S

Scale-Free Nature of Social Networks

Piotr Fronczak
Faculty of Physics, Warsaw University of Technology, Warsaw, Poland

Synonyms

Complex networks; Network models; Scale-free distributions; Universal scaling

Glossary

Degree	The degree of a node in a network is the number of edges or connections to that node.
ER graph	The network model in which edges are set between nodes with equal probabilities.
Fat-tailed distributions	Have tails that decay more slowly than exponentially. All power-law distributions are fat tailed, but not all fat-tailed distributions are power laws (e.g., the lognormal distribution is fat tailed but is not a power-law distribution).
Node degree distribution	The distribution function *P(k)* that gives the probability that a node selected at random has exactly *k* edges.
Power-law distribution	Has a probability function of the form $P(x) \sim x^{-a}$.
Scale-freeness	Feature of objects or laws that does not change if length scale is multiplied by a common factor; also known as scale invariance.
SF network	The network with power-law distribution of node degrees.

Definition

The notion of scale-freeness and its prevalence in both natural and artificial networks have recently attracted much attention. In physics and mathematics, scale-freeness (or more formally – scale invariance) is a feature of objects or laws that does not change if length scale is multiplied by a common factor. The term gained large popularity in 1999 when Barabasi and Albert used it as a descriptor of networks in which node degrees (vertex connectivity) follow a power-law distribution (Barabasi and Albert 1999). Since most large complex networks are characterized by the distributions which at least partially are reminiscent of power functions, the term "scale-free," applied to networks, lost its formal meaning and nowadays is widely used (albeit erroneously) to describe the network with fat-tailed node degree distribution.

The overwhelming number of studies conducted in the last decade made it clear that the scale-free network topology can have a strong

R. Alhajj, J. Rokne (eds.), *Encyclopedia of Social Network Analysis and Mining*,
DOI 10.1007/978-1-4614-7163-9_248-1

impact on the dynamical processes taking place on these networks such as opinion formation (Aleksiejuk et al. 2002), diffusion of information (Cohen et al. 2000), and epidemic spreading (Pastor-Satorras and Vespignani 2001). Nowadays, the recently acquired knowledge about the network structure revolutionizes not only many fields of science, like biology, computer science, and economics, but also the society and its perception of the ubiquitous networks.

Introduction

Scale-freeness is the property which is fascinating especially for physicists, since most phenomena studied by physicists are not scale invariant. Among seminal exceptions are phase transitions in thermodynamic systems which are associated with the emergence of power-law distributions of certain quantities (Yeomans 2002). Similarly, the phenomenon known as self-organized criticality (a property of dynamical systems which have a critical point as an attractor) displays the spatial and/or temporal scale-free nature of the critical point of a phase transition, but without the need to tune control parameters to precise values (Bak 1996).

In mathematics, scale invariance is an exact form of self-similarity where at any magnification there is a smaller piece of the object that is similar to the whole. Self-similarity is a typical property of fractals.

A common aspect of both, phase transitions and self-similar fractals, is universality, i.e., the observation that there are properties for a large class of different systems that are independent of the dynamical details of the particular system.

These reasons (universality and criticality) explain the excitement of scientists, when the power-law character of node degree distribution has been observed in drastically increasing number of real networks. The promise of the discovery of the universal character of surrounding us social, technological, and natural networks made the notion of scale-freeness frequently misused. Nevertheless, it is a notion that has clearly taken root with today's society effectively guiding the communicative patterns of different scientific communities. In the following paragraphs, we will use this notion in its less formal meaning as a shortcut of the networks with fat-tailed (or almost power-law) distribution of node degrees.

Despite the pure mathematical differences, the properties of idealized (scale-free) and realistic (almost scale-free) networks have the same implications for real-world applications.

Key Points

To understand the scale-free architecture of the networks, it is useful to contrast it with the other network model which dominated the network research for decades, namely, the model of network developed by Erdos and Renyi in 1959 (ER graph) (Erdos and Renyi 1960). The importance of the ER graph for modeling real-world networks is currently diminished; however, it is still a fundamental model in the random network theory. In the following we will briefly introduce ER graphs and emphasize differences between them and SF networks. We will present the methods of detection of the scale-free character of the node degree distribution in networks. We will discuss the most popular model in which the growing network evolves into scale-free state. Finally, we will discuss the vulnerability of SF networks to epidemics and intentional attacks and their extreme tolerance on random failures.

