Taming complexity

ALBERT-LÁSZLÓ BARABÁSI

is at the Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Harvard University, Boston, Massachusetts 02115, USA, and Center for Complex Networks Research and Department of Physics, University of Notre Dame, Indiana 46556, USA.

e-mail: alb@nd.edu

The science of networks is experiencing a boom. But despite the necessary multidisciplinary approach to tackle the theory of complexity, scientists remain largely compartmentalized in their separate disciplines. Can they find a common voice?

The traditional physics curriculum has little room for networks. Most students are exposed to the impressive variety of crystal lattices and their impact on a solid's electronic properties. Graduate students may encounter some simple networks in the context of percolation, neural networks or string theory. But most networks emerging in physical systems offer few mysteries or scientific challenges. Therefore, the recent fever developing within the physics community for networks may come as a surprise.

As a measure of this boom, informal statistics show that over 200 papers with networks in the title were submitted to *Physical Review E* during the past year. Some physics journals have embraced this trend by starting a new section on the subject, whereas others find room for them in the 'interdisciplinary physics' section. These papers address a wide range of problems, from the structure of social and cellular networks to the Internet or the World Wide Web. Why this exploding interest? And why are these papers in physics journals?

We are surrounded by complex systems — from a biological cell, made of thousands of different molecules that seamlessly work together, to our society, a collection of six billion mostly cooperating individuals — which display endless signatures of order and self-organization. Understanding and quantifying this complexity is a grand challenge for science.

Physicists have a particularly long fascination with the subject. Gas theory set the stage at the end of the nineteenth century, demonstrating that the measurable properties of gases can be reduced to the random motion of billions of atoms and molecules. In the 1960s and 70s, the theory of critical phenomena enabled systematic approaches to quantify the transition from disorder to order in material systems, such as magnets or liquids. Chaos theory, with its seductive message that complex and unpredictable behaviour emerge from the nonlinear interactions of a few components, has dominated our quest to understand complex temporal behaviour in the 1980s. The 1990s were the decade of fractals, quantifying the geometry of patterns emerging in self-organized systems, from surfaces to snowflakes. Yet, despite these important conceptual advances, nobody seriously believed that we have a current theory of complexity.

When we try to characterize the many complex systems around us, the available tools fail for increasingly obvious reasons. First, most complex systems are not made of identical and undistinguishable components, as gases or magnets are — each gene in a cell or individual in a country has its own characteristic behaviour. Second, and more importantly, the components obey neither the extreme disorder of gases, in which a molecule can collide with any other molecule, nor the extreme order of magnets, where spins interact only with their immediate neighbours in a nicely periodic lattice. Rather, in complex systems the interactions form exquisite networks, each component being in contact with selected interaction partners.

Although each discipline has developed some rudimentary models to capture the relevant networks, the most sophisticated of these is the random network theory explored in the 1960s by the mathematicians Paul Erdős and Alfréd Rényi[1]. In a random network, each pair of nodes is connected with a link with probability p. But do we seriously believe that real networks are random? No. Successful functioning of cells or society must be governed by laws and organizing principles that should be reflected in their architecture as well.

In the absence of large maps, we could only speculate about the structure of real networks. This changed rather abruptly at the end of 1990s, thanks to the Internet (Fig. 1). Maps of World Wide Web documents connected to each other by URLs[2,3], of Hollywood actors linked by movies[3,4], or of metabolites held together by reactions[5,6] suddenly could be extracted from various Internet-based databases. These maps not only catalysed the emergence of network science, but forced on it a methodology involving simultaneous data collection, model building and analytical work. This is where physicists have had a clear advantage over mathematicians, who have concentrated on the analytical side, and biologists, who have meticulously catalogued a huge amount of data. In contrast, the most cited contributions to network theory have not embraced such specialization, but combine at least two of the measurement/modelling/theory components[3,4,7–9].

