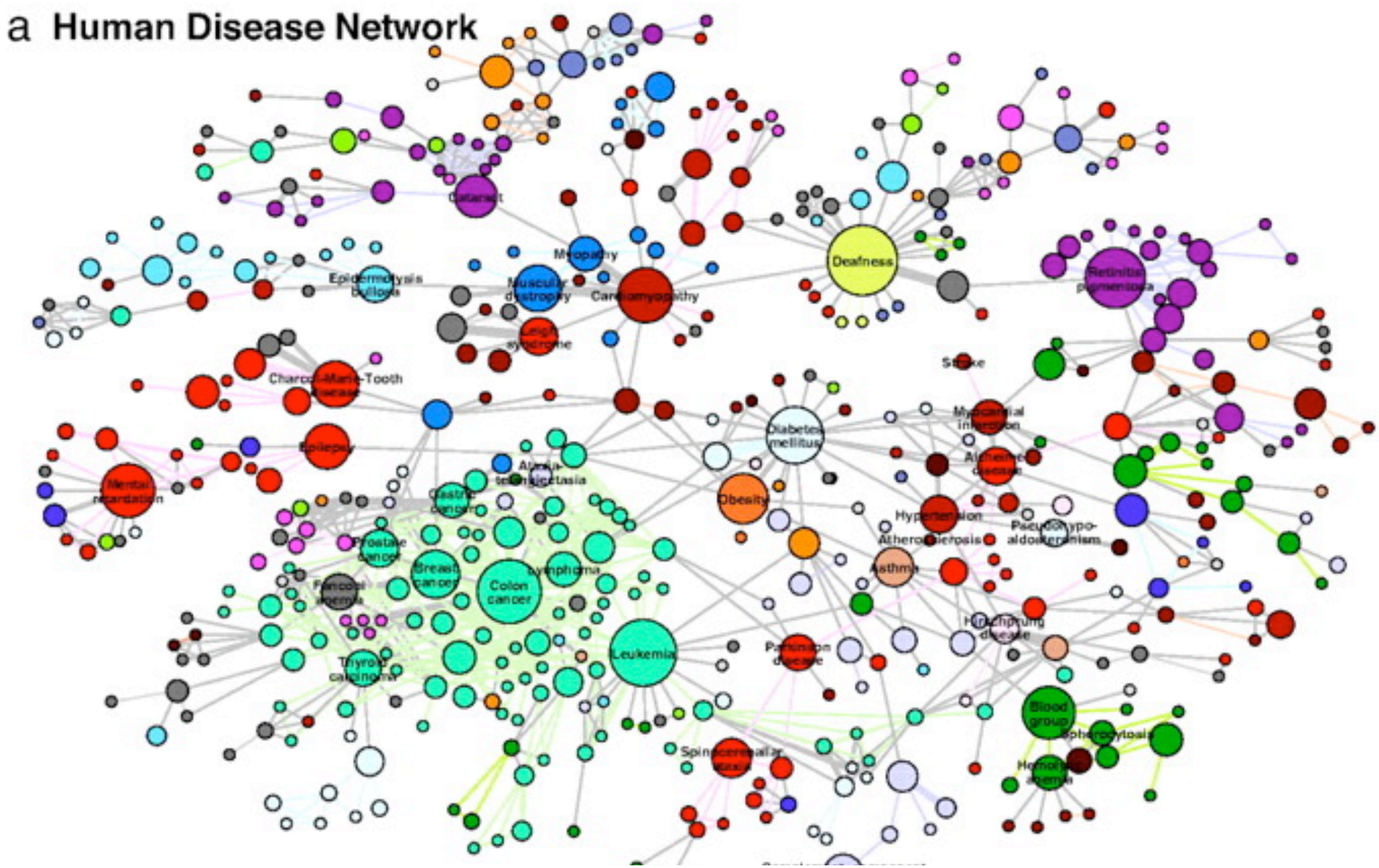
gene duplication



- When a gene is duplicated
  - every gene that had a connection to it, now has connection to 2 genes
  - preferential attachment at work…

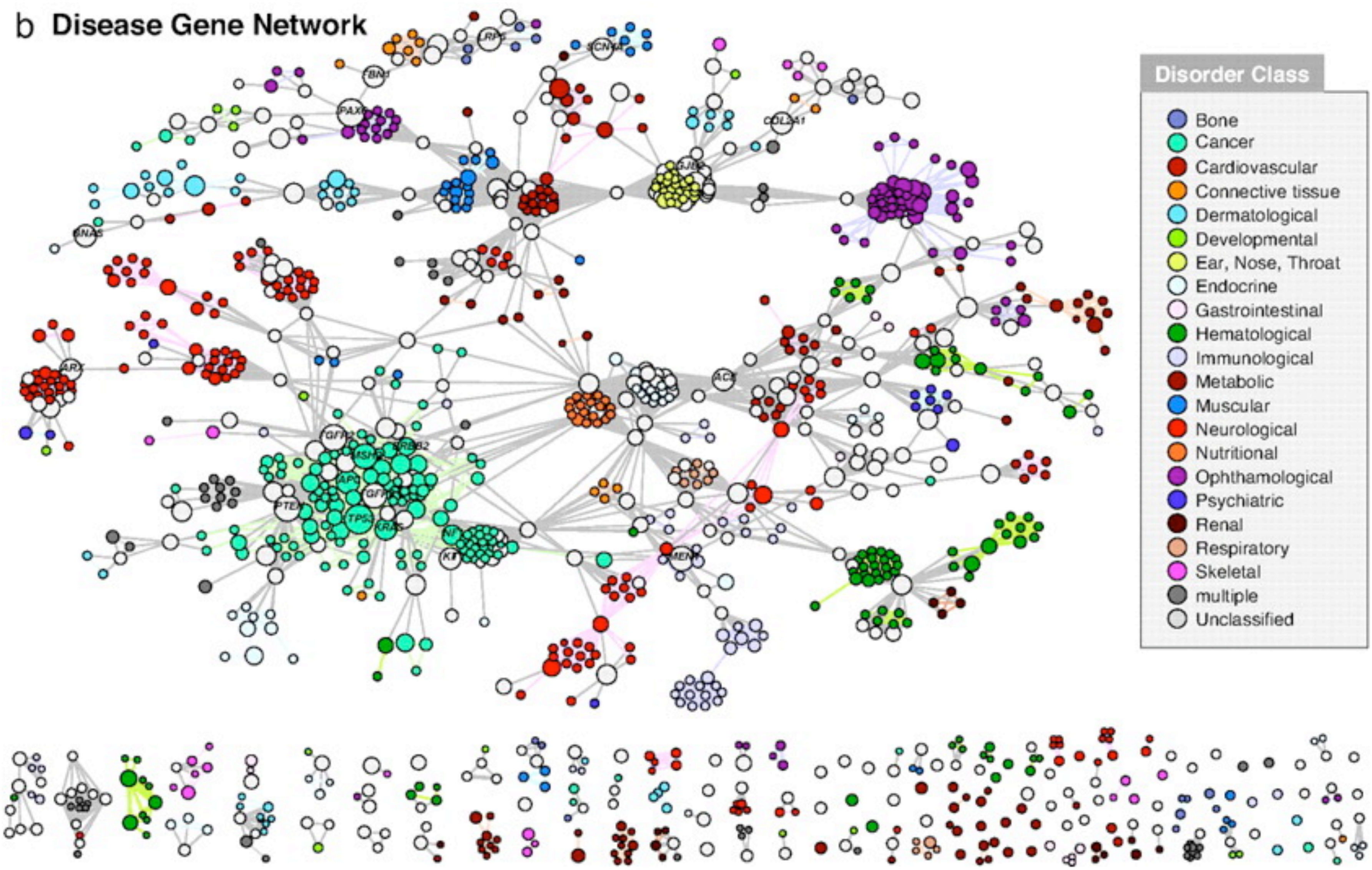
Source: Barabasi & Oltvai, *Nature Reviews* 2003

Q: do you expect disease genes to be the essential genes?



