Graph Theory in the Information Age

Fan Chung

In the past decade, graph theory has gone through a remarkable shift and a profound transformation. The change is in large part due to the humongous amount of information that we are confronted with. A main way to sort through massive data sets is to build and examine the network formed by interrelations. For example, Google's successful Web search algorithms are based on the WWW graph, which contains all Web pages as vertices and hyperlinks as edges. There are all sorts of information networks, such as biological networks built from biological databases and social networks formed by email, phone calls, instant messaging, etc., as well as various types of physical networks. Of particular interest to mathematicians is the collaboration graph, which is based on the data from *Mathematical Reviews*. In the collaboration graph, every mathematician is a vertex, and two mathematicians who wrote a joint paper are connected by an edge.

Figure 1 illustrates a portion of the collaboration graph consisting of about 5,000 vertices, representing mathematicians with Erdős number 2 (i.e., mathematicians who wrote a paper with a coauthor of Paul Erdős).

Graph theory has two hundred years of history studying the basic mathematical structures called graphs. A graph G consists of a collection V of vertices and a collection E of edges that connect pairs of vertices. In the past, graph theory has been used in a wide range of areas. However, never before have we confronted graphs of not only such tremendous sizes but also extraordinary richness and complexity, both at a theoretical and a practical level. Numerous challenging problems have attracted the attention and imagination of researchers from physics, computer science, engineering, biology, social science, and mathematics. The new area of "network science" emerged, calling for a sound scientific foundation and rigorous analysis for which graph theory is ideally suited. In the other direction, examples of real-world graphs

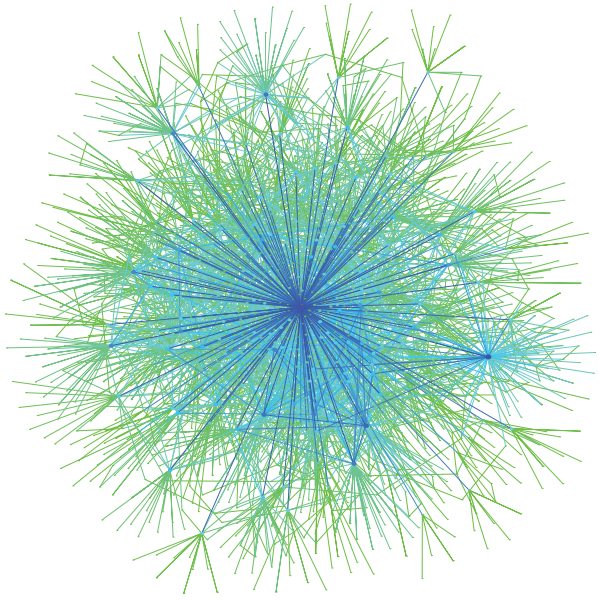

Figure 1. An induced subgraph of the collaboration graph.

Fan Chung is professor of mathematics at the University of California, San Diego. Her email address is fan@ucsd.edu.

* *This article is based on the Noether Lecture given at the AMS-MAA-SIAM Annual Meeting, January 2009, Washington D. C.*

lead to central questions and new directions for research in graph theory.

These real-world networks are massive and complex but illustrate amazing coherence. Empirically, most real-world graphs have the following properties:

- sparsity—The number of edges is within a constant multiple of the number of vertices.
- small world phenomenon—Any two vertices are connected by a short path. Two vertices having a common neighbor are more likely to be neighbors.
- power law degree distribution—The degree of a vertex is the number of its neighbors. The number of vertices with degree j (or having j neighbors) is proportional to $j^{-\beta}$ for some fixed constant β.

To deal with these information networks, many basic questions arise: What are basic structures of such large networks? How do they evolve? What are the underlying principles that dictate their behavior? How are subgraphs related to the large (and often incomplete) host graph? What are the main graph invariants that capture the myriad properties of such large graphs?