Historical Background

Power-law distributions in nature and society were already known in the nineteenth century. Italian economist, Vilfredo Pareto, in 1897, was the first to discover that the distribution of income in society follows the power law (Barabasi 2002). In 1925, George Udny Yule proposed a stochastic process (later called the Yule process but is now better known as preferential attachment – see the next section) that leads to a distribution with a power-law tail – in this case, the distribution of species and genera (Yule 1925). In 1965, Derek John de Solla Price demonstrated a power-law distribution of links in a network of scientists linked by citation (Price 1965). Although D. Price was a physicist, his discovery was totally ignored in physical sciences. In physics, the lattices and random networks like ER graph were the main objects of study until late 1990s, when Barabasi and others rediscovered the importance of SF networks in technology, nature, and society.

Properties of Scale-Free Networks

Two Opposing Models of Random Networks

The definition of ER graph is simple: in a graph with N nodes, each possible pair of distinct nodes is connected with an edge with probability p. In that model the majority of nodes have a degree that is close to the average degree of the whole network, and this average has small variance (the number of nodes with a given degree decays exponentially fast away from the mean degree). In Fig. 1a, we show a typical representative of this model. As one can see, the sizes of all nodes (which reflect node degrees) are similar. For large N and infinitesimal p (i.e., for large and sparse networks), the node degree distribution follows a Poisson law

$$P(k) = e^{-pN}\frac{(pN)^k}{k!},$$

where k is a node degree and the average node degree $\langle k \rangle = pN$ (Newman et al. 2002). The characteristic bell shape of *P(k)* around the average node degree is visible in Fig. 2a.

As we stated previously, recent studies show that most large complex networks are characterized by a connectivity distribution different to a Poisson distribution (among the exceptions are train networks or electrical power grids). For example, the World Wide Web, Internet, e-mail, and collaboration networks have a degree distribution that follows (at least in some range) a power-law relationship defined by

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

where γ, called *scale-free exponent*, ranges usually between 2 and 3 in real networks. Such networks have a very uneven distribution of connections. There are many nodes with only a few links and a few nodes with a large number of links. The difference between this type of network and a Poisson-like one is clearly visible in Fig. 1b, where some nodes act as "highly connected hubs" while the most of them have only one connection. The fat tail of this distribution, shown in Fig. 2b, is an evidence of an extreme heterogeneity of connections in the network.

Scale-Freeness of Networks with Power-Law Distribution of Node Degrees

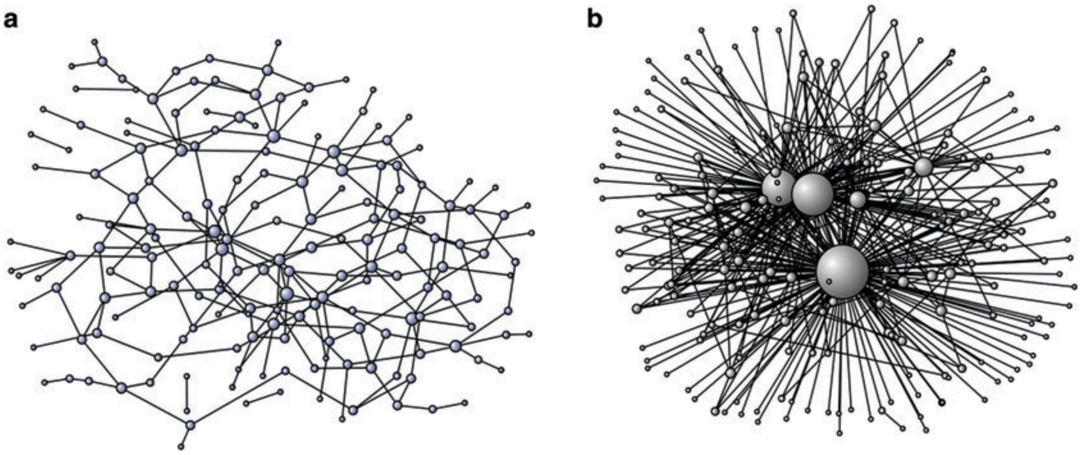
Why the networks with the power-law distributions of node degrees are called scale-free? It is because a power-law distribution is scale invariant. If we rescale a measure of connectivity (e.g., counting how many tens of connections a node has, instead of counting all its connections), the connectivity distribution $P(10\,k)$ will be still proportional to the original *P(k)*. Mathematically, multiplying degree k by a constant c, the distribution remains the same and only scales the function: $P(ck) = c^{-\gamma}P(k)$, where $P(k) = ck^{-\gamma}$. To show that power-law distribution is the only one, which fulfills this condition, we take the logarithm of its both sides:

