WHY THE ASIAN CAMEL HAS TWO HUMPS ?

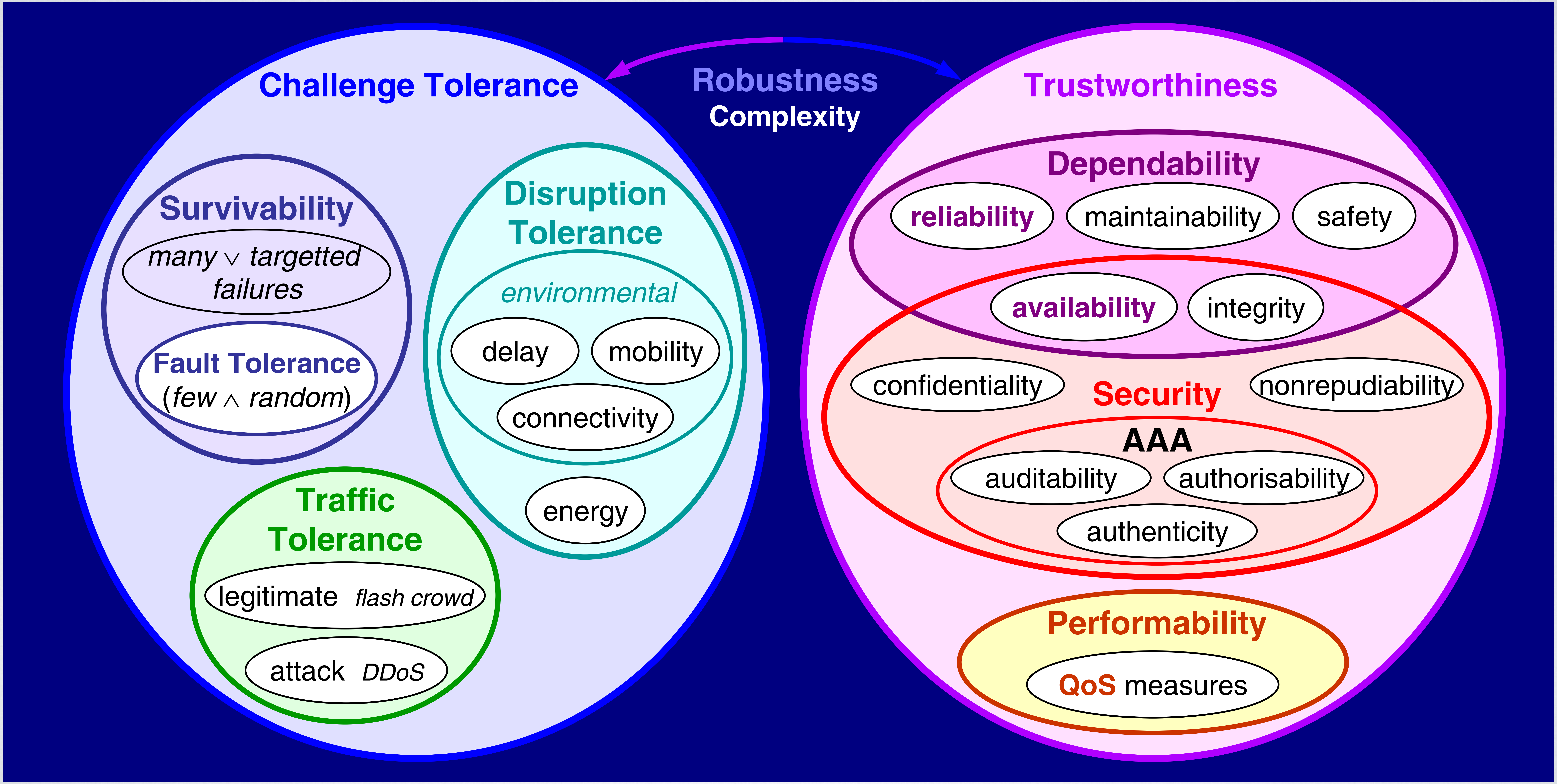
A peripatetic journey on the basis of redundancy and resilience.

Kavé Salamatian
LISTIC, University of Savoie

BIOLOGICAL SYSTEMS

- Robust
  - Maintaining its functions despite external and internal perturbation.
    - Function in unpredictable environments with unreliable components.
- Evolvable
  - Adapt in ways that exploit new resources or allow them to persist under unprecedented environmental regime shift
- Resilient
  - Provide and maintain acceptable activity in the face of faults and challenges to normal operation
    - Bouncing back

RESILIENCE IN NETWORKED SYSTEMS



James Sterbenz course on « Resilient and Survivable Networking »

ROBUSTNESS ?

- Robustness of what?
  - Robustness is only meaningful for a specific set of feature
    - A (dynamic) state, a process, a function
      - feature persistence or feature reproducibility?
- Robustness with respect to what?
  - Robustness is meaningful with respect to a specified set of perturbations
    - Bounded and likely
      - transient versus permanent perturbations
      - large versus small perturbations
      - changes in the system versus changes in its environment
      - changes in the system parameters versus changes in its constitution (e.g. removal of a link or a node in a network)
      - additive noise vs multiplicative noise

EX: PROTEIN FOLDING

- A vey small fraction of possible amino-acid sequences corresponds to functional proteins.
- A functional protein exhibits a primary folded structure, the native structure and several metastable ones. Protein function involves transitions between these conformations.
    - The conformation and the transition is believed to follow a free energy landscape
    - Proper functioning requires
        - the conformations are structurally robust
        - the transitions between them occur in a controlled way.
    - For several proteins, this basic mechanism is supplemented by chaperone, *i.e*, specific auxiliary proteins binding the misfolded ones
    - The very existence of chaperones and their functions
        - Results from the co-evolution

ROBUSTNESS IN NETWORKS

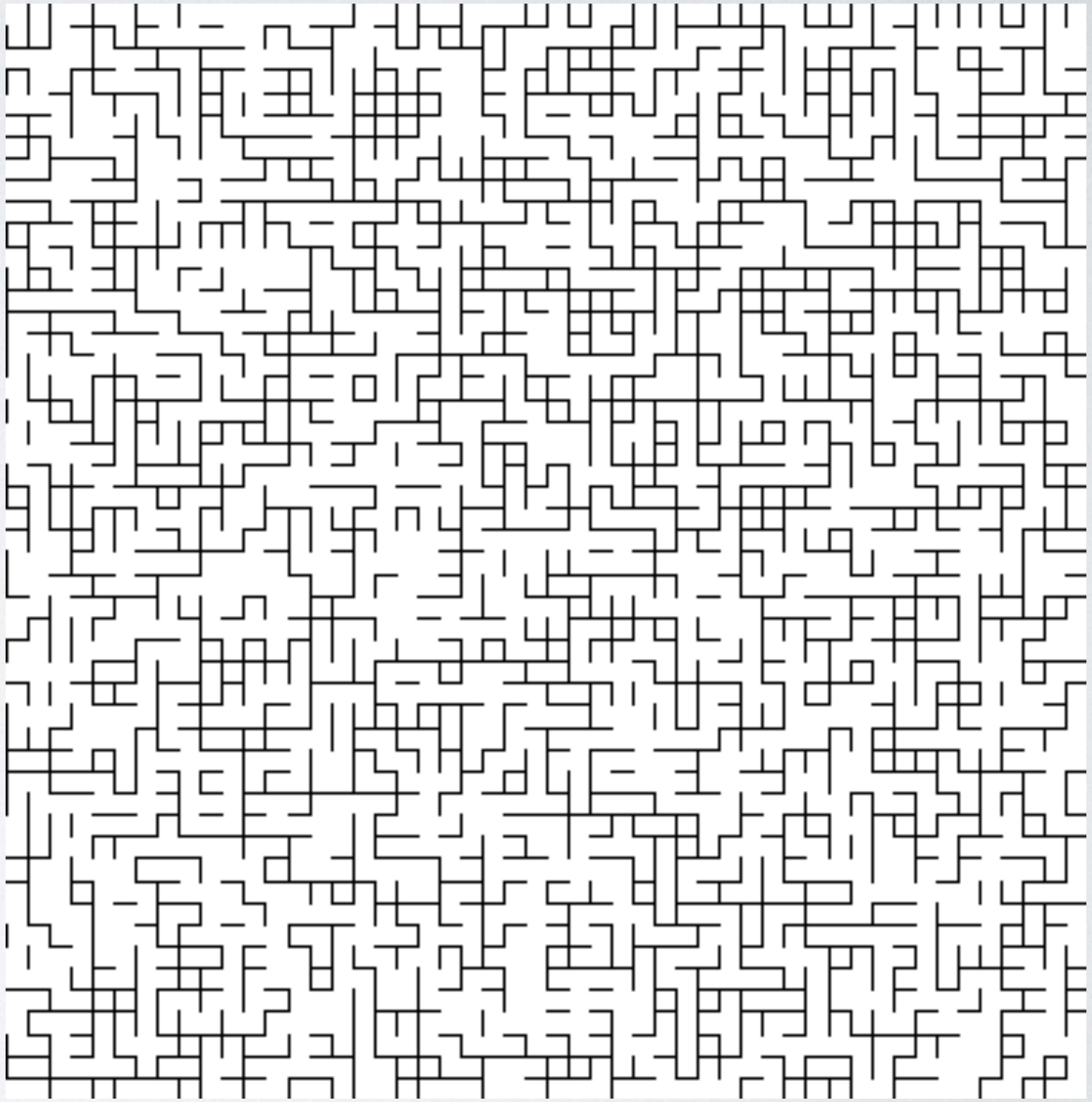
- The robustness of a network refers to the robustness of the phenomenon that the network captures
  - Connectivity Robustness
    - Percolation approaches
  - Robustness of the dynamics,
    - Persistence of the large-scale behaviour and properties after local perturbation
    - Persistence of some local properties despite a global change
    - Robustness of the transfer (information or other) between two nodes or two regions of the network despite the presence perturbations.
  - Mechanisms
    - Diversion
      - Existence of alternative paths
    - Plasticity
      - The possibility of rewiring some connections;

