Wednesday, March 8, 2006

Complex Networks

Presenter:

Jirakhom Ruttanavakul

CS 790R, University of Nevada, Reno

Presented Papers

- Emergence of scaling in random networks, Barabási & Bonabeau (2003)
- Scale-free networks, Barabási & Albert (1999)
- Scale-free and hierarchical structures in complex networks, Barabási et al. (2002)

Random vs. Scale-Free Networks

- Random Networks
  - The number of vertices is fixed from the beginning and edges can be randomly connected or reconnected
  - The probability that two vertices are connected is random and uniform
- Scale-Free Networks
  - Vertices can be added or removed from the network, thus the size of the network varies over time
  - Higher probability of connection to already popular vertices
  - Network contains important nodes that have connections to many other nodes and are called “hubs”

Examples of Scale-Free Networks & Hubs

- WWW: Yahoo!, Google, etc.
- Physical Structure of the Internet: routers
- Sexual relationships: Sweden
- People connected by e-mail
- Hollywood: Kevin Bacon
- Scientific papers connected by citations: Erdős papers
- Business Partnerships: Genzyme, Chiron, Genentech
- Etc.

Random vs. Scale-Free Networks

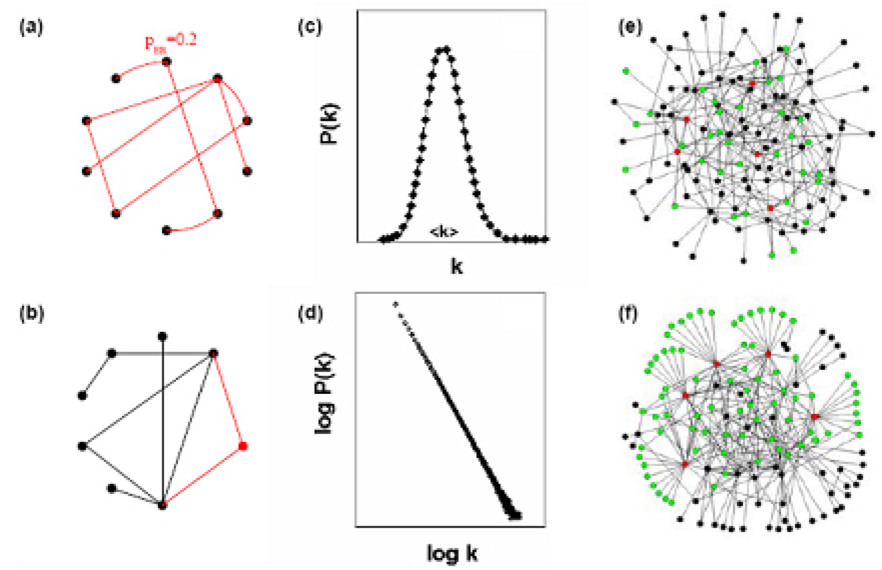


FIGURE 1. **(a)** The Erdős-Rényi random network model is constructed by laying down N nodes and connecting each pair of nodes with probability p. The figure shows a particular realization of such a network for $N = 10$ and $p = 0.2$. **(b)** The scale-free model assumes that the network continually grows by the addition of a new nodes. The figure shows the network at time t (black nodes and links) and after the addition of a new node at time $t + 1$ (red). The probability that the new node chooses a node with k links follows (2), favoring highly connected nodes, a phenomenon called preferential attachment. **(c)** For the random graph generated by the Erdős-Rényi model the degree distribution, $P(k)$, is strongly peaked at $k = < k >$ and decays exponentially for large k. **(d)** $P(k)$ for a scale-free network does not have a peak, and decays as a power-law, $P(k) \sim k^{-\gamma}$. **(e)** The random network generated by the Erdös-Rényi model is rather homogeneous, i.e. most nodes have approximately the same number of links. **(f)** In contrast, a scale-free network is extremely inhomogeneous: while the majority of the nodes have one or two links, a few nodes have a large number of links, guaranteeing that the system is fully connected. To show this, we colored with red the five nodes with the highest number of links, and with green their first neighbors. While in the exponential network only 27% of the nodes are reached by the five most connected nodes, in the scale-free network more than 60% are, demonstrating the key role hubs play in the scale-free network. Note that both networks contain the same number of nodes and links, 130 and 430, respectively. After [26].

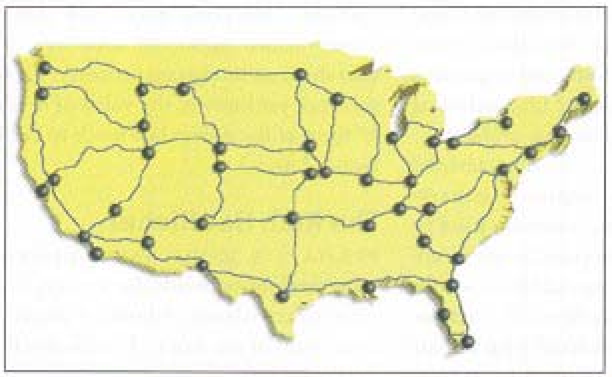
Random vs. Scale-Free Networks

RANDOM VERSUS SCALE-FREE NETWORKS

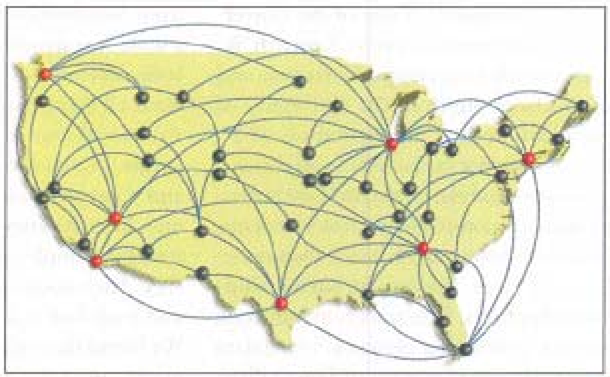
RANDOM NETWORKS, which resemble the U.S. highway system (*simplified in left map*), consist of nodes with randomly placed connections. In such systems, a plot of the distribution of node linkages will follow a bell-shaped curve (*left graph*), with most nodes having approximately the same number of links.

