Models of random graphs and a note on eigenvalues

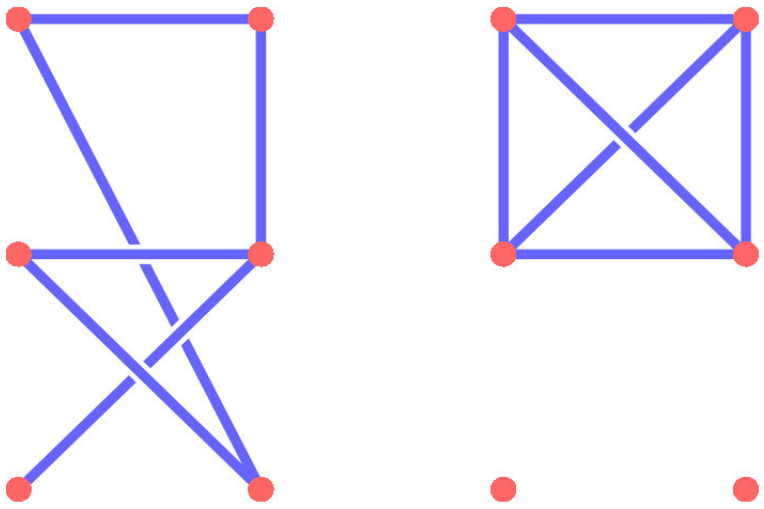
Prepared by Linh V. Tran

October 17, 2005

1 Random graph models

1.1 Erdős Renyi model

There are several ways to think about random graphs. We will first consider the Erdős-Renyi model $G(n,p)$. In this model, the random graph $G(n,p)$ is a probability space where the elements are all the graphs on n vertices. We usually consider labelled graphs to avoid worrying about possible automorphism. The probability distribution is given so that each edge appears independently with probability p.
If p is large then the graph G tends to be dense, i.e., with many edges, conversely if p is small G tends to be sparse. Note by changing p we do not change the underlying space but the assocciated probability distribution.
We should get rid of our "mental image" about random graphs. For example, these following two graphs of six edges on six vertices are perfectly equivalent in the sense of random graphs, i.e., they have the same chance to appear, though one looks much more "structured" than the other.

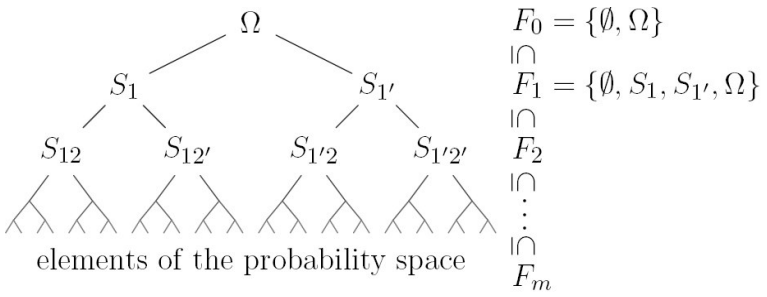


Our notation of randomness are not completely without merit. It could be good to be able to detect "randomness" as we think of it. This leads to the subject of quasi-randomness. This is a collection of measurable properties (which tends to be related to our intuitive notions

of randomness) which are equivalent in the sense that if a graph satisfied one quasi- random property it satisfies them all. See [2] for details.

1.2 Martingale model

When you want to "pick" a random graph, it is easier to give a construction of a "typical" graph then to construct the whole probability space. Think about growing a graph randomly by deciding whether to include a given edge at each step with probability p. There is a suitable probabilistic tool for controlling properties of random graphs: the martingales. Generally, given a probability space Ω, a martingale is a filtration which refines Ω by a sequence of σ-fields of subsets of Ω: $F_0 \subset F_1 \subset \cdots \subset F_m$. As an example, Ω could be the set of all graphs on n vertices, and then at each stage we refine the space according to whether or not we add an edge. So $F_0 = \{\emptyset, \Omega\}$, $F_1 = \{\emptyset, S_1, S_{1'}, \Omega\}$ where S_1 consists of all graphs containing edge 1 and $S_{1'}$ consists of all graphs not containing edge 1, $F_2 = \{\emptyset, S_1, S_{1'}, S_{12}, S_{12'}, S_{1'2}, S_{1'2'}, \Omega\}$ where S_{12} consists of all graphs containing edges 1 and 2, $S_{12'}$ consists of all graphs containing edge 1 but not containing edge 2, and so on.



Let X denotes a random variable on our probability space and define

$$X_k = E(X|F_k)$$

Then $\{X_k, F_k\}_{k=0}^m$ is a martingale. Note that we have some basic properties of martingale:

$$X_0 = E(X|F_0) = EX,$$

$$X_k = E(X_{k+1}|F_k).$$

Note that martingales need not to be binary as in our example here, they could have multiple branches. Look in [1], Chapter 7 for some detailed examples of martingales for random graph.

