

COMPLEX NETWORKS

Lies, damned lies and statistics

Statistical physics can reveal the fabric of complex networks, for example, potential oligarchies formed by its best-connected members. But care has to be taken to avoid jumping to conclusions.

**LUIS A. NUNES AMARAL AND
ROGER GUIMERA**

are in the Department of Chemical and Biological Engineering and Northwestern Institute on Complex Systems, Northwestern University, Evanston, Illinois 60208, USA.

e-mail: amaral@northwestern.edu; rguimera@northwestern.edu

Old boys networks, fraternities, free-mason shops — many of us look at them with suspicion. The worry about the dangers of oligarchies is not new. Adam Smith¹ noted that “people of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices.” However, do the powerful indeed purposefully organize into such ‘rich-club’ structures or do the connections we observe arise merely as a natural consequence of a stochastic process — if you have many connections, you are quite likely to connect ‘by chance’ to another well-connected member of the community. On page 110 of this issue, Colizza and colleagues² tackle the intricate problem of quantifying the existence of systematically created oligarchies within networks, and apply their findings to cases from many different areas. Surprisingly, they find that the Internet, which had been reported to display an oligarchic structure³, in fact does not.

At its core, the question that Colizza *et al.*² address is no different from the problem of deciding if, for example, someone who is 6 foot 5 inches — or 196 cm — is tall or not. Obviously, the question only makes sense if one defines the group of people with whom the person in question is being compared. A height of 196 cm is significantly larger than the height of the average adult female, but it is significantly smaller than the height of the average centre player in the US national basketball association (NBA). The challenge clearly is in identifying the appropriate group for making the comparison. An NBA scout will not decide if a prospect is tall enough to play centre by comparing his height to the heights of the general population. In the same spirit, Colizza and her co-workers demonstrate that the first measure³ introduced to quantify the presence of oligarchies in complex networks is prone to

misinterpretation: it will take increasing values as the number of connections of the nodes in the network increases. Thus, an oligarchy will always appear to be present, even if the network is random.

This shortcoming, which casts doubt on the conclusions that can be drawn from such an analysis, can be circumvented. Colizza *et al.* show that the value of the so-called rich-club coefficient needs to be compared with its expected value for a random network whose nodes have the exact same number of connections as the network one is interested in. In other words, an appropriate null model has to be found for comparison. Colizza and colleagues have now found the correct null model for the rich-club coefficient and, when applied to real networks, it reveals some surprises; they uncover that the Internet does not have an oligarchic structure whereas, for example, scientific collaborations do. That is, a social network, such as the one created by the scientific collaborations among researchers displays a markedly rich-club structure. This tells us that our worries about oligarchies within social systems are not unwarranted. The ultra-rich and powerful do not congregate at the Davos meetings by pure chance. On the other hand, rather unexpectedly, local hubs in the Internet, while providing rich connectivity to other nodes, are not tightly interconnected among themselves.

The importance of the findings reported by Colizza *et al.* is not restricted to the detection of oligarchies in networks. They demonstrate that — to prove the saying “there are three kinds of lies: lies, damned lies, and statistics” wrong — appropriate null models are needed to provide a suitable normalization before drawing conclusions from absolute measurements. Indeed, similar considerations apply to the analysis of one of the most important (and intensively studied) properties of complex networks — their modular structure^{4,5}. Most real networks are organized into modules; think, for example, of groups of friends or collaborators within social networks, or of pathways within the metabolism. The relevant questions are: Can we determine whether a network has a significantly modular structure? Is modularity enhanced during the