In contrast, scale-free networks, which resemble the U.S. airline system (*simplified in right map*), contain hubs (*red*)—nodes with a very high number of links. In such networks, the distribution of node linkages follows a power law (*center graph*) in that most nodes have just a few connections and some have a tremendous number of links. In that sense, the system has no "scale." The defining characteristic of such networks is that the distribution of links, if plotted on a double-logarithmic scale (*right graph*), results in a straight line.

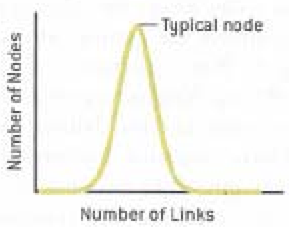
Random Network



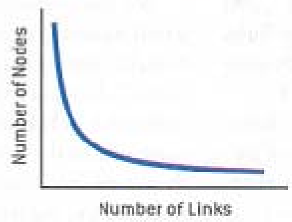
Scale-Free Network

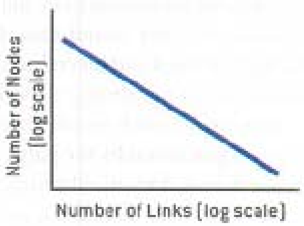


Bell Curve Distribution of Node Linkages



Power Law Distribution of Node Linkages





Examples of Scale-Free Networks & Hubs

TABLE 1. The scaling exponents characterizing the degree distribution of several scale-free networks, for which $P(k)$ follows a power-law (1). We indicate the size of the network and its average degree $\langle k \rangle$. For directed networks we list separately the indegree (γ_{in}) and outdegree (γ_{out}) exponents, while for the undirected networks, marked with a star, these values are identical. Expanded after Ref. [1].

Network	Size	$\langle k \rangle$	γ_{out}	γ_{in}	Reference
WWW	325,729	4.51	2.45	2.1	[11]
WWW	4×10^7	7	2.38	2.1	[27]
WWW	2×10^8	7.5	2.72	2.1	[14]
WWW, site	260,000			1.94	[28]
Internet, domain*	3,015 - 4,389	3.42 - 3.76	2.1 - 2.2	2.1 - 2.2	[23]
Internet, router*	3,888	2.57	2.48	2.48	[23]
Internet, router*	150,000	2.66	2.4	2.4	[29]
Movie actors*	212,250	28.78	2.3	2.3	[30]
Coauthors, SPIRES*	56,627	173	1.2	1.2	[31]
Coauthors, neuro.*	209,293	11.54	2.1	2.1	[32]
Coauthors, math*	70,975	3.9	2.5	2.5	[32]
Sexual contacts*	2810		3.4	3.4	[24]
Metabolic, E. coli	778	7.4	2.2	2.2	[7]
Protein, S. cerev.*	1870	2.39	2.4	2.4	[8]
Ythan estuary*	134	8.7	1.05	1.05	[33]
Silwood park*	154	4.75	1.13	1.13	[33]
Citation	783,339	8.57		3	[15]
Phone-call	53×10^6	3.16	2.1	2.1	[34]
Words, conccurence*	460,902	70.13	2.7	2.7	[20]
Words, synonyms*	22,311	13.48	2.8	2.8	[19]
Protein, S. Cerev*	9,85	1.83	2.5	2.5	[35]
Comic Book Characters	6,486	14.9	0.66	3.12	[36]
E-mail	59,912	2.88	2.03	1.49	[37]
Protein Domains*	876	9.32	1.6	1.6	[38]
Prot. Dom. (PromDom)*	5995	2.33	2.5	2.5	[39]
Prot. Dom. (Pform)*	2478	1.12	1.7	1.7	[39]
Prot. Dom. (Prosite)*	13.60	0.77	1.7	1.7	[39]

Why Scale-Free Networks are Important

- Contemporary science cannot describe systems composed of non-identical elements that have diverse and non-local interactions (elements = vertices, interactions = edges).
  - Living systems: vertices = proteins & genes, or nerve cells; edges = chemical interactions, or axons
  - Social sciences: vertices = individuals or organizations; edges = social interactions between them
  - WWW: vertices = HTML documents; edges = hyperlinks
  - Language: vertices = words; edges = syntactic relationships
- The topology of real networks is mostly unknown, because these networks are very large, and interactions are very complex
- Researchers have little understanding of network structures and properties