PERCOLATION THEORY

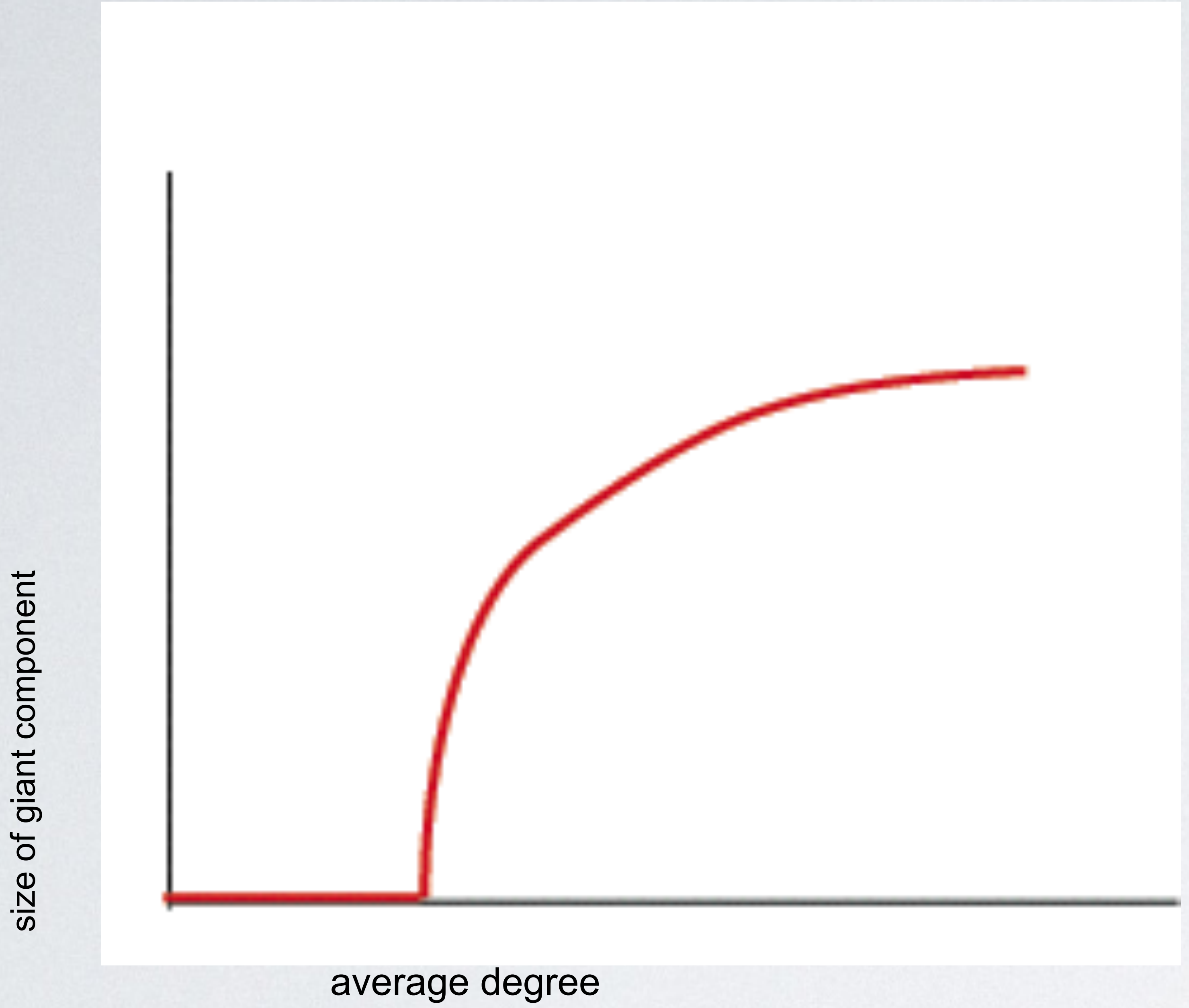
- Q: If a given fraction of nodes or edges are removed…
  - how large are the connected components?
  - what is the average distance between nodes in the components

EDGE REMOVAL

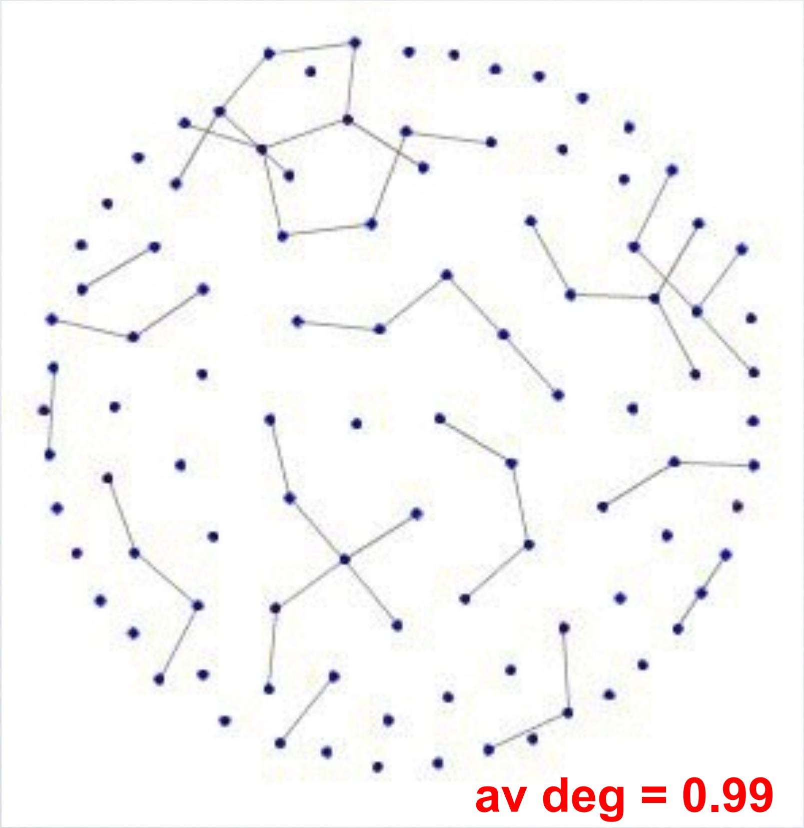
- Bond percolation
    - edges are removed with probability (1-p)
        - corresponds to random failure of links
- targeted attack
    - causing the most damage to the network with the removal of the fewest edges
    - strategies: remove edges that are most likely to break apart the network or lengthen the average shortest path