$$\ln P(ck) = -\gamma \ln c + \ln P(k)$$

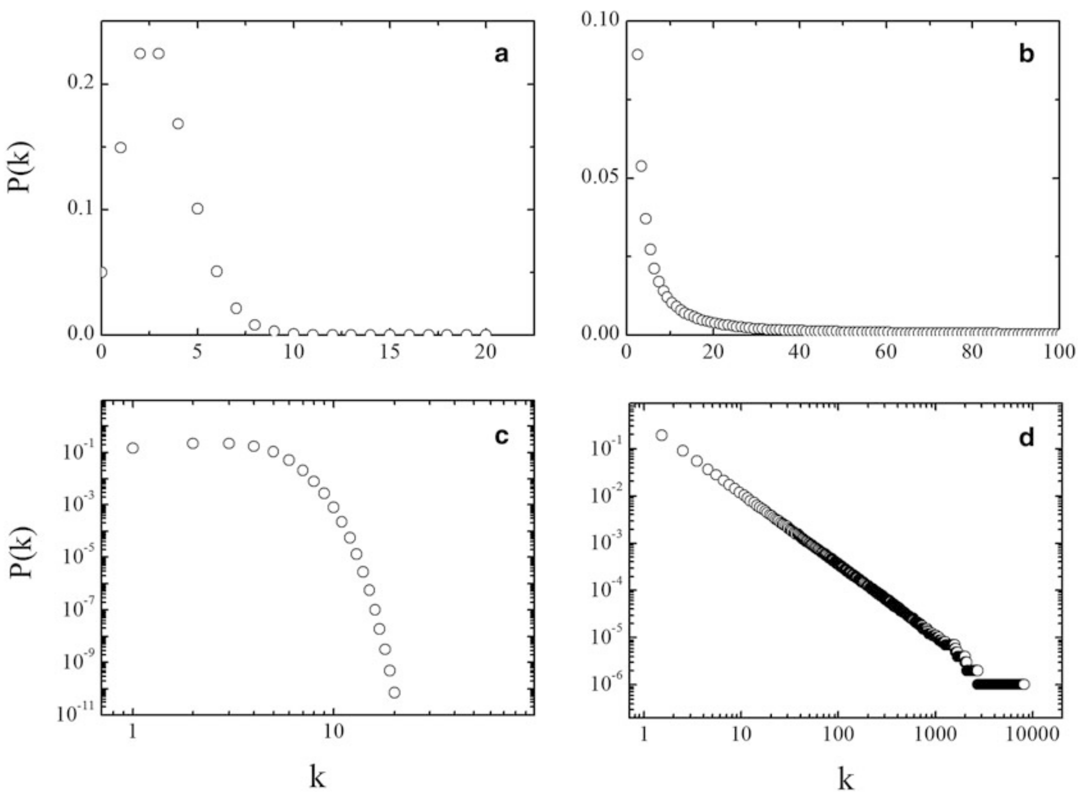
Now, we introduce a new function $R(k)$ defined as $R(\ln k) = P(k)$. This gives

$$\ln R(\ln ck) = -\gamma \ln c + \ln R(\ln k)$$

and, after rearrangement,



Scale-Free Nature of Social Networks, Fig. 1 Two realizations of a ER graph (**a**) and SF network (**b**), both with the same number of nodes and edges. Size of the nodes is proportional to their degrees