To answer these problems, we first delve into the wealth of knowledge from the past, although it is often not enough. In the past thirty years there has been a great deal of progress in combinatorial and probabilistic methods, as well as spectral methods. However, traditional probabilistic methods mostly consider the same probability distribution for all vertices or edges while real graphs are uneven and clustered. The classical algebraic and analytic methods are efficient in dealing with highly symmetric structures, whereas real-world graphs are quite the opposite. Guided by examples of real-world graphs, we are compelled to improvise, extend and create new theory and methods. Here we will discuss the new developments in several topics in graph theory that are rapidly developing. The topics include a general random graph theory for any given degree distribution, percolation in general host graphs, PageRank for representing quantitative correlations among vertices, and the game aspects of graphs.

Random Graph Theory for General Degree Distributions

The primary subject in the study of random graph theory is the classical random graph $G(n,p)$, introduced by Erdős and Rényi in 1959 [38, 39] (also independently by Gilbert [44]). In $G(n,p)$, every pair of a set of n vertices is chosen to be an edge with probability p. In a series of papers, Erdős and Rényi gave an elegant and comprehensive analysis describing the evolution of $G(n,p)$ as p increases. Note that a random graph in $G(n,p)$ has the same expected degree at every vertex, and therefore $G(n,p)$ does not capture some of the main behaviors of real-world graphs. Nevertheless, the approaches and methods in classical random graph theory provide the foundation for the study of random graphs with general degree distributions.

Many random graph models have been proposed in the study of information network graphs, but there are basically two different approaches. The "online" model mimics the growth or decay of a dynamically changing network, and the "offline" model of random graphs consists of specified families of graphs as the probability spaces together with some specified probability distribution.

One online model is the so-called *preferential attachment scheme*, which can be described as "the rich get richer". The preferential attachment scheme has been receiving much attention in the recent study of complex networks [11, 57], but its history can be traced back to Vilfredo Pareto in 1896, among others. At each tick of the clock (so to speak), a new edge is added, with each of its endpoints chosen with probability proportional to their degrees. It can be proved [15, 31, 57] that the preferential attachment scheme leads to a power law degree distribution. There are several other online models, including the duplication model (which seems to be more feasible for biological networks, see [35]), as well as many recent extensions, such as adding more parameters concerning the "talent" or "fitness" of each node [50].

There are two main offline graph models for graphs with general degree distribution—the configuration model and random graphs with expected degree sequences. A random graph in the configuration model with degree sequences $d_1, d_2, \ldots, d_n$ is defined by choosing a random matching on $\sum_i d_i$ "pseudo nodes", where the pseudo nodes are partitioned into parts of sizes d_i, for $i = 1, \ldots, n$. Each part is associated with a vertex. By using results of Molloy and Reed [58, 59], it can be shown [2] that under some mild conditions, a random power law graph with exponent β almost surely has no giant component if $\beta \geq \beta_0$ where β_0 is a solution to the equation involving the Riemann zeta function $\zeta(\beta - 2) - 2\zeta(\beta - 1) = 0$.

The general random graph model $G(\mathbf{w})$ with expected degree sequence $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ follows the spirit of the Erdős-Rényi model. The probability of having an edge between the ith and jth vertices is defined to be $w_i w_j / \mathrm{Vol}\,(G)$, where $\mathrm{Vol}\,(G)$ denotes $\sum_i w_i$. Furthermore, in $G(\mathbf{w})$ each edge is chosen independently of the others, and therefore the analysis can be carried out. It was proved in [28] that if the expected average degree is strictly greater than 1 in a random graph in $G(\mathbf{w})$, then there is a giant component (i.e., a connected component of volume a positive fraction of that of the whole graph). Furthermore,

the giant component almost surely has volume $\delta \mathrm{Vol}\,(G) + O(\sqrt{n}\log^{3.5} n)$, where δ is the unique nonzero root of the following equation [29]:

$$\sum_{i=1}^{n} w_i e^{-w_i \delta} = (1-\delta)\sum_{i=1}^{n} w_i. \tag{1}$$

Because of the robustness of the $G(\mathbf{w})$ model, many properties can be derived. For example, a random graph in $G(\mathbf{w})$ has average distance almost surely equal to $(1+o(1))\frac{\log n}{\log \tilde{w}}$, and the diameter is almost surely $\Theta(\frac{\log n}{\log \tilde{w}})$, where $\tilde{w} = \sum_i w_i^2 / \sum_i w_i$ provided some mild conditions on $\mathbf{w}$ are satisfied [27]. For the range $2 < \beta < 3$, where the power law exponents β for numerous real networks reside, the power law graph can be roughly described as an "octopus" with a dense subgraph having small diameter $O(\log\log n)$ as the core, while the overall diameter is $O(\log n)$ and the average distance is $O(\log\log n)$ (see [31]).

