ANALYSIS OF RANDOM GRAPHS WITH METHODS OF MARTINGALE THEORY

Summary of PhD thesis

Ágnes Backhausz

Supervisor:

Tamás Móri CSc, associate professor

Department of Probability Theory and Statistics, Institute of Mathematics

Faculty of Science, Eötvös Loránd University

Faculty of Science, Eötvös Loránd University, Budapest

Doctoral School of Mathematics (Leader: Laczkovich Miklós)

Mathematics Doctoral Program (Leader: Szűcs András)

2012

1 Preliminaries

The first famous applications of the probabilistic method to the problems of graph theory are due to Pál Erdős and Alfréd Rényi [11, 12] and Edgar Gilbert [14] from the end of the fifties. However, random graph models had already appeared earlier, for example in the paper of Yule from 1925 [23] or in the work of Simon from 1955 [21]. The number of vertices is given both models (i.e. the Erdős–Rényi and the Gilbert model), and the edges are chosen randomly. The asymptotic behaviour of the connectivity of the random graph and the size of its connected components were examined as the number of vertices tended to infinity. The models were used to prove certain theorems of combinatorics, for example, to give a lower bound on Ramsey numbers.

Several decades later, when it was possible to collect data from large real networks, random graphs had a new role in modelling large networks. In 1999 Albert-László Barabási and Réka Albert analysed real networks, for example, that of the links between pages of the internet [5]. It turned out that for large d the ratio of vertices of degree d is approximately $C \cdot d^{-\gamma}$, where the value of the characteristic exponent γ is usually between 2 and 3. This showed that the models of Erdős–Rényi type are not the best for modelling real networks, because the decay of this so called asymptotic degree distribution is faster than exponential in this case, and polynomial is needed. Hence they proposed a family of models evolving randomly in time. Step by step a new vertex is added to the system. We choose its neighbours randomly from the old vertices. The probability that a given old vertex of degree d gets a new edge is proportional to d. They conjectured that these models have the scale free property, i.e., for large d the ratio of vertices of degree d decays polynomially [5]. The precise formulation of the model and the mathematical proofs are in the paper of Bollobás, Riordan, Spencer and Tusnády from 2001 [8].

Later on, many random graph models were constructed, partly for modelling real networks, and the scale free property turned out to be important in several cases. We mention the works of Cooper and Frieze [9], Deijfen, van den Esken, van der Hofstad and Hooghiemstra [10], Katona, Móri [16, 19], and finally, Sridharan, Gao, Wu and Nastos [22].

The main goal of the research was understanding the local behaviour of certain scale free random graph models and analysing some new models using the methods of martingale theory. We included interactions of more than two vertices in the new models. In most of the already known models the evolution of the graph is based on the degrees of the old vertices, and it is not taken into account which groups of vertices get new edges in the same step.

2 Methods

Most of our questions concerning random graphs were about the asymptotic behaviour of the degrees or the degree distribution. On the other hand, we assumed in all the models that only one new vertex is born in a given step, and old vertices of the same degree (or weight) are connected to it with the same probability (but naturally not always independently). This yields that we may construct martingales using the degree of a given vertex or the number of vertices of a given degree: the degree of the new vertex is the sum of certain indicators, and hence the conditional expectations may be written as a sum.

Therefore the methods of the proofs often originate from martingale theory. The basic results that we used may be found in the book of Neveu [20]. Some additional tools are in some papers of Tamás Móri [16, 19, 17, 18]. We used the following proposition most frequently.

Proposition 2.3 *Suppose that* $(M_n, \mathcal{G}_n)$ *is a square integrable nonnegative submartingale, and*

$$A_n = EM_1 + \sum_{i=2}^{n} \left(E\left(M_i \middle| \mathcal{G}_{i-1} \right) - M_{i-1} \right), \quad B_n = \sum_{i=2}^{n} D^2 \left(M_i \middle| \mathcal{G}_{i-1} \right).$$

If $B_n^{1/2} \log B_n = O\left(A_n\right)$*, then* $M_n \sim A_n$ *holds as* $n \to \infty$ *almost everywhere on the event* $\{A_n \to \infty\}$*.*

In addition, we used some results on urn models. However, these are often proved by martingales, e.g. in the paper of Gouet [15].

