Emergence of Scaling in Random Networks
Albert-László Barabási, et al.
Science 286, 509 (1999);
DOI: 10.1126/science.286.5439.509

This copy is for your personal, non-commercial use only.

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by clicking here.

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines here.

The following resources related to this article are available online at www.sciencemag.org (this information is current as of April 21, 2010):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/286/5439/509

A list of selected additional articles on the Science Web sites related to this article can be found at:
http://www.sciencemag.org/cgi/content/full/286/5439/509#related-content

This article cites 11 articles, 3 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/286/5439/509#otherarticles

This article has been cited by 3777 article(s) on the ISI Web of Science.

This article has been cited by 86 articles hosted by HighWire Press; see:
http://www.sciencemag.org/cgi/content/full/286/5439/509#otherarticles

This article appears in the following subject collections:
Physics
http://www.sciencemag.org/cgi/collection/physics
Emergence of Scaling in Random Networks

Albert-László Barabási* and Réka Albert

Systems as diverse as genetic networks or the World Wide Web are best described as networks with complex topology. A common property of many large networks is that the vertex connectivities follow a scale-free power-law distribution. This feature was found to be a consequence of two generic mechanisms: (i) networks expand continuously by the addition of new vertices, and (ii) new vertices attach preferentially to sites that are already well connected. A model based on these two ingredients reproduces the observed stationary scale-free distributions, which indicates that the development of large networks is governed by robust self-organizing phenomena that go beyond the particulars of the individual systems.

The inability of contemporary science to describe systems composed of nonidentical elements that have diverse and nonlocal interac-

Department of Physics, University of Notre Dame, Notre Dame, IN 46556, USA.

*To whom correspondence should be addressed. E-mail: alb@nd.edu
ing systems form a huge genetic network whose vertices are proteins and genes, the chemical interactions between them representing edges (2). At a different organization-al level, a large network is formed by the nervous system, whose vertices are the nerve cells, connected by axons (3). But equally complex networks occur in social science, where vertices are individuals or organizations and the edges are the social interactions between them (4), or in the World Wide Web (WWW), whose vertices are HTML documents connected by links pointing from one page to another (5, 6). Because of their large size and the complexity of their interactions, the topology of these networks is largely unknown.

Traditionally, networks of complex topology have been described with the random graph theory of Erdős and Rényi (ER) (7), but in the absence of data on large networks, the predictions of the ER theory were rarely tested in the real world. However, driven by the computerization of data acquisition, such topological information is increasingly available, raising the possibility of understanding the dynamical and topological stability of large networks.

Here we report on the existence of a high degree of self-organization characterizing the large-scale properties of complex networks. Exploring several large databases describing the topology of large networks that span fields as diverse as the WWW or citation patterns in science, we show that, independent of the system and the identity of its constituents, the probability $P(k)$ that a vertex in the network interacts with $k$ other vertices decays as a power law, following $P(k) \sim k^{-\gamma}$. This result indicates that large networks self-organize into a scale-free state, a feature unpredicted by all existing random network models. To explain the origin of this scale invariance, we show that existing network models fail to incorporate growth and preferential attachment, two key features of real networks. Using a model incorporating these two ingredients, we show that they are responsible for the power-law scaling observed in real networks. Finally, we argue that these ingredients play an easily identifiable and important role in the formation of many complex systems, which implies that our results are relevant to a large class of networks observed in nature.

Although there are many systems that form complex networks, detailed topological data is available for only a few. The collaboration graph of movie actors represents a well-documented example of a social network. Each actor is represented by a vertex, two actors being connected if they were cast together in the same movie. The probability that an actor has $k$ links (characterizing his or her popularity) has a power-law tail for large $k$, following $P(k) \sim k^{-3.7}$ (Fig. 1A). A more complex network with over 800 million vertices (8) is the WWW, where a vertex is a document and the edges are the links pointing from one document to another. The topology of this graph determines the Web’s connectivity and, consequently, our effectiveness in locating information on the WWW (5). Information about $P(k)$ can be obtained using robots (6), indicating that the probability that $k$ documents point to a certain Web page follows a power law, with $\gamma_{WWW} = 2.1 \pm 0.1$ (Fig. 1B) (9). A network whose topology reflects the historical patterns of urban and industrial development is the electrical power grid of the western United States, the vertices being generators, transformers, and substations and the edges being to the high-voltage transmission lines between them (10). Because of the relatively modest size of the network, containing only 4941 vertices, the scaling region is less prominent but is nevertheless approximated by a power law with an exponent $\gamma_{power} = 4$ (Fig. 1C). Finally, a rather large complex network is formed by the citation patterns of the scientific publications, the vertices being papers published in refereed journals and the edges being links to the articles cited in a paper. Recently Redner (11) has shown that the probability that a paper is cited $k$ times (representing the connectivity of a paper within the network) follows a power law with exponent $\gamma_{cite} = 3$.