Properties of Scale-Free Networks

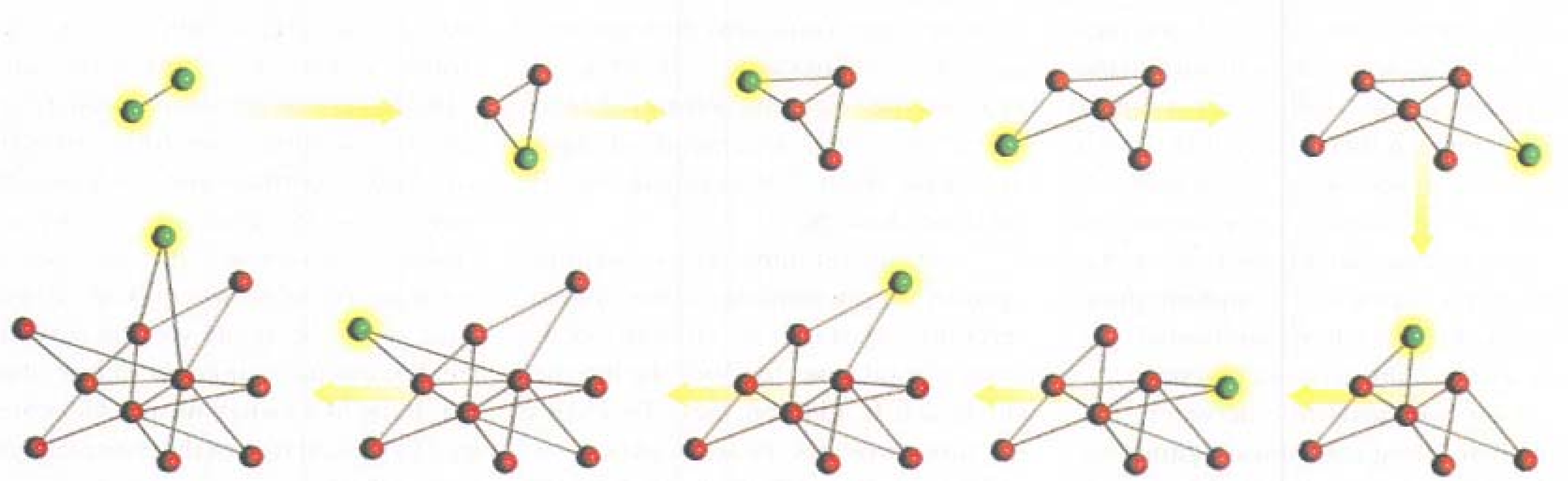
- Network can be freely expanded – Adding new vertices (Growth)
- New vertices usually are connected to already well connected vertices (Preferential Attachment)
- The probability of a vertex to interact with other *k* vertices decays as a "Power Law":

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

- Surprisingly, all examples given earlier shared the same power-law and γ tends to fall between 2 and 3
- The power-law distribution implies that nodes with only a few links are numerous, but few nodes have a large number of links

BIRTH OF A SCALE-FREE NETWORK

A SCALE-FREE NETWORK grows incrementally from two to 11 nodes in this example. When deciding where to establish a link, a new node (*green*) prefers to attach to an existing node (*red*) that already has many other connections. These two basic mechanisms—growth and preferential attachment—will eventually lead to the system's being dominated by hubs, nodes having an enormous number of links.



Networks following a Power Law

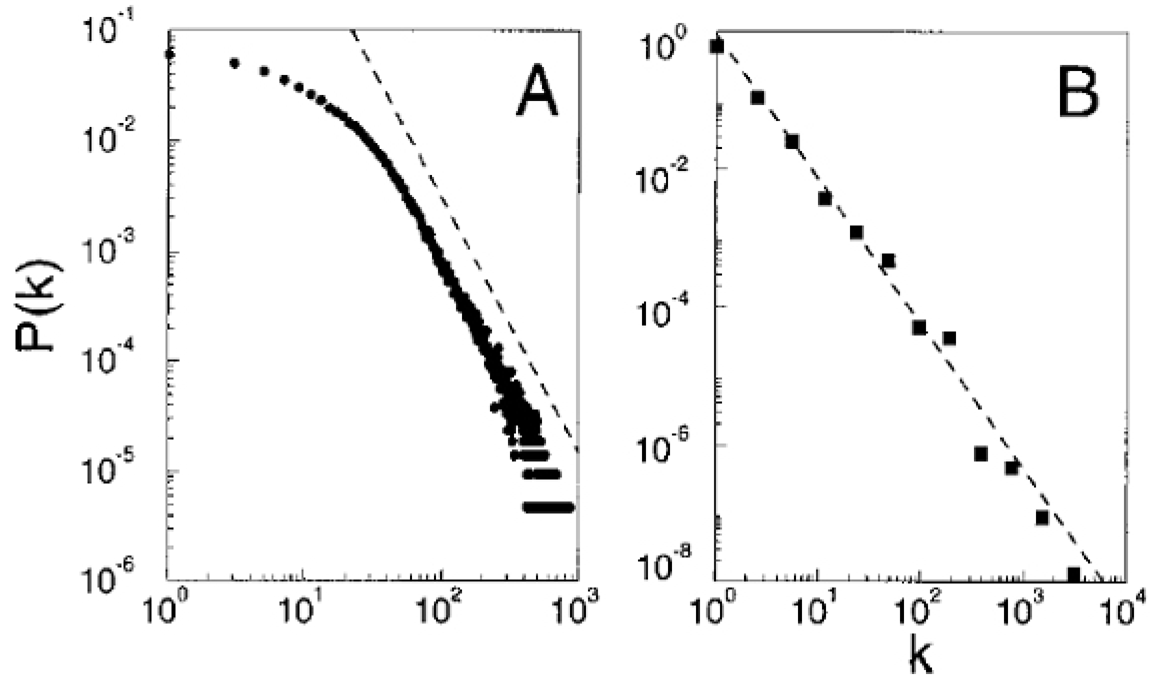


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). (**C**) Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{\text{actor}} = 2.3$, (B) $\gamma_{\text{www}} = 2.1$ and (C) $\gamma_{\text{power}} = 4$.