Scale-Free Nature of Social Networks, Fig. 2 Node degree distributions of ER graph (*left column*) and SF network (*right column*) in normal (*top row*) and double logarithmic (*bottom row*) scale

$$\frac{\ln R(\ln c + \ln k) - \ln R(\ln k)}{\ln c} = -\gamma.$$

In the limit $\ln c \to 0$, the left side becomes a derivative

$$\frac{d\ln R(\ln k)}{d\ln k} = -\gamma.$$

Since the right side is constant, integrating the equation gives

$$\ln R(\ln k) = -\gamma \ln k + \text{const},$$

and finally

$$R(\ln k) = P(k) = \text{const} \cdot k^{-\gamma}.$$

An important difference between fat-tailed and Poisson-like distributions is that moments of the former (i.e., mean $\langle k \rangle$ and variance $\delta^2(k)$) poorly characterize the distribution (in fact, they are undefined for certain power-law distributions). The moments μ_m of order m are defined as follows:

$$\mu_m = \sum_{k=0}^{k_{\max}} k^m P(k).$$

From the definition, in infinite networks (i.e., when $k_{\max} \to \infty$), all higher moments of order $m \geq \gamma - 1$ of the power-law distribution diverge. Since a mean and a variance are the moments of the first and the second order, respectively, a variance is infinite for γ in the range of typical real-world networks ($2 \leq \gamma \leq 3$):

$$\delta^2 = \mu_2 \sim \sum_{k=0}^{\infty} k^2 k^{-\gamma} = \infty, \quad \text{for } \gamma \leq 3.$$

Although the real networks are finite, the variance can be still several orders larger than the mean. Since a variance describes the error of measured mean node degree, its enormously large value questions the quality of the measurement, and assigning the scale (related to the mean degree) to the network is a misuse.

Plotting Scale-Free Distributions

Since many fat-tailed distributions look similarly as in Fig. 2b (e.g., log-normal or stretched exponential distributions), to better expose the power-law nature of the node degree distribution, one usually plots the data on a double logarithmic scale. In that case the power law transforms into a straight line with a slope of $-\gamma$ (see Fig. 2d and compare it with Poisson distribution shown in Fig. 2c), as follows:

$$\begin{aligned} P(k) &= a \cdot k^{-\gamma} \\ \ln P(k) &= \ln\left(a \cdot k^{-\gamma}\right) \\ \ln P(k) &= \ln a + \ln\left(k^{-\gamma}\right) \\ \ln P(k) &= \ln a - \gamma \ln k \\ Y &= A - \gamma X \end{aligned}$$

where X and Y are transformed variables and A is a transformed constant.

In practice, measuring a slope directly from Fig. 2d is usually very erroneous, due to the poor statistics at the tail of the distribution. Direct histograms are almost always noisy in this region. The solution is to construct a histogram in which the bin sizes increase exponentially with degree. The number of samples in each bin is then divided by the width of the bin to normalize the measurement. Plotting histogram in a logarithmic degree scale, one obtains the even widths of the bins.

An even more discriminating method to verify potential power-law character of the node degree distribution is to plot the complementary cumulative distribution function

$$P_C(k) = \int_k^{\infty} P(k)dk \sim k^{-(\gamma+1)},$$

which is the probability that the degree of a randomly chosen node is greater than or equal to k. Such a plot has the advantage that all the original data are represented. When we make a conventional histogram by binning, any differences between the values of data points that fall in the same bin are lost. The cumulative distribution function does not suffer from this problem. The cumulative distribution also reduces the noise in the tail, which is clearly illustrated in Fig. 3.

Seminal Model of Preferential Attachment

Soon after the discovery of the scale-free structure of the World Wide Web, it has been realized that many other real networks also show power-law distribution of node degrees. This feature has been observed in the Internet, communication, and transportation networks (Albert et al. 1999; Guimera et al. 2005), as well as in many social networks, such as networks of scientific citations (Redner 1998), e-mail networks (Ebel et al. 2002), or even sexual contact networks (Liljeros et al. 2001). The initial surprise of omnipresence of SF networks quickly turned into a question: Why so many networks have the same scale-free character of connections? When a feature appears in many systems that do not have an obvious connection to each other, you should suspect that there is a common causal principle, which can be described in the most general terms, without reference to the details of this or any other system. Is a scale invariance of complex networks a result of some universal rules that govern the dynamics of these systems?