The above examples (12) demonstrate that many large random networks share the common feature that the distribution of their local connectivity is free of scale, following a power law for large $k$ with an exponent $\gamma$ between 2.1 and 4, which is unexpected within the framework of the existing network models. The random graph model of ER (7) assumes that we start with $N$ vertices and connect each pair of vertices with probability $p$. In the model, the probability that a vertex has $k$ edges follows a Poisson distribution $P(k) = e^{-\lambda} \lambda^k / k!$, where

$$\lambda = N \left( \frac{N-1}{k} \right) p \left( 1 - p \right)^{N-1-k}$$

In the small-world model recently introduced by Watts and Strogatz (WS) (10), $N$ vertices form a one-dimensional lattice, each vertex being connected to its two nearest and next-nearest neighbors. With probability $p$, each edge is reconnected to a vertex chosen at random. The long-range connections generated by this process decrease the distance between the vertices, leading to a small-world phenomenon (13), often referred to as six degrees of separation (14). For $p = 0$, the probability distribution of the connectivities is $P(k) = 6(k - z)$, where $z$ is the coordination number in the lattice; whereas for finite $p$, $P(k)$ still peaks around $z$, but it gets broader (15). A common feature of the ER and WS models is that the probability of finding a highly connected vertex (that is, a large $k$) decreases exponentially with $k$; thus, vertices with large connectivity are practically absent. In contrast, the power-law tail characterizing $P(k)$ for the networks studied indicates that highly connected (large $k$) vertices have a large chance of occurring, dominating the connectivity.

There are two generic aspects of real networks that are not incorporated in these models. First, both models assume that we start with a fixed number ($N$) of vertices that are then randomly connected (ER model), or reconnected (WS model), without modifying $N$. In contrast, most real world networks are open and they form by the continuous addition of new vertices to the system, thus the number of vertices $N$ increases throughout the lifetime of the network. For example, the actor network grows by the addition of new actors to the system, the WWW grows exponentially over time by the addition of new Web pages (8), and the research literature constantly grows by the publication of new papers. Consequently, a common feature of

![Fig. 1. The distribution function of connectivities for various large networks. (A) Actor collaboration graph with $N = 212,250$ vertices and average connectivity ($k$) = 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) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.](image)
these systems is that the network continuously expands by the addition of new vertices that are connected to the vertices already present in the system.

Second, the random network models assume that the probability that two vertices are connected is random and uniform. In contrast, most real networks exhibit preferential connectivity. For example, a new actor will be cast with an established one is much more likely than that the new actor will be cast with other less-known actors. Similarly, a newly created Web page will be more likely to include links to well-known popular documents with already-high connectivity, and a new manuscript is more likely to cite a well-known and thus much-cited paper than its less-cited and consequently less-known peer. These examples indicate that the probability with which a new vertex connects to the existing vertices is not uniform; there is a higher probability that it will be linked to a vertex that already has a large number of connections.

We next show that a model based on these two ingredients naturally leads to the observed scale-invariant distribution. To incorporate the growing character of the network, starting with a small number \( m_0 \) of vertices, at every time step we add a new vertex with \( m(t) \leq m_0 \) edges that link the new vertex to \( m \) different vertices already present in the system. To incorporate preferential attachment, we assume that the probability \( \Pi \) that a new vertex will be connected to vertex \( i \) depends on the connectivity \( k_i \) of that vertex, so that \( \Pi(k_i) = k_i / \sum k_i \). After time steps, the model leads to a random network with \( t + m \) vertices and \( m(t) \) edges. This network evolves into a scale-invariant state with the probability that a new vertex is connected with equal probability to any vertex in the system [that is, \( \Pi(k) = C/(m_0 + t + 1) \)]. Such a model (Fig. 2B) leads to \( P(k) \propto \exp(-bk) \), indicating that the absence of preferential attachment eliminates the scale-free feature of the distribution. In model B, we start with \( N \) vertices and no edges. At each time step, we randomly select a vertex, and connect it with probability \( \Pi(k_i) = k_i / \sum k_i \), to vertex \( i \) in the system. Although at early-time the model exhibits power-law scaling, \( P(k) \) is not stationary; because \( N \) is constant and the number of edges increases with time, after \( T = N^2 \) time steps the system reaches a state in which all vertices are connected. The failure of models A and B indicates that both ingredients—growth and preferential attachment—are needed for the development of the stationary power-law distribution observed in Fig. 1.

