

Discussion



Cite this article: Barabási A-L. 2013 Network science. *Phil Trans R Soc A* 371: 20120375. <http://dx.doi.org/10.1098/rsta.2012.0375>

One contribution of 15 to a Discussion Meeting Issue 'Web science: a new frontier'.

Subject Areas:

systems theory

Keywords:

networks, web infrastructure,
complex systems

Author for correspondence:

Albert-László Barabási
e-mail: a.barabasi@neu.edu

Network science

Albert-László Barabási^{1,2}

¹Department of Physics, Northeastern University, 360 Huntington Avenue, Boston, MA 02115, USA

²Department of Medicine, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA

Professor Barabási's talk described how the tools of network science can help understand the Web's structure, development and weaknesses. The Web is an information network, in which the nodes are documents (at the time of writing over one trillion of them), connected by links. Other well-known network structures include the Internet, a physical network where the nodes are routers and the links are physical connections, and organizations, where the nodes are people and the links represent communications.

As the Web began to grow in the 1990s, it was thought that it most probably had the properties of a *random network*, a type of network characterized mathematically by Erdős & Rényi [1], in which the probability of two nodes being connected is a constant. In such a network, the *degree distribution* is of Poisson form, so that most nodes have a degree (number of connections) that is close to the average. Furthermore, because the Poisson curve falls away from its peak faster than exponentially, it implies that there will be extremely few nodes that are either very highly connected or hardly connected at all.

The work by Albert *et al.* [2] has shown that the World Wide Web does not have this random network structure. In fact, the degree distribution is a *power law*, where the probability of a node having degree k is proportional to $k^{-\gamma}$, where γ is about 2. Such a network has very different properties from a random network because a power law decays much less quickly than a Poisson curve; rather than virtually all nodes having more or less the same degree, we should expect a few nodes to be very highly connected, and the vast majority to have smaller degree than the average. In particular, we cannot pick a 'typical' node; for this reason such networks are called *scale-free* [2]. This is not a unique property of the Web; the Internet is a scale-free network [3], as are many online communities [4,5], as well as networks in the natural world, such as the metabolic network of organisms (where each node is a metabolite and the links are the

reactions between them). Such networks are in effect held together by a small number of highly connected hubs. This scale-free property is an important organizing principle for the Web.

A second important organizing principle is the *small world* property, which says that two nodes are likely to be connected, even in such a very large and sparse scale-free network as the Web, by a relatively short path of nodes—in the case of the Web, the path length is about 19 [2].

The range of scale-free networks is so wide as to raise the question of how such a diverse group of networks could converge on a single type of structure. This led Barabási to describe a third organizing principle about the evolution and growth of the Web through *preferential attachment* [6]. The Web expands by the addition of new documents. New documents are more likely to link to older documents or well-known sites that are highly connected than otherwise. The probability that a new node will connect to a node with k links is proportional to k . A network that grows with preferential attachment will exhibit scale-free properties, as hubs emerge and become larger. Intuitively, one would expect preferential attachment to be a common mechanism, which goes some way to explaining why scale-free networks are so common.

One implication is that the nodes that appeared earlier in the network will be more likely to become hubs, because they have a greater opportunity to increase their degree (degree increases proportionally to the square root of time). But we know from the Web that this is not the only factor, as large hubs, such as Google and Facebook, have appeared relatively late in its history. To accommodate this phenomenon, Barabási introduced the notion of *fitness*, a fourth organizing principle. This is a measure of how likely a link will be made to a node once it has been found, and this, combined with its degree, enables a realistic model to be developed of how its degree grows over time given preferential attachment, and how some nodes' degrees increase more rapidly than others [7,8]. Growth occurs over time, but the rate of growth is controlled by the fitness; this provides competition in networks, and the nodes with a greater fitness will tend to 'win out' and become very highly connected.

One important property of scale-free networks is that they are resilient against random error or decay. If one removes nodes from a network at random, then most types of network will eventually fragment into a set of smaller networks or individual nodes. But a scale-free network remains robust against random decay; it will shrink but not fall apart [9]. If the power-law distribution has a degree exponent of under 3 (as does the Web), then the scale-free network can continue to remain connected indefinitely; the critical point at which the network is expected to fall apart is 1—that is, we need to effectively remove all nodes to destroy such a network. The flip side of this is that, in the event of a targeted attack, where the most connected nodes are deliberately removed first, then the network will be destroyed very quickly. This is the Achilles' heel of a scale-free network.

As a result of studying these networks, Barabási argued that we have seen the emergence of *network science* [10], which overlaps with Web science. Network science is an attempt to understand networks emerging in nature, technology and society using a unified set of tools and principles. Despite apparent differences, many networks emerge and evolve, driven by a fundamental set of laws and mechanisms, and these are the province of network science.

References

1. Erdős P, Rényi A. 1959 On random graphs. *Publicationes Mathematicae* **6**, 290–297.
2. Albert R, Jeong H, Barabási A-L. 1999 Diameter of the world-wide web. *Nature* **401**, 130–131. (doi:10.1038/43601)
3. Faloutsos M, Faloutsos P, Faloutsos C. 1999 On power-law relationships of the Internet topology. In *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (ACM SIGCOMM'99)*, Cambridge, MA, 30 August–3 September 1999. New York, NY: Association for Computing Machinery. (See 1999 *Comput. Commun. Rev.* **29**(4), 251–262.)
4. Ebel H, Mielsch L, Bornholdt S. 2002 Scale-free topology of e-mail networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **66**, 035103. (doi:10.1103/PhysRev.66.035103)

5. Holme P, Edling CR, Liljeros F. 2004 Structure and time evolution of an Internet dating community. *Soc. Netw.* **26**, 155–174. (doi:10.1016/j.socnet.2004.01.007)
6. Barabási A-L, Albert R. 1999 Emergence of scaling in random networks. *Science* **286**, 509–512. (doi:10.1126/science.286.5439.509)
7. Bianconi G, Barabási A-L. 2001 Bose–Einstein condensation in complex networks. *Phys. Rev. Lett.* **86**, 5632–5635. (doi:10.1103/PhysRevLett.86.5632)
8. Bianconi G, Barabási A-L. 2001 Competition and multiscaling in evolving networks. *Europhys. Lett.* **54**, 436–442. (doi:10.1209/epl/i2001-00260-6)
9. Albert R, Jeong H, Barabási A-L. 2000 Error and attack tolerance of complex networks. *Nature* **406**, 378–382. (doi:10.1038/35019019)
10. Barabási A-L. 2002 *Linked: the new science of networks*. Cambridge, MA: Perseus.