For the spectra of power law graphs, there are basically two competing approaches. One is to prove analogues of Wigner's semicircle law (which is the case for $G(n,p)$), while the other predicts that the eigenvalues follow a power law distribution [40]. Although the semicircle law and the power law have very different descriptions, both assertions are essentially correct if the appropriate matrices associated with a graph are considered [33, 34]. For $\beta > 2.5$, the largest eigenvalue of the adjacency matrix of a random power law graph is almost surely $(1+o(1))\sqrt{m}$, where m is the maximum degree. Moreover, the k largest eigenvalues have power law distribution with exponent $2\beta - 1$ if the maximum degree is sufficiently large and k is bounded above by a function depending on β, m and w. When $2 < \beta < 2.5$, the largest eigenvalue is heavily concentrated at $cm^{3-\beta}$ for some constant c depending on β and the average degree. Furthermore, the eigenvalues of the (normalized) Laplacian satisfy the semicircle law under the condition that the minimum expected degree is relatively large [34].

The online model is obviously much harder to analyze than the offline model. One possible approach is to couple the online model with the offline model of random graphs with a similar degree distribution. This means to find the appropriate conditions under which the online model can be sandwiched by two offline models within some error bounds. In such cases, we can apply the techniques from the offline model to predict the behavior of the online model (see [30]).

Random Subgraphs in Given Host Graphs

Almost all information networks that we observe are subgraphs of some host graphs that often have sizes prohibitively large or with incomplete information. A natural question is to attempt to deduce the properties of a random subgraph from the host graph and vice versa. It is of interest to understand the connections between a graph and its subgraphs. What invariants of the host graph can or cannot be translated to its subgraphs? Under what conditions can we predict the behavior of all or any subgraphs? Can a sparse subgraph have very different behavior from its host graph? Here we discuss some of the work in this direction.

Many information networks or social networks have very small diameters (in the range of $\log n$), as dictated by the so-called small world phenomenon. However, in a recent paper by Liben-Nowell and Kleinberg [52], it was observed that the tree-like subgraphs derived from some chain-letter data seem to have relatively large diameter. In the study of the Erdős-Rényi graph model $G(n,p)$, it was shown [60] that the diameter of a random spanning tree is of order $\sqrt{n}$, in contrast with the fact that the diameter of the host graph K_n is 1. Aldous [4] proved that in a regular graph G with a certain spectral bound σ, the diameter of a random spanning tree T of G, denoted by $diam(T)$, has expected value satisfying

$$\frac{c\sigma\sqrt{n}}{\log n} \le \mathrm{E}(diam(T)) \le \frac{c\sqrt{n}\log n}{\sqrt{\sigma}}$$

for some absolute constant c. In [32], it was shown that for a general host graph G, with high probability the diameter of a random spanning tree of G is between $c\sqrt{n}$ and $c'\sqrt{n}\log n$, where c and c' depend on the spectral gap of G and the ratio of the moments of the degree sequence.

One way to treat random subgraphs of a given graph G is as a (bond) percolation problem. For a positive value $p \le 1$, we consider G_p, which is formed by retaining each edge independently with probability p and discarding the edge with probability $1-p$. A fundamental problem of interest is to determine the critical probability p for which G_p contains a giant connected component. In the applications of epidemics, we consider a general host graph being a contact graph, consisting of edges formed by pairs of people with possible contact. The question of determining the critical probability then corresponds to the problem of finding the epidemic threshold for the spreading of the disease.