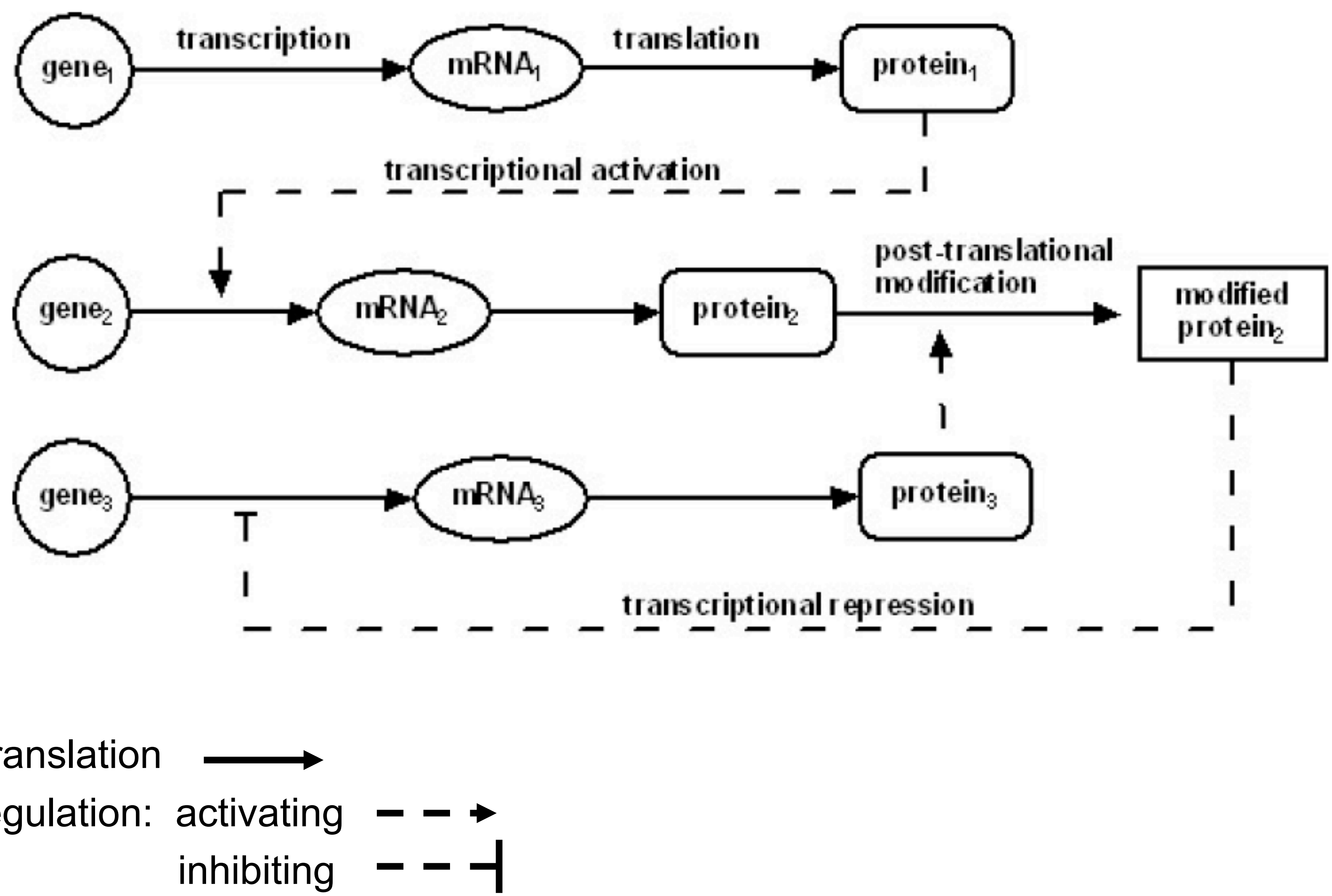
source: Goh et al. PNAS May 22, 2007 vol. 104 no. 21 8685-8690 10.1073/pnas.0701361104

Q: where do you expect disease genes to be positioned in the gene network



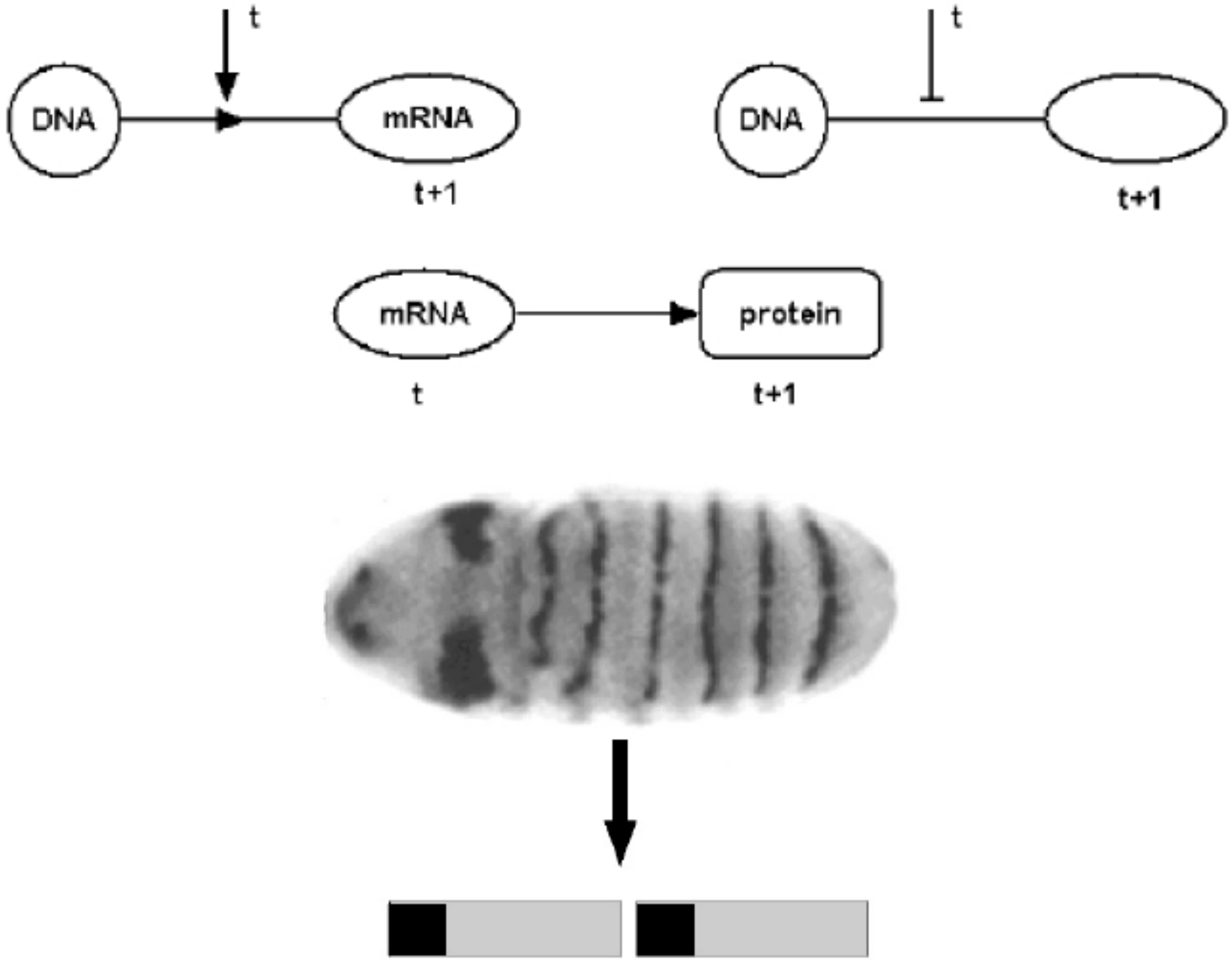
source: Goh et al. PNAS May 22, 2007   vol. 104 no. 21 8685-8690 10.1073/pnas.0701361104