Because of the preferential attachment, a vertex that acquires more connections than another one will increase its connectivity at a higher rate; thus, an initial difference in the connectivity between two vertices will increase further as the network grows. The rate at which a vertex acquires edges is \( \dot{k} = k / 2t \), which gives \( k\!(t) = m(t)/t^{0.5} \), where \( t \) is the time at which vertex \( i \) was added to the system (see Fig. 2C), a scaling property that can be directly tested once time-resolved data on network connectivity becomes available. Thus older (with smaller \( t \)) vertices increase their connectivity at the expense of the younger (with larger \( t \)) ones, leading over time to some vertices that are highly connected, a “rich-get-richer” phenomenon that can be easily detected in real networks. Furthermore, this property can be used to calculate \( \gamma \) analytically. The probability that a vertex \( i \) has a connectivity smaller than \( k \), \( P(k_i < k) \), can be written as \( P(t > m^2/t^2) \). Assuming that we add the vertices to the system at equal time intervals, we obtain \( P(t > m^2/t^2) = 1 - P(t_i \leq m^2/t^2) = 1 - m^2/t^2(t + m) \). The probability density \( P(k) \) can be obtained from \( P(k) = \delta P(k_i < k)/\delta k \), which over long time periods leads to the stationary solution

\[
P(k) = \frac{2m^2}{k^3}
\]

giving \( \gamma = 3 \), independent of \( m \). Although it reproduces the observed scale-free distribution, the proposed model cannot be expected to account for all aspects of the studied networks. For that, we need to model these systems in more detail. For example, in the model we assumed linear preferential attachment; that is, \( \Pi(k) \propto k \). However, although in general \( \Pi(k) \) could have an arbitrary non-linear form \( \Pi(k) \propto k^n \), simulations indicate that scaling is present only for \( \alpha = 1 \). Furthermore, the exponents obtained for the different networks are scattered between 2.1 and 4. However, it is easy to modify our model to account for exponents different from \( \gamma = 3 \). For example, if we assume that a fraction \( p \) of the links is directed, we obtain \( \gamma(p) = 3 - p \), which is supported by numerical simulations (16). Finally, some networks evolve not only by adding new vertices but by adding (and sometimes removing) connections between established vertices. Although these and other system-specific features could modify the exponent \( \gamma \), our model offers the first successful mechanism accounting for the scale-invariant nature of real networks.

Growth and preferential attachment are mechanisms common to a number of complex systems, including business networks (17, 18), social networks (describing individuals or organizations), transportation networks (19), and so on. Consequently, we expect that the scale-invariant state observed in all systems for which detailed data has been available to us is a generic property of many complex networks, with applicability reaching far beyond the quoted examples. A better description of these systems would help in understanding other complex systems.

---

**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.
as well, for which less topological information is currently available, including such important examples as genetic or signaling networks in biological systems. We often do not think of biological systems as open or growing, because their features are genetically coded. However, possible scale-free features of genetic and signaling networks could reflect the networks’ evolutionary history, dominated by growth and aggregation of different constituents, leading from simple molecules to complex organisms. With the fast advances being made in mapping out genetic networks, answers to these questions might not be too far away. Similar mechanisms could explain the origin of the social and economic disparities governing competitive systems, because the scale-free inhomogeneities are the inevitable consequence of self-organization due to the local decisions made by the individual vertices, based on information that is biased toward the more visible (richer) vertices, irrespective of the nature and origin of this visibility.