Network Models of ER & WS

- ER (Erdős and Rényi)
  - Start with N vertices; the probability of connection is unformly p
  - Probability of a vertex to be connected to k other vertices is

$$P(k) = \frac{e^{-\lambda}\lambda^k}{k!} \qquad \text{where} \qquad \lambda = N\binom{N-1}{k}p^k(1-p)^{N-1-k}$$

- WS (Watts and Strogatz)
  - Start with N vertices forming a 1-D lattice: each vertex is connected to its nearest and next nearest neighbors
  - Then each edge can be rewired to another vertex randomly chosen with probability p
  - If $p = 0$, z = coordination number in the lattice

$$P(k) = \delta(k - z)$$

- In these two models, nodes with a large number links (hubs) are absent

Incorporating Two Major Factors

- Two major factors – Growth and Preferential Attachment
- Growth : Start with m_o nodes and add new nodes with $m \leq m_o$ edges linked to different existing vertices
- Preferential Attachment: Assume probability (Π) that a new node will be connected to an existing node i depends on the connectivity k_i of that node
  - $\Pi(k_i) = k_i / \Sigma_j k_j$
- After t time steps, this model will lead to a random network with $t+m_o$ nodes and mt edges
- Follows a power law with $\gamma_{\text{model}} = 2.9 \pm 0.1$ (correct model should have a distribution whose features are independent of time)

Why These Two Factors are Important

- To prove that these two factors are important in the development of the network, the authors investigate two variants of the model
- Model A: keep the growth but eliminate preferential attachment
  - Instead, a new vertex is connected with equal probability to any vertex in the system $\Pi(k_i) = 1 / (m_o + t - 1)$
  - This leads to $P(k) \sim \exp(-\beta k)$ and eliminates the scale-free property

Why These Two Factors are Important

- Model B: The number of vertices is fixed, and preferential attachment is integrated into the network $\Pi(k_i) = k_i / \Sigma_j k_j$
  - At first, the system follows as power-law, but after N^2 time steps, all the nodes are connected
- In the development of power-law (scale-free) distribution network, both factors are needed

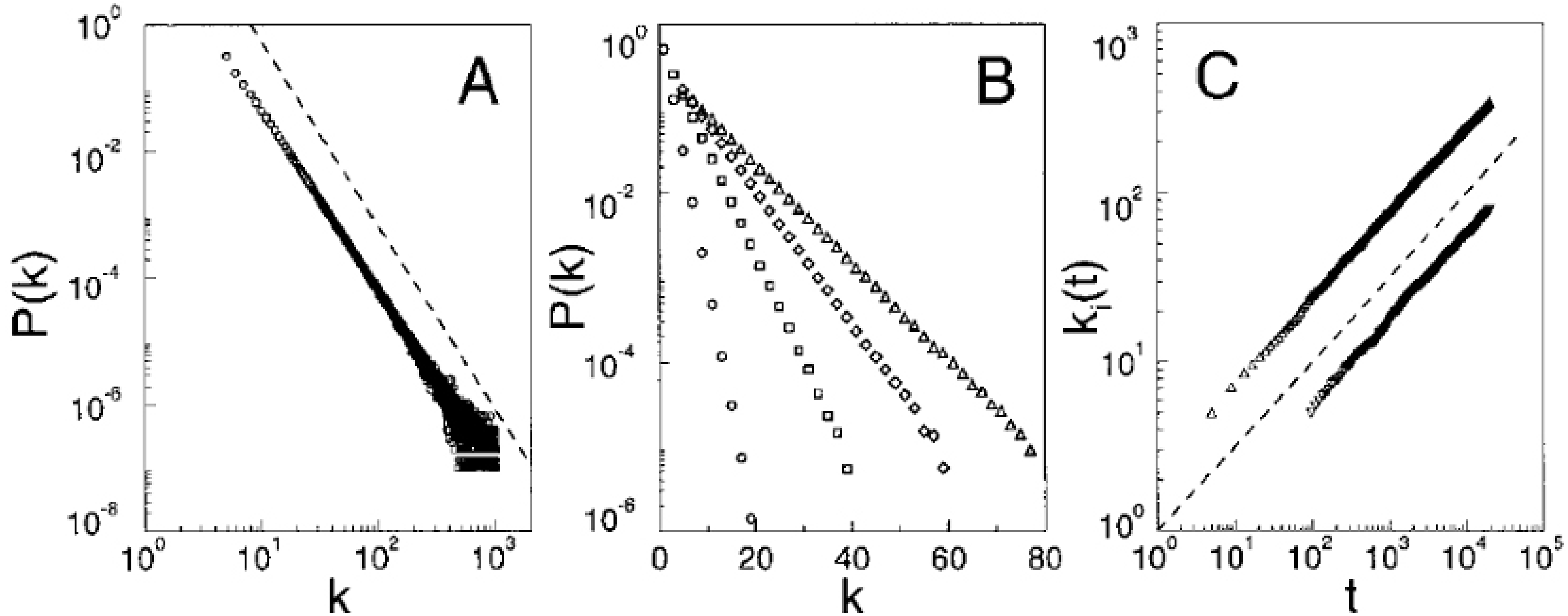


Fig. 2. (**A**) The power-law connectivity distribution at t = 150,000 (○) and t = 200,000 (□) as obtained from the model, using $m_0 = m = 5$. The slope of the dashed line is $\gamma = 2.9$. (**B**) The exponential connectivity distribution for model A, in the case of $m_0 = m = 1$ (○), $m_0 = m = 3$ (□), $m_0 = m = 5$ (◇), and $m_0 = m = 7$ (△). (**C**) Time evolution of the connectivity for two vertices added to the system at $t_1 = 5$ and $t_2 = 95$. The dashed line has slope 0.5.