Results on the asymptotic behaviour of slowly and regularly varying sequences were also applied. These are from the papers of Bojanić and Seneta [7], Galambos and Seneta [13], and from the book of Bingham [6].

3 Main results

We deal with random graph models in which a new vertex is born at each step, and it is connected to some of the old vertices randomly. In many cases old vertices with larger degree get new edges with larger probability. This is the so called preferential attachment property.

All of our models has asymptotic degree distribution. That is, the proportion of vertices of degree d is convergent almost surely as the number of steps goes to infinity. In many of the models the sequence of the limits is polynomially decaying, so we found or supposed the scale free property. In one of the models not the degrees, but weights of the vertices determine the dynamics.

3.1 Local asymptotic degree distribution

In Chapter 2 we looked for local asymptotic degree distribution in random graph models with asymptotic degree distribution or with scale free property. More precisely, if we select some vertices in a certain way and calculate the proportion of vertices of degree d, then this proportion may asymptotically differ from the limit of the proportion of vertices of degree d in the whole graph. We mean almost sure convergence here again. The degrees are always counted in the whole graph.

For example, the set of selected vertices may consist of the neighbours of a given vertex, or of the vertices that are at a given distance of a given vertex, or of the common neighbours of some fixed vertices. In Chapter 2 we gave sufficient conditions for the existence of the local asymptotic degree distribution defined above. Wa also showed how it is determined by the quantities describing the graph model and the set of selected vertices. This is a model free result: neither the model nor the selection method is needed exactly. We found sufficient conditions for the following.

Theorem 2.1 *[3, Theorem 1] Suppose that the conditions hold for the sequence of random graphs and set of selected vertices* (G_n, S_n)*. Denote by* $X^*[n,d]$ *the number of vertices of degree* d *in* S_n*, which is the set of selected vertices after* n *steps. Let* m *denote the minimal possible degree of the new vertex. Then the limits*

$$\lim_{n\to\infty} \frac{X^*[n,d]}{|S_n|} = x_d$$

exist and are positive for all $d \geq m$*-re almost surely.*

The sequence (x_d) *satisfies the following recurrence equation:*

$$x_m = \frac{\alpha\, q_m}{\alpha + \frac{p_m - c_m}{c_m}}\,, \qquad x_d = \frac{x_{d-1}\cdot \frac{k_{d-1}}{c_{d-1}} + \alpha \cdot q_d}{\alpha + \frac{k_d}{c_d}} \quad (d \geq m+1). \tag{1}$$

The sequence (x_d) *is a probability distribution, that is, it sums up to* 1*. Moreover,* $x_d \sim L \cdot d^{-\gamma^*}$ *holds as* $d \to \infty$ *with some* $L > 0$ *and the following exponent:*

$$\gamma^* = \alpha\,(\gamma - 1) + 1.$$

Here the sequences (c_d) and (p_d) and γ belong to the graph sequence, while (q_d) and α are related to the selection method.

After that we showed how the theorem may be applied in particular cases (Sections 2.6.1 – 2.6.4.), and we proved the necessity of some of the conditions (Section 2.6.5.). The main results may be found in papers [2, 2011] and [3, 2011].

2.2 Limit distribution of degrees in random trees

In Chapter 3 we examined a vertex at a fixed position in the random recursive tree with linear weights. We may consider the third child of the root (the first vertex) for example, in spite of the fact that we do not know in which step it is born. The model is the following. At each step a new vertex is born, and the probability that it is connected to a given old vertex of degree d is proportional to $d + \beta$, where$\beta > -1$ is a parameter of the model. The question is the asymptotic behaviour of the degree of a vertex in a fixed position.

