(3) A random graph model for massive graphs

Many graphs that arise in Internet computing or large data sets have sizes prohibitively or only partial information are available. As a result, new problems and research directions emerge. One of such examples is to model massive graphs using random graph techniques. Random graphs share many similarities with large, sparse graphs arising in various applications. For example, suppose we consider a telephony graph made of vertices as users and each new call is associated with an edge between the users. Such an evolving process has a similar flavor as the evolution of random graphs. Furthermore, a subgraph of such a graph is still somewhat random-like, which is another characteristic of random graphs. However, there are many other distinct features of such "realistic" graphs. For example, they often have vertices of large degrees while random graphs are almost regular. The usual techniques can not be easily applied.

Recently, however, some structure of the web graphs has come to light which may enable us to describe graph properties of the massive graphs qualitatively. Several research groups [1, 4, 10, 13, 12] independently measured the degree sequences of the Web and showed that it is well approximated by a power law distribution. That is, the number of nodes, y, of a given degree x is proportional to $x^{-\beta}$ for some constant $\beta > 0$. In fact, the power law distribution of the degree sequence may be a ubiquitous characteristic, applying to many massive real world graphs.

In [2], we investigated a random graph model for graphs with a power law degree sequence. Our *power law* random graph model also has two parameters, α (called logsize) and β (called log growthrate). Let y be the number of nodes with degree x. $P(\alpha, \beta)$ denote a random power law graph, chosen with equal probability from all graphs with $y = e^{\alpha}/x^{\beta}$ (where self loops are allowed).

In [2], we study the connectivity properties of $P(\alpha, \beta)$ as a function of the power β which hold almost surely for sufficiently large graphs. Briefly, we show that when $\beta < 1$, the graph is almost surely connected. For $1 < \beta < 2$ there is a giant component, i.e., a component of size $\Theta(n)$. Moreover, all smaller components are of size $O(1)$. For $2 < \beta < \beta_0 = 3.4785$ there is a giant component and all smaller components are of size $O(\log n)$. For $\beta = 2$ the smaller components are of size $O(\log n / \log \log n)$. For $\beta > \beta_0$ the graph almost surely has no giant component. In addition we derive several results on the sizes of the second largest component. For example, we show that for

$\beta > 4$ the numbers of components of given sizes can be approximated by a power law as well.

Our model is a special case of random graphs with a given degree sequence for which there have studied in a number of papers. For example, Wormald [19] studied the connectivity of graphs whose degrees are in an interval $[r, R]$, where $r \geq 3$. Łuczak [15] considered the asymptotic behavior of the largest component of a random graph with given degree sequence as a function of the number of vertices of degree 2. His result was further improved by Molloy and Reed [16, 17]. They consider a random graph on n vertices with the following degree distribution. The number of vertices of degree $0, 1, 2, \ldots$ are about $\lambda_0 n, \lambda_1 n, \ldots$ respectively, where the λ's sum to 1. It is shown in [16] that if $Q = \sum_i i(i-2)\lambda_i > 0$ and the maximum degree is not too large, then such random graphs have a giant component with probability tending to 1 as n goes to infinity, while if $Q < 0$ then all components are small with probability tending to 1 as $n \to \infty$. They also examined the threshold behavior of such graphs. In this paper, we will apply these techniques to deal with the special case that applies to our model. However, since the power law graphs have vertices with large degree, the above results and methods can not be directly applied.

There are numerous questions that remain to be studied. For example, what is the effect of time scaling? Indeed, massive graphs are usually dynamically evolved. How does it correspond with the evolution of β? What are the structural behaviors of the call graphs? What are the correlations between the directed and undirected graphs? It is of interest to understand the phase transition of the giant component, the diameters and other invariants in the massive graphs.

References

[1] J. Abello, A. Buchsbaum, and J. Westbrook, A functional approach to external graph algorithms, *Proc. 6th European Symposium on Algorithms,* pp. 332–343, 1998.

[2] W. Aiello, F. Chung, L. Lu, A random graph model for massive graphs, *Proceedings of the Thirtysecond Annual ACM Symposium on Theory of Computing*, (2000), to appear.

[3] W. Aiello, F. Chung, L. Lu, Random evolution of power law graphs, manuscript.

[4] R. Albert, H. Jeong and A. Barabási, Diameter of the World Wide Web, *Nature*, **401**, September 9, 1999.

[5] N. Alon and J. H. Spencer, *The Probabilistic Method*, Wiley and Sons, New York, 1992.

[6] A. Barabási, and R. Albert, Emergence of scaling in random networks, *Science*, **286**, October 15, 1999.

[7] A. Barabási, R. Albert, and H. Jeong Scale-free characteristics of random networks: the topology of the world wide web, *Elsevier Preprint* August 6, 1999.

[8] P. Erdős and A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960), 17–61.

[9] P. Erdős and A. Rényi, On the strength of connectedness of random graphs, *Acta Math. Acad. Sci. Hungar.* **12** (1961), 261-267.

[10] M. Faloutsos, P. Faloutsos, and C. Faloutsos, On power-law relationships of the internet topology, *Proceedings of the ACM SIGCOM Conference*, Cambridge, MA, 1999.

[11] Brian Hayes, Graph theory in practice: Part II, *American Scientists*, **88**, (March-April, 2000), 104-109.

[12] J. Kleinberg, S. R. Kumar, P. Raphavan, S. Rajagopalan and A. Tomkins, The web as a graph: Measurements, models and methods, *Proceedings of the International Conference on Combinatorics and Computing*, July 26–28, 1999.

[13] S. R. Kumar, P. Raphavan, S. Rajagopalan and A. Tomkins, Trawling the web for emerging cyber communities, *Proceedings of the 8th World Wide Web Conference*, Edinburgh, Scotland, May 15–19, 1999.

[14] S. R. Kumar, P. Raghavan, S. Rajagopalan and A. Tomkins, Extracting large-scale knowledge bases from the web, *Proceedings of the 25th VLDB Conference*, Edinburgh, Scotland, September 7–10, 1999.

[15] Tomasz Łuczak, Sparse random graphs with a given degree sequence, *Random Graphs*, vol 2 (Poznań, 1989), 165-182, Wiley, New York, 1992.

[16] Michael Molloy and Bruce Reed, A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, Vol. **6**, no. 2 and 3 (1995). 161-179.

[17] Michael Molloy and Bruce Reed, The size of the giant component of a random graph with a given degree sequence, *Combin. Probab. Comput.* **7**, no. (1998), 295-305.

[18] P. Raghavan, personal communication.

[19] N. C. Wormald, The asymptotic connectivity of labelled regular graphs, *J. Comb. Theory* (B) **31** (1981), 156-167.

[20] N. C. Wormald, Models of random regular graphs, *Surveys in Combinatorics*, 1999 (LMS Lecture Note Series 267, Eds J.D.Lamb and D.A.Preece), 239–298.

[21]