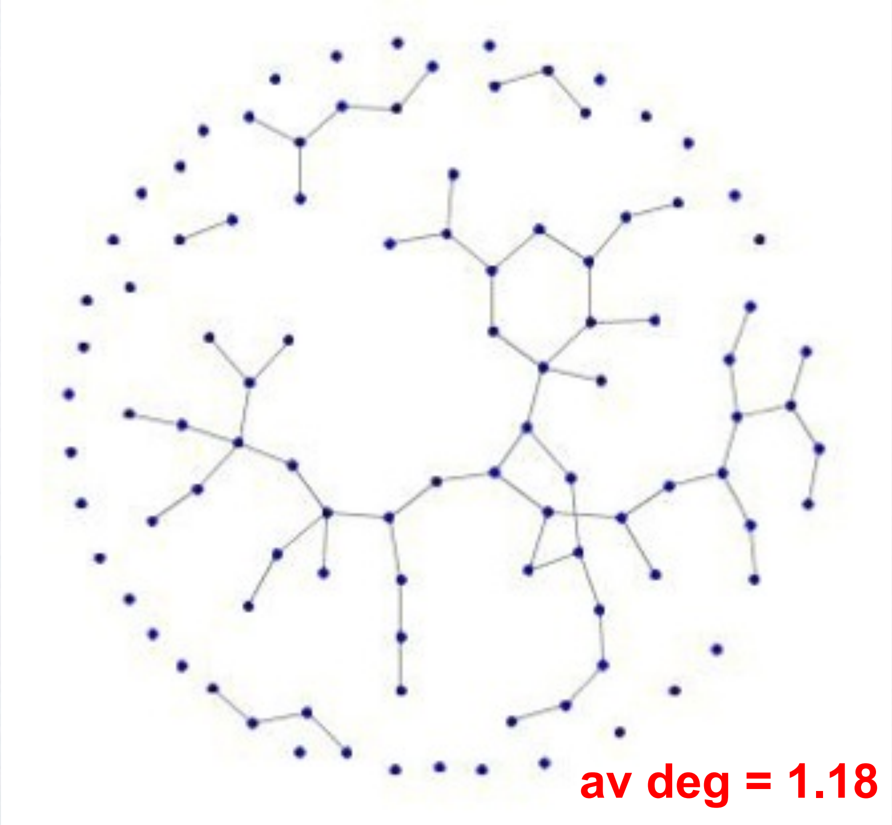


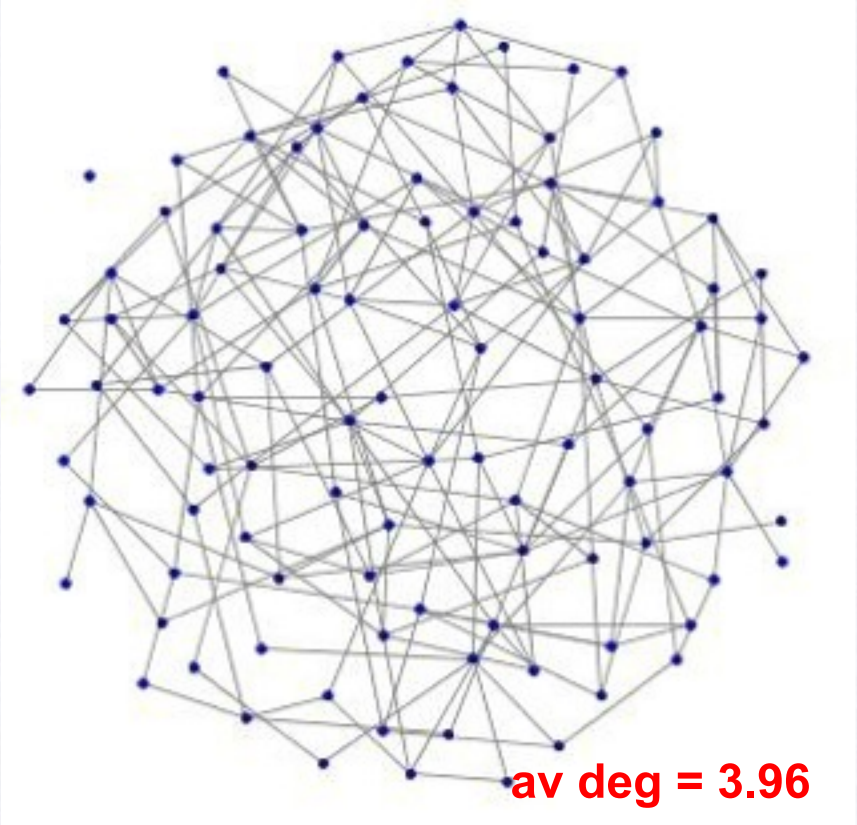
PERCOLATION THRESHOLD



- Percolation threshold
  - the point at which the giant component emerges
- As the average degree increases to p = 1, a giant component suddenly appears
- Edge removal is the opposite process As the average degree drops below 1 the network becomes disconnected

