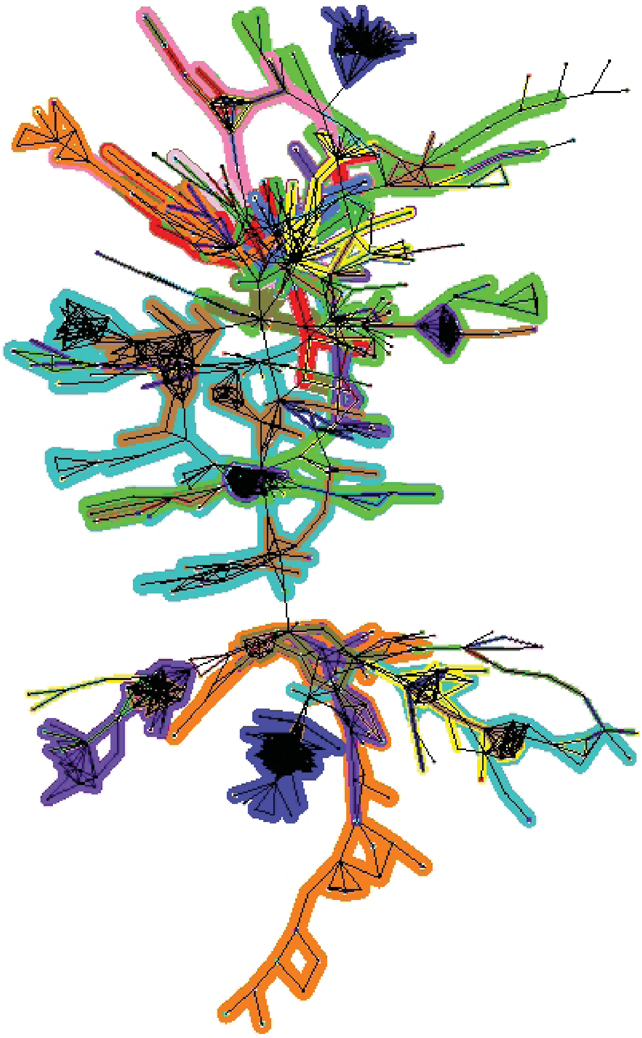

This is a typical complex network representing interactions between proteins. Fractal theory can help us identify groups of proteins which have similar functionality (shown by similar colors). See Professor Makse's featured paper on page 4.

Networks and Fractal Networks

HERNÁN MAKSE
Professor of Physics, Benjamin Levich Institute for Physico-Chemical Hydrodynamics

It's a world of laughter, a world or tears
It's a world of hopes, it's a world of fears
There's so much that we share
That it's time we're aware
It's a small world after all
"It's a Small World" by Richard M. Sherman and Robert B. Sherman

NETWORKS AROUND US

At some point, we have all agreed that "it's a small world" to express our surprise over unexpected common acquaintances. This simple phrase has become the symbol of a quiet revolution in our understanding of how complex interactions give rise to simple universal laws. Some decades ago, sociologists determined experimentally that the above expression holds true, but only recently have we been able to understand how it is possible to be within a distance of six handshakes with every person in the world (with an earth population in the billions). More importantly, we have now realized that this proximity is only one of the many counter-intuitive manifestations of a unifying underlying pattern. Such laws are not restricted in social systems only, but seem to cross the borders of many disciplines.

When scientists deal with a problem including a small number of interacting elements they are usually able to provide a solution and predict the evolution of the system. How, though, can we predict the emerging behavior of a system usually including millions? The field of complexity, including sub-fields such as chaos theory, self-organization, fractal theory, etc, provides a number of tools that help us predict the properties of the system if the properties of the system's parts are known. The most recent advance in the field is *network theory*. According to this approach, the complex mesh of interactions between the constituent elements carries all the necessary information on the system that we need to uncover and understand the network structure and dynamics.

Why should we need a new approach, when similar problems have been addressed through standard Physics methods? The field of Statistical Physics, after all, has been solving problems over an enormous number of interacting particles and still has a tremendous predictive power. The answer lies in the fact that interactions in Physics practically always depend on the distance between two particles. However, this condition is too restrictive for, e.g., a social system where any kind of information can be rapidly disseminated through e-mail, phone etc. In network theory, every system is composed of only two elements: a) *nodes*, which represent an element that acts towards establishing relations, and can be a person in society, a computer in the Internet, a protein in a protein interaction network, etc, and b) *links* which denote an interaction between the two connected nodes. This interaction can be, e.g., acquaintance in a social system, network cables in the Internet, or protein binding.

