RANDOM GRAPHS

Svante Janson

Once upon a time, I began as an analyst. I studied function spaces such as BMO and operators on function spaces, see for example [6,7], but gradually I became more and more interested in probability theory. Some years ago I worked on random coverings and other problems in geometrical probability [8,9], and at present most of my time is devoted to the related field of combinatorial probability, in particular random graphs.

This may seem to be very different from the harmonic and functional analysis that I once worked on (and still continue with sometimes), but the difference in methods is not so great. I study random graphs as a probabilist dealing with some combinatorial structures, and my methods are probabilistic and based on analysis, using for example integration theory, functional analysis, martingales and stochastic integration.

In this presentation I will give a survey over some recent results on random graphs where I have been at least partly involved. The systematic study of random graphs was started by Erdős and Rnyi [4] in 1960, and the theory has expanded rapidly during the last decade. For a fuller historical account, and for many other results on random graphs, I refer to Bollobs's book [3].

Definitions

A random graph is a graph generated by some random procedure. There are many (non-equivalent) ways to define random graphs. The simplest, denoted by $G_{n,m}$ (or one of several common similar notations), where n and m are two integers with $0 \le m \le \binom{n}{2}$, is obtained by taking a set of n elements as the set of vertices, for definiteness we may take the integers $1, \dots, n$, and then randomly selecting (by drawing without replacement) m of the $\binom{n}{2}$ possible edges. A closely related model, denoted by for example $G_{n,p}$, where $0 \le p \le 1$, is obtained by taking the same vertex set but now selecting every possible edge with probability p, independently of all other edges. (In particular, $p = 1/2$ gives the uniform distribution over all (labelled) graphs on n vertices.) We are mainly interested in the case when n, the number of vertices, is very large, and especially in asymptotic results when $n \to \infty$ and m or p is a given function of n.

Note that there is *a priori* no structure in these models. All vertices and pairs of vertices are equivalent, and there is for example no 'distance' given between the vertices before the random selection of edges. This is in contrast to percolation models and models in statistical physics such as the Ising model, where a lattice structure is given and a random graph then is defined using the given structure. (The usual homogeneous random graphs are related to mean-field models in statistical mechanics, which also lack an external structure.)

These standard models can be modified in many ways, for example by allowing loops, multiple edges or directed edges, by introducing random labels on vertices or edges, or by making them less homogeneous, for example in order to adapt them to graphs observed

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

in some application, but the two models just defined suffice for most of my work. Moreover, although there are many applications of random graphs, both deep and shallow, to mathematics and to other sciences such as computer science, sociology and ecology, I have never worked on any of them. I study the random graphs as interesting mathematical objects, regardless of whether the results have any direct implications for applications or not, and I concentrate on the simplest models and try to analyse their behaviour in great detail. Extensions to wider classes of random graphs may come as a second step once the simple models are well understood.

The model $G_{n,p}$ has a random number of edges (with a binomial distribution), but if the number of vertices is large, the distribution of the number of edges is well concentrated about its mean $\binom{n}{2}p$, and for many problems (but not all) we obtain the same asymptotic results for $G_{n,m}$ as for $G_{n,p}$ with $p = m/\binom{n}{2}$. The model $G_{n,p}$ is usually easier to analyse that $G_{n,m}$, since there is a lot of independence built into its construction. On the other hand, the model $G_{n,m}$ is often more natural, and does sometimes exhibit phenomena not seen in $G_{n,p}$.

It is often useful to regard the random graph as evolving in time by a stochastic process, starting with a vertex set without edges and then adding edges until the complete graph is obtained. One version is to add new edges at random at times 1,2, ... ; this gives a process which gives a random graph of type $G_{n,m}$ at time m. Another version, which I find more powerful, is obtained by adding each edge at some random time, for example uniformly distributed in (0,1) and independent of the other edges. This gives a random graph $G_{n,p}$ at a fixed time $p \in (0,1)$ and a graph $G_{n,m}$ at the random time at which the m:th edge appears. Results for both models may then be obtained from results for the process, see [13]. It is also of interest to study properties of the whole process, such as the first time something happens or the maximum of some variable during the evolution.

The structure and evolution of the components

Consider the evolution of a random graph with a large number n vertices as the number m of edges grows. As was shown already by Erdős and Rnyi [4], the following happens (at least with probability tending to 1 as $n \to \infty$). Let ε denote any positive number $\varepsilon(n) \to 0$ as $n \to \infty$.

In the beginning, we have only isolated vertices. As we add edges, first components with two vertices appear; then some of these grow further and we obtain trees with 3 and more vertices. As long as m is small, $m = o(n)$, the graph is a forest, consisting of a large number of mostly small trees which grow (by merging) as m increases.