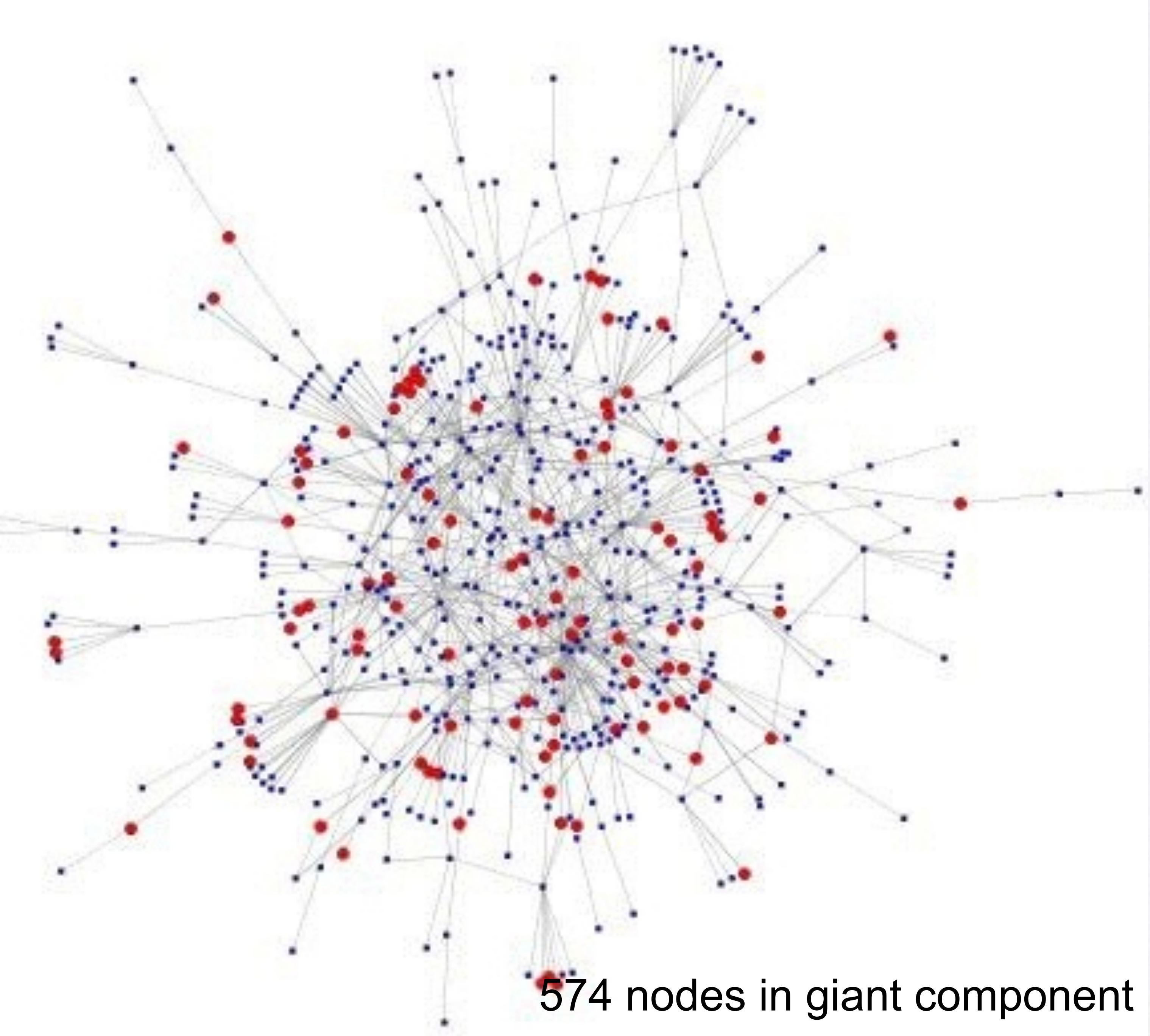


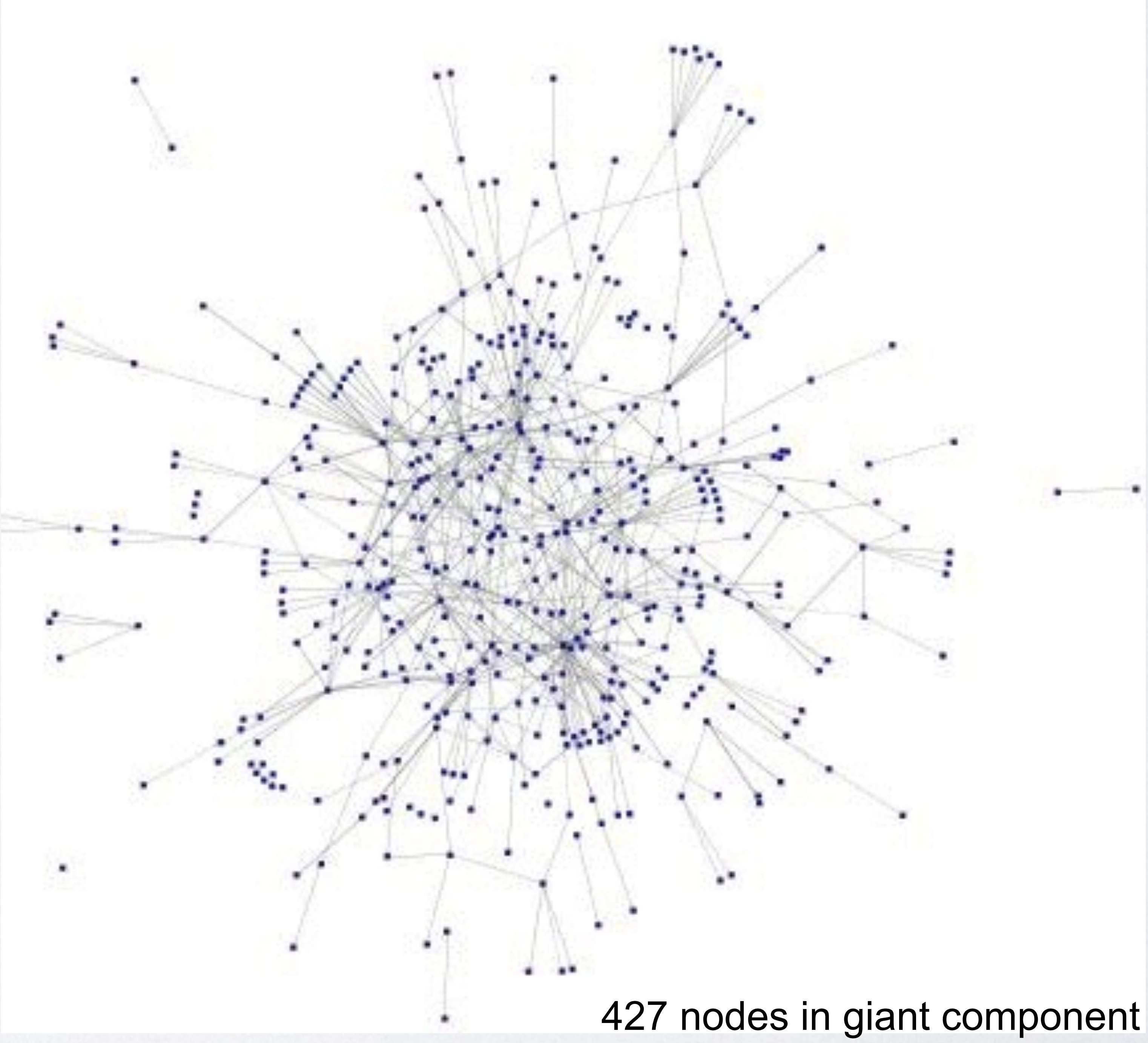




SCALE-FREE NETWORK RANDOM NETWORK

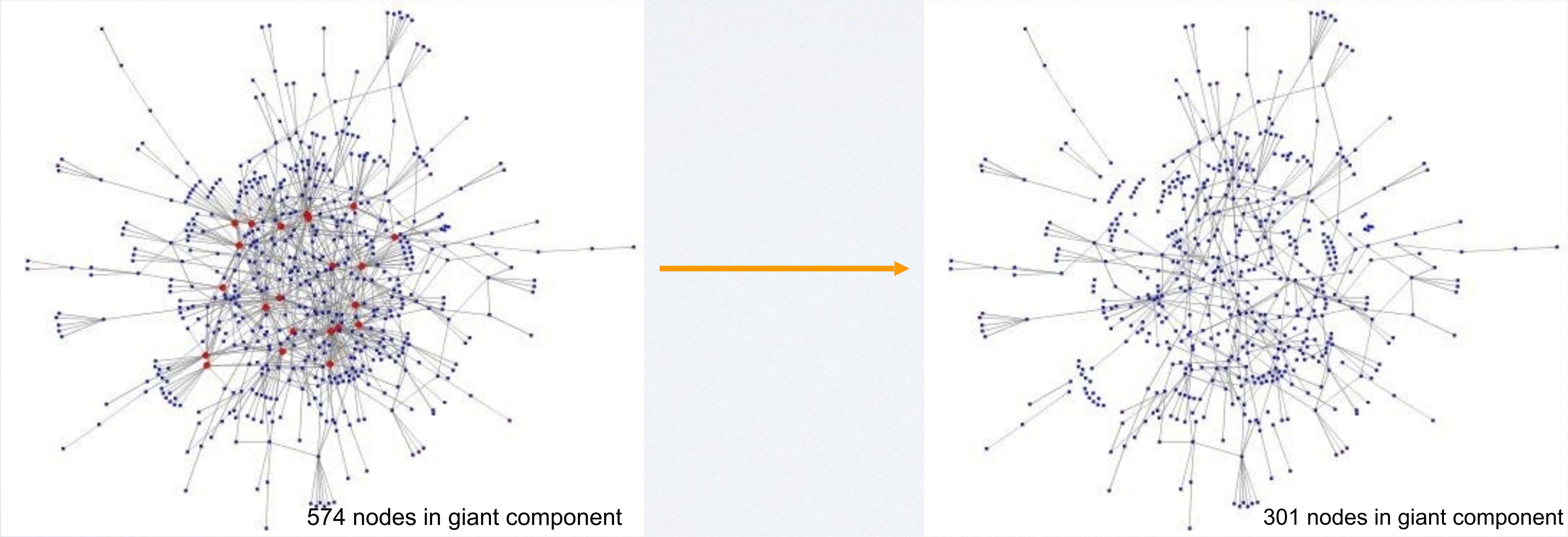
- gnutella network
  - 20% of nodes removed