Percolation problems have long been studied [45, 49] in theoretical physics, especially with the host graph being the lattice graph $\mathbb{Z}^k$. Percolation problems on lattices are known to be notoriously difficult even for low dimensions and have only been resolved very recently by bootstrap percolation [8, 9]. In the past, percolation problems have been examined for a number of special host graphs. Ajtai, Komlós, and Szemerédi considered the percolation on hypercubes [3]. Their work was further extended to Cayley graphs [16, 17, 18, 55] and regular graphs [42]. For expander graphs with degrees bounded by d, Alon, Benjamini, and Stacey

[5] proved that the percolation threshold is greater than or equal to $1/(2d)$. In the other direction, Bollobás, Borgs, Chayes, and Riordan [13] showed that for dense graphs (where the degrees are of order $\Theta(n)$), the giant component threshold is $1/\rho$ where ρ is the largest eigenvalue of the adjacency matrix. The special case of having the complete graph K_n as the host graph concerns the Erdős-Rényi graph $G(n,p)$, which is known to have the critical probability at $1/n$, as well as the "double jump" near the threshold.

For general host graphs, the answer has been elusive. One way to address such questions is to search for appropriate conditions on the host graph so that percolations can be controlled. Recently it has been shown [26] that if a given host graph G satisfies some (mild) conditions depending on its spectral gap and higher moments of its degree sequence, for any $\epsilon > 0$, if $p > (1+\epsilon)/\tilde{d}$, then asymptotically almost surely the percolated subgraph G_p has a giant component. In the other direction, if $p < (1-\epsilon)/\tilde{d}$, then almost surely the percolated subgraph G_p contains no giant component. We note that the second order average degree $\tilde{d}$ is $\tilde{d} = \sum_v d_v^2/(\sum_v d_v)$, where d_v denotes the degree of v.

In general, subgraphs can have spectral gaps very different from those of the host graph. However, if a graph G has all its nontrivial eigenvalues of the (normalized) Laplacian lying in the range within σ from the value 1, then it can be shown [25] that almost surely a random subgraph G_p has all its nontrivial eigenvalues in the same range (up to a lower-order term) if the degrees are not too small.

PageRank and Local Partitioning

In graph theory there are many essential geometrical notions, such as distances (typically, the number of hops required to reach one vertex from another), cuts (i.e., subsets of vertices/edges that separate a part of the graph from the rest), flows (i.e., combinations of paths for routing between given vertices), and so on. However, real-world graphs exhibit the small world phenomenon, so any pair of vertices are connected through a very short path. Therefore the usual notion of graph distance is no longer very useful. Instead, we need a quantitative and precise formulation to differentiate among nodes that are "local" from "global" and "akin" from "dissimilar". This is exactly what PageRank is meant to achieve.

In 1998 Brin and Page [19] introduced the notion of PageRank for Google's Web search algorithm. Different from the usual methods in pattern matching previously used in data retrieval, the novelty of PageRank relies entirely on the underlying Web graph to determine the "importance" of a Web page. Although PageRank is originally designed for the Web graph, the concept and definitions work well for any graph. Indeed, PageRank has become a valuable tool for examining the correlations of pairs of vertices (or pairs of subsets) in any given graph and hence leads to many applications in graph theory.

The starting point of the PageRank is a typical random walk on a graph G with edge weights w_{uv} for edge u,v. The probability transition matrix P is defined by: $P(u,v) = \frac{w_{uv}}{d_u}$, where $d_u = \sum_v w_{u,v}$. For a *preference vector* s, and a *jumping constant* $\alpha > 0$, the PageRank, denoted by $\mathrm{pr}(\alpha, s)$ as a row vector, can be expressed as a series of random walks as follows:

$$\mathrm{pr}_{\alpha,f} = \alpha \sum_{k=0}^{\infty} (1-\alpha)^k s P^k. \tag{2}$$

Equivalently, $\mathrm{pr}(\alpha, s)$ satisfies the following recurrence relation:

$$\mathrm{pr}(\alpha, s) = \alpha s + (1-\alpha)\mathrm{pr}(\alpha, s)P. \tag{3}$$

In the original definition of Brin and Page [19], s is taken to be the constant function with value $1/n$ at every vertex motivated by modeling the behavior of a typical surfer who moves to a random page with probability α and clicks a linked page with probability $1-\alpha$.