The first cycle appears when $\varepsilon n < m < (\frac{1}{2} - \varepsilon)n$ and then the graph consists of trees and a few unicyclic components until $m \approx n/2$.

At about this number of edges, a 'phase transition' occurs and the structure of the graph changes dramatically. For example, let $\delta > 0$ be fixed. When $m = (1 - \delta)n/2$ there are many trees and possibly some unicyclic components, and the largest component has order $O(\log n)$, but when $m = (1 + \delta)n/2$ there is a unique giant component with $\sim c(\delta)n$ vertices. This component has many more edges than vertices, but all other components are trees or unicyclic, and the second largest component has order $O(\log n)$.

Note that the phase transition occurs when the mean degree of a vertex $2m/n \approx 1$.

As more edges are added, the giant grows and swallows other components, but the picture remains qualitatively the same as long as $m = O(n)$. When $m \gg n$, there are no cycles left outside the giant component, and no new will have time to form. There are still some trees outside the giant component, but the last of them (an isolated vertex) disappears when $m \approx \frac{1}{2}n \log n$, more precisely when $m = \frac{1}{2}n \log n + O_p(n)$. From this point on, the random graph is connected.

Let us look closer at two phases of this evolution.

First, what is the size of the first cycle that appears? It turns out that the length converges in distribution as $n \to \infty$, without any normalization.

Theorem 1 [10]. *Let L_n be the length of the first cycle. Then*

$$\mathrm{P}(L_n = k) \to p_k = \tfrac{1}{2} \int_0^1 t^{k-1}(1-t)^{1/2} e^{t/2+t^2/4} dt, \qquad k \geq 3.$$

The limit distribution has infinite mean, in fact $p_k \asymp k^{-3/2}$, and thus $\mathrm{E}\, L_n \to \infty$. More precisely:

Theorem 2 [5,15]. $\mathrm{E}\, L_n \sim K n^{1/6}$, *where* $K = 2^{-1/6} 3^{-2/3} e^{3/4} \pi^{1/2} \Gamma(1/3)$.

Secondly, let us study the phase transition more closely. It turns out that the correct scale for doing this is to let $m = \frac{1}{2}n + xn^{2/3}$, $-\infty < x < \infty$. In this range, the largest component has order $\approx n^{2/3}$; moreover, there are several components of about this size. The largest components may be either trees, unicyclic or multicyclic.

Before this critical range of m, there are no components with more than one cycle, and after it there is always exactly one, the giant component. During this range, however, it is possible that there are several components with more than one cycle, and we have the following precise result.

Theorem 3 [12,15]. *The probability that an evolving random graph never has more than one multicyclic component during the entire evolution approaches* $\frac{5\pi}{18} \approx 0.8727$ *as* $n \to \infty$.

Subgraph counts

Another class of results give asymptotics for the number of times a certain substructure appears in a random graph. For example, let H be some fixed graph and let the random variable X_n be the number of copies of H in one of our random graphs, say $G_{n,p}$ for some $p = p(n)$. (In some sense, these results deal mainly with the local structure of a random graph, whereas the results above concern the global structure.)

Note that X_n can be written as a sum of indicator variables $\sum_{H'} I(H' \subset G_{n,p})$, where H' ranges over all subgraphs of the complete graph K_n that are isomorphic to H. These indicators are not completely independent, but for $G_{n,p}$ most of them are independent and for $G_{n,m}$ most of the dependencies are weak. Various standard methods from probability theory for dealing with sums of dependent variables may thus be used, and have been used by several authors, to obtain asymptotic results for X_n; for example the method of moments, projection methods, Stein's method and martingale methods. (For Poisson limits, I recommend Stein's method, see the recent book [1] for details.)

The typical results, for this and many similar variables for $G_{n,p}$ or $G_{n,m}$, are that if p or m is such that the expected number converges to some finite value, then the variable has asymptotically a Poisson distribution, and if the expectation tends to infinity, then the distribution is asymptotically normal. There are some exceptions, however; for the variable defined above, some restriction on the structure of H is needed for the Poisson result and for the normal result when p is small, and we need $1 - p \gg n^{-2}$; see [3,18,17].

If p is fixed, there are no problems. Let v and $e \geq 1$ be the number of vertices and edges in H. Then $\mathrm{Var}\, X_n \asymp n^{2v-2}$, and

$$\frac{X_n - \mathrm{E}\, X_n}{(\mathrm{Var}\, X_n)^{1/2}} \xrightarrow{d} \mathrm{N}(0,1). \tag{1}$$

If we, however, redefine X_n and count only *induced* subgraphs of the random graph, something interesting happens. (An induced subgraph is obtained by selecting a subset of the vertices together with all edges between them in the graph; a general subgraph on the same vertices may contain fewer edges.)