Fix $k \in \mathbb{Z}_+$ and $(x_1, \dots, x_k) \in \mathbb{Z}_+^k$. The root's x_1th child's x_2th child's etc. x_kth child has label x, and its degree after n steps is denoted by $\deg(x, G_n)$.

Theorem 3.1 *[1, Theorem 1] For $k \in \mathbb{Z}_+$ and $(x_1, \dots, x_k) \in \mathbb{Z}_+^k$ we have*

$$\frac{\deg(x, G_n)}{n^{1/(2+\beta)}} \to \zeta_x$$

almost surely as $n \to \infty$ for some positive random variable ζ_x. The distribution of ζ_x is identical to the distribution of $\zeta_\emptyset \cdot \xi_1 \cdot \ldots \cdot \xi_k$, where

- *$\zeta_\emptyset, \xi_1, \dots, \xi_k$ are independent random variables;*
- *$\zeta_\emptyset$ corresponds to the case when the vertex is the root;*
- *ξ_1 has distribution Beta$(1 + \beta, x_1 - 1)$, if $x_1 > 1$; $\xi_1 \equiv 1$, if $x_1 = 1$;*
- *ξ_s has distribution Beta$(1 + \beta, x_s)$ for all integers $2 \le s \le k$.*

The results, similar statements for some other models (Proposition 3.2 – 3.4.) and a functional limit theorem (Theorem 3.5.) may be found in [1, 2011].

3.3 The model of 3-interactions

In Chapter 4 we introduced and analysed a model with dynamics based on interactions of three vertices. Every vertex, edge and triangle has a nonnegative weight, which corresponds to the number of interactions of this object. At each step three vertices interact; either a new one with two old vertices, or three old vertices. These may be chosen uniformly at random, or with probabilities proportional to the weight of the edge or that of the triangle.

We may also consider the degree of a fixed vertex; that is, the number of distinct vertices that has interacted with it. We determined the asymptotic joint distribution of the weight and the degree, i.e. the limit of the proportion of vertices with given weight and degree.

Theorem 4.1 *For integers $1 \le w$ and $2 \le d \le 2w$ denote by $X[n,d,w]$ the number of vertices of degree d and weight w after n steps, and by V_n the total number of vertices after n steps. Then*

$$\frac{X[n,d,w]}{V_n} \to x_{d,w}$$

holds almost surely as $n \to \infty$ with a positive constant $x_{d,w}$. We have the following recurrence equation:

$$\begin{aligned}
x_{2,1} &= \frac{1}{\alpha+\beta+1};\\
x_{d,w} &= \frac{1}{\alpha w+\beta+1}\big[\alpha_1(w-1)x_{d,w-1}+\alpha_2(w-1)x_{d-1,w-1}+\beta x_{d-2,w-1}\big],\\
\alpha_1 &= (1-p)q, \quad \alpha_2 = \frac{2pr}{3}, \quad \alpha = \alpha_1+\alpha_2,\\
\beta &= \frac{1}{p}[2(1-r)+3(1-p)(1-q)].
\end{aligned}$$

Here p is the probability that a new vertex is born at a given step; r is the probability that we choose the other vertices with probabilities proportional to edge weights; q is the probability that we choose three old vertices with probabilities proportional to triangle weights, if no new vertex is born at this step.

After that we proved the scale free property both for weights and degrees (Theorem 4.5. and 4.7.). We also described the asymptotic behaviour of the weight and degree of a given vertex, and the maximal weight and degree (Theorem 4.8 – 4.11.). Some of these results can be found in [4, 2012].

5 Conclusions

In the thesis we examined scale free random graph models, where the proportion of vertices of degree d tends to a constant c_d almost surely as the number of vertices goes to infinity, and $c_d \sim Kd^{-\gamma}$ holds as $d \to \infty$. If we would like to estimate c_d or γ, but the network is to large, we have to be careful with the sampling method. For example, if we consider the proportion of vertices of degree d in the neighbourhood of a given vertex (the degree is counted in the whole network), then the almost sure limit may differ from c_d. Moreover, the exponent of the new asymptotic degree distribution is less than γ. With certain sufficient conditions the new exponent is $\gamma(\alpha - 1) + 1$, where α is the exponent of the regular growth of the number of selected vertices.