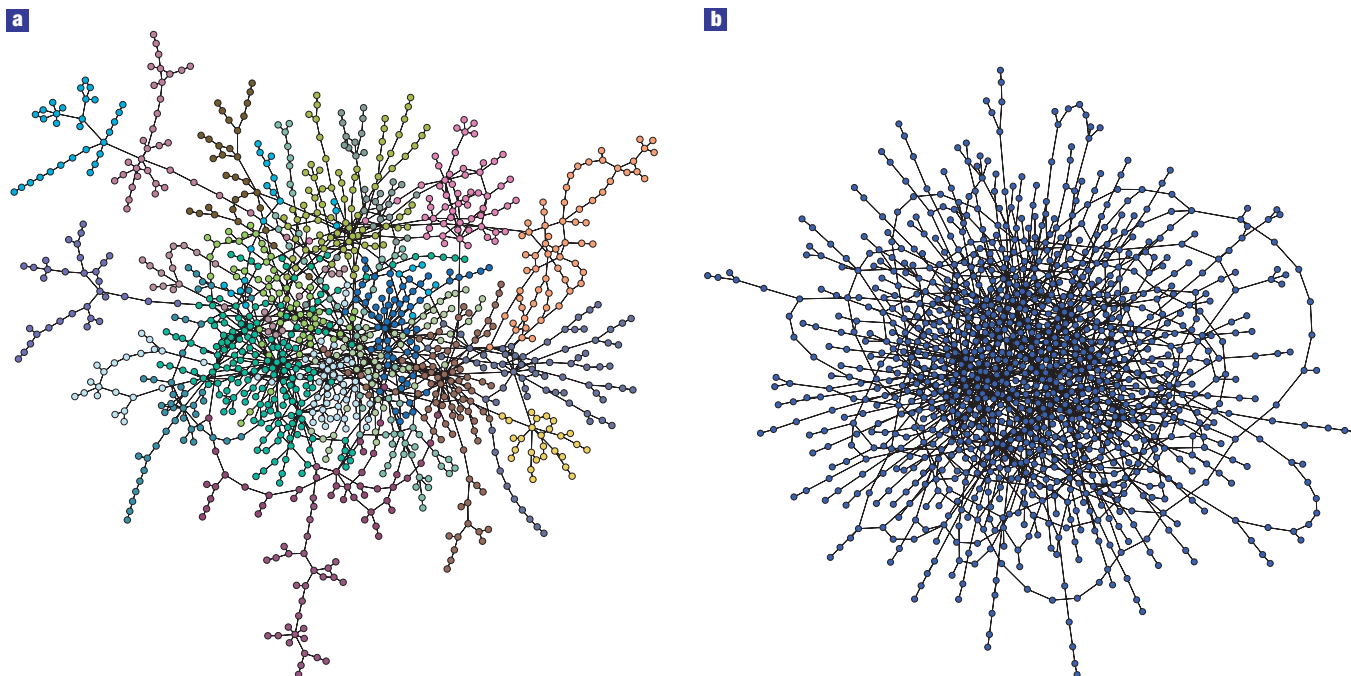


Figure 1 Cautionary tale for network analysis. **a**, In the metabolic network of *Escherichia coli*, each node represents a metabolite and two metabolites are connected if there is a biochemical reaction that transforms one into the other. Different colours represent different modules⁵. **b**, Randomization of the same network. Even though the degree distributions of the two networks are identical, they have quite different structural properties. For example, the metabolic network is modular whereas its randomization is not.

formation and evolution of a particular network? As happens when attempting to uncover the presence of a rich-club oligarchy, one needs to be careful and compare any measured values to the appropriate null model. Even within completely random networks, sub-graphs exist that are much denser than the graph as a whole. But these denser regions arise entirely due to fluctuations, that is, due to chance⁴.

At a more general level, the significance of reported observations on the global properties of complex networks, such as 'scale-free' (that is, power-law) distributions of the number of links per node, must be taken with care⁶. Consider the metabolic network⁵ in Fig. 1a. The overall degree distribution of the network is relatively well approximated by a power law. However, different modules have quite distinct degree distributions. For comparison, a 'randomized' version of the same network, although displaying the same overall degree distribution, lacks modular structure. In fact, a plausible explanation for the structure and

emergence of modules may have to do with attaining specific goals⁷ rather than with preferential attachment⁸.

Methods developed in the framework of statistical physics provide a powerful tool to analyse the organization of complex networks, and can reveal similar organization principles for networks from such dissimilar areas as biology, technology or social sciences. But surely, great care has to be taken to put the analysis on solid statistical grounds. Otherwise, our analysis will only tell us what we want to hear.

REFERENCES

1. Smith, A. *An Inquiry into the Nature and Causes of the Wealth of Nations* (Edinburgh, 1776).
2. Colizza, V., Flammini, A., Serrano, M. A. & Vespignani, V. *Nature Phys.* **2**, 110–115 (2006).
3. Zhou, S. & Mondragon, R. J. *IEEE Comm. Lett.* **8**, 180–182 (2004).
4. Guimera, R., Sales-Pardo, M. & Amaral, L. A. N. *Phys. Rev. E* **70**, 025101 (2004).
5. Guimera, R. & Amaral, L. A. N. *Nature* **433**, 895–900 (2005).
6. Amaral, L. A. N., Scala, A., Barthelemy, M. & Stanley, H. E. *Proc. Natl Acad. Sci. USA* **97**, 11149–11152 (2000).
7. Kashtan, N. & Alon, U. *Proc. Natl Acad. Sci. USA* **102**, 13773–13778 (2005).
8. Barabási, A.-L. & Albert, R. *Science* **286**, 509–512 (1999).