The Rich get Richer

- All nodes are not equal, the more connected nodes tend to acquire new connections from the new nodes added to the system
  - more connected actors tend to be chosen for a new role
- With preferential attachment, a vertex that acquires more connections than another one tends to increase its connectivity at a higher rate (earlier nodes are favored, becoming popular nodes and more favored, etc.)
  - $\partial k_i / \partial t = k_i / 2t$, which gives $k_i(t) = m(t/t_i)^{0.5}$, where t_i is the time vertex i was added

How to Model a Network

- Use "rich-get-richer" properties to calculate γ analytically, by defining $P[k_i(t) < k]$, or $P[t_i > m^2 t/k^2]$
  - Assume the vertices are added to the system at the same time
  - Over a long period of time, the system will reach $P(k) = 2m^2/k^3$ giving $\gamma = 3$, independently of m
  - This model can't be expected to account for all aspects of the studied networks
- Based on the authors' simulations, scaling is present only for $\Pi(k) \sim k$. If the mechanism is faster than linear, the topology will be star-shaped.
- The model can be easily modified to account for exponents different from $\gamma = 3$, for example a fraction p of the links can be redirected, yielding $\gamma(p) = 3 - p$

Advantages & Disadvantages

- Advantages of scale-free networks
  - Robust against accidental failures
  - Understanding the characteristics of the scale-free networks can prevent disasters
    - Computer viruses
    - Epidemic of diseases
- Disadvantages of scale-free networks
  - Vulnerable to coordinated attacks
  - Can't easily eradicate the viruses or diseases already in the system

Stopping Viruses in Scale-Free Networks

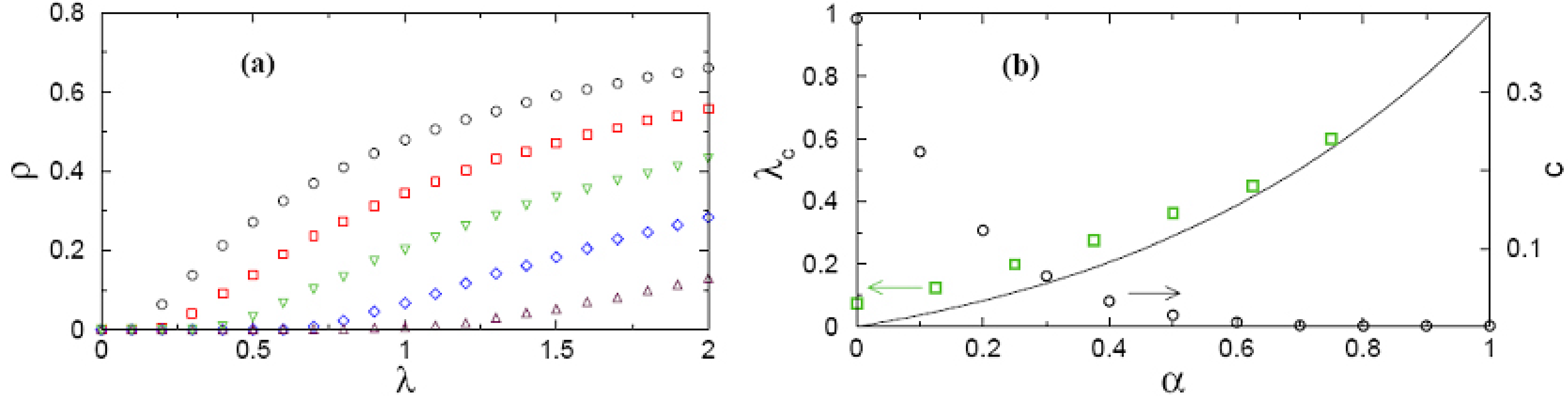
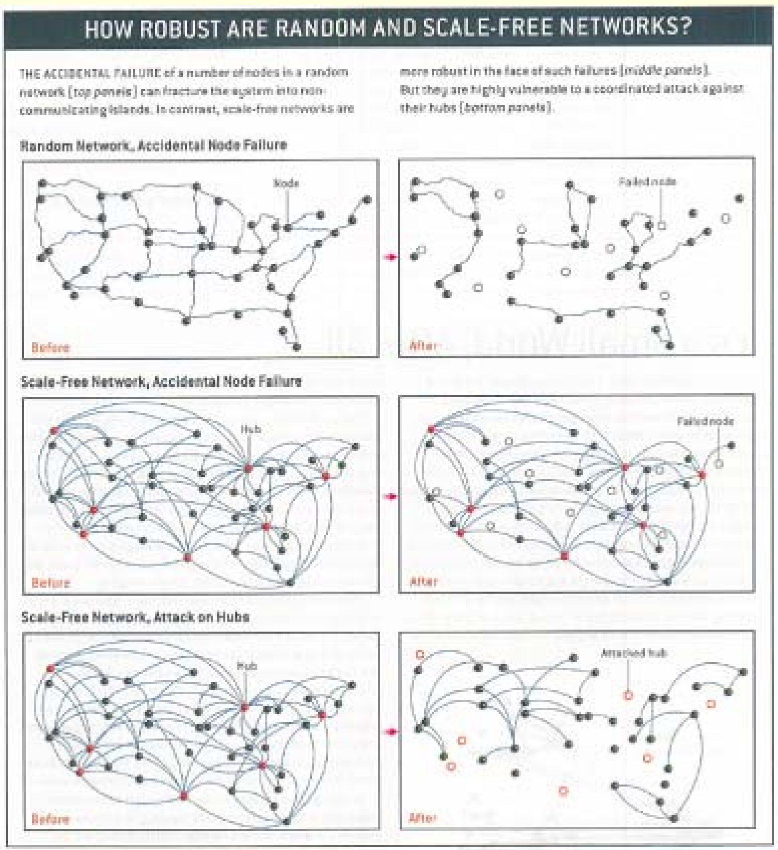
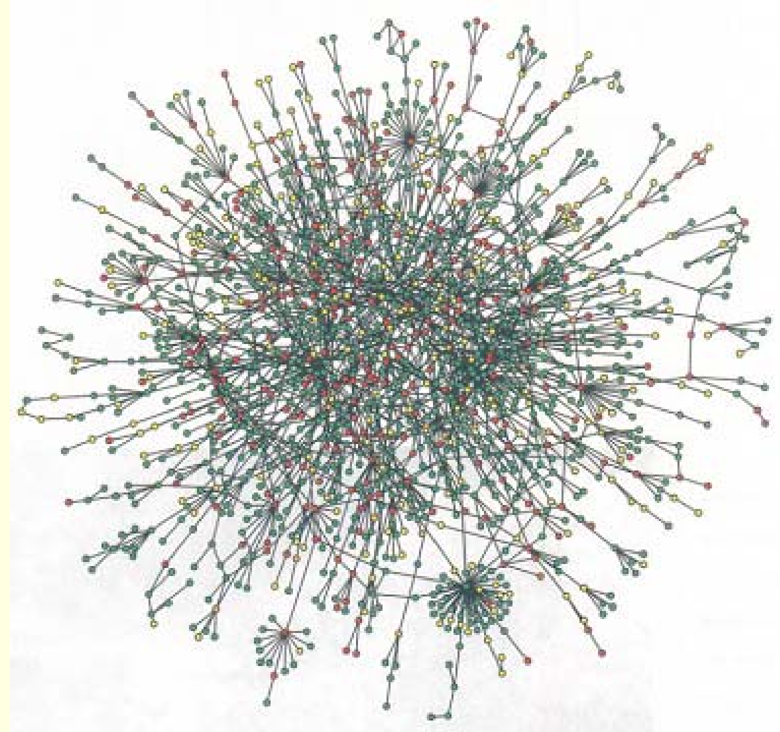