Because of the close connection of PageRank with random walks, there are very efficient and robust algorithms for computing and approximating PageRank [6, 12, 47]. This leads to numerous applications, including the basic problem of finding a "good" cut in a graph. A quantitative measure for the "goodness" of a cut that separates a subset S of vertices is the Cheeger ratio:

$$h(S) = \frac{|E(S, \bar{S})|}{\mathrm{vol}\,(S)},$$

where $E(S,\bar{S})$ denotes the set of edges leaving S and $\mathrm{vol}\,(S) = \sum_{v \in S} d_v$. The *Cheeger constant* h_G of a graph is the minimum Cheeger ratio over all subsets S with $\mathrm{vol}\,(S) \le \mathrm{vol}\,(G)/2$. The traditional divide-and-conquer strategy in algorithmic design relies on finding a cut with small Cheeger ratio. Since the problem of finding any cut that achieves the Cheeger constant of G is NP-hard [43], one of the most widely used approximation algorithms was a spectral partitioning algorithm. By using eigenvectors to line up the vertices, the spectral partitioning algorithm reduces the number of cuts under consideration from an exponential number of possibilities to a linear number of choices. Nevertheless, there is still a performance guarantee provided by the Cheeger inequality:

$$2h_G \ge \lambda \ge \frac{h_f^2}{2} \ge \frac{h_G^2}{2},$$

where h_f is the minimum Cheeger ratio among subsets that are initial segments in the order

determined by the eigenvector f associated with the spectral gap λ.

For large graphs with billions of nodes, it is not feasible to compute eigenvectors. In addition, it is of interest to have *local cuts* in the sense that for given seeds and the specified size for the parts to be separated, it is desirable to find a cut near the seeds separating a subset of the desired size. Furthermore, the cost/complexity of finding such a cut should be proportional to the specified size of the separated part but independent of the total size of the whole graph. Here, PageRank comes into play. Earlier, Spielman and Teng [63] introduced local partitioning algorithms by using random walks with the performance analysis using a mixing result of Lovász and Simonovitz [54] (also see [56]). As it turns out, by using PageRank instead of random walks, there is an improved partitioning algorithm [6] for which the performance is supported by a *local Cheeger inequality* for a subset S of vertices in a graph G:

$$h_S \geq \lambda_S \geq \frac{h_g^2}{8 \log \mathrm{vol}\ (S)} \geq \frac{h_S^2}{8 \log \mathrm{vol}\ (S)},$$

where λ_S is the Dirichlet eigenvalue of the induced subgraph on S, h_S is the local Cheeger constant of S defined by $h_S = \min_{T \subseteq S} h(T)$, and h_g is the minimum Cheeger ratio over all PageRank g with the seed as a vertex in S and α appropriately chosen depending only on the volume of S. This approximation partition algorithm can be further improved using the fact that the set of seeds for which the PageRank leads to the Cheeger ratio satisfying the above local Cheeger inequality is quite large (about half of the volume of S). We note that the local partitioning algorithm can also be used as a subroutine for finding balanced cuts for the whole graph.

Note that PageRank is expressed as a geometric sum of random walks in (2). Instead, we can consider an exponential sum of random walks, called heat kernel pagerank, which in turn satisfies the heat equation. The heat kernel pagerank leads to an improved local Cheeger inequality [23, 24] by removing the logarithmic factor in the lower bound. Numerous problems in graph theory can possibly take advantage of PageRank and its variations, and the full implications of these ideas remain to be explored.

Network Games

In morning traffic, every commuter chooses his/her most convenient way to get to work without paying attention to the consequences of the decision to others. The Internet network can be viewed as a similar macrocosm that functions neither by the control of a central authority nor by coordinated rules. The basic motivation for each individual can only be deduced by greed and selfishness. Every player chooses the most convenient route and uses strategies to maximize possible payoff. In other words, we face a combination of game theory and graph theory for dealing with large networks both in quantitative analysis and algorithm design. Many questions arise. Instead of just proving the existence of *Nash equilibrium*, we would like to design algorithms to effectively compute or approximate the Nash equilibrium. How rapidly can such algorithms converge? There has been a great deal of progress in the computational complexity of Nash equilibrium [22, 37].

The analysis of *selfish routing* comes naturally in network management. How much does uncoordinated routing affect the performance of the network, such as stability, congestion, and delay? What are the trade-offs for some limited regulation? The so-called *price of anarchy* refers to the worst-case analysis to evaluate the loss of collective welfare from selfish routing. There has been extensive research done on selfish routing [62]. The reader is referred to several surveys [41, 51] and some recent books on this topic [61].