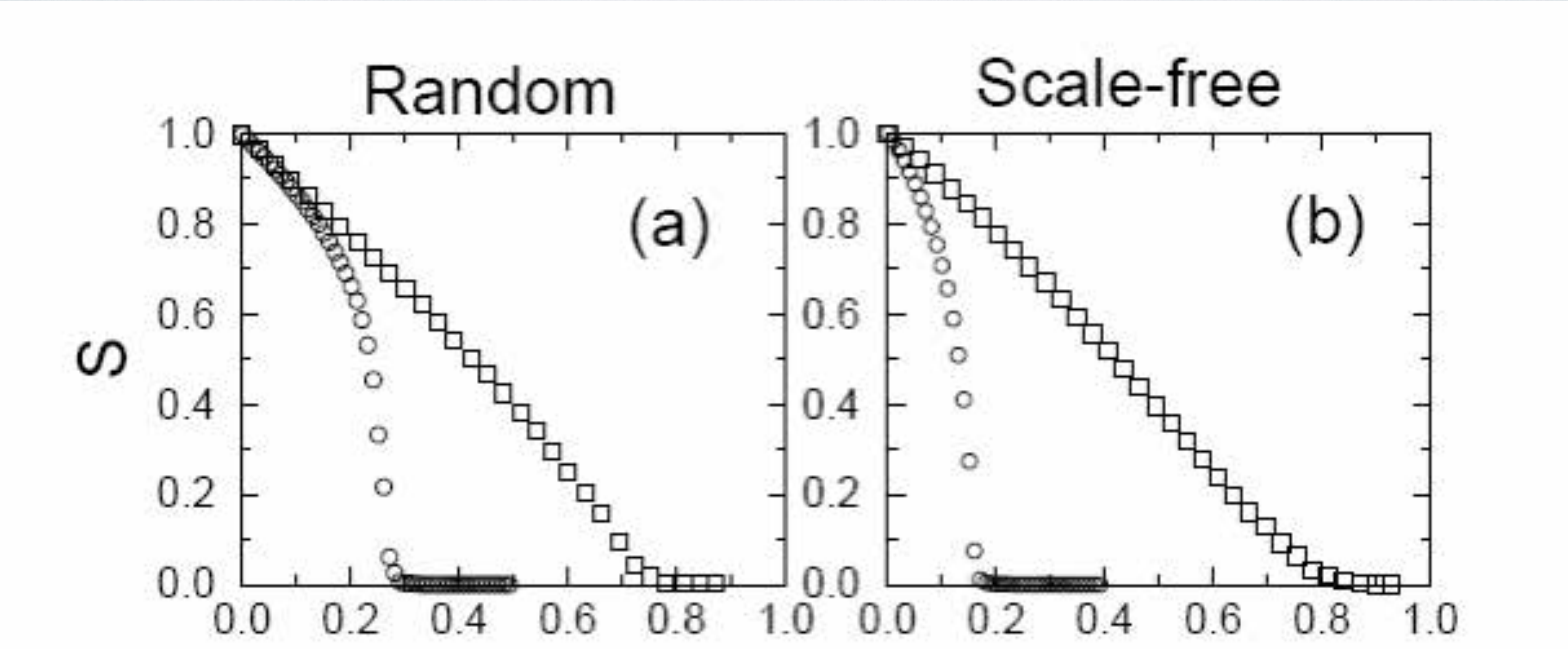


SCALE FREE NETWORK TARGETED ATTACKS

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



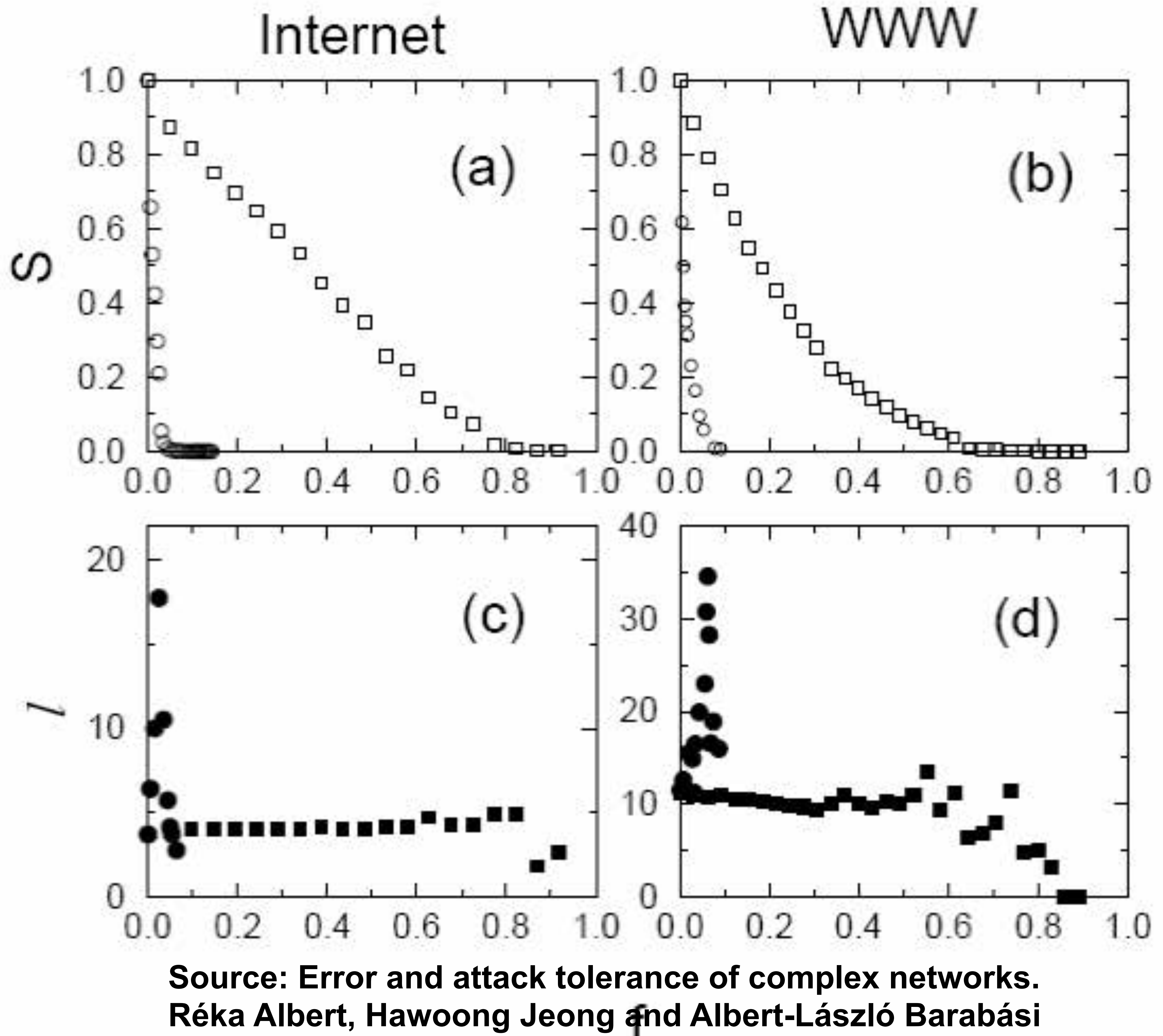
RANDOM VS. SCALE FREE NETWORKS



random failure

targeted attack

REAL NETWORKS



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

INFORMATION THEORY

$s_i = -k \log p_i$

$$S(x) = -k \log P(x)$$

- if P(x) is the weight we give to the presence of something, then H(x) becomes a measure of its absence !
  - If P(x) is very small, the large H(x) reflects that most of the time we do not see the event.
  - The indeterminacy of event

$$h_i = -kp_i \log p_i$$

    - 0 if p=0 or p=1
    - Entropy is therefore the indeterminacy of an ensemble $H(X) = -k \sum_i p_i \log p_i$
      - A metric of the total capacity of the ensemble to undergo change

INFORMATION THEORY

- Let's extend to relationships between events
  - Mutual information
  - Conditional entropy
  - Fundamental property
  - the capacity for evolution or self-organization (H) toward perturbation can be decomposed into two components.
    - Ascendency: I(X;Y) quantifying all that it regular, orderly, coherent and efficient.
    - Reserve : H(X|Y) representing the irregular, disorderly, incoherent and inefficient behaviors.

$$s_{ij} = -k \log p_{ij}$$

$$t_{i|j} = k \log(p_i p_j) - [-k \log(p_{ij})] = k \log \left(\frac{p_{ij}}{p_i p_j} \right)$$

