Theoretical Computer Science special issue on Complex Networks

Ravi Kumar [a], Matthieu Latapy [b]

[a] *Yahoo! Research, 701 First Ave, Sunnyvale, CA 94089, USA*

[b] *LIAFA – CNRS and Université Paris 7, 2, place Jussieu, 75005 Paris, France*

Complex networks are large networks encountered in practice that seem to lack any apparent structure. Typical complex networks include internet topologies, web graphs, peer-to-peer networks, social networks (cooperation between people, economical exchanges, friendships, and others), biological networks (protein/gene interactions, brain, food webs), linguistic networks (co-occurrence graphs, synonyms), and many more. Precisely defining a complex network is a tricky proposition since it is hard to set up a definition that is broad enough to capture the above examples, yet restrictive enough to make mathematical sense.

Notice that the above networks are very different in nature. Some are submitted to physical constraints (e.g., the internet or the brain) while others are purely virtual (e.g., social friendship networks); some come from natural sciences (e.g., biological networks) and others arise from technology (e.g., the internet); some are explicitly optimized with a notion of efficiency (e.g., economic networks or the internet) while others are not (e.g., the web). All these networks, however, have several properties in common [56]. In particular, they have

- low global density—the number of links grows roughly linearly with the number of nodes, leading to a constant average degree;
- low average distance, typically logarithmic or sub-logarithmic in the number of nodes;
- heavy-tailed degree distributions, often modeled by power laws, Zipf, or log-normal distributions;
- high local density—the probability that two neighbors of any nodes are themselves connected (i.e., the clustering coefficient) is large compared to the global density.

* Corresponding author: latapy@liafa.jussieu.fr

The study of real-world complex networks emerged as an active area of research in the late nineties and has continued to be so ever since.

As evident from the examples, the study of complex networks is multi-disciplinary by nature. The tools to study these networks come from various disciplines as well: combinatorics and graphy theory, computer science, and statistical physics. Computer science, in particular theoretical computer science, has recently focused a lot of attention to try to understand and analyze these networks, especially so since many real-world complex networks are of prime importance in computer science. But computer science's interest in complex networks is not just because of the objects that they represent. More importantly, problems underlying the study of complex networks, irrespective of computing considerations, fit in two important areas of computer science research: graph algorithms and networking.

We identify five primary research topics in complex networks. Our classification is only suggestive and by no means exhaustive/exclusive; it is natural that many problems can be assigned to more than one topic.

Measurement.

Most real-world complex networks, like those we stated earlier, are not directly accessible. In many cases, one has only a limited and often expensive access to the underlying network; the measurement method used to access the network highly depends on the actual network itself. For instance, in case of the world-wide web, we can easily obtain the outlinks of a given node (since they are the hyperlinks on a given page), while the inlinks are much harder to obtain. Another example is the case of the internet, where the main access mechanism is the `traceroute` tool, which provides (mostly, the shortest) path from a source to a destination [9,23,15,50,52]. In protein networks, one generally obtains all the links of some chosen nodes, but not all [53]. In short, these measurement procedures are intricate and rely on a precise knowledge of the underlying network.

Going further, in most cases, the obtained view of the network is not only partial, but also biased by the actual measurement. For instance, if we explore the internet using the `traceroute` tool from one machine, we will obtain a tree-like structure; using this infer global properties of the internet can be misleading. The impact of measurement procedures on the observed properties has been studied in [29,16,1,34,28,46,48,49].

Much work has been done to design measurement methods for various complex networks of interest; see for instance [39,33,11,23,53]. These methods are in general designed to collect as much data as possible and they do not address the quality of the data. It is important to design measurement procedures with provable guarantees on the quality of the measurement. The development of accurate, efficient, and rigorous measurement methods are interesting research topics.

Analysis.

Given a large graph obtained by measuring a real-world complex network, we often want to extract interesting information from it. This is the goal of the analysis of complex networks.

To achieve this goal, we generally consider statistical properties of the network. Basic properties include size, density, average distance, degree distributions, and clustering coefficient. Beyond this, we can study correlations between in and out degrees, correlations between degree and clustering coefficient, betweenness centrality of nodes and links, connectivity structures, community structures, and so on.

These properties not only help us compare two complex networks but also validate theoretical models. Besides shedding light on the global structure of the network, they also identify portions of the network that may play particular and interesting roles. For instance, in social networks, the betweenness centrality of links (number of shortest paths in the graph that use the given link) reveals bridges between communities.