Many classical problems in graph theory can be reexamined from the perspective of game theory. One popular topic on graphs is chromatic graph theory. For a given graph G, what is the minimum number of colors needed to color the vertices of G so that adjacent vertices have different colors? In addition to theoretical interests, the graph coloring problem has numerous applications in the setting of conflict resolution. For example, each faculty member (as a vertex) wishes to schedule classes in a limited number of classrooms (as colors). Two faculty members who have classes with overlapping time are connected by an edge, and then the problem of classroom scheduling can be viewed as a graph coloring problem. Instead of having a central agency to make assignments, we can imagine a game-theoretic scenario that the faculty members coordinate among themselves to decide a nonconflicting assignment. Suppose there is a payoff of 1 unit for each player (vertex) if its color is different from all its neighbors. A proper coloring is then a Nash equilibrium, since no player has an incentive to change his/her strategy.

Kearns et al. [48] conducted an experimental study of several coloring games on specified networks. Many examples were given to illustrate the difficulties in analyzing the dynamics of large networks in which each node takes simple but selfish steps. This calls for rigorous analysis, especially along the line of the combinatorial probabilistic methods and generalized Martingale approaches that have been developed in the past ten years [20]. Some work in this direction has been done on a multiple round model of graph coloring games [20], but more work is needed.

Summary

It is clear that we are at the beginning of a new journey in graph theory, emerging as a central part of the information revolution. It is a long way from the "seven bridges of Königsberg", a problem posed by Leonhard Euler in 1736. In contrast to its origin in recreational mathematics, graph theory today uses sophisticated combinatorial, probabilistic, and spectral methods with deep connections with a variety of areas in mathematics and computer science. In this article, some vibrant new directions in graph theory have been selected and described to illustrate the richness of the mathematics involved, as well as the utilization through major threads of current technology. The list of the sampled topics is by no means complete, since these areas of graph theory are still rapidly developing. Abundant opportunities in research, theoretical and applied, remain to be explored.

References

[1] D. Achlioptas, R. M. D'Souza, and J. Spencer, Explosive percolation in random networks, *Science* **323**, no. 5920, (2009), 1453–1455.

[2] W. Aiello, F. Chung, and L. Lu, A random graph model for massive graphs, *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, ACM Press, New York, 2000, pp. 171–180.

[3] M. Ajtai, J. Komlós, and E. Szemerédi, Largest random component of a k-cube, *Combinatorica* **2** (1982), 1–7.

[4] D. Aldous, The random walk construction of uniform spanning trees, *SIAM J. Discrete Math.* **3** (1990), 450–465.

[5] N. Alon, I. Benjamini, and A. Stacey, Percolation on finite graphs and isoperimetric inequalities, *Annals of Probability* **32**, no. 3 (2004), 1727–1745.

[6] R. Andersen, F. Chung, and K. Lang, Local graph partitioning using pagerank vectors, *Proceedings of the 47th Annual IEEE Symposium on Foundation of Computer Science (FOCS'2006)*, pp. 475–486.

[7] ———, Detecting sharp drops in PageRank and a simplified local partitioning algorithm, *Theory and Applications of Models of Computation, Proceedings of TAMC 2007, LNCS* **4484**, Springer Berlin/Heidelberg, (2007), pp. 1–12.

[8] J. Balogh, B. Bollobás, and R. Morris, Bootstrap percolation in three dimensions, *Annals of Probability*, to appear.

[9] ———, The sharp threshhold for r-neighbour bootstrap percolation, preprint.

[10] Y. Bao, G. Feng, T.-Y. Liu, Z.-M. Ma, and Y. Wang, Describe importance of websites in probabilistic view, *Internet Math.*, to appear.

[11] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, *Science* **286** (1999), 509–512.

[12] P. Berkhin, Bookmark-coloring approach to personalized pagerank computing, *Internet Math.*, to appear.

[13] B. Bollobás, C. Borgs, J. Chayes, and O. Riordan, Percolation on dense graph sequences, preprint.

[14] B. Bollobás, Y. Kohayakawa, and T. Łuczak, The evolution of random subgraphs of the cube, *Random Structures and Algorithms* **3**(1), 55–90, (1992).