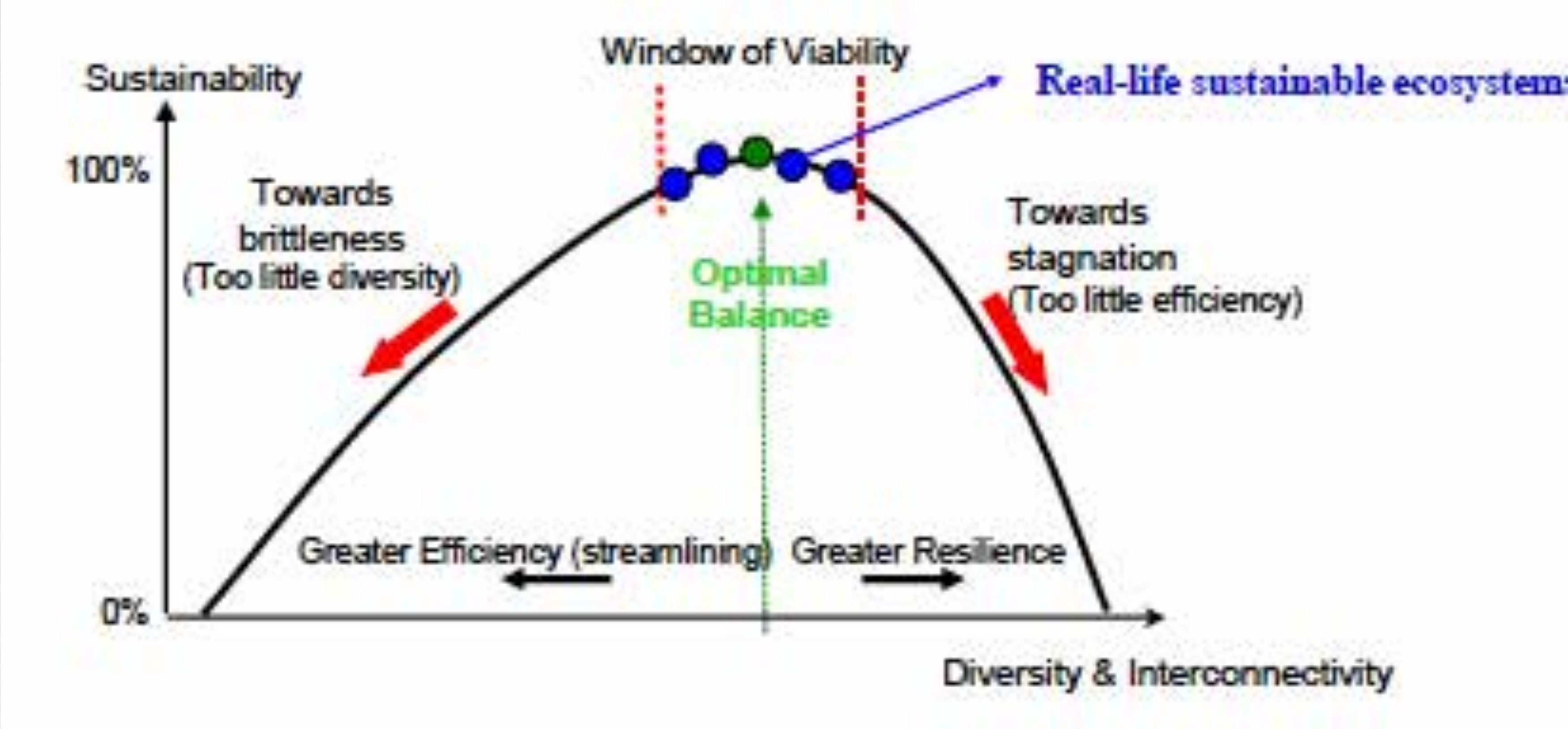
$$I(X;Y) = \sum_{i,j} p_{i,j} \log \frac{p_{ij}}{p_i p_j}$$

$$H(X) \geq I(X;Y)$$

$$H(X|Y) = H(X) - I(X;Y)$$

$$H(X) = H(X|Y) + I(X;Y)$$

EVOLUTION FITNESS



$$J(X;Y) = \frac{I(X;Y)}{H(X)}$$

- Systems with small ascendency or reserve are not survivable
  - Systems that endure lie somewhere between these extremes.
- We define the evolution fitness of a system by

$$F = -2J(X;Y)^{\beta} \log J(X;Y)^{\beta}, 1 \geq F \geq 0$$

$$F_{\max} = 1 \; at \; J(X;Y) = 2^{1/\beta}$$

KÖRNER GRAPH ENTROPY

- A subset S of the vertices V of a graph G = (V, E) is independent if no edge in the graph has both endpoints in S.
- Given a graph G, define the graph entropy of G $H(G) = \min_{X,Y} I(X;Y)$
    - where the minimum is taken over all pairs of random variables X, Y such that
        - X is a uniformly random vertex in G.
        - Y is an independent set containing X.
    - Ex: for an unbalanced complete bipartite graph Km,n. $H(G) \leq H(\frac{n}{m+n})$
- Property
    - (Disjoint union). If G1, . . . , Gk are the connected components of G, and for each i, ρi =|V (Gi)|/|V (G)| is the fraction of vertices in Gi, then $H(G) = \sum_i \rho_i H(G_i)$

GIBBS GRAPH ENTROPY

- Coming from statistical physics
  - A network ensemble is formed by the set of network satisfying a given number of constraints.
  - A partition function Z counts the number of networks in the ensemble

$$Z = \sum_{\{a_{ij}\}} \prod_k \delta(\mathrm{constraint}_k(\{a_{ij}\})) e^{-\sum_{i<j} \sum_\alpha h_{ij}(\alpha) \delta_{a_{ij},\alpha}},$$

  - Gibbs Entropy $\Sigma = \frac{1}{N} \log Z \ s.t. \ h_i j(\alpha) = 0; \forall (i, j, \alpha)$
  - Link probability $\pi_{ij}(\alpha) = \frac{\partial \log Z}{\partial h_{ij}(\alpha)}$
  - Graph Shannon Entropy $\sum_{i,j} \pi_{ij}(\alpha) \log \pi_{ij}(\alpha)$

$$\lim_{N \to \infty} \Sigma \to \sum_{i,j} \pi_{ij}(\alpha) \log \pi_{ij}(\alpha)$$

SPECTRAL GRAPH THEORY

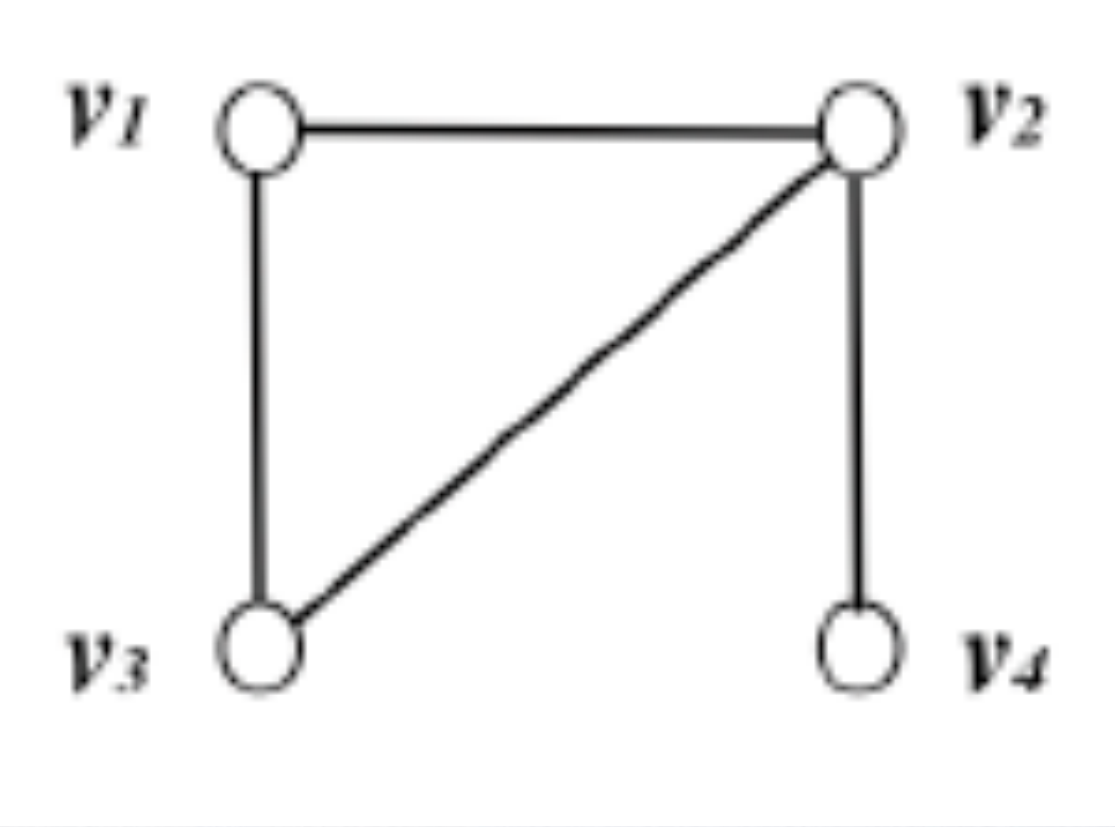
- Study the properties of graphs via the eigenvalues and eigenvectors of their associated graph matrices
  - The adjacency matrix, the graph Laplacian and their variants.
    - These matrices have been extremely well studied from an algebraic point of view.
- The Laplacian allows a natural link between discrete representations (graphs), and continuous representations, such as metric spaces and manifolds.
- Laplacian embedding consists in representing the vertices of a graph in the space spanned by the smallest eigenvectors of the Laplacian
  - A geodesic distance on the graph becomes a spectral distance in the embedded (metric) space.