gene regulatory networks



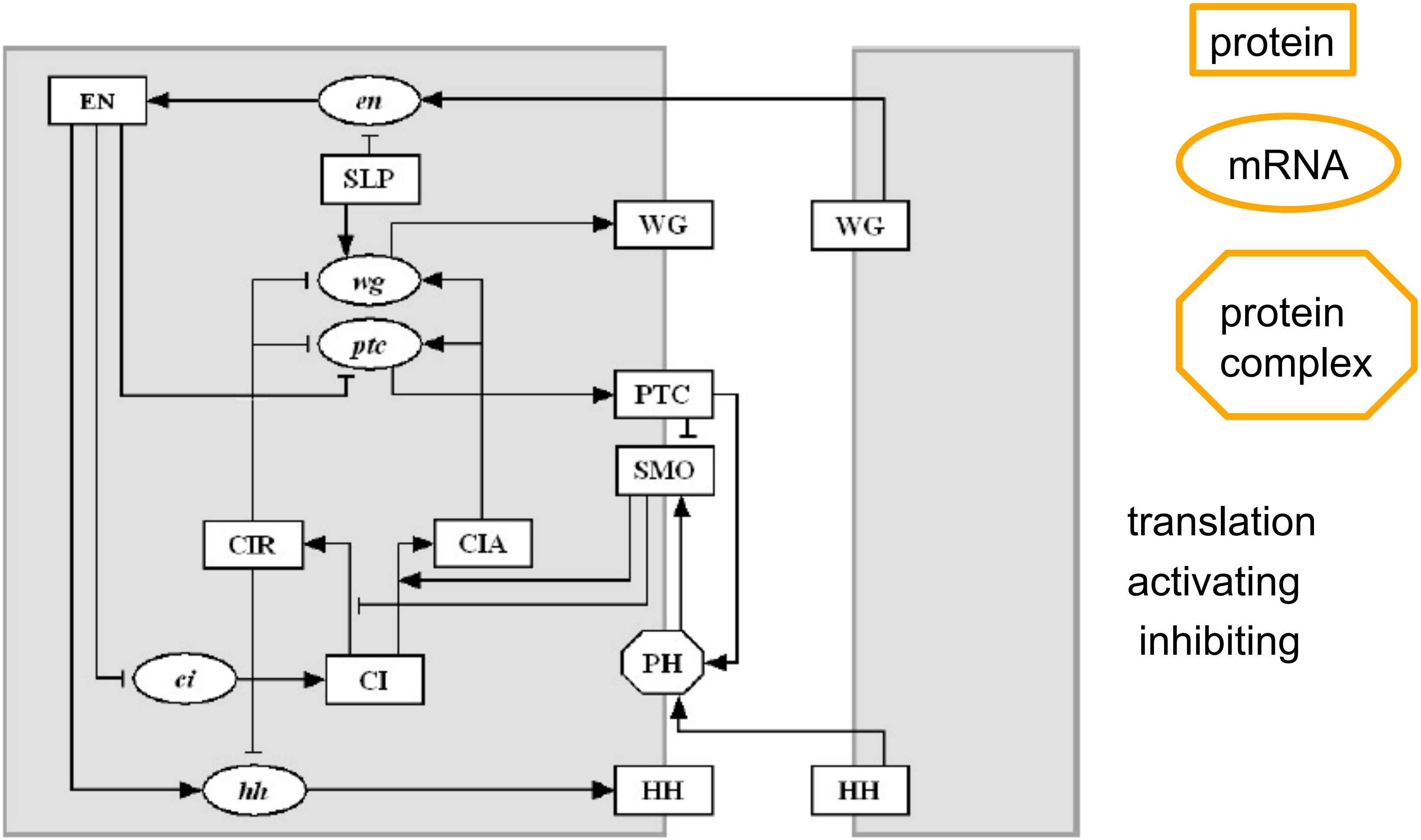
slide after Reka Albert

simple model of ON/OFF gene dynamics



Source: Albert and Othmer, Journal of Theoretical Biology 23(1), p. 1-18, 2003. doi:10.1016/S0022-5193(03)00035-3

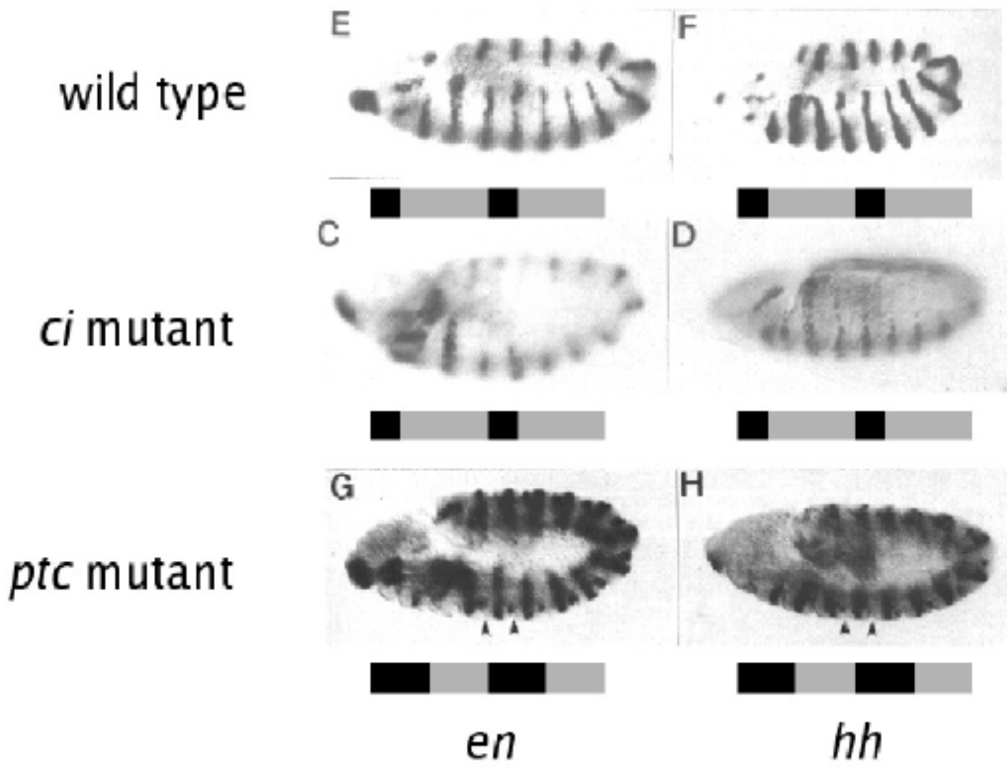
network interactions between segment polarity genes



Source: Albert and Othmer, Journal of Theoretical Biology 23(1), p. 1-18, 2003. doi:10.1016/S0022-5193(03)00035-3