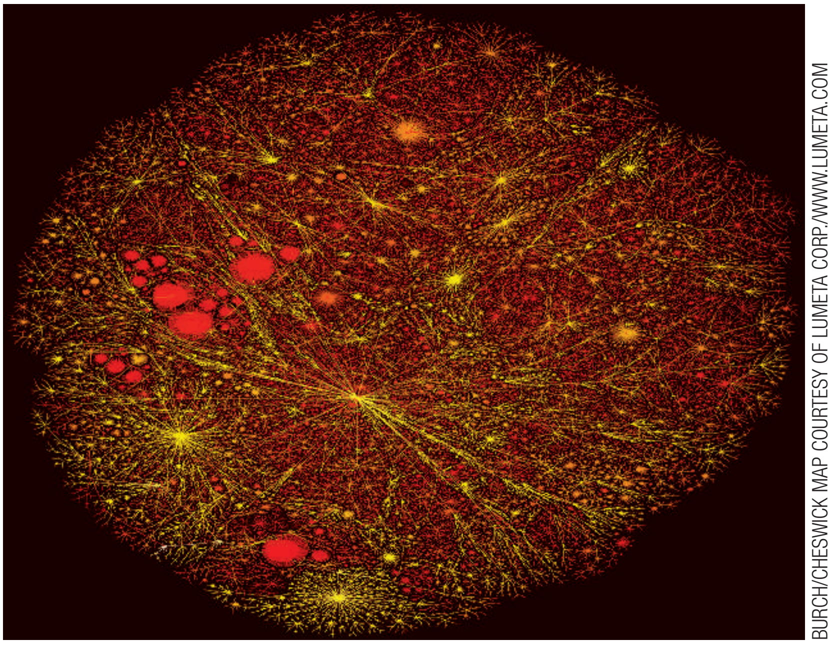
UNCOVERING THE UNDERLYING ORDER

Statistical mechanics taught us that intensive, system-size-independent quantities offer some of the best measures of a system's internal characteristics. This applies to complex networks as well: the most revealing measure of a network's overall structure is hidden in the degree distribution, which is the probability that a node has k links (degree k). In a random poissonian network in which most nodes have roughly the same degree, deviations from the average degree are rare[10].

Large network maps offered a chance to test the validity of this prediction, giving several surprises. First, for the World Wide Web[2,3], the Internet[11], the cell[5,6] or the social network defined by e-mail communication[12], the degree distribution follows a power law, in contrast with the random model. Such networks are 'scale-free'. The power law means that real networks are not as democratic as the random model suggests, but a few highly connected nodes, or hubs, hold together a large number of small nodes[3]. Second, the likelihood that a node's neighbours are connected to each other is higher in real networks than in a random universe[4].

The scale-free nature of networks has some important consequences. For example, if nodes randomly fail in a random network, the network should fall apart after a critical fraction of them is eliminated. In scale-free networks, however, the network is not destroyed by the removal of any finite fraction of nodes[8,13]. This extreme robustness is accompanied by a fragility to attacks: the systematic removal of a few hubs would easily destroy the network. Similarly, in a random network, electronic or biological viruses spread only if their spreading rate exceeds a critical threshold. In a scale-free network the threshold is zero, so that even weakly virulent viruses spread and persist[7,14]. Real networks are full of communities, corresponding to groups of nodes more tightly connected to each other than to other nodes[15–17]. Finally, the collective behaviour of most processes taking place on scale-free networks is drastically different from their behaviour on random or regular networks, forcing us to re-think network-based dynamical processes.

The ubiquitous scale-free property in real networks indicates that drastically different networks follow common organizing principles. But what is the origin of this universality? A few simple microscopic laws govern the network's evolution. For example, for a scale-free topology to emerge, the network must grow by the addition of new nodes, and links prefer to attach to the more-connected nodes — a process called preferential attachment[3]. The discovery and classification of such growth mechanisms not only led to detailed models of specific networks but also to the recognition that rate equations offer a viable tool for predicting the large-scale topology of a wide range of real networks[18,19]. This has closed the discovery loop, from the empirical observation of scale-free and clustered topologies, to microscopic models that can reproduce the observations, to continuum theories that offer exact predictions for the scaling exponents and their dependence on the microscopic parameters[3,7,18–21].