**References and Notes**

9. In addition to the distribution of incoming links, the WWW displays a number of other scale-free features characterizing the organization of the Web pages within a domain [B. A. Huberman and L. A. Adamic, Nature 401, 131 (1999)]; the distribution of searches [B. A. Huberman, P. L. T. Piroli, J. E. Pitkow, R. J. Lukose, Science 280, 95 (1998)], or the number of links per Web page (6).
12. We also studied the neural network of the worm *Caenorhabditis elegans* (3, 10) and the benchmark diagram of a computer chip (see http://vlsicad.cs. ucla.edu/~cheese/spd98.html). We found that $P(k)$ for both was consistent with power-law tails, despite the fact that for *C. elegans* the relatively small size of the worm (306 vertices) severely limits the data quality, whereas for the wiring diagram of the chips, vertices with over 200 edges have been eliminated from the database.
16. For most networks, the connectivity $m$ of the newly added vertices is not constant. However, choosing $m$ randomly will not change the exponent $\gamma$ (Y. Tu, personal communication).
18. Preferential attachment was also used to model evolving networks (L. A. N. Amaral and M. Barthélemy, personal communication).
20. We thank D. J. Watts for providing the *C. elegans* and power grid data, B. C. Tjaden for supplying the actor data, H. Jeong for collecting the data on the WWW, and L. A. Amaral for helpful discussions. This work was partially supported by NSF Career Award DMI-9710998.

24 June 1999; accepted 2 September 1999

---

**Osmium Isotope Constraints on Ore Metal Recycling in Subduction Zones**

Brent I. A. McInnes, Jannene S. McBride, Noreen J. Evans, David D. Lambert, Anita S. Andrew

Veined peridotite xenoliths from the mantle beneath the giant Ladolam gold deposit on Lihir Island, Papua New Guinea, are 2 to 800 times more enriched in copper, gold, platinum, and palladium than surrounding depleted arc mantle. Gold ores have osmium isotope compositions similar to those of the underlying subduction-modified mantle peridotite source region, indicating that the primary origin of the metals was the mantle. Because the mantle is relatively depleted in gold, copper, and palladium, tectonic processes that enhance the advective transport and concentration of these fluid soluble metals may be a prerequisite for generating porphyry-epithermal copper-gold deposits.

The tectonic relationship between subduction-related magmatism at convergent margins and porphyry-copper-gold (Cu-Au) ore formation has long been recognized (1). However, the physical and chemical processes that govern Cu-Au metallogeny and the ultimate source(s) of the metals in these ore deposits are poorly understood. The rhenium-osmium (Re-Os) isotopic system (based on the $\beta^{-}$ decay of $^{187}$Re to $^{187}$Os) is a tracer of metallogenic processes at convergent margins because both elements have geochemical properties similar to metals that occur in porphyry ore deposits (2, 3). Because Re is highly concentrated in crustal rocks and Os is concentrated in the mantle (4, 5), this isotopic system is particularly useful for quantifying the flux of ore elements in island arc settings where the two principal reservoirs for metals are subducted crust and mantle wedge peridotite.

Os isotope studies in subduction zones are currently limited because of the rarity of deep-seated suprasubduction samples (rocks overlying a subducted slab). Previous studies (6, 7) have demonstrated that radiogenic Os is introduced into the subarc mantle by hydrous, oxidizing fluids derived during slab dehydration. We report on results from a suprasubduction xenolith locality, the Tubaf seamount in the Lihir island group of the Tabar-Lihir-Tanga-Feni island arc in Papua New Guinea (Fig. 1). This xenolith locality is important for the following reasons: (i) it contains samples that represent a complete section of oceanic lithosphere at an intraoceanic convergent margin, (ii) it is located adjacent to one of the world’s largest and youngest volcano-hosted Au deposits, and (iii) it contains metasomatized mantle peridotite xenoliths with Au-enriched vein minerals that crystallized in the mantle from oxidizing, alkalii- and sulfur-rich hydrous fluids.

During the oceanographic investigation of submarine hydrothermal systems in Papua New Guinea (8), a submarine cinder cone (Tubaf volcano, 1280 m below sea level; 3° 15.25’ S, 152° 32.50’ E) was discovered 14 km southwest of the giant Ladolam gold mine (>40 million oz contained Au) on Lihir Island. Dredge and video-grab sampling of the 1-km-diameter volcanic cone returned 130 ultramafic, mafic, and sedimentary xenoliths. The study of these samples has provided an unprecedented view of the source region of an island arc magmatic system with a propensity to produce giant porphyry-epithermal ore deposits. The xenolith assemblage includes spinel lherzolite, harzburgite, websterite, orthopyroxenite, clinopyroxenite, syenite, serpentinite, gabbro, hornblende gabbro, diabase, basalt, pelagic deep-sea sediment, and shallow-water volcanoclastic sediment as well as coralline and coralgal limestone. These lithologies represent a cross section of the suprasubduction assemblage and can be reassembled into an “ophiolite-type” model of oceanic lithosphere (Fig. 1) (9).