FIGURE 6. Curing the hubs. (a) Prevalence, ρ, measured as the fraction of infected nodes in function of the effective spreading rate λ for $\alpha = 0$ (circle), 0.25 (square), 0.50 (triangle down), 0.75 (diamond) and 1 (triangle up), as predicted by Monte-Carlo simulations using the SIS model on a scale-free [25] network with N=10,000 nodes. While for $\alpha = 0$ the epidemic threshold is zero, a nonzero α leads to the emergence of a finite epidemic threshold. **(b)** The dependence of the epidemic threshold λ_c on α as predicted by our calculations (continuous line) based on the continuum approach described in Ref. [68], and by the numerical simulations based on the SIS model (green boxes). The small deviation between the numerical results and the analytical prediction is due to the uncertainty in determining the precise value of the threshold in Monte-Carlo simulations. The vertical axis on the r.h.s. labels the number of cures, c, administered in an unit time per node for different values of α, shown as black circles on the figure. The rapidly decaying c indicates that more successful is a policy in selecting and curing hubs (larger is α), fewer cures are required for a fixed spreading rate ($\lambda = 0.75$). The data points in (a) and (b) are averaged over 10 independent runs. After [72].





MAP OF INTERACTING PROTEINS in yeast highlights the discovery that highly linked, or hub, proteins tend to be crucial for a cell's survival. Red denotes essential proteins (their removal will cause the cell to die). Orange represents proteins of some importance (their removal will slow cell growth). Green and yellow represent proteins of lesser or unknown significance, respectively.