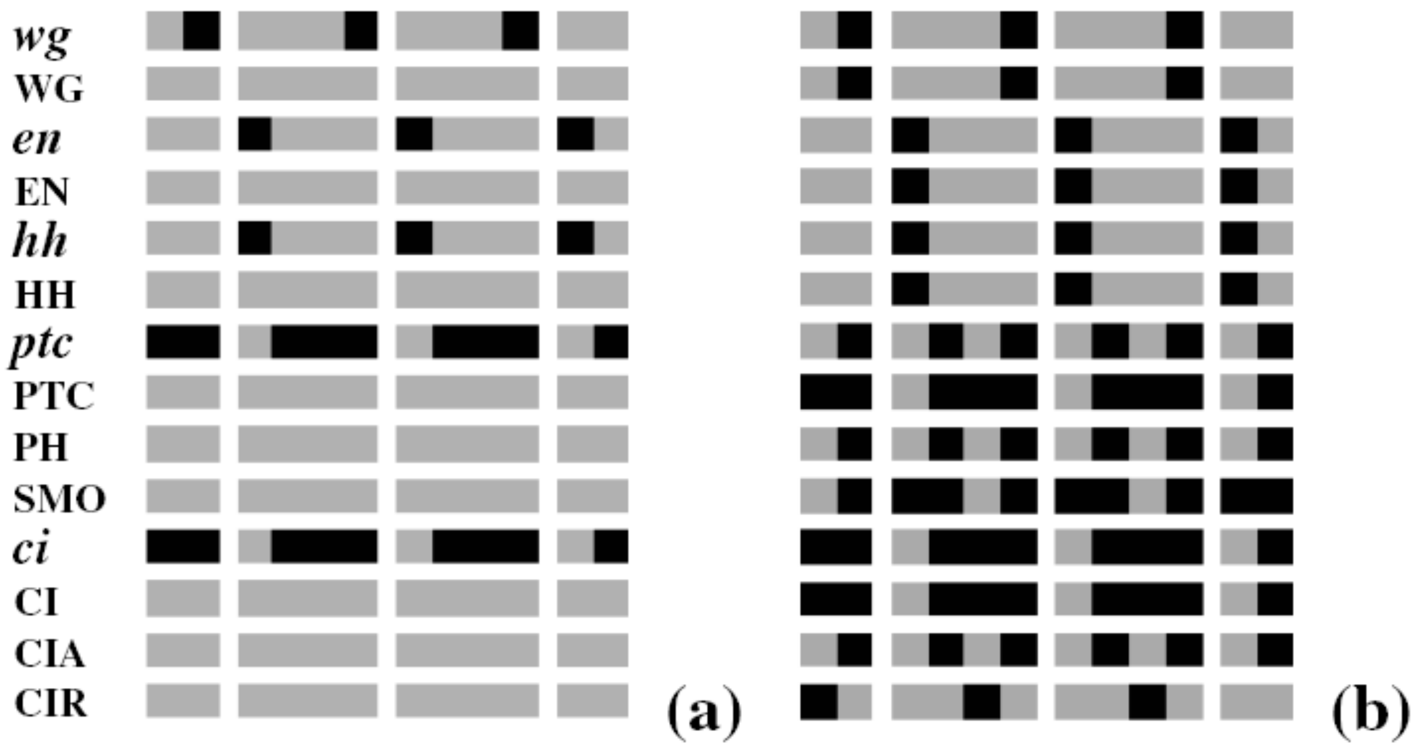
excellent agreement between model and observed gene expression patterns

- test by observing the effect of gene mutation in specimen and in model



Source: Albert and Othmer, Journal of Theoretical Biology 23(1), p. 1-18, 2003. doi:10.1016/S0022-5193(03)00035-3

predicting drosophila gene expression patterns with a boolean model



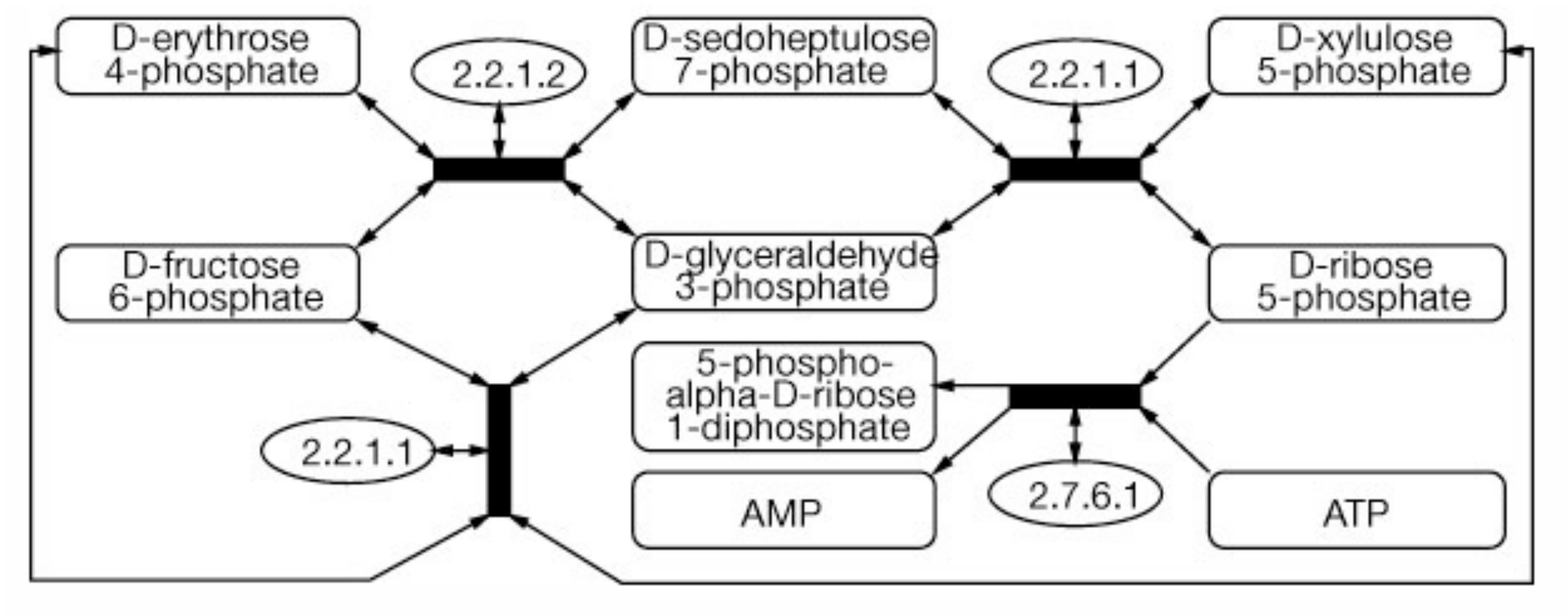
initial state

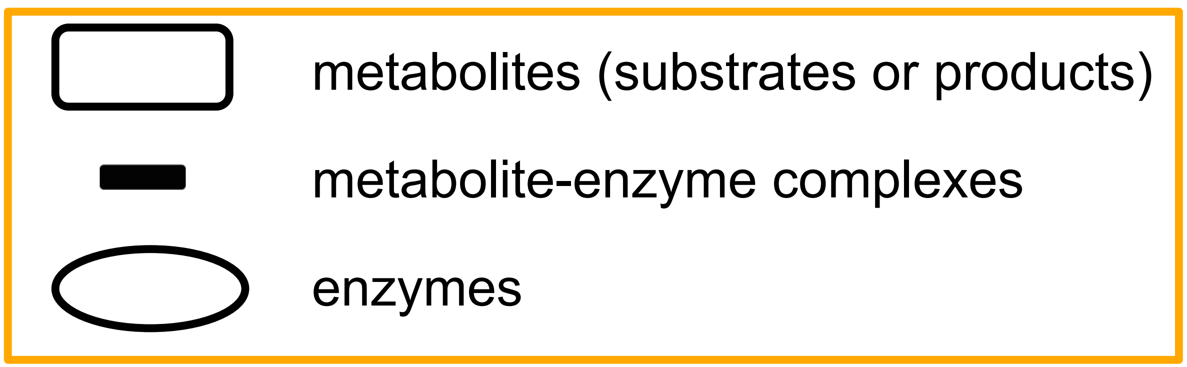
predicted by model

Source: Albert and Othmer, Journal of Theoretical Biology 23(1), p. 1-18, 2003. doi:10.1016/S0022-5193(03)00035-3