There are several papers that deal with complex network analysis. Most of them are case studies, see for instance [8,33,22,42,31,56,32,24,25]. Papers that discuss general analytic tools and methods include [41,56,4]. For surveys of complex network analysis, see [55,51,2,43,18,17].

We note that a lot of work until now have dealt with classical undirected graphs. Most real-world complex networks, however, are more complicated. They can be directed or weighted or heterogeneous (nodes and/or links are of several kinds), or hybrid (several networks defined on a same set of nodes). It is important to develop rigorous tools to analyze these graphs as well. Some preliminary work in this direction include [5,35,36]; much remains to be done, however.

Modeling.

The aim of modeling is to capture salient properties of the network in an analyzable framework. An important use of such a model would be to generate synthetic networks that resemble real-world counterparts but with different parameters (e.g., size); this is crucial for simulation studies. Models can also be used to predict the evolutionary behavior of the network. Models come in handy in the design of efficient algorithms that can exploit critical properties of the network. Last but not least, capturing a complex network through a simple model is an important step in understanding the network.

Two main approaches have been developed to model complex networks. The first is to sample a graph uniformly at random from the set of all graphs with given properties [21,7,6,40,41]. For instance, a well-studied class of random graphs are those generated by the Erdös–Rényi $G_{n,p}$ random graph model [21,7]; the parameters of the model are the size n and the global density p. Even if these parameters are set to match those of real-world complex networks, graphs generated by this model possess Poisson degree distribution and a vanishing clustering coefficient [7]. This clearly suggests that complex networks cannot be modeled by $G_{n,p}$ random graphs.

An alternate random graph model, called the configuration model, chooses a graph uniformly at random from all graphs with a prescribed degree distribution [6,40,41]; the degree distribution, which is the parameter of this model, can be set to match practical observations. While the average distance grows sub-logarithmically with the size of the graph [37,14,19], the clustering coefficient is still very low [41].

The second approach is to use a generative model to describe complex networks. A typical example is the preferential attachment model [4]: nodes arrive one by one, and link themselves to a pre-existing node with probability proportional to the degree of the latter. This "rich get richer" principle can be analytically shown to induce power-law degree distributions. These generative models are intuitively convincing but are often harder to analyze.

Much work needs to be done in the modeling of complex networks. For instance, there is currently no consensus on a model that would produce graphs with given density, degree distributions, and clustering coefficient [18,27,26].

Algorithms.

As we saw earlier, complex networks enjoy many properties in common. This entails the possibility of developing dedicated algorithms that can exploit these properties. There is a growing need for such algorithms and this is because of two reasons. First, these networks are central to many practical applications, such as internet routing, search engines, and bioinformatics. Second, they are very large in size, ranging from several hundreds of thousands to billions of nodes, thereby making most classical algorithms less appealing. Notice that the memory representation and storage of these graphs is a challenge in itself. Indeed, in many cases, their sheer size makes it impossible to store them in main memory.

The development of algorithms dedicated to large real-world complex networks is fairly preliminary [54,20,25,44,47]. This constitutes a very rich and promising area of investigation, especially in theoretical computer science.

Phenomena.

While we have been focusing on complex networks from a topological point of view, we should keep in mind that they can support a variety of important phenomena that can occur in practice. We give two examples. Consider a scenario where the connectivity of a network is compromised by an adversarial removal of nodes. This resilience and robustness question has been studied in [38,3,10,12,13,30]. Consider a diffusion process where an infection spreads through the network and at a given time step, a node can be infected based on its "state" and that of its neighbors; This occurs in epidemiology, marketing, and information sciences. This has been studied in [45,57,58].

While analysis and modeling of complex networks (and phenomena occurring on them) have received quite a lot of attention, measurement and algorithms are relatively less studied topics. The papers in this special issue address these two topics. The first one deals with measurement:

- *Exploring networks with traceroute-like probes: theory and simulations*, by Luca Dall'Asta, Ignacio Alavrez-Hamelin, Alain Barrat, Alexei Vazquez and Alessandro Vespignani

The others, listed below, deal with algorithms:

- *Broadcasting in unstructured peer-to-peer overlay networks*, by Fred Annexstein, Kenneth Berman and Mijhalo Jovanovic
- *Efficiently covering complex networks with cliques of similar vertices*, by Michael Behrisch and Anusch Taraz
- *Scalable percolation search on complex networks*, by Nima Sarshar, Oscar Boykin and Vwani Roychowdhury
- *D2B: a de Bruijn based content-addressable network*, by Pierre Fraigniaud and Philippe Gauron
- *Local heuristics and the emergence of spanning subgraphs in complex networks*, by Alexandre O. Stauffer and Valmir C. Barbosa
- *Could any graph be turned into a small-world?*, by Philippe Duchon, Nicolas Hanusse, Emmanuelle Lebhar and Nicolas Schabanel

We received 27 papers for this special issue, out of which we chose 7. We thank all the authors who submitted their papers. We also thank the reviewers for the excellent work in evaluating the submissions.