Theorem 4 [2,11,13]. *For the number X_n of induced subgraphs isomorphic to H in $G_{n,p}$, $p \in (0,1)$ fixed, the following holds.*

(i) *If $p \neq p_H = e/\binom{n}{2}$, then $\operatorname{Var} X_n \asymp n^{2v-2}$ and (1) holds.*
(ii) *If $p = p_H$, then for most H the variance is of a smaller order due to some cancellation, $\operatorname{Var} X_n \asymp n^{2v-3}$, but (1) still holds.*
(iii) *For some H and $p = p_H$, the variance is of still smaller order due to further cancellation, $\operatorname{Var} X_n \asymp n^{2v-4}$. In this case, the distribution is not asymptotically normal, in fact*

$$\frac{X_n - \operatorname{E} X_n}{(\operatorname{Var} X_n)^{1/2}} \xrightarrow{d} c_1\xi - c_2\eta, \tag{2}$$

where $\xi \sim \mathrm{N}(0,1)$ and $\eta \sim \chi^2(1)$ are independent and $c_2 > 0$.

The exceptional graphs that give case (iii) have a combinatorial description and are called *proportional*. The smallest examples have 8 vertices, for example the wheel W_8. Jan Krrman [16] has found an example with 64 vertices where furthermore $c_1 = 0$.

For induced subgraph counts in $G_{n,m}$, with $m = [\binom{n}{2}p]$ for a fixed p, the situation is similar but not identical. There are now four cases. The standard case is $\operatorname{Var} X_n \asymp n^{2v-3}$ and there are exceptional cases where the variance has order n^{2v-4} or n^{2v-5}; in all three cases (1) holds. In the fourth case there is even more cancellation, $\operatorname{Var} X_n \asymp n^{2v-6}$ and the distribution has a non-normal limit similar to (2) above. It is, however, still not known whether the fourth case actually can occur. The combinatorial conditions on H are rather complicated, and no example is known, although it seems likely that such an H with 256 vertices may exist.

Finally, let us consider a case where H is not fixed but grows with n, viz. the number of Hamilton cycles in a random graph. In this case, the number turns out to be asymptotically normal for $G_{n,m}$, for a wide range of m, but asymptotically *log-normal* for $G_{n,p}$ [14].

References

1. A. D. Barbour, L. Holst and S. Janson, *Poisson Approximation*, Oxford Univ. Press, Oxford, 1992.
2. A. D. Barbour, M. Karoński and A. Ruciński, *A central limit theorem for decomposable random variables with applications to random graphs*, J. Combin. Th. Ser. B **47** (1989), 125–145.
3. B. Bollobs, *Random Graphs*, Academic Press, London, 1985.
4. P. Erdős and A. Rnyi, *On the evolution of random graphs*, Magyar Tud. Akad. Mat. Kut. Int. Kzl. **5** (1960), 17–61.
5. P. Flajolet, D. E. Knuth and B. Pittel, *The first cycles in an evolving graph*, Discrete Math. **75** (1989), 167–215.
6. S. Janson, *On functions with conditions on the mean oscillation*, Ark. Mat. **14** (1976), 189–196.
7. S. Janson, *On BMO and related spaces*, Ph. D. Thesis, Uppsala 1977.
8. S. Janson, *Random coverings and related problems*, Ph. D. Thesis, Uppsala 1984.
9. S. Janson, *Random coverings in several dimensions*, Acta Math. **156** (1986), 83–118.
10. S. Janson, *Poisson convergence and Poisson processes with applications to random graphs*, Stoch. Proc. Appl. **26** (1987), 1–30.
11. S. Janson, *A functional limit theorem for random graphs with applications to subgraph count statistics*, Random Struct. Alg. **1** (1990), 15–37.
12. S. Janson, *Multicyclic components in a random graph process*, Random Struct. Alg. **4** (1993), 71–84.
13. S. Janson, *Orthogonal decompositions and functional limit theorems for random graph statistics*, Memoirs Amer. Math. Soc. (to appear).

14. S. Janson, *The numbers of spanning trees, Hamilton cycles and perfect matchings in a random graph* (to appear).
15. S. Janson, D. E. Knuth, T. Łuczak and B. Pittel, *The birth of the giant component*, Random Struct. Alg. **4** (1993) (to appear).
16. J. Krrman, *An example of a superproportional graph*, Random Graphs '91, Proceedings, Poznań 1991, Wiley (to appear).
17. A. Ruciński, *When are small subgraphs of a random graph normally distributed?*, Prob. Th. Rel. Fields. **78** (1988), 1–10.
18. A. Ruciński and A. Vince, *Balanced graphs and the problem of subgraphs of random graphs*, Congressus Numerantum **49** (1985), 181–190.

Department of Mathematics, Uppsala University, P.O. Box 480, 751 06 Uppsala, Sweden
E-mail address: svante.janson@math.uu.se