Metabolic networks

- metabolic reaction networks (tri-partite)

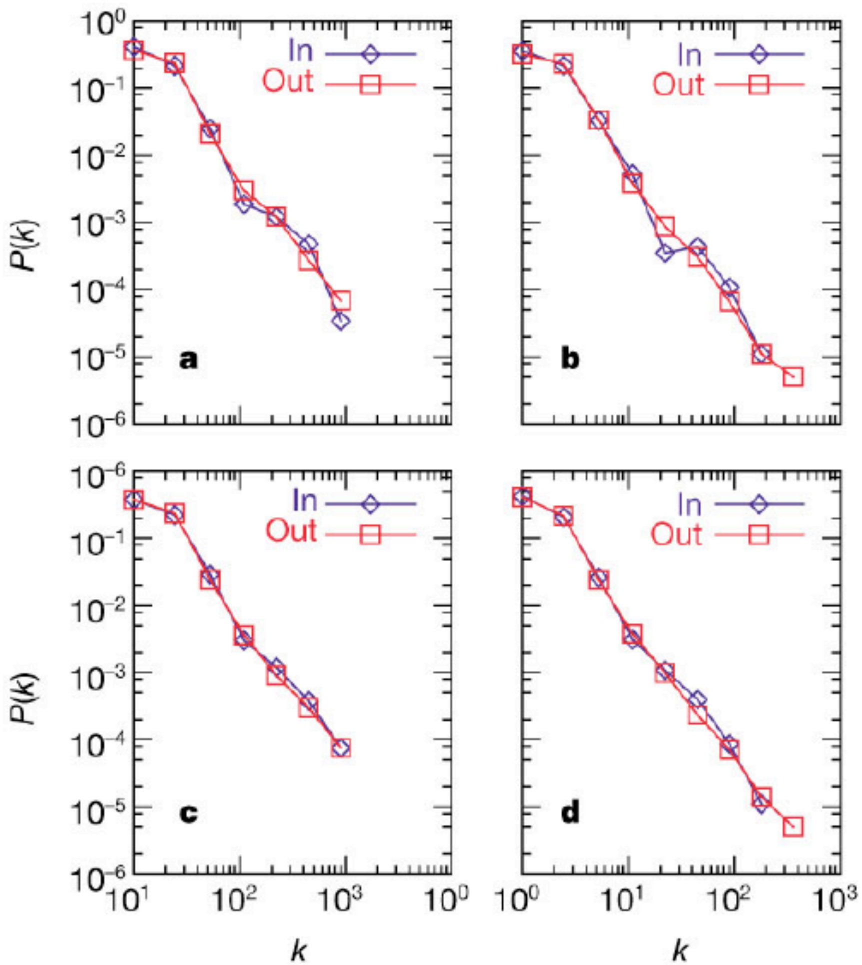




Source: Jeong et al., Nature 407, 651-654 (5 October 2000) | doi:10.1038/35036627

Metabolic networks are scale-free

- In the bi-partite graph:
  - the probability that a given substrate participates in k reactions is $k^{-\alpha}$
    - indegree: $\alpha = 2.2$
    - outdegree: $\alpha = 2.2$

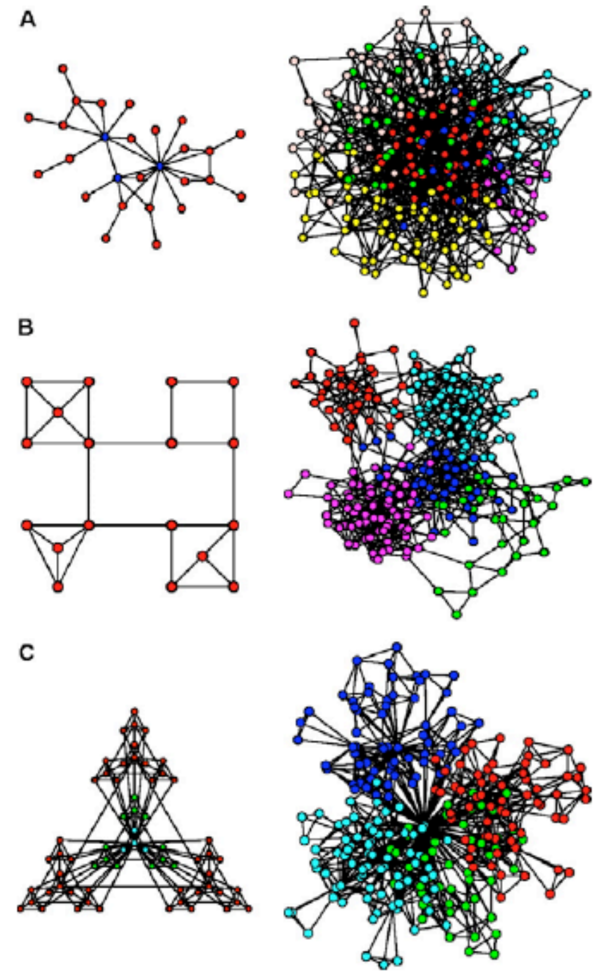


(**a**) *A. fulgidus* (Archae) (**b**) *E. coli* (Bacterium) (**c**) *C. elegans* (Eukaryote), (**d**) averaged over 43 organisms

Source: Jeong et al., Nature 407, 651-654 (5 October 2000) | doi:10.1038/35036627

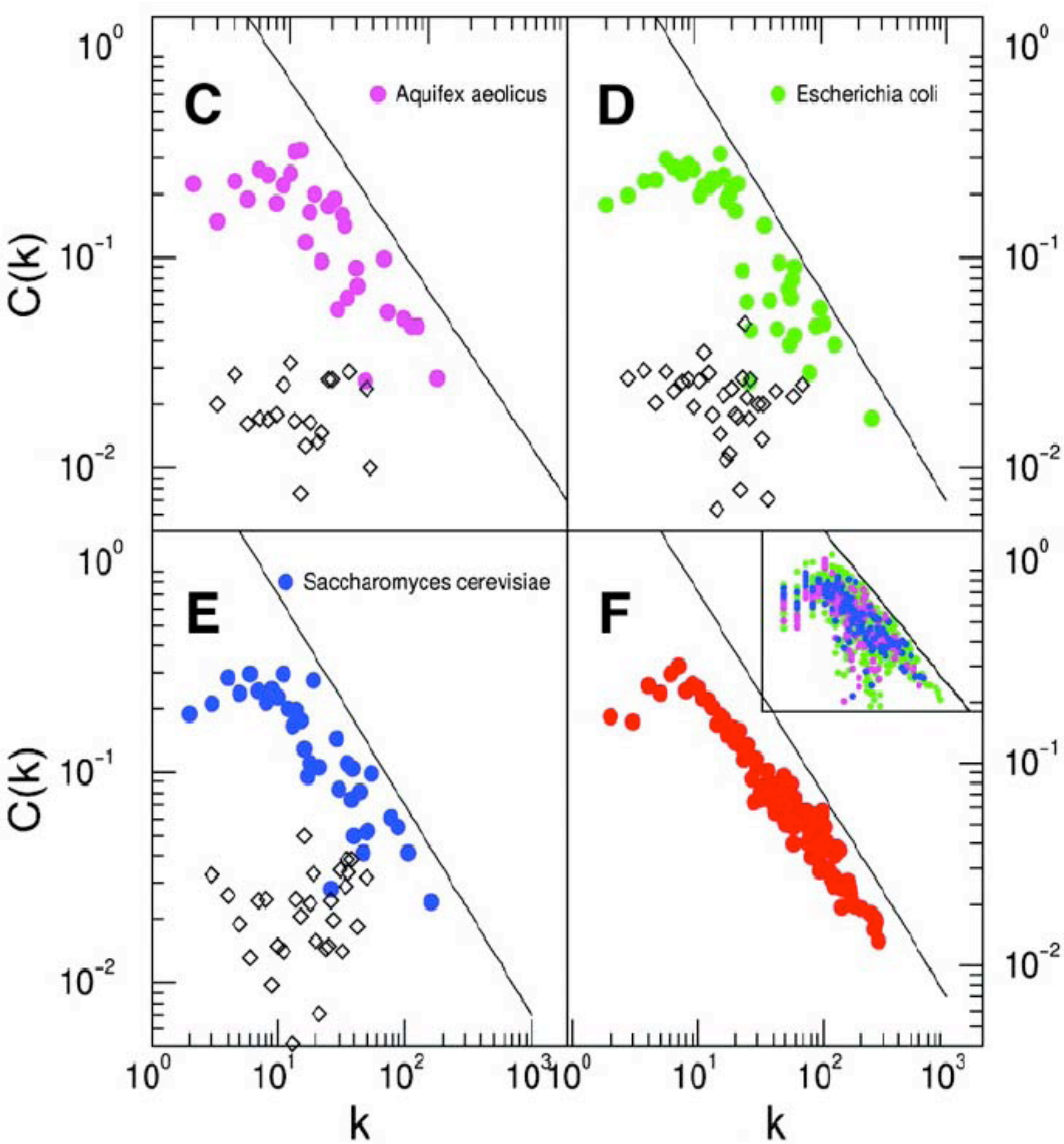
Is there more to biological networks than degree distributions?