References

[1] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore. On the bias of traceroute sampling, or: Why almost every network looks like it has a power law. In *ACM Symposium on Theory of Computing (STOC2005)*, 2005.

[2] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 47, 2002.

[3] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance in complex networks. *Nature*, 406:378–382, 2000.

[4] A.-L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[5] A. Barrat, M. Barthlemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States*, 101(11):3747–375, 2004.

[6] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1:311–316, 1980.

[7] B. Bollobas. *Random Graphs.* Cambridge University Press, 2001.

[8] A.Z. Broder, S.R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. L. Wiener. Graph structure in the web. *WWW9 / Computer Networks*, 33(1-6):309–320, 2000.

[9] CAIDA. skitter. `http://www.caida.org/tools/measurement/skitter/`.

[10] D.S. Callaway, M.E.J Newman, S.H. Strogatz, and D.J. Watts. Network robustness and fragility: percolation on random graphs. *Physics Review Letters*, 85:5468–5471, 2000.

[11] J. Cho and H. Garcia-Molina. Estimating frequency of change. Technical report, Stanford University, 2000.

[12] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin. Resilience of the internet to random breakdown. *Physics Review Letters*, 85, 2000.

[13] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin. Breakdown of the internet under intentional attack. *Physics Review Letters*, 86:3682–3685, 2001.

[14] R. Cohen and S. Havlin. Scale free networks are ultrasmall. *Phys. Rev. Lett.*, 90, 2003.

[15] J.-J. Pansiot D. Magoni. Analysis of the autonomous system network topology. *ACM SIGCOMM Computer Communication Review*, 31(3):26 – 37, July 2001.

[16] L. Dall'Asta, J.I. Alvarez-Hamelin, A. Barrat, A. Vzquez, and A. Vespignan. A statistical approach to the traceroute-like exploration of networks: theory and simulations. In *Workshop on Combinatorial and Algorithmic Aspects of Networking*, 2004.

[17] S.N. Dorogovtsev and J.F.F. Mendes. *Evolution of Networks: From Biological Nets tou the Internet and WWW*. Oxford University Press, 2000.

[18] S.N. Dorogovtsev and J.F.F. Mendes. Evolution of networks. *Advances in Physics*, 51, 2002.

[19] S.N. Dorogovtsev, J.F.F. Mendes, and A.N. Samukhin. Metric structure of random networks. *Nucl. Phys. B 653, 307*, 2003.

[20] Kemal Efe, Vijay Raghavan, C. Henry Chu, Adrienne L. Broadwater, Levent Bolelli, and Seyda Ertekin. The shape of the web and its implications for searching the web, 2000.

[21] P. Erdös and A. Rényi. On random graphs I. *Publications Mathematics Debrecen*, 6:290–297, 1959.

[22] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *Proceedings og the internatonal ACM conference SIGCOMM*, pages 251–262, 1999.

[23] R. Govindan and H. Tangmunarunkit. Heuristics for internet map discovery. In *Proc. IEEE Infocom*, 2000.

[24] Jean-Loup Guillaume, Stevens Le Blond, and Matthieu Latapy. Statistical analysis of a p2p query graph based on degrees and their time-evolution. In *Lecture Notes in Computer Sciences (LNCS), proceedings of the 6-th International Workshop on Distributed Computing (IWDC)*, 2004.

[25] Jean-Loup Guillaume, Stevens Le Blond, and Matthieu Latapy. Clustering in p2p exchanges and consequences on performances. In *Lecture Notes in Computer Sciences (LNCS), proceedings of the 4-th international workshop on Peer-to-Peer Systems (IPTPS)*, 2005.

[26] Jean-Loup Guillaume and Matthieu Latapy. Bipartite graphs as models of complex networks. In *Lecture Notes in Computer Sciences (LNCS), proceedings of the 1-st International Workshop on Combinatorial and Algorithmic Aspects of Networking (CAAN)*, 2004.

[27] Jean-Loup Guillaume and Matthieu Latapy. Bipartite structure of *all* complex networks. *Information Processing Letters (IPL)*, 90(5):215–221, 2004.