The background of this phenomenon may be the following. A fixed vertex gets new neighbours more and more rarely. Hence its neighbours are typically older than a typical vertex of the whole graph. Older vertices usually have larger degree, especially because of the preferential attachment property. Therefore the degrees are typically large in the neighbourhood of a fixed vertex (or more generally, in the set of selected vertices), and the exponent of the asymptotic degree distribution decreases.

Examining randomly growing trees with linear weights, we may conclude that the degree of a fixed vertex and the degree of a vertex in a fixed position have the same asymptotic behaviour, as expected. The structure of the limiting random variable shows an embedded urn model.

Finally we analysed a random model based on interactions of three vertices, and found that the degree is strongly concentrated for large weights. However, there is no deterministic connection between these quantities.

To sum up, we showed how the methods of martingale theory may be used efficiently for problems about the degrees in scale free random graph models.

References

[1] Á. Backhausz. Limit distribution of degrees in random family trees. *Electron. Commun. Probab.*, 16:29–37, 2011.

[2] Á. Backhausz. Local degree distributions: examples and counterexamples. *Period. Math. Hungar.*, 16:1465–1488, 2011.

[3] Á. Backhausz and T. F. Móri. Local degree distribution in scale free random graphs. *Electron. J. Probab.*, 16:1465–1488, 2011.

[4] Á. Backhausz and T. F. Móri. A random graph model based on 3-interactions. *Ann. Univ. Sci. Budapest. Sect. Comput.*, 36:41–52, 2012.

[5] A-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[6] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation.* Cambridge University Press, Cambridge, 1987.

[7] R. Bojanić and E. Seneta. Slowly varying functions and asymptotic relations. *J. Math. Anal. Appl.*, 34:302–315, 1971.

[8] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Structures Algorithms*, 18(3):279–290, 2001.

[9] C. Cooper and A. Frieze. A general model of web graphs. *Random Structures Algorithms*, 22(3):311–335, 2003.

[10] M. Deijfen, H. van den Esker, R. van der Hofstad, and G. Hooghiemstra. A preferential attachment model with random initial degrees. *Ark. Mat.*, 47(1):41–72, 2009.

[11] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.

[12] P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

[13] J. Galambos and E. Seneta. Regularly varying sequences. *Proc. Amer. Math. Soc.*, 41:110–116, 1973.

[14] E. N. Gilbert. Random graphs. *Ann. Math. Statist.*, 30:1141–1144, 1959.

[15] R. Gouet. Martingale functional central limit theorems for a generalized Pólya urn. *Ann. Probab.*, 21(3):1624–1639, 1993.

[16] Zs. Katona and T. F. Móri. A new class of scale free random graphs. *Statist. Probab. Lett.*, 76(15):1587–1593, 2006.

[17] T. F. Móri. A surprising property of the Barabási–Albert random tree. *Studia Sci. Math. Hungar.*, 43(2):265–273, 2006.

[18] T. F. Móri. Degree distribution nearby the origin of a preferential attachment graph. *Electron. Commun. Probab.*, 12:276–282, 2007.

[19] T. F. Móri. Random multitrees. *Studia Sci. Math. Hungar.*, 47:59–80, 2010.

[20] J. Neveu. *Discrete-parameter martingales.* North-Holland Publishing Co., Amsterdam, 1975.

[21] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42:425–440, 1955.

[22] A. Sridharan, Yong Gao, Kui Wu, and J. Nastos. Statistical behavior of embeddedness and communities of overlapping cliques in online social networks. *INFOCOM, 2011 Proceedings IEEE*, pages 546–550, 2011.

[23] G. U. Yule. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. *Philos. Trans. R. Soc. Lond. Ser. B.*, 213:402410, 1925.