BURCH/CHESWICK MAP COURTESY OF LUMETA CORP./WWW.LUMETA.COM

Figure 1 **Connected. Full Internet maps such as this one vividly highlight the complexity of our information network. But more importantly, they enable network theories to be tested and refined.**

THE EMERGING NETWORK SCIENCE

The closure of this loop, as well as the simultaneous awakening to the importance of networks of a number of fields, has led to an explosion of network research. As discussed in the recent National Research Council (NRC) report on *Network Science*[22], this is not limited to physics, but captures biology, computer science, mathematics and the social sciences. Some papers on the subject have had more than 1,000 citations in the past five years. The most prominent scientific journals, from *Nature* and *Science* to *PNAS*, have devoted special issues, reviews, or editorials to networks. During the past five years, over twenty international conferences, workshops and summer schools have focused exclusively on network science. Three general-interest books have brought network science to the public, making the bestseller lists in several countries. Most major US universities have developed network science courses, offered by a variety of departments.

Network science has shown immediate economic benefits as well. The poster child of early network thinking is Google, whose phenomenal success is rooted in its algorithm that uses the topology of the Web to rank the search results. Today, more than twenty new companies exploit the increasing knowledge of social networks, with applications from discovering business contacts to dating services.

Yet the road to a fundamental and comprehensive understanding of networks is still rather rocky. First of all, our funding system, while often paying lip service to the crucial importance of networks in many areas of science and technology, at the same time is not ready to acknowledge the existence of an emerging research area. As a result, scientists interested in the fundamental properties of networks and complex systems need to sell themselves as single-discipline-based players, convincing either the physics community that their research is physics or the biology establishment that network science is all about biology. But as the NRC survey has found, most of these practitioners are deeply interdisciplinary, identifying on average 3.6 research fields in which they are active, with 80% of the 600-plus respondents mentioning more than one field.

Advances will also need those in network science to find a common voice and understand the similarity of the intellectual challenges they face. For instance, the epidemiologist's desire to predict the extent of the next influenza outbreak, the computer security expert's need to contain the next worm, or the social scientist's need to explore the emergence of ideas are all limited by the need to quantify how the structure of a network impacts the spreading processes taking place on it. Despite these commonalities, the big picture is often obstructed by disciplinary fences. This is well illustrated by the NRC survey in which 95% of the respondents classify their own work as potentially belonging to network science, yet only 70% claim that such a field exists!

Even those that understand the importance of networks are often driven by the need for quick fixes, expecting solutions overnight. It took well over a half a century for the first glimpses of the quantum world to translate to transistors and computers. Despite its early commercial successes, network science is more basic than applied today, and it will take decades to bring the understanding it provides to full fruition. The desire to design a better Internet or a disease-resistant cell is also fuelled by a common misunderstanding of how networks emerge and evolve: there is no central design behind the Internet or the cell that would enable us to redesign them overnight. These systems follow common laws precisely because the decisions of where to connect are distributed, delegated to the individual nodes. We can already design topologies that are far more tolerant of attack and error than those seen today in nature and technology. But who can force the cell or the World Wide Web to follow them? Humility must win over hubris: we must first observe and understand, in order to improve and design.

THE ROAD TO COMPLEXITY

But are we getting any closer to an answer to complexity? As it stands, network theory is not a proxy for a theory of complexity — it only addresses the emergence and structural evolution of the skeleton of a complex system. The overall behaviour of a complex system, which we ultimately need to understand and quantify, is as much rooted in its architecture as it is in the nature of the dynamical processes taking place on these networks.

We are, however, at the threshold of unravelling the characteristics of these dynamical processes. Data-collection capabilities can now capture not only a system's topology, but also the simultaneous dynamics of its components, such as the communication and travel patterns of millions of individuals, or the expression level of all genes in a cell. These offer hope for a systematic program, starting from a measurement-based discovery process and potentially ending in a theory of complexity with predictive power.