ADJACENCY MATRIX

- For a graph with n vertices, the entries of the adjacency matrix are defined by:

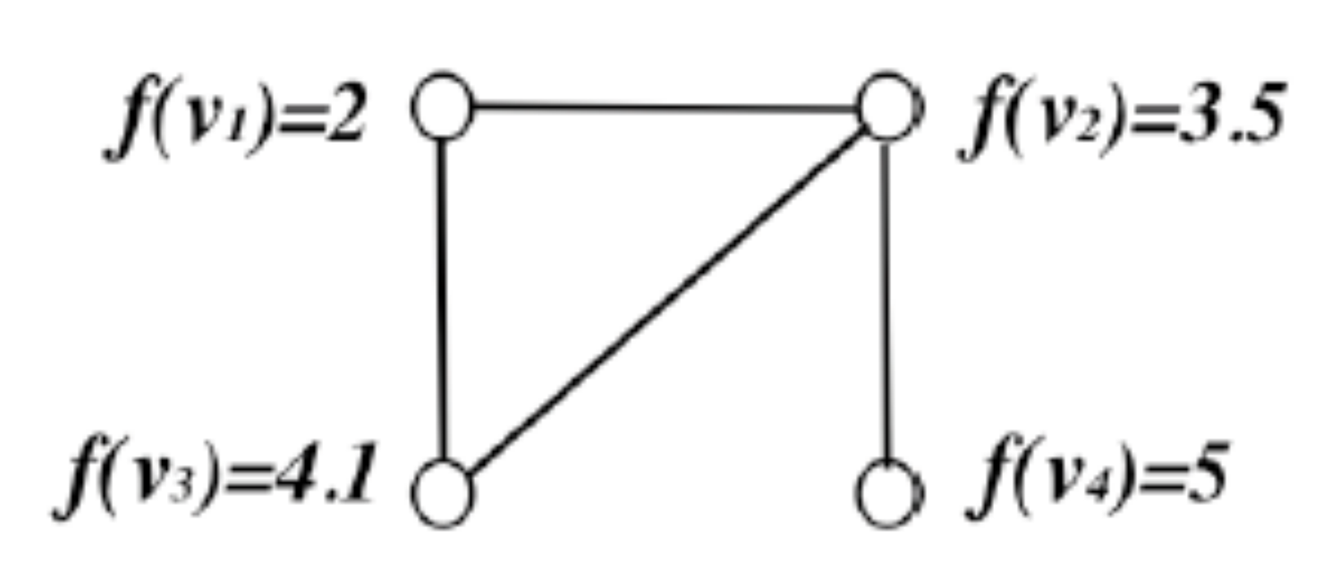
$$A = \begin{cases} a_{ij} = 1 & \text{if there is an edge } e_{ij} \\ a_{ij} = 0 & \text{if there is not an edge } e_{ij} \\ a_{ii} = 0 & \end{cases}$$

$$\begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$



REAL VALUED FUNCTION OVER GRAPHS

- We consider real-valued functions $f : \mathcal{V} \to \mathbb{R}$ on the set of the graph's vertices $f = (f(v_1, \ldots, f(v_n)) = (f_1, \ldots, f_n) \in \mathbb{R}^n$
  - Assigns a real number to each graph node.
    - Notation: $g = \mathbf{A}f,\ g(i) = \sum_{i \to j} f(j)$



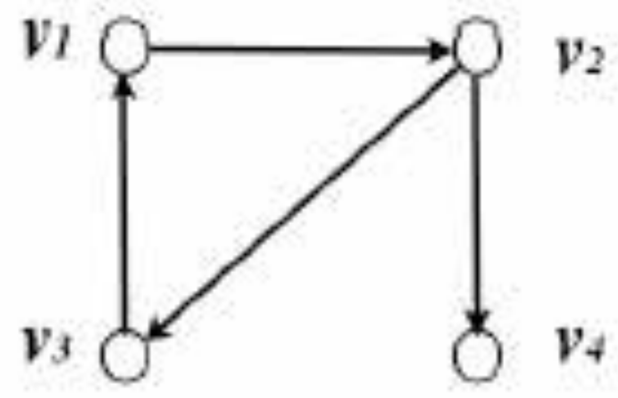
- The eigenvectors of the adjacency matrix, can be viewed as eigenfunctions. $\mathbf{A}X = \lambda X$
  - Quadratic form $f^T \mathbf{A} f = \sum_{i \to j} f(i) f(j)$

INCIDENCE MATRIX

- Dual matrix of adjacency
- Matrix defined on the edge of the graph

$$\nabla = \begin{cases} \nabla_{ev} = -1, & \text{if } v \text{ is the initial vertex of edge } e \\ \nabla_{ev} = +1, & \text{if } v \text{ is the terminal vertex of edge } e \\ \nabla_{ev} = 0, & \text{if } v \text{ is not in } e \end{cases}$$

$$\nabla = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & +1 \end{bmatrix}$$



- The mapping $f \to \nabla f$ is known as the co-boundary mapping of f

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & +1 \end{bmatrix} \begin{pmatrix} f(1) \\ f(2) \\ f(3) \\ f(4) \end{pmatrix} = \begin{pmatrix} f(2) - f(1) \\ f(1) - f(3) \\ f(3) - f(2) \\ f(4) - f(2) \end{pmatrix}$$