Hierarchical Network Model

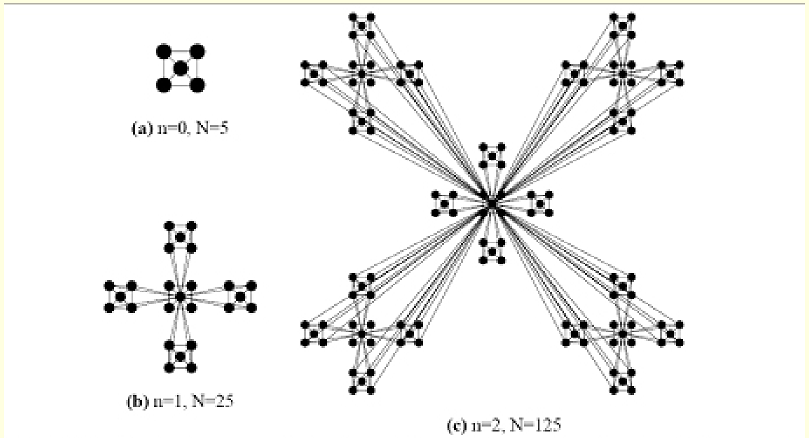
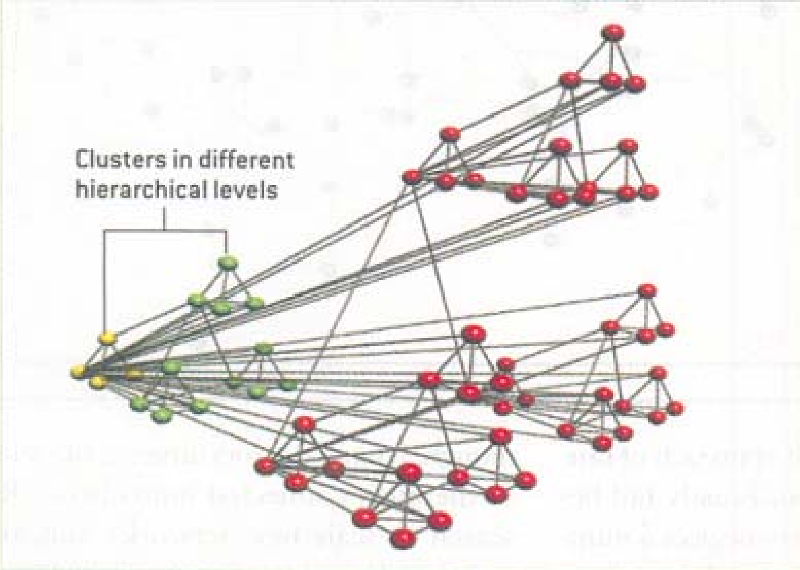


FIGURE 3. The iterative construction leading to a hierarchical network. Starting from a fully connected cluster of five nodes shown in **(a)** (note that the diagonal nodes are also connected – links not visible), we create four identical replicas, connecting the peripheral nodes of each cluster to the central node of the original cluster, obtaining a network of $N = 25$ nodes **(b)**. In the next step we create four replicas of the obtained cluster, and connect the peripheral nodes again, as shown in **(c)**, to the central node of the original module, obtaining a $N = 125$ node network. This process can be continued indefinitely. After [50].



HIERARCHICAL CLUSTERS, shown schematically, could include, say, Web pages on the Frank Lloyd Wright home Fallingwater (*yellow*), which could be linked to other clusters (*green*) focusing on Wright, famous homes or Pennsylvania's attractions. Those sites, in turn, could be connected to clusters (*red*) on famous architects or architecture in general.

Why Hierarchical Networks

- The architecture of hierarchical networks is significantly different from scale-free and random networks
- Can't be described using scale-free or random network models
- Rather follow a scaling law:

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

  - Where: C is the Clustering Coefficient of a node with k links
    - $C = 2n_i/k_i(k_i\text{-}1)$; n_i is the number of links between the k_i neighbors of i. Random Network: $C(N) \sim N^{-1}$; Scale-Free Network : $C(N) \sim N^{-0.75}$
- Ex. of hierarchical networks:
  - 5 nodes : $C = 1$, $k = 5$
  - 25 nodes : $C = 3/19$, $k = 20$
  - 125 nodes : $C = 3/83$, $k = 84$