The knowledge of the nodes in a network are not enough to uniquely characterize this network. If we consider rumour spreading in a given population we need to know which individuals communicate with each other, independently of physical proximity. A different process, such as virus spreading for which a physical contact is required, will give rise to a network with different connections. So, a network does not necessarily yield the same picture, even when it refers to the same population, because the interpretation of what a link represents can drastically change the network form.

Network theory can be considered a part of graph theory, which has long been developed in mathematics, with important contributions from some well-known mathematicians. For many decades, the focus was on random networks, where any node has the same probability of being connected to any other node. The idea was based on simple observational facts. For example, when the Internet started expanding there was no central authority to regulate its growth and routers would connect to each other without having to obey specific rules. Thus, there is no reason to assume that any non-random structure would emerge. One of the main quantities used to characterize a network structure is the degree distribution $P(k)$, where the degree k is the number of links for a given node. In random graphs, the distribution $P(k)$ assumes a gaussian form around a well-defined average value.

HERNÁN MAKSE is a member of City College's Levich Institute, as well as the Department of Physics, and holds a Ph.D. in physics from Boston University. His research interests lie in complex systems, granular materials, and "jamming" in soft-condensed matter systems. Hernán Makse's research has attracted some attention, recently. In addition to a NSF-CAREER award on soft-materials, he has received NSF grants to study the dynamics of social networks, and the mathematical properties of biological networks and he is a co-recipient of the 2005 New York City Mayor's Award for Excellence in Science and Technology.

Dr. Makse deals in theories relating to granular materials and soft condensed matter. His work has helped in the understanding of glasses and other disordered systems, such as sand dunes. Dr. Makse's research has already played an important role in placing thermodynamics of granular materials on a firm footing, testing and developing a unifying thermodynamic framework, which promises to lead to a common understanding of a wide range of systems which are inherently out of equilibrium.

Hernán is a native of Argentina, and began studying maths early. He studied physics at Boston University, where his interest for grains first asserted itself. "During my Ph.D. studies at Boston University I was interested in the patterns that appear in sedimentary rocks," he said. "Sedimentary rocks are composed of layers of different grains, a pattern that is called stratification. I was intrigued by these patterns." He found a similarity between those patterns and the patterns of bread crumbs, not in the lab, but in the kitchen. "One day I was playing with grains - actually I was cooking Milanesas, a typical Argentinian dish, and I poured bread crumbs on the table," Mr. Makse said. "I found that by pouring grains of different sizes, like in a pile of sand, the grains separate in layers, just like the ones observed in sedimentary rocks. This led me to investigate this phenomenon further, and eventually I ended up being an 'expert' in granular materials." Today, he attempts the development of statistical mechanics theories to understand not only granular matter, but also the so-called soft materials, which include colloids, such as paints, emulsions, blood or milk and glassy and amorphous materials.

Dr. Makse is also pursuing a theoretical understanding of complexity. He contends that the principles of statistical mechanics used to explain the organization of condensed matter can be applied to complex networks from biological systems, to the Internet, to social networks. His work could lead to strategies for protecting Internet networks from attacks and algorithms for improving immunization strategies that take into account the modular nature of society.

His research experience so far has shown the potential of working at the interface of diverse fields in order to solve problems of fundamental importance in physics, engineering, biology, sociology and economics. From his early work as a Ph.D. student in Boston University his research has had a unifying theme: namely, the emergence of complexity in physical, biological and social systems. Early on in his career he developed models of urban economics using physical concepts from phase transitions in liquids. This initial work, published fourteen years ago and featured in many specialized and divulgation journals, has many of the ingredients that still appear in his research today (Yes, he has not moved on very much...) He uses paradigms from the physical sciences (e.g. phase transitions) to explain the emergence of complexity in entirely different set of natural phenomena (e.g. the morphology of cities and towns, the structure of society and the economy or of living systems shapes by evolution).

These skills learned in his first formative years are still transmitted to his students today. Students are an integral part of Dr. Makse's research team. "At CCNY, I am able to draw on an excellent pool of undergraduate and graduate students from physics and engineering," he says. "I could not do my work without them." Hernán has been lucky to attract a set of highly motivated students to his lab. The students' performance is outstanding, despite the challenges that they may face, such as enduring long subway commutes, having to take classes in a second language, and learning how to successfully navigate through bureaucracy. The challenging environment at CCNY simulates evolution, brings out the best adaptations in students and helps them develop a strong educational foundation. At the end of their journeys, typical CCNY students are better prepared than those from other universities. Hernán hopes that he can show students how their academic world of classes and labs are interconnected with systems and communities in the world at large.