At the level of the nodes and links, apparently random and unpredictable events seem to be driving each network. How would I know whom you will meet tomorrow, what page would you link your webpage to, or which molecule will react with an ATP molecule in your nerve cell? The true intellectual thrill for a physicist studying complex networks comes from the recognition that despite this microscopic randomness, a few fundamental laws and organizing principles can explain the topological features of such diverse systems as the cell, the Internet or society. Are we going to find generic organizing principles that are just as intriguing and powerful as those uncovered in the past few years in our quest to understand network topology? This is hard to predict, but I expect that such laws and principles do exist. And as long as we believe in their existence, it is worth searching for them.

Although for many decades questions related to complexity were driven by statistical physics, in the new era of interdisciplinary and multidisciplinary science, the circle is much wider. Given its set of tools ideally suited to the problem, physics will continue to play a leading role. Yet the stakes and challenges are too high for a single community to face them all, requiring a grand coalition of many disciplines to tame complexity. Will it take five years or five decades to make significant advances? Hard to know for sure, but if the history of network theory is any guide, once we get a first glimpse of some universal order, it will take no time to unfold the whole construction. At that point we will have a chance to understand the key to nature's secret code for multitasking — the one that orchestrates the actions of uncountable components into a magic dance of order and ultimate elegance.

REFERENCES

1. Erdős, P. & Rényi, A. *Publ. Math. Debrecen* **6,** 290–291 (1959).
2. Albert, R., Jeong, H. & Barabási, A.-L. *Nature* **401,** 130–131 (1999).
3. Barabási, A.-L. & Albert, R. *Science* **286,** 509–512 (1999).
4. Watts, D. J. & Strogatz, S. H. *Nature* **393,** 440–442 (1998).
5. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A.-L. *Nature* **407,** 651–654 (2000).
6. Wagner, A. & Fell, D. *Proc. R. Soc. Lond. B* **268,** 1803–1810 (2001).
7. Pastor-Satorras, R. & Vespignani, A. *Phys. Rev. Lett.* **86,** 3200–3203 (2001).
8. Albert, R., Jeong, H. & Barabási, A.-L. *Nature* **406,** 378–382 (2000).
9. Amaral, L. A. N., Scala, A., Barthelemy, M. & Stanley, H. E. *Proc. Natl Acad. Sci.* **97,** 11149–11152 (2001).
10. Bollobás, B. *Random Graphs* (Cambridge Univ. Press, 2001).
11. Faloutsos, M., Faloutsos, P. & Faloutsos, C. *Comput. Commun. Rev.* **29,** 251–262 (1999).
12. Ebel, H., Mielsch, L. I. & Bornholdt, S. *Phys. Rev. E* **66,** 035103 (2002).
13. Cohen, R., Reez, K., Ben-Avraham, D. & Havlin, S. *Phys. Rev. Lett.* **85,** 4626–4628 (2000).
14. Eubank, S. *et al. Nature* **429,** 180–184 (2004).
15. Girvan, M. & Newman, M. E. J. *Proc. Natl Acad. Sci.* **99,** 7821–7826 (2002).
16. Palla, G., Derenyi, I., Farkas, I. & Vicsek, T. *Nature* **435,** 814–818 (2005).
17. Milo, R. *et al. Science* **298,** 824–827 (2002).
18. Dorogovtsev, S. N. & Mendes, J. F. F. *Advances in Physics* **51,** 1079–1187 (2002).
19. Krapivsky, P. L., Redner, S. & Leyvraz, F. *Phys. Rev. Lett.* **85,** 4629–4632 (2000).
20. Caldarelli, G., Capocci, A., De Los Rios, P. & Munoz, M. A. *Phys. Rev. Lett.* **89,** 258702 (2002).
21. Bianconi, G. & Barabási, A.-L. *Phys. Rev. Lett.* **86,** 5632–5635 (2001).
22. *Network Science* (National Research Council, National Academy Press, Washington DC, 2005).