[28] Jean-Loup Guillaume and Matthieu Latapy. Complex network metrology. *Complex Systems*, 2005. To appear.

[29] Jean-Loup Guillaume and Matthieu Latapy. Relevance of massively distributed explorations of the internet topology: Simulation results. In *Proceedings of the 24-th IEEE international conference INFOCOM*, 2005.

[30] Jean-Loup Guillaume, Matthieu Latapy, and Clmence Magnien. Comparison of failures and attacks on random and scale-free networks. In *Lecture Notes in Computer Sciences (LNCS), proceedings of the 8-th international conference On Principles Of Distributed Systems (OPODIS)*, 2004.

[31] Ramon Ferrer i Cancho and Ricard V. Sol. The small-world of human language. Technical report, Santa Fe Working paper 01-03-016, 2001.

[32] H. Jeong, B. Tombor, R. Albert, Z. Oltvai, and A. Barabasi. The large-scale organization of metabolic networks. *Nature, 407, 651*, 2000.

[33] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins. The Web as a graph: Measurements, models, and methods. In T. Asano, H. Imai, D. T. Lee, S. Nakano, and T. Tokuyama, editors, *Proc. 5th Annual Int. Conf. Computing and Combinatorics, COCOON*, number 1627. Springer-Verlag, 1999.

[34] A. Lakhina, J. Byers, M. Crovella, and P. Xie. Sampling biases in IP topology measurements. In *Proc. IEEE Infocom*, 2003.

[35] Matthieu Latapy, Nicolas Gast, and Nathalie Del Vecchio. Basic notions for the analysis of large affiliation networks / bipartite graphs. 2005. Submitted.

[36] Matthieu Latapy, Clmence Magnien, Mahendra Mariadassou, and Didier Sverin. A basic toolbox for the analysis of the dynamics of growing networks. 2005. In preparation.

[37] L. Lu. The diameter of random massive graphs. In *Proceedings of the 12-th Annual Symposium on Discrete Algorithms (SODA)*, pages 912–921, 2001.

[38] Clmence Magnien, Matthieu Latapy, and Jean-Loup Guillaume. Impact of random failures and attacks on poisson and power-law random networks. 2005. Submitted.

[39] Damien Magoni and Mickaël Hoerdt. Internet core topology mapping and analysis. *Computer Communications*, 28:494–506, 2005.

[40] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 1995.

[41] Mark E. J. Newman, Stevens H. Strogatz, and Duncan J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physics Reviews E*, 64, 2001. ArXiV preprint `cond-mat/0007235`.

[42] M.E.J. Newman. Scientific collaboration networks: Ii. shortest paths, weighted networks, and centrality. *Phys. Rev. E*, 64, 2001.

[43] M.E.J. Newman. The structure and function of complex networks. *SIAM Review*, 45, 2:167–256, 2003.

[44] M.E.J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 69(2), 2004.

[45] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physcs Review Letters*, 86:3200–3203, 2001.

[46] T. Petermann and P. De Los Rios. Exploration of scale-free networks. *To appear in Eur. Phys. J. B*, 2004.

[47] Pascal Pons and Matthieu Latapy. Computing communities in large networks using random walks. In *Lecture Notes in Computer Sciences (LNCS), proceedings of the 20-th International Symposium on Computer and Information Sciences (ISCIS)*, 2005. To appear.

[48] P. De Los Rios. Exploration bias of complex networks. In *Proceedings of the 7th Conference on Statistical and Computational Physics Granada*, 2002.

[49] Garry Robins, Philippa Pattison, and Jodie Woolcock. Missing data in networks: exponential random graph (p*) models for networks with non-respondents. *Social Networks*, 26(3):257–283, 2004.

[50] N. Spring, R. Mahajan, and D. Wetherall. Measuring ISP topologies with Rocketfuel. In *Proc. ACM SIGCOMM*, 2002.

[51] S.H. Strogatz. Exploring complex networks. *Nature*, 410, 2001.

[52] A. Vazquez, R. Pastor-Satorras, and A. Vespignani. Internet topology at the router and autonomous system level. [cond-mat/0206084].

[53] C. von Mering, R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields, and P. Bork. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417:399–403, 2002.

[54] T. Walsh. Search in a small world. In *Proc. IJCAI*, 1999.

[55] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994. Revised, reprinted edition, 1997.

[56] D. Watts and S. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

[57] H. Peyton Young. The diffusion of innovations in social networks, 2002. Santa Fe Institue working paper 02-04-018.

[58] D. H. Zanette. Critical behavior of propagation on small-world networks. *Physcs Review E*, 64, 2001.