Although there are many different processes which can give rise to the same power-law structure of complex networks, the one deserves particular attention at least for the two reasons. Firstly, its universal character allows to adapt the process to many social but also technological and natural networks. Secondly, it has been independently rediscovered several times in different fields and ages. The process is currently known as Matthew effect, Yule process, Dulbecco's law, Rich gets richer, or preferential attachment (Barabasi and Albert 1999). Since it is used so widely across domains, the claim about its universality is reasonable.

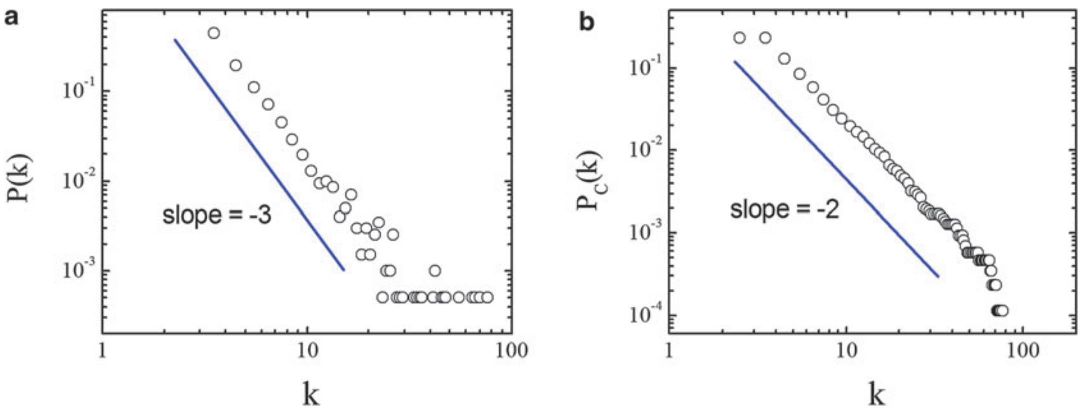
The process, adopted to networks, comprises of two complementary mechanisms: network growth and preferential rules of joining nodes. Barabasi and Albert, who introduced the process to the modern science of complex networks, stated that real networks are not formed as a result of purely random process, in which completely randomly selected nodes are connected by the edges. Most of the social and technological networks grow and change over time – they evolve. In networks, the newly added nodes prefer to create connections with such ones that already have a lot of other connections. The mechanisms underlying this preference can be different. For example, new actors are more likely to play supporting roles in films with established stars, than in those where there are only other unknown actors. Thus, the more famous you are, the more probably that you will attract new connections. The same principle seems to govern the structure of citation network. Preferential attachment corresponds to the feature that a publication with a large number of citations continues to be well cited in the future merely by virtue of being well-cited now. In the network of acquaintances, my friends introduce me to their friends. The more friends I have, the more recognized I am and the more chances to meet new people I have. In WWW, the more pages linked to a web page, the more Internet users visit that site and the greater the likelihood that they will place a link to this page on their own website.

The algorithm of the discussed process consists of two steps:

1. Starting with a small number m_0 of nodes, at every time step, we add a new node with $m \leq m_0$ edges that link the new node to m different nodes already present in the system.
2. When choosing the nodes to which the new node connects, we assume that the probability $\prod$ that a new node will be connected to node i depends on the degree k_i of node i, such that

$$(k_i) = \frac{k_i}{\sum_j k_j}.$$

After t time steps, this algorithm results in a network with $N = t + m_0$ nodes and mt edges. In Fig. 4 first steps of the network evolution have been shown. Already after several steps, the hubs in the network become clearly visible.



Scale-Free Nature of Social Networks, Fig. 3 Node degree distribution of SF network (**a**) and its cumulative distribution (**b**)

Mathematical derivations show that the node degree distribution of the network evolves into a scale-free one with the scale-free exponent $\gamma = 3$ independently of m, the only parameter in the model.

One has to keep in mind that the presented model does not share all properties observed in the real-world networks, e.g., it is less clustered. Soon, after this model was introduced, a large number of similar models, all based on some type of connecting preference, emerged, all leading to a power-law distribution of node degrees but also demonstrating a better agreement with real networks with reference to other network metrics.