It was, thus, a striking surprise when the connectivity of the Internet was measured in the late 90s and it was found that the degree distribution obeys a power law form

$$P(k) \sim k^{-\gamma} . \quad (1)$$

The properties of such a distribution are drastically different than a gaussian. These networks were called *scale-free*, to denote the absence of a characteristic scale for the degree in the system. The wide-tailed distribution reveals that most nodes have a small connectivity, but there exist a few nodes (hubs) whose connectivity is many orders of magnitude higher. Evidence was quickly gathered that this behavior is universal and most real-life net2 works follow this distribution: the Internet at the Router Level and at the Autonomous System Level, the World-Wide-Web, the airport network, collaboration networks between actors or co-authors, email networks, social networks, protein interaction networks, metabolic networks, and many more. Another common property is the *small-world* effect, since it was shown that the average distance in these networks is usually a small number, of the order of the logarithm of the network size.

An explosion in the literature followed this discovery. Network scientists have tried hard to understand the properties of these structures. Many surprises were met along the way: for example, these networks are extremely robust to any attempt of randomly destroying nodes, and are in practice indestructible. Still, their long-range connectivity is rapidly destroyed when the most-connected nodes (hubs) can be identified and eliminated.

FRACTAL NETWORKS

One key property, though, was missing. All these networks were annoyinly reminiscent of fractal objects, which were heavily studied in the 80s and 90s, but were evading their classification as such. Self-similarity is a property of fractal structures, a concept introduced by Mandelbrot and one of the fundamental mathematical results of the 20th century. A fractal object exhibits selfsimilarity in the sense that any part of the whole looks similar to the whole (compare, for example, part of a cloud with the cloud itself). The mathematical description of a fractal object requires that the mass M(l) included in a box of size l should scale as a power-law with the size, i.e.

$$M(l) \sim l^{-d_f} , \quad (2)$$

where d_f is the fractal dimension of the object. Typically, though, real world scale-free networks exhibit the small-world property, which implies that the number of nodes increases exponentially with the diameter of the network, rather than the power-law behavior expected for self-similar

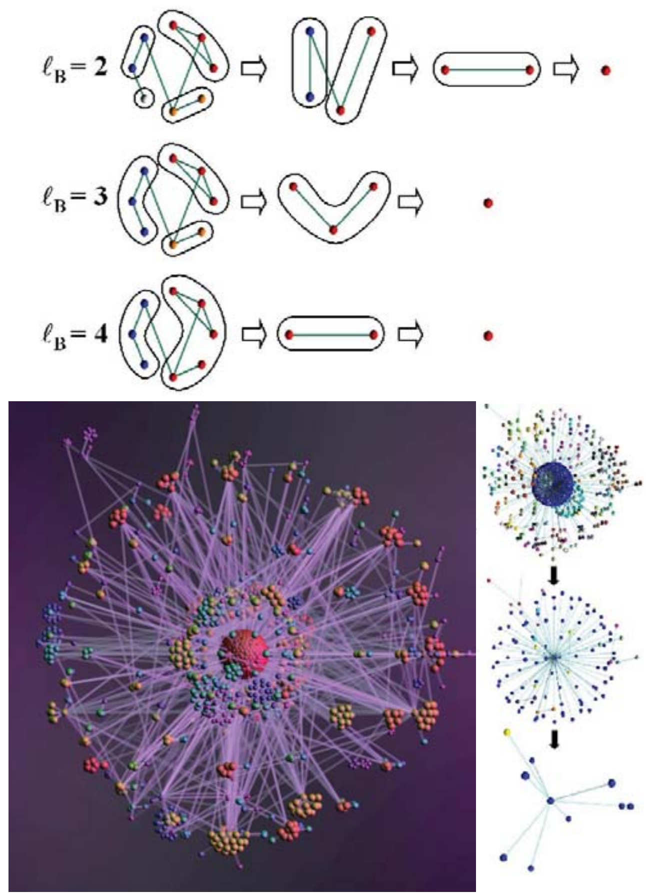

FIG. 1: The renormalization procedure in complex networks. a) Demonstration of the method for different l_B and different stages in a network demo. The first column depicts the original network. The system is tiled with boxes of size l_B (different colors correspond to different boxes). All nodes in a box are connected by a minimum distance smaller than the given l_B. For instance, in the case of $l_B = 2$, one identifies four boxes which contain the nodes depicted with colors red, orange, white, and blue, each containing 3, 2, 1, and 2 nodes, respectively. Then each box is replaced by a single node; two renormalized nodes are connected if there is at least one link between the unrenormalized boxes. Thus we obtain the network shown in the second column. The renormalization procedure is applied again and repeated until the network is reduced to a single node (third and fourth columns for different l_B). b) Three stages in the renormalization scheme applied to the entire WWW. We fix the box size to $l_B = 3$ and apply the renormalization for four stages. This corresponds, for instance, to the sequence for the network demo depicted in the second row in part a of this figure. We color the nodes in the web according to the boxes to which they belong.

structures. For this reason complex networks were believed to not be length-scale invariant or self-similar.

