

NETWORK SCIENCE

Luck or reason

The concept of preferential attachment is behind the hubs and power laws seen in many networks. New results fuel an old debate about its origin, and beg the question of whether it is based on randomness or optimization.

ALBERT-LÁSZLÓ BARABÁSI

Often a field's most profuse concept is also its most mysterious. Think wavefunctions in quantum mechanics, dark energy in astrophysics and non-coding DNA in genomics. Network science has its own: preferential attachment, which states that the more connected a network node is, the more links it will acquire in the future. The impact of preferential attachment is hard to miss — the principle is responsible for the omnipresent network hubs, from Facebook and Google on the World Wide Web to protein p53, the 'cancer hub', in human cells. However, its origins remain a source of constant wonder and speculation. The latest attempt to shed light on its roots is presented by Papadopoulos *et al.*¹ in a paper published on *Nature's* website today.

Preferential attachment made its first appearance in 1923 in the celebrated urn model of the Hungarian mathematician György Pólya², and it has reappeared repeatedly over the past century, particularly in the social sciences. Although Robert Merton named it the Matthew effect³ in 1968 after the Gospel of Matthew, "For everyone who has will be given more, and he will have an abundance", its current usage emerged only in 1999, with the discovery that it accounts for the power-law distributions observed in several real networks⁴.

A new node joining a network, such as a new web page or a new protein, can in principle connect to any pre-existing node. However, preferential attachment dictates that its choice will not be entirely random, but linearly biased by the degree of the pre-existing nodes — that is, the number of links that the nodes have with other nodes. This induces a rich-get-richer effect, allowing the more-connected nodes to gain more links at the expense of their less-connected counterparts. Hence, the large-degree nodes turn into hubs and the network becomes scale-free — the probability distribution of the degrees over the entire network follows a power law. This is a frail set-up, as any nonlinearity in preferential attachment will either eliminate the hubs

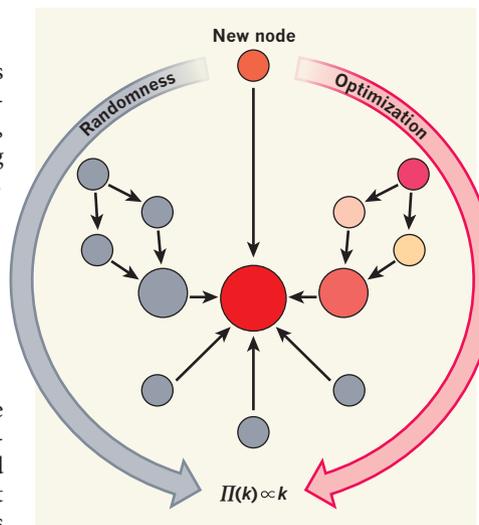


Figure 1 | Randomness or optimization? Two families of models could explain the origin of preferential attachment in networks, according to which the probability $\Pi(k)$ that a new node links to a pre-existing node that has degree k (the number of links that the node has with other nodes) is proportional to k . One family of models, to which the model introduced by Papadopoulos *et al.*¹ belongs, assumes that preferential attachment is rooted in an optimization framework (right side). In these models, a new node will connect to the node that is most similar to it (most similar colour) but also has the largest degree. The central node offers the best balance between these two options. The other family of models relies on randomness (left side). In this case, the new node is colour-blind, so it randomly selects a link and connects to its target. Once again, the central node, which has the most links pointing to it, has the highest chance of being selected.

or generate super-hubs, leading to the loss of the scale-free property⁵. However, in every system in which it has been possible to measure preferential attachment, a linear form has been detected^{6,7}.

The centuries-old proverb 'birds of a feather flock together' captures the idea that humans tend to hang out with those who are similar to them. Sociologists call this homophily, and it is perhaps one of the best documented concepts in the social sciences. Papadopoulos *et al.*¹ propose that homophily might also contribute to preferential attachment. They

introduce a model in which each node is assigned a randomly chosen position along a circle that serves as a 'homophily space': the closer two nodes are to each other on the circle (that is, the smaller the angle θ spanned by the nodes when measured from the circle's centre), the more similar they are (see Fig. 1 of the paper¹). The network expands through the addition of new nodes, such that a node added at time $t = 1, 2, \dots$ will choose to connect to a pre-existing node added at time s only if node s offers the smallest of all possible products $s\theta_{st}$, where θ_{st} is the angular distance between nodes s and t . Hence the new node optimizes its choice between two often conflicting interests: the node it will link to should be the most connected (the oldest, with the smallest s) and the most similar to it (the smallest θ_{st}).