- No modularity
- Modularity
- Hierarchical modularity



Source: E. Ravasz et al., Science 297, 1551 -1555 (2002)

How do we know that metabolic networks are modular?



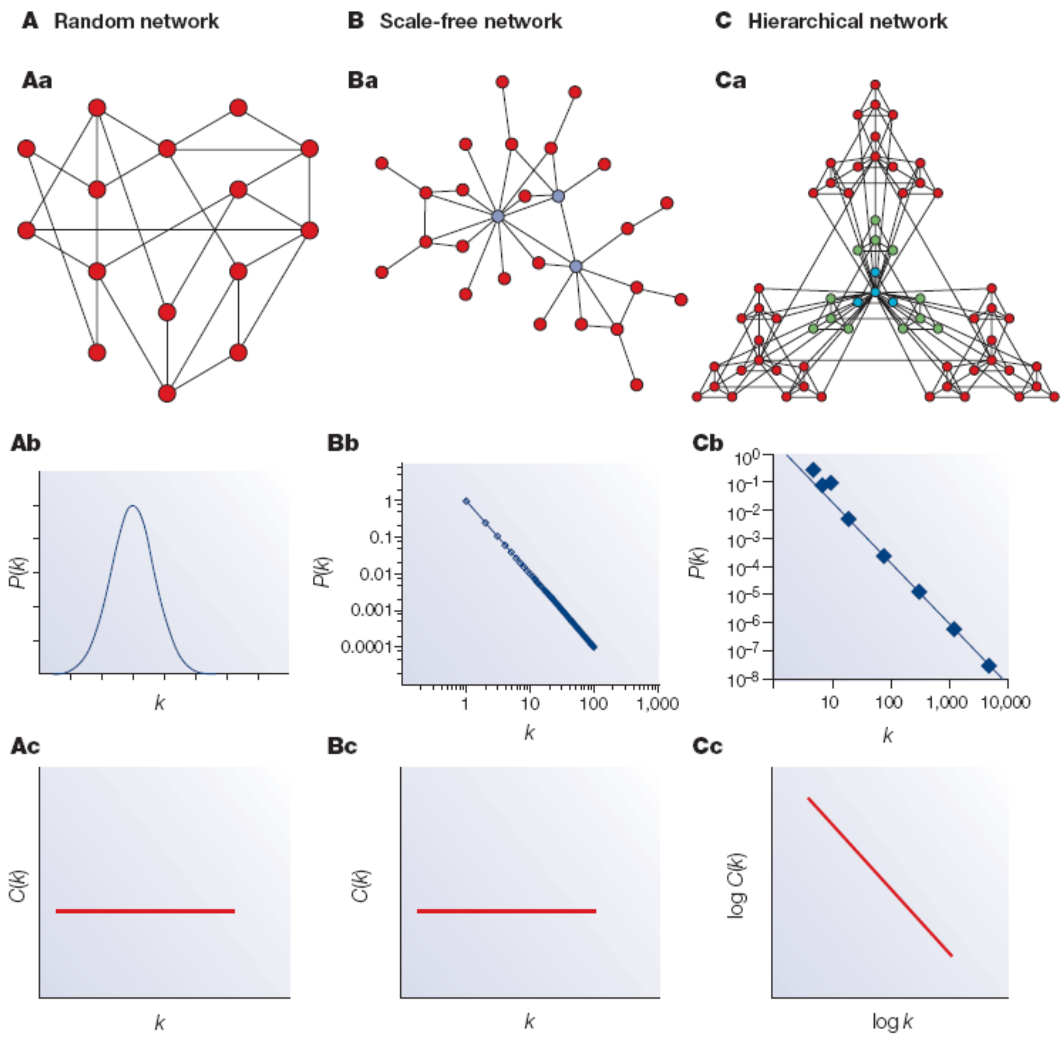
clustering decreases with degree as

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

randomized networks (which preserve the power law degree distribution) have a clustering coefficient independent of degree

Source: E. Ravasz et al., Science 297, 1551 -1555 (2002)

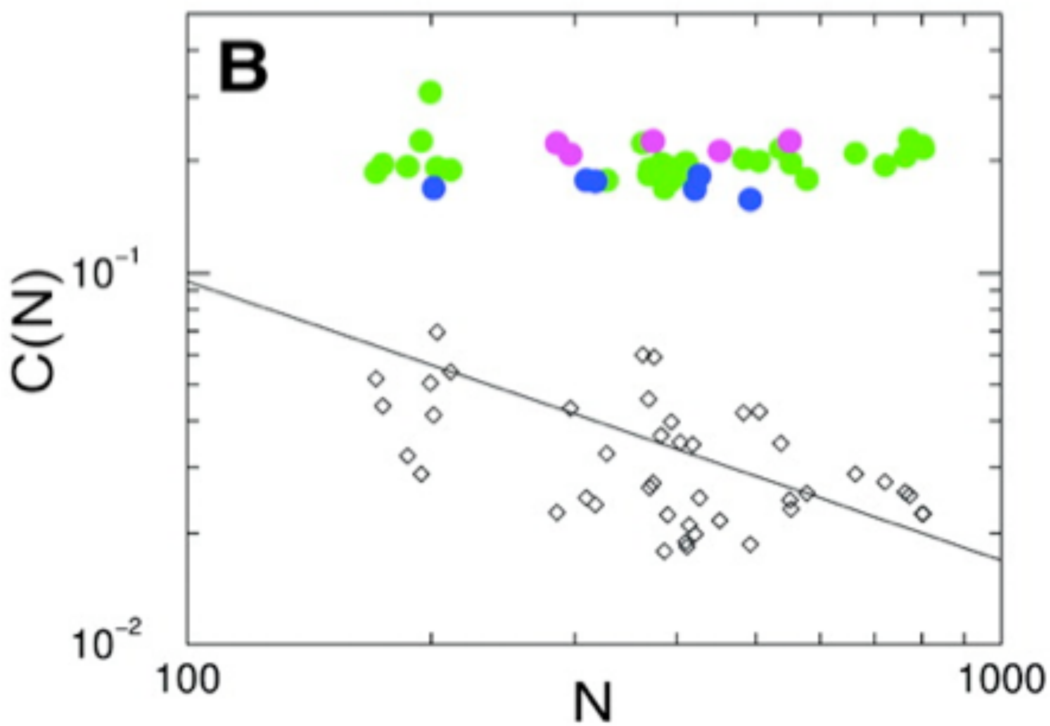
clustering coefficients in different topologies



Source: Barabasi & Oltvai, *Nature Reviews* 2003

How do we know that metabolic networks are modular?