Recently, our lab presented an approach to analyze complex networks, that reveals their underlying selfsimilarity. This result is achieved by the application of a renormalization procedure which coarse-grains the system into boxes containing nodes within a given size. As a result, a power-law relation between the number of boxes needed to cover the network and the size of the box is found, defining a finite self-similar exponent. These fundamental properties help us to understand the emergence of the scale-free property in complex networks. They suggest common self-organization dynamics of diverse networks at different scales into a critical state and in turn bring together previously unrelated fields: the statistical physics of complex networks with renormalization group, fractals and critical phenomena.

In our approach, the network is covered with the minimum possible number N_B of non-overlapping boxes, where the maximum distance between any two nodes in a box is less than l_B (finding the optimum covering is a hard computational problem that requires special techniques). The resulting relation

$$N_B \sim l_B^{-d_B} \qquad (3)$$

defines the fractal dimension of the given network. The coexistence of the small-world property and the fractality can be intuitively understood as follows: in a pure fractal network the length of a path between any pair of nodes scales as a power-law with the number of nodes in the network. Therefore, the diameter L also follows a power-law, $L \sim N^{1/d_B}$. If we add a few shortcuts (links between randomly chosen nodes), many paths in the network are drastically shortened and the small-world property emerges as $L \sim \log N$. In spite of this fact, for shorter scales, $l_B << L$, the network still behaves as a fractal. In this sense, we can say that globally the network is small-world, but locally (for short scales) the network behaves as a fractal.

The idea for renormalizing the network emerges naturally from the concept of fractality. Renormalization is a procedure where smaller replicas of a given object are continuously created, retaining at the same time the essential structural features, and hoping that the coarsegrained copies will be more amenable to analytic treatment. If a network is self-similar, then it will look more or less the same under different scales. The criterion that we use to decide on whether a renormalized structure retains its form is the invariance of the main structural features, expressed mainly through the degree distribution.

The method works as follows. We start by fixing the value of l_B and apply the box-covering algorithm in order to cover the entire network with boxes. In the renormalized network each box is replaced by a single node and two nodes are connected if there existed at least one connection between the two corresponding boxes in the original network. The resulting structure represents the first stage of the renormalized network. We can apply the same procedure to this new network, as well, resulting in the second renormalization stage network, and so on until we are left with a single node. Using this technique we are able to construct an analytic theory and calculate important topological properties, characterized mainly through scaling exponents.

CURRENT INTERESTS

Today, research in the field shifts focus towards applying the accumulated theoretical knowledge to the case of real-systems and understanding the implications of the specific structure. In our group, we try to understand how we can extract behavioral information through the actions of members in online communities. Analyzing the databases of exchanged messages or favorites lists between members it is possible to reconstruct the social web of connectivities, at least in the online world. The analysis of this network can then shed light on our motives behind our online actions, such as for example whether we prefer to communicate with friends, or friends of friends, or even complete strangers. It is possible to quantify the relative strength of each behavioral mechanism, and we have discovered that our behavior changes drastically as we become more involved in a community.

Another important line of research in our lab concerns biological networks, and how they have evolved over time. A large number of protein interaction networks for dif4 ferent species have been shown to be fractal. Based on this, we were able to partially reconstruct these networks at earlier stages of evolution. We try to understand the growth of such networks from some earlier simpler state to their present fractal form. Has fractality always been there or has it emerged as an intermediate stage obeying certain evolutionary drive forces? Is fractality a stable condition or growing networks will eventually fall into a non-fractal form?

The study of networks has offered a unified view of many systems, otherwise completely unrelated. The idea of using common methods for solving different problems has always been very appealing, and as such, network theory is a prominent example of inter-disciplinary research. A deeper understanding of self-similarity and fractality in complex network will help us better understand many fundamental properties of real-world networks.

ACKNOWLEDGMENTS

The author thanks Lazaros Gallos for his contributions.