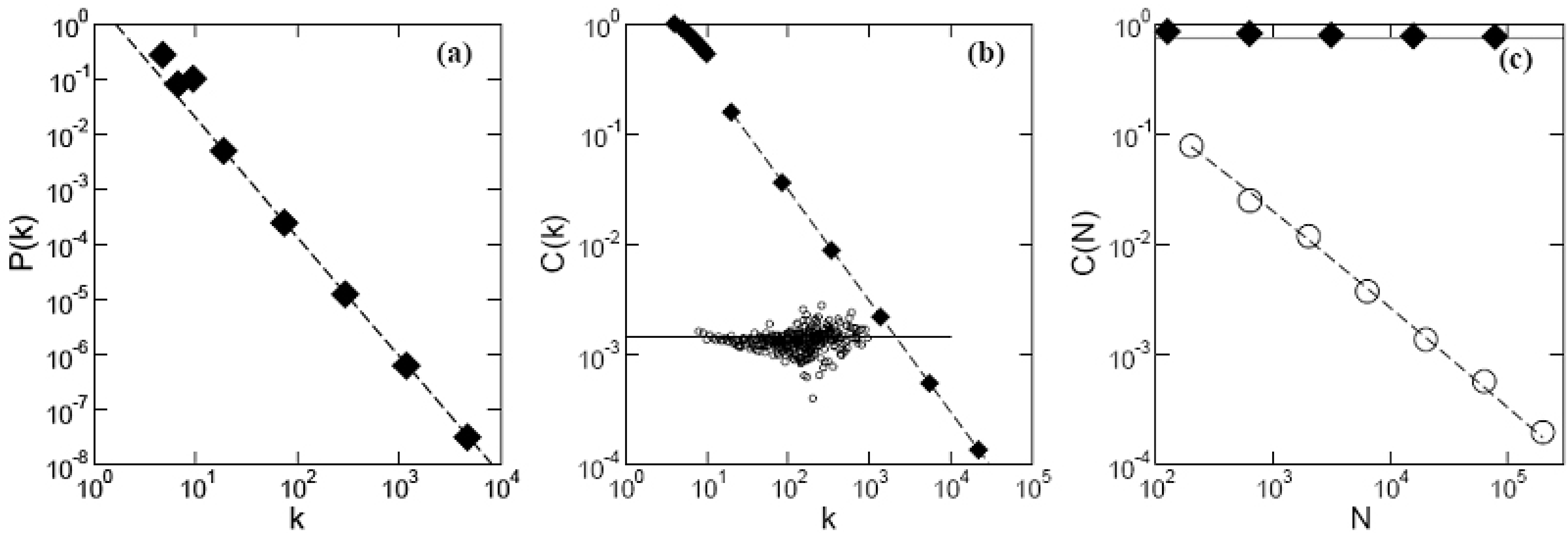


FIGURE 4. Scaling properties of the hierarchical model shown in Fig. 3 ($N = 5^7$). **(a)** The numerically determined degree distribution. The assymptotic scaling, with slope $\gamma = 1 + \ln 5 / \ln 4$, is shown as a dashed line. **(b)** The $C(k)$ curve for the model, demonstrating that it follows Eq. (11). The open circles show $C(k)$ for a scale-free model [25] of the same size, illustrating that it does not have a hierarchical architecture. **(c)** The dependence of the clustering coefficient, C, on the size of the network N. While for the hierarchical model C is independent of N (diamond), for the scale-free model $C(N)$ decreases rapidly (circle). After [50].

Real-World Hierarchical Networks

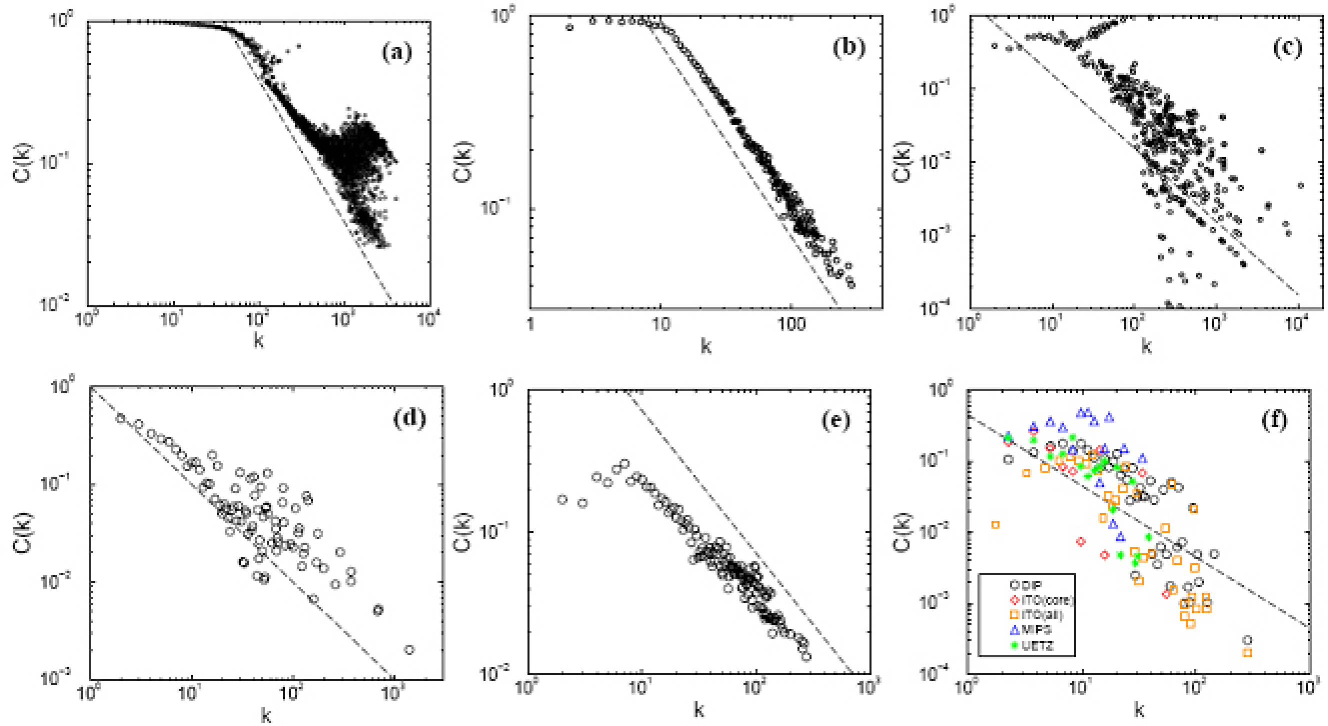


FIGURE 5. The scaling of $C(k)$ with k for four large networks: **(a)** Actor network, two actors being connected if they acted in the same movie according to the `www.IMDB.com` database. **(b)** The semantic web, connecting two English words if they are listed as synonyms in the Merriam Webster dictionary [19]. **(c)** The World Wide Web, based on the data collected in Ref. [11]. **(d)** Internet at the Autonomous System level, each node representing a domain, connected if there is a communication link between them. **(e)** The metabolic networks of 43 organisms with their averaged $C(k)$ curves. **(f)** The protein-protein physical interaction networks using four different databases [56, 57, 58, 59]. The dashed line in each figure has slope -1, following Eq. (11). After [50, 60, 61].

Conclusion

- Complex networks whose number of vertices is known in advance and fixed can be described by random network models
- Expandable networks that have preferential attachment follow a power law and can be described by scale-free network models
- In hierarchical networks, the clustering coefficient follows a scaling law

Comments & Questions