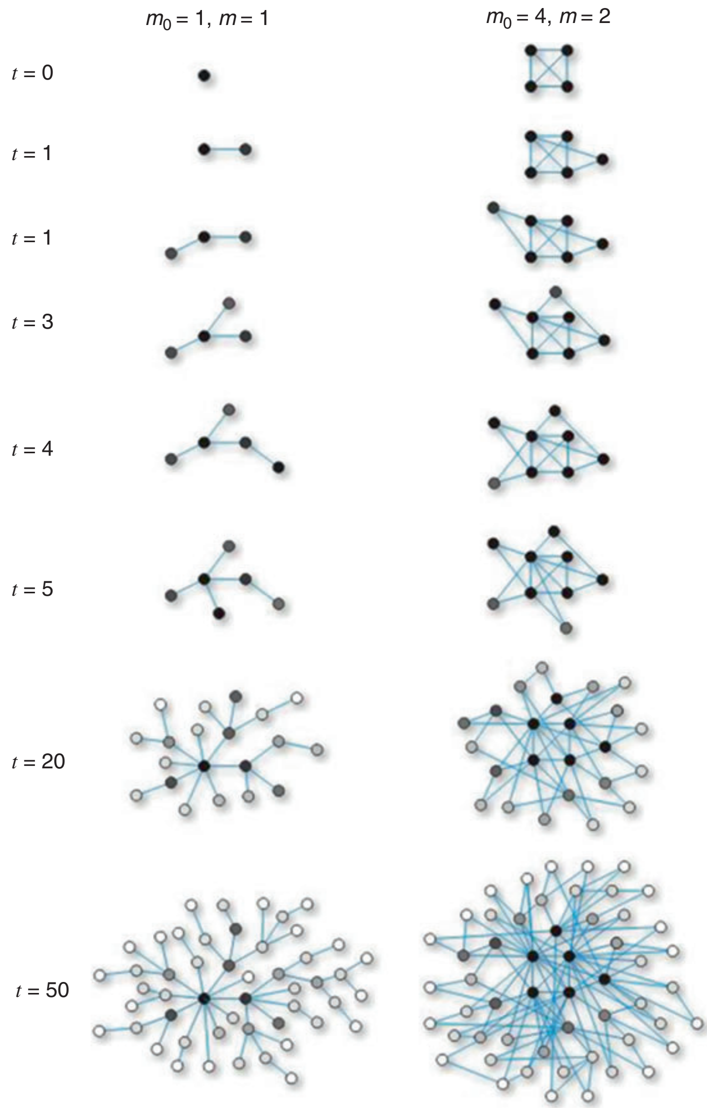
Preferential attachment is not the only possible explanation for the formation of scale-free structure of connections in complex networks. Among others there are rewiring processes (Aiello et al. 2002), optimization-based models (Valverde et al. 2002), and also static constructions (Park and Newman 2004; Goh et al. 2001).

Resilience and Vulnerability of SF Networks to Failures and Attacks

It has appeared recently that scale-free structure of complex networks has an important influence on their resilience to failures and attacks. In particular, SF networks seem much more robust than ER graphs in case of failures (modeled by a random removal of nodes or links) (Cohen et al. 2000), while they are more sensitive to attacks (modeled by the targeted removal of selected nodes or links) (Cohen et al. 2001). By the resilience we understood that despite removed nodes the main part of the network (so called giant component) is still interconnected (i.e., any two nodes in that part are connected to each other by paths). If the node elimination proceeds, then at some critical moment, the network breaks apart into small disconnected parts. The moment when this dramatic breakdown occurs strongly depends on the network structure as well as on the method of node's elimination (random or targeted). In the random removal case, the critical moment of destruction occurs much earlier in ER graphs, in opposite to SF networks (see Fig. 5). It means that SF network is much more resilient to accidental damages. However, in case of intentional attack, when the nodes of the network are removed in decreasing order of their degree, SF network appears to be much more vulnerable than ER graph (since the removal of the hubs results in the largest possible damage; see Fig. 6). This vulnerability of SF networks to intentional attacks has been described as their Achilles' heel.