[15] B. Bollabás and O. Riordan, Robustness and vulnerability of scale-free random graphs, *Internet Math.* **1** (2003) no. 1, 1–35.

[16] C. Borgs, J. Chayes, R. van der Hofstad, G. Slade, and J. Spencer, Random Subgraphs of Finite Graphs: I. The scaling window under the triangle condition, *Random Structures and Algorithms* **27** (2005), 137–184.

[17] ———, Random subgraphs of finite graphs: II. The lace expansion and the triangle condition, *Annals of Probability* **33** (2005), 1886–1944.

[18] ———, Random subgraphs of finite graphs: III. The phase transition for the n-cube, *Combinatorica* **26** (2006), 395–410.

[19] S. Brin and L. Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems* **30** (1–7), (1998), 107–117.

[20] K. Chaudhuri, F. Chung, and M. S. Jamall, A network game, *Proceedings of WINE 2008*, *Lecture Notes in Computer Science*, Vol. 5385, Springer Berlin/Heidelberg (2008), 522–530.

[21] J. Cheeger, A lower bound for the smallest eigenvalue of the Laplacian, *Problems in Analysis* (R. C. Gunning, ed.), Princeton Univ. Press, Princeton, NJ, 1970, pp. 195–199.

[22] X. Chen and X. Deng, Settling the complexity of 2-player Nash-equilibrium, *The 47th Annual IEEE Symposium on Foundations of Computer Science*, FOCS 2006, pp. 261–272.

[23] F. Chung, The heat kernel as the pagerank of a graph, *Proc. Nat. Acad. Sciences* **105** (50) (2007), 19735–19740.

[24] ———, A local graph partitioning algorithm using heat kernel pagerank, *WAW 2009*, *Lecture Notes in Computer Science*, vol. 5427, Springer, 2009, pp. 62–75.

[25] F. Chung and P. Horn, The spectral gap of a random subgraph of a graph, Special issue of WAW2006, *Internet Math* **2** (2007), 225–244.

[26] F. Chung, P. Horn, and L. Lu, The giant component in a random subgraph of a given graph, *Proceedings of WAW2009*, *Lecture Notes in Computer Science*, vol. 5427, Springer, 2009, 38–49.

[27] F. Chung and L. Lu, The average distances in random graphs with given expected degrees, *Proceedings of National Academy of Sciences* **99** (2002), 15879–15882.

[28] ———, Connected components in random graphs with given expected degree sequences, *Annals of Combinatorics* **6** (2002), 125–145.

[29] ———, The volume of the giant component of a random graph with given expected degrees, *SIAM J. Discrete Math.* **20** (2006), 395–411.

[30] ———, Coupling online and offline analyses for random power law graphs, *Internet Math.* **1** (2004), 409–461.

[31] ———, *Complex Graphs and Networks*, CBMS Lecture Series, No. 107, AMS Publications, 2006, vii + 264 pp.

[32] F. Chung, P. Horn, and L. Lu, Diameter of random spanning trees in a given graph, preprint.