Interestingly, by placing each node at distance $r_t = \ln t$ from the centre of the homophily circle, the authors find that the network evolves not on the circle but in a hyperbolic space, a geometrical space that is familiar mainly to those well versed in cosmology and general relativity. In this space, strange things can happen, such as parallel lines meeting each other and triangles that have zero-degree angles. Yet the model has its simplest interpretation in this peculiar space, where new nodes simply connect to the nodes closest to them. The authors show that the resulting network is scale-free and that a linear preferential attachment is the model's emerging feature.

The new model fuels a slowly evolving debate — is preferential attachment rooted in pure chance or in some form of optimization? Indeed, the most accepted mechanisms of preferential attachment rely on dumb luck. The simplest one is this: first randomly select a link in a directed network, for example the links of the World Wide Web that point to a document; then connect the new node to the selected link's target⁸. The more connected nodes have an advantage here, as the chance that a new node connects to them is proportional to their degree. Variants of this simple mechanism lie behind the popular copying model proposed to explain the scale-free nature of the web⁹ and the emergence of hubs in protein-interaction networks through gene duplication^{10,11}. According to these mechanisms, preferential attachment does not require human agency, but is rather a consequence of purely random actions. By contrast, Papadopoulos and colleagues' model calls for clear agency, as each new node seeks to link to the closest and oldest node. In this respect, the model supports earlier mechanisms, developed in the context of the Internet^{12,13}, proposing that preferential attachment

is rooted in a wish to balance distance to the target node with some utility, such as access to bandwidth.

Both approaches are tempting. Random models ask little of us, and demonstrate how random actions can result in outcomes that are not so random. Yet we do not think that the choices we make are ever random, fuelling the attractiveness of models that invoke some form of optimization.

This tension between two equally attractive but apparently opposing alternatives is by no means new. In the 1960s, the economist Herbert Simon and the mathematician Benoît Mandelbrot fought a fierce public dispute, with Simon defending the role of randomness and preferential attachment in explaining the power-law distribution of word frequencies in text, and Mandelbrot arguing for an optimization framework¹⁴. In the past decade, experimental evidence for preferential attachment in the context of networks has tilted the argument

in Simon's favour. And now the debate is shifting to a deeper question — whether preferential attachment is the outcome of random actions or optimization (Fig. 1).

This debate helps us to understand how preferential attachment emerges in an identical form in such widely different systems. The fact that the effect is widespread suggests that it probably derives from both agency and random actions. Most complex systems have a bit of both, so we do not need to choose between them. Luck or reason, preferential attachment wins either way. And so do we, gaining a deeper understanding of this puzzling yet ubiquitous force. ■

Albert-László Barabási is at the Center for Complex Network Research, Northeastern University and in the Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA.
e-mail: barabasi@gmail.com

1. Papadopoulos, F., Kitsak, M., Serrano, M. A., Boguñá, M. & Krioukov, D. *Nature* <http://dx.doi.org/10.1038/nature11459> (2012).
2. Eggenberger, F. & Pólya, G. *J. Appl. Math. Mech. (ZAMM)* **3**, 279–289 (1923).
3. Merton, R. K. *Science* **159**, 56–63 (1968).
4. Barabási, A.-L. & Albert, R. *Science* **286**, 509–512 (1999).
5. Krapivsky, P. L., Redner, S. & Leyvraz, F. *Phys. Rev. Lett.* **85**, 4629–4632 (2000).
6. Jeong, H., Néda, Z. & Barabási, A.-L. *Europhys. Lett.* **61**, 567–572 (2003).
7. Newman, M. E. J. *Phys. Rev. E* **64**, 025102(R) (2001).
8. Dorogovtsev, S. N., Mendes, J. F. F. & Samukhin, A. N. *Phys. Rev. Lett.* **85**, 4633–4636 (2000).
9. Kumar, R. et al. in *Proc. 19th Symp. Princ. Database Syst.* (eds Vianu, V. & Gottlob, G.) 1–10 (ACM, 2000).
10. Pastor-Satorras, R., Smith, E. & Solé, R. V. *J. Theor. Biol.* **222**, 199–210 (2003).
11. Vazquez, A., Flammini, A., Maritan, A. & Vespignani, A. *ComplexUs* **1**, 38–44 (2003).
12. D'Souza, R. M., Borgs, C., Chayes, J. T., Berger, N. & Kleinberg, R. D. *Proc. Natl Acad. Sci. USA* **104**, 6112–6117 (2007).
13. Fabrikant, A., Koutsoupias, E. & Papadimitriou, C. in *Proc. 29th Int. Colloq. Automata, Languages and Programming* (eds Widmayer, P. et al.) 110–122 (Springer, 2002).
14. Kornai, A. *Mathematical Linguistics* 71 (Springer, 2008).