- clustering coefficient is the same across metabolic networks in different species with the same substrate
- corresponding randomized scale free network: C(N) ~ $N^{-0.75}$ (simulation, no analytical result)



bacteria

archaea (extreme-environment single cell organisms)

eukaryotes (plants, animals, fungi, protists)

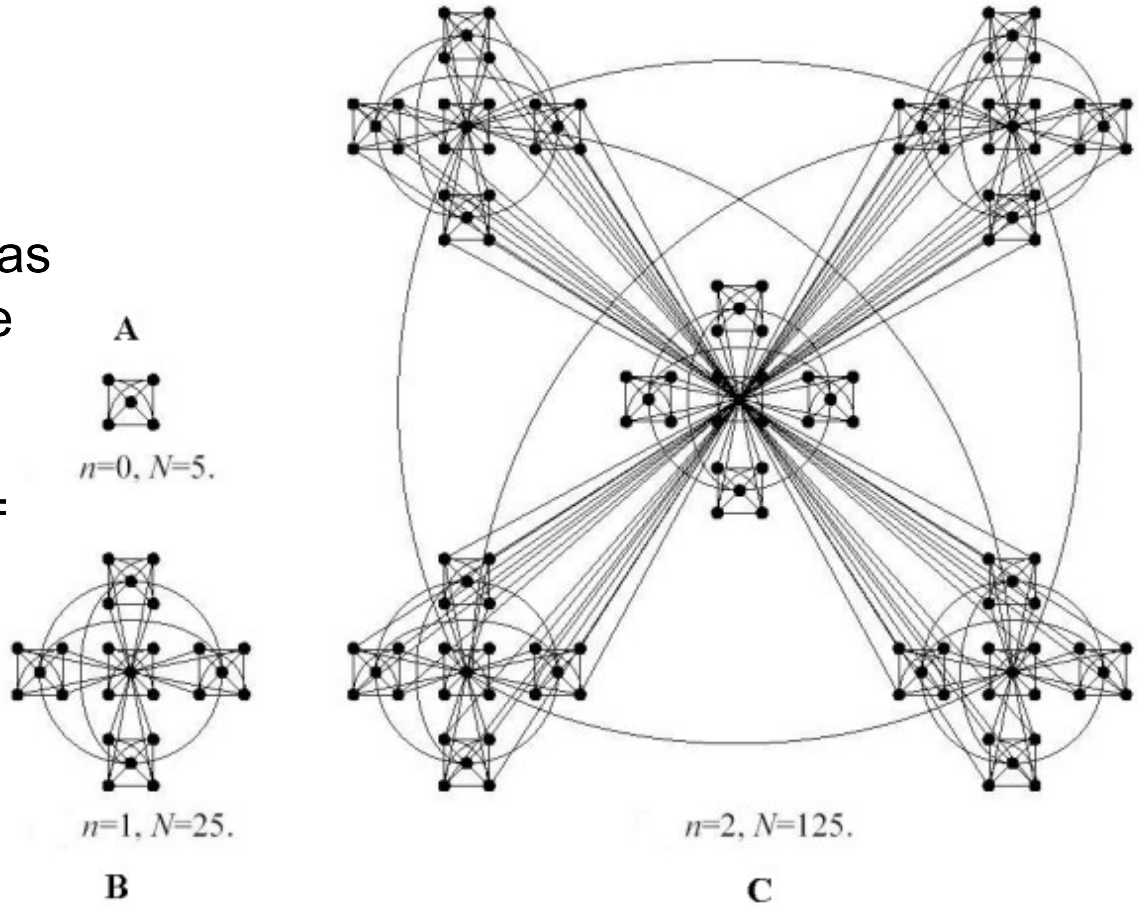
scale free network of the same size

Source: E. Ravasz et al., Science 297, 1551 -1555 (2002)

Constructing a hierarchically modular network

RSMOB model

- Start from a fully connected cluster of nodes
- Create 4 identical replicas of the cluster, linking the outside nodes of the replicas to the center node of the original (N = 25 nodes)



- This process can repeated indefinitely
- (initial number of nodes can be different than 5)

Source: Ravasz and Barabasi, PRE 67 026112, 2003, doi: 10.1103/PhysRevE.67.026112

Properties of the hierarchically modular model

RSMOB model

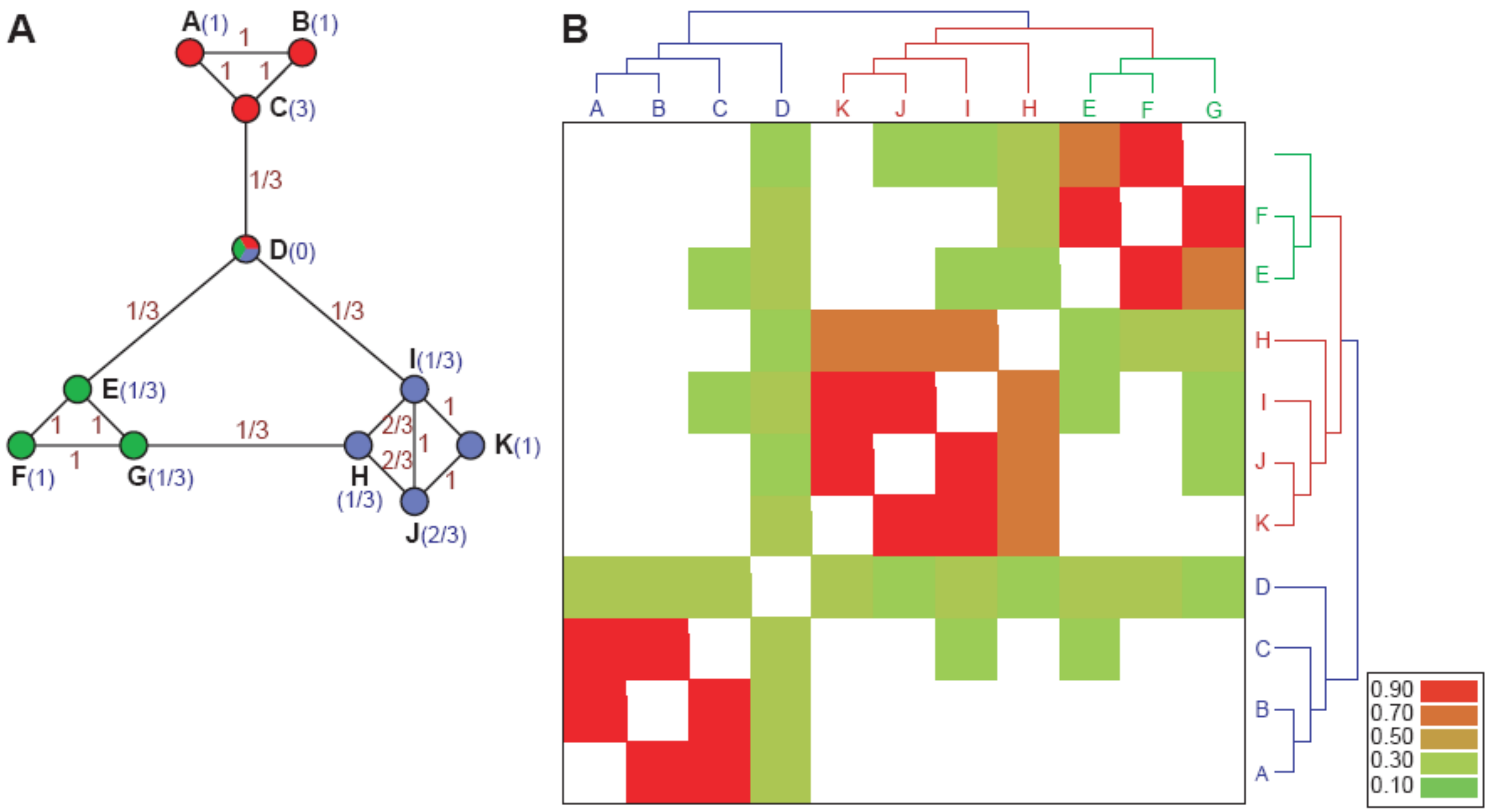
- Power law exponent $\gamma = 2.26$ (in agreement with real world metabolic networks)
- $C \approx 0.6$, independent of network size (also comparable with observed real-world values)
- $C(k) \approx k^{-1}$, as in real world network