Epidemic Spreading in SF Networks

Scale-free nature of social networks has a great implication for understanding the spread of information, diseases, opinions, and innovations in

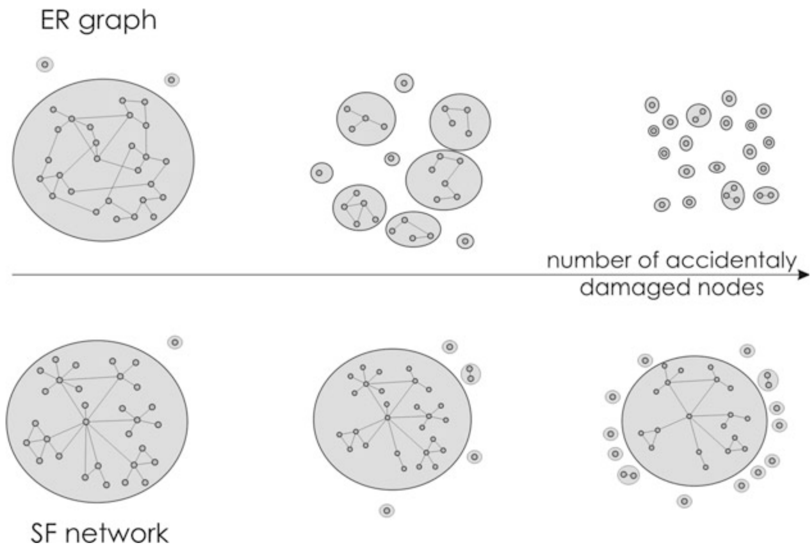


Scale-Free Nature of Social Networks, Fig. 4 Example of realization of two different growing networks in preferential attachment model. The *colors* of the nodes represent their age

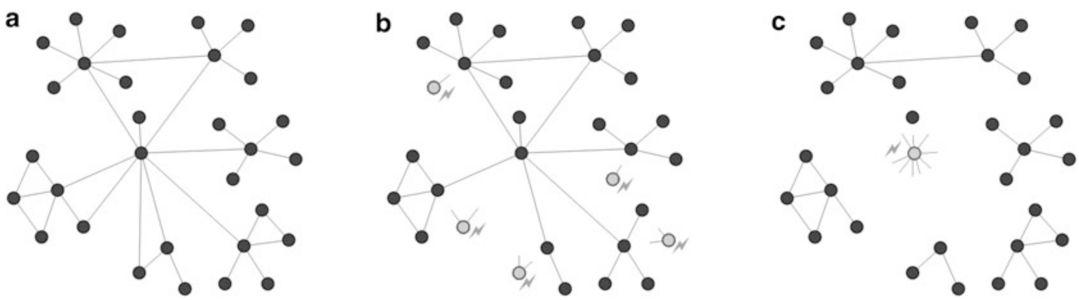
society. Standard epidemiological models usually consider networks with the well-defined average node degree, such as ER graphs. In those networks the models predict a critical threshold for the propagation of a contagion throughout a population. This epidemic threshold is determined by the virulence of the infection. In other words, if the spreading rate is larger than the threshold, the infection spreads and becomes persistent. Below the threshold, the infection dies out.

It turns out that in SF networks the above statement is no longer correct. In 2001, Pastor-Satorras and Vespignani found that in that case the threshold is zero (Pastor-Satorras and Vespignani 2001). It means that all viruses, even those that are weakly contagious, will spread and persist in the system. The main reason is that the presence of hub nodes can facilitate epidemic spreading due to the large numbers of neighbors. Infected hub passes the infection to numerous other nodes, faster than the typical node recovers.

Specifically, in SF network, the traditional random immunization could easily fail because nearly everyone would have to be treated to ensure that



Scale-Free Nature of Social Networks, Fig. 5 ER graphs break apart into small disconnected parts much faster than SF networks if the nodes are removed accidentally



Scale-Free Nature of Social Networks, Fig. 6 Random and targeted elimination of nodes. Original SF network (**a**), the network with randomly damaged nodes (**b**), and the network with the damaged hub (**c**)

the hubs were not missed. New immunization strategies have to be developed to recover the epidemic threshold. It turns out that one of the most efficient approaches is to selectively immunize hub nodes. Such a strategy is known as targeted immunization (Pastor-Satorras and Vespignani 2002).

Key Applications

The paradigm of SF networks has numerous applications to problems in different areas including epidemiology (Colizza et al. 2007), human mobility (Gonzalez et al. 2008), social networks (Huberman et al. 2009), life sciences (Guimera and Nunes Amaral 2005), information flow (Helbing et al. 2006), and ecology (Montoya et al. 2006).