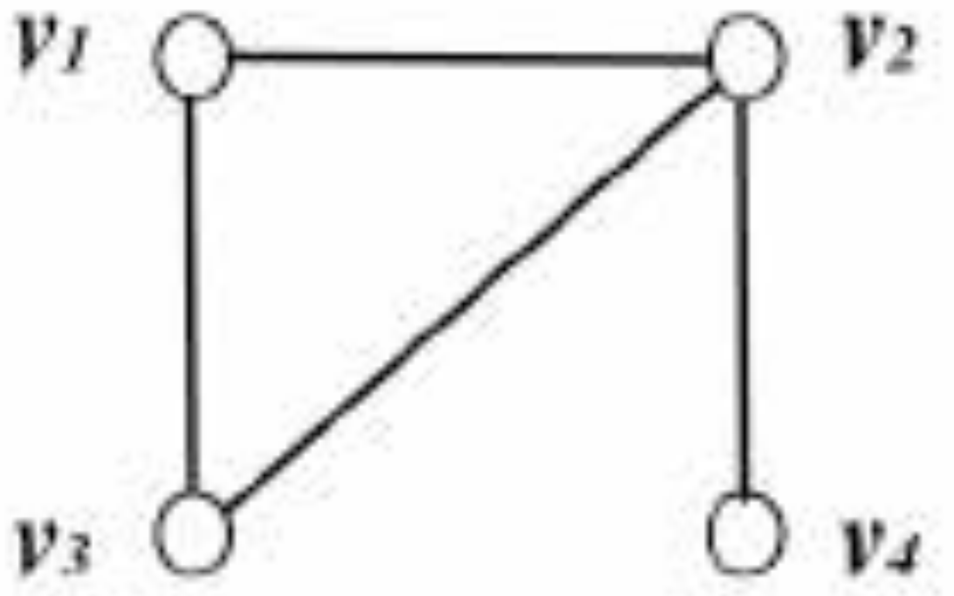
THE LAPLACIAN MATRIX

$$L = \nabla^T \nabla$$

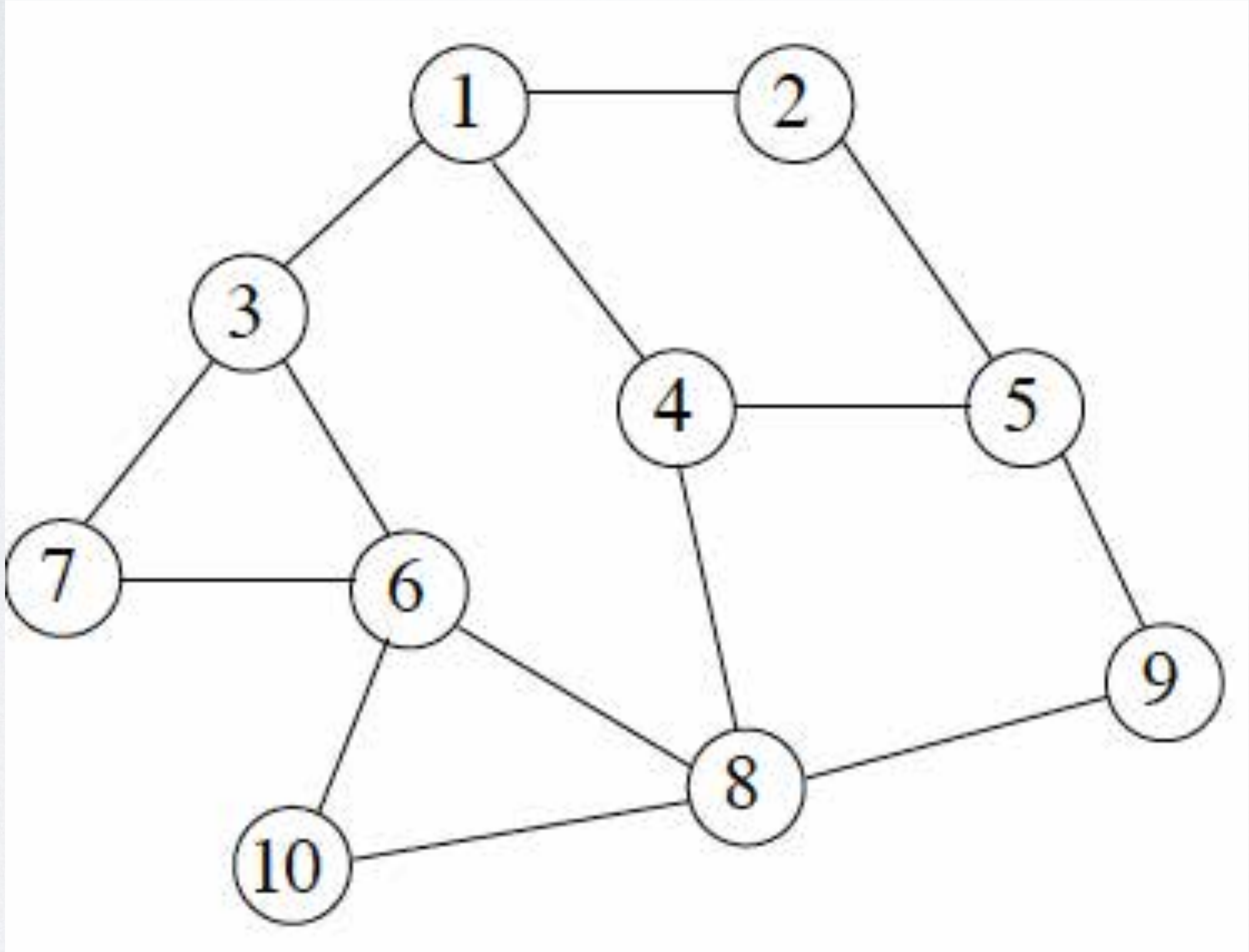
$$(Lf)(v_i) = \sum_{v_j \to v_i} (f(v_i) - f(v_j))$$

- Connection between Laplacian and Adjcency matrix
  - D degree matrix $L = D - W$

$$\mathbf{L} = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 2 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}$$



EXAMPLE



EXAMPLE

$$\mathbf{L} = \begin{bmatrix} 3 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 3 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 3 & -1 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 3 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 4 & -1 & -1 & 0 & -1 \\ 0 & 0 & -1 & 0 & 0 & -1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & -1 & 0 & 4 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 2 \end{bmatrix}$$

$$\mathbf{\Lambda} = [\; 0.0000 \quad 0.7006 \quad 1.1306 \quad 1.8151 \quad 2.4011 \quad 3.0000 \quad 3.8327 \quad 4.1722 \quad 5.2014 \quad 5.7462 \;]$$

EXTENSION

- Laplacian for a weighted graph is defined as $L = D - W$
  - W is the weight matrix, $w_{ij} = w(x_i, x_j)$
  - D is a diagonal matrix with $d_{ii} = \sum_j w_{ij}$
  - Laplacian regularization $N(f) = f^T L f$
  - Normalized Laplacian $L' = (I - D^{-1}W)$

FIEDLER VALUE OF A GRAPH

- The first non-null eigenvalue λ_{k+1} is called the Fiedler value.
  - The corresponding eigenvector is called the Fiedler vector.
  - The Fiedler value is the algebraic connectivity of a graph, the further from 0, the more connected.
  - The Fiedler vector has been extensively used for spectral partioning

VON NEUMANN ENTROPY

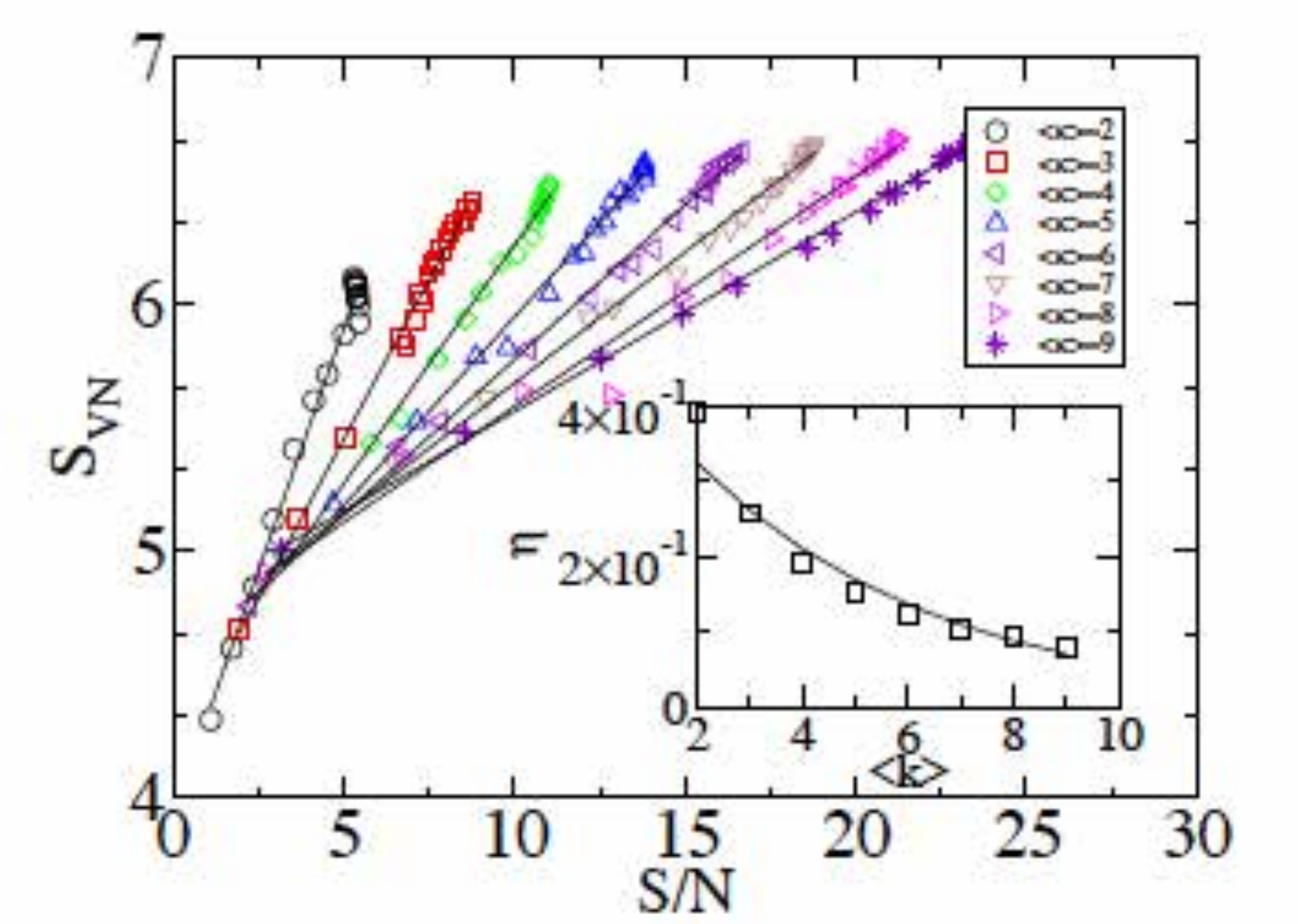
- Strongly related to spectral properties of graph
  - Based on the normalized laplacian L

$$S_{VN} = -\langle Tr(L)\log(L)\rangle_{\Pi}$$

$$\log(L) = V\log(\Lambda)V'$$

  - Relations with Shannon Entropy

$$S_{VN} = \eta\frac{S}{N} + \beta$$



CONCLUSION

- We tried to formalize the trade-off between ascendency/reserve, performance/plasticity, etc…
  - Korner Graph Entropy seems to go toward the first term
  - Von Neumann one toward the second one


FINALLY WHY THE ASIAN CAMEL HAS TWO HUMPS ?