- How to test for hierarchically arranged modules in real world networks
  - perform hierarchical clustering on the topological overlap map
  - can be done with Pajek

Discovering hierarchical structure using topological overlap

- A: Network consisting of nested modules
- B: Topological overlap matrix

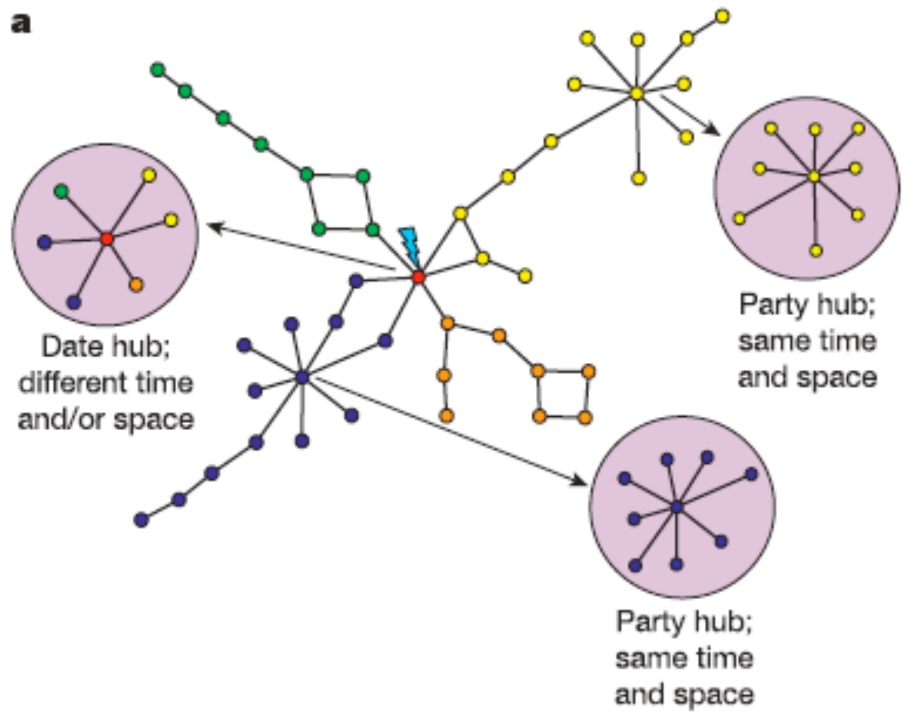
hierarchical clustering



Source: E. Ravasz et al., Science 297, 1551 -1555 (2002)

Modularity and the role of hubs

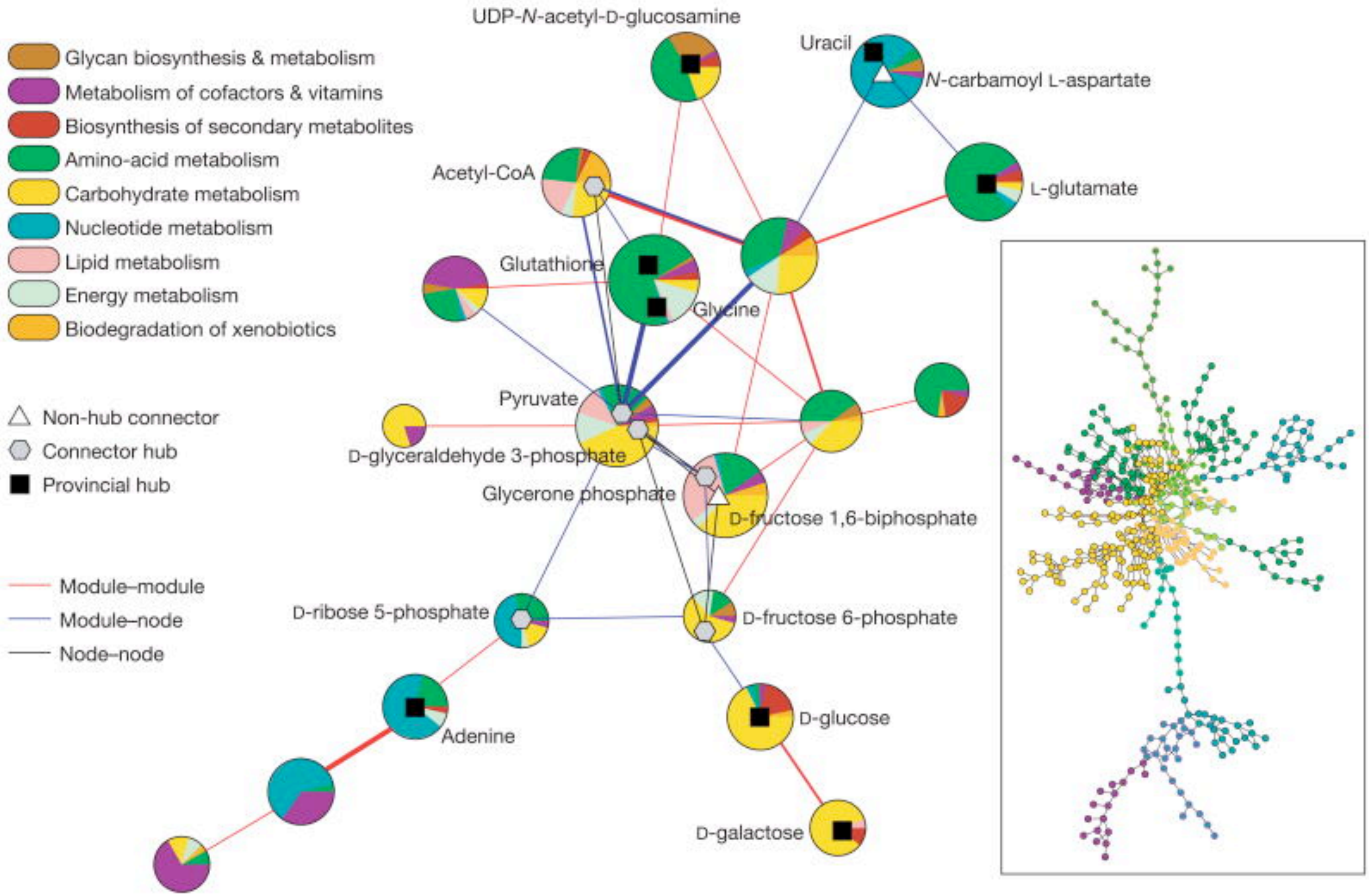
- Party hub:
  - interacts simultaneously within the same module
- Date hub:
  - sequential interactions
  - connect different modules – connect biological processes



- Q:
  - which type of hub is more likely to be essential?

Source: Han et al, Nature 443, 88 (2004)

metabolic network of *e. coli*



Source: Guimera & Amaral, Nature. 2005 February 24; 433(7028): 895–900. doi: 10.1038/nature03288.

summing it up

- 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 be start being produced or cease to do so
- in biological networks, more central nodes cannot be done without