Future Directions

The structure, topological properties, and appropriate measures were the main research topics in complex network domain in recent years. Currently, dynamical processes taking place in the networks are quite intensively studied. It is believed that further understanding of dynamics on complex networks is the general direction of the field. There is a continuous shift from studies of networks in general and features that are common to most of them to more application-driven studies of increasingly narrow classes of networks. After a decade of mostly descriptive studies and just potential applications, there is a final need to transfer an acquired knowledge into concrete market applications. Complex networks research society should provide the manageable solutions to global challenges, like vaccination campaigns against serious viruses, risk reduction of financial crises, and preventing cascading bankruptcies among interlinked economies.

Cross-References

▶ Exponential Random Graph Models
▶ Models of Social Networks
▶ Network Models

References

Aiello W, Chung F, Lu L (2002) Random evolution of massive graphs. In: Pardalos PM, Abello J, Resende MGC (eds) Handbook of massive data sets. Kluwer, Dordrecht, pp 97–122
Albert R, Jeong H, Barabasi A-L (1999) Diameter of the World Wide Web. Nature 401:130–131
Aleksiejuk A, Holyst JA, Stauffer D (2002) Ferromagnetic phase transition in Barabasi-Albert networks. Phys A 310:260–266
Bak P (1996) How nature works: the science of self-organized criticality. Copernicus, New York
Barabasi A-L (2002) Linked: the new science of networks. Perseus Books Group, Reading
Barabasi A-L, Albert R (1999) Emergence of scaling in random networks. Science 286:509–512
Cohen R, Erez K, ben Avraham D, Havlin S (2000) Resilience of the Internet to random breakdowns. Phys Rev Lett 85:4626
Cohen R, Erez K, ben Avraham D, Havlin S (2001) Breakdown of the Internet under intentional attack. Phys Rev Lett 86:3682
Colizza V, Barrat A, Barthelemy M, Valleron AJ, Vespignani A (2007) Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. PLoS Med 4:e13
Ebel H, Mielsch LI, Bornholdt S (2002) Scale-free topology of e-mail networks. Phys Rev E 66:035103
Erdos P, Renyi A (1960) On the evolution of random graphs. Pub Math Inst Hung Acad Sci 5:17–61
Goh K-I, Kahng B, Kim D (2001) Universal behaviour of load distribution in scale-free networks. Phys Rev Lett 87:278701
Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding individual human mobility patterns. Nature 453:779–782
Guimera R, Nunes Amaral LA (2005) Functional cartography of complex metabolic networks. Nature 433(7028):895–900
Guimerà R, Mossa S, Turtschi A, Amaral LAN (2005) The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles. Proc Natl Acad Sci 102:7794–7799
Helbing D, Armbruster D, Mikhailov AS, Lefeber E (2006) Information and material flows in complex networks. Physica A 363:11–16
Huberman BA, Romero DM, Wu F (2009) Social networks that matter: twitter under the microscope. First Monday 14:1–5
Liljeros F, Edling CR, Amaral LAN, Stanley HE, Aberg Y (2001) The web of human sexual contacts. Nature 411:907–908
Montoya J, Pimm SL, Solé RV (2006) Ecological networks and their fragility. Nature 442(7100):259–264
Newman ME, Watts DJ, Strogatz SH (2002) Random graph models of social networks. Proc Natl Acad Sci U S A 99(Suppl 1):2566–2572
Park J, Newman MEJ (2004) Statistical mechanics of networks. Phys Rev E 70:066117
Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. Phys Rev Lett 86:3200
Pastor-Satorras R, Vespignani A (2002) Immunization of complex networks. Phys Rev E 65:036104
Price DJ d S (1965) Networks of scientific papers. *Science* 149:510–515
Redner S (1998) How popular is your paper? An empirical study of the citation distribution. Eur Phys J B 4:131–134
Valverde S, Ferrer Cancho R, Solé RV (2002) Scale-free networks from optimal design. Europhys Lett 60:512
Yeomans JM (2002) Statistical mechanics of phase transitions. Oxford University Press, New York
Yule UG (1925) A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. Philosophical transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character, vol 213. The Royal Society, pp 21–87