[33] F. CHUNG, L. LU, and V. VU, Eigenvalues of random power law graphs, *Annals of Combinatorics* **7** (2003), 21–33.
[34] ———, The spectra of random graphs with given expected degrees, *Proceedings of National Academy of Sciences* **100**, no. 11 (2003), 6313–6318.
[35] F. CHUNG, L. LU, G. DEWEY, and D. J. GALAS, Duplication models for biological networks, *J. Computational Biology* **10**, no. 5 (2003), 677–687.
[36] F. CHUNG and S.-T. YAU, Coverings, heat kernels and spanning trees, *Electronic Journal of Combinatorics* **6** (1999), #R12.
[37] C. DASKALAKIS, P. GOLDBERG and C. PAPADIMITRIOU, Computing a Nash equilibrium is PPAD-complete, to appear in *SIAM J. on Computing*.
[38] P. ERDŐS and A. RÉNYI, On random graphs, I. *Publ. Math. Debrecen* **6** (1959), 290–297.
[39] ———, On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kutató Int. Közl.* **5** (1960), 17–61.
[40] M. FALOUTSOS, P. FALOUTSOS, and C. FALOUTSOS, On power-law relationships of the Internet topology, *Proceedings of the ACM SIGCOM Conference*, ACM Press, New York, 1999, 251–262.
[41] R. FELDMANN, M. GAIRING, T. LUCKING, B. MONIEN, and M. RODE, Selfish routing in noncooperative networks: A survey, *Lecture Notes in Computer Science*, vol. 2746, Springer, Berlin/Heidelberg, 2003, pp. 21–45.
[42] A. FRIEZE, M. KRIVELEVICH, AND R. MARTIN, The emergence of a giant component of pseudo-random graphs, *Random Structures and Algorithms* **24**, ACM New York (2004), 42–50.
[43] M. R. GAREY and D. S. JOHNSON, *Computers and Intractability, A Guide to the Theory of NP-Completeness*, W. H. Freeman and Co., San Francisco, 1979, x + 338 pp.
[44] E. N. GILBERT, Random graphs, *Annals of Mathematical Statistics* **30** (1959), 1141–1144.
[45] G. GRIMMETT, *Percolation*, Springer, New York, 1989.
[46] J. HOPCROFT and D. SHELDON, Manipulation-resistant reputations using hitting time, *WAW 2007*, *Lecture Notes in Computer Science*, vol. 4863, Springer, Berlin/Heidelberg, 2007, pp. 68–81.
[47] G. JEH and J. WIDOM, Scaling personalized web search, *Proceedings of the 12th World Wide Web Conference WWW*, ACM Press, New York, 2003, pp. 271–279.
[48] M. KEARNS, S. SURI, and N. MONTFORT, An experimental study of the coloring problem on human subject networks, *Science* **322**, no. 5788 (2006), 824–827.
[49] H. KESTEN, *Percolation theory for mathematicians*, volume 2 of *Progress in Probability and Statistics*, Birkhäuser, Boston, MA., 1982.
[50] J. S. KONG, N. SARSHAR, and V. P. ROYCHOWDHURY, Experience versus talent shapes the structure of the Web, *PNAS* **105**, no. 37 (2008), 13724–13729.
[51] S. KONTOGIANNIS and P. SPIRAKIS, Atomic selfish routing in networks: a survey, *Lecture Notes in Computer Science*, vol. 3828, Springer, Berlin/Heidelberg, 2005, pp. 989–1002.
[52] DAVID LIBEN-NOWELL and JON KLEINBERG, Tracing information flow on a global scale using Internet, *PNAS* **105**, no. 12 (2008), 4633–4638.
[53] L. LOVÁSZ, Random walks on graphs, *Combinatorics, Paul Erdős Is Eighty*, Vol. 2, Bolyai Society Mathematical Studies, 2, Keszthely (Hungary), 1993, 1–46.
[54] L. LOVÁSZ and M. SIMONOVITS, Random walks in a convex body and an improved volume algorithm, *Random Structures and Algorithms* **4** (1993), 359–412.
[55] C. MALON and I. PAK, Percolation on finite Cayley graphs, *Lecture Notes in Computer Science* **2483**, Springer, Berlin, 2002, pp. 91–104.
[56] M. MIHAIL, Conductance and convergence of Markov chains: A combinatorial treatment of expanders, *Foundation of Computer Science*, (1989), 526–531.
[57] M. MITZENMACHER, A brief history of generative models for power law and lognormal distributions, *Internet Math.* **1** (2004), 226–251.
[58] M. MOLLOY and B. REED, A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms* **6** (1995), 161–179.
[59] M. MOLLOY and B. REED, The size of the giant component of a random graph with a given degree sequence, *Combin. Probab. Comput.* **7** (1998), 295–305.
[60] A. RÉNYI and G. SZEKERES, On the height of trees, *J. Austral. Math. Soc.* **7** (1967), 497–507.
[61] T. ROUGHGARDEN, *Selfish Routing and the Price of Anarchy*, MIT Press, 2006.
[62] T. ROUGHGARDEN and E. TARDOS, How bad is selfish routing? *Journal of the ACM* **49** (2002), 236–259.
[63] D. SPIELMAN and S.-H. TENG, Nearly-linear time algorithms for graph partitioning, graph sparsification, and solving linear systems, *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, ACM Press, New York 2004, pp. 81–90.