1.3 Almost surely

We are interested in properties of random graphs. However we must be careful since every graph is a random graph and a property satisfied by every graph is not interesting. So we will consider properties that are almost surely true.
We say a random graph $G(n,p)$ has a property Q *almost surely* (a.s.) if

$$P(G(n,p) \text{ has property } Q) \to 1 \text{ as } n \to \infty.$$

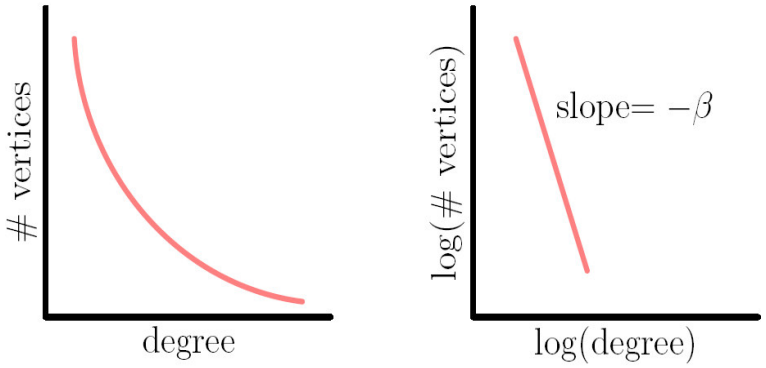
Example 1. Suppose $p > 0$ is fixed, then $G(n,p)$ is connected a.s. Indeed, we have

$$\begin{aligned} P(G(n,p) \text{ is connected}) &> P(G(n,p) \text{ has diameter } \leqslant 2) \\ &\geqslant P(\forall (x,y), \exists w \text{ s.t. } x \sim w \sim y \text{ (2-step path)}) \\ &= 1 - P(\exists (x,y), \forall w \text{ not } x \sim w \sim y) \\ &\geqslant 1 - n^2(1-p^2)^n \to 1 \text{ as } n \to \infty. \end{aligned}$$

This show that when p is fixed we have lots of edges and high conectivity. These graphs tend to be uninteresting, though they have the advantage of being easy to verify properties on. More interesting are random graphs which are sparse, these tend to have p as a function of n, i.e., $p(n) = c/n$ or $p(n) = \log n/n$.

1.4 Expected degree model

The graphs $G(n,p)$ also are less interesting because they do not appeared to accurately model real world graphs. For instance the "internet graph" (here vertices are webpages and edges are links), the "telephone graph" (here vertices are phones and edges are calls), and the "Hollywood graph" (here vertices are actors and edges are present for actors who have been in the same movie) all appear to exhibit a power law distribution of its vertices. That is the probability of having a vertex of degree K is propotional to $K^{-\beta}$ (most graphs empirically have $2 < \beta < 3$). This says these graphs have lots of vertices of low degree and few degree of high degree.



For example, in the internet graph there are lots of pages with few links (i.e., most homepages) and a few vertices of vertices very high degree (i.e., Amazon, Google, Yahoo). We would like a random graph model which favors graphs which have a power law distribution. We consider another model to present random graph which we denote by $G(\mathbf{w})$ (see [3]). Let $\mathbf{w} = (w_1, w_2, .., w_n)$ be expected values of degree of n vertices. Then we set up a probability distribution so that for the edges we have

$$P(i \sim j) = \frac{w_i w_j}{\sum_1^n w_k}$$

Note that the w_i's are not required to be integer. Also note that in this model we will allow loops, this is done mainly for notational convenience. To assure that all probabilities are less

than 1 we need the assumption

$$(w_{\max})^2 \leqslant \sum_1^n w_k.$$

So we are likely to connect vertices with high weights and unlikely to connect vertices with low weights. This random graph model allows for preferred vertices. In particular, by giving $\mathbf{w}$ a power law distribution our graphs will favor power law distribution.

Recall that we denote $vol(G) = \sum_1^n d_i$ to be *actual* volume of G, and we will let $Vol(G) = \sum_1^n w_i$ to be the *expected* volume of G. As one would expext, the expectation of degree d_i is equal to expected value w_i:

$$E(d_i) = \sum_j E(i \sim j) = \sum_j \frac{w_i w_j}{Vol(G)} = w_i \frac{\sum_j w_j}{Vol(G)} = w_i.$$

Note that the Edős-Renyi model $G(n,p)$ is a special case with $w_i = np$ for all $i = 1, \dots, n$.

2 A note on largest eigenvalue of the adjacency matrix

Let G be a undirected graph on n vertices and A be the adjacency matrix of G, i.e.

$$A(i,j) = \begin{cases} 1 & \text{if } v_i \sim v_j; \\ 0 & \text{otherwise.} \end{cases}$$

where A has eigenvalues $\alpha_1 \geqslant \alpha_2 \geqslant ... \geqslant \alpha_n$, with all α_i's are real since A is symmetric.

Proposition. *Let $d_{\max}$ be the maximum degree of the vertices of G. Then $\alpha_1 \geqslant \sqrt{d_{\max}}$.*

We'll need one technical tool which will play an important role in directed graphs.

Perron-Frobenius Theorem.(See [4]) *If A is non-negative and irreducible then the spectral radius $\rho(A)$ is*

1. *an eigenvalue of A;*
2. *the eigenvector of $\rho(A)$ has all possitive entries.*

For graphs the condition that the matrix A is irreducible corresponds to G being connected (for undirected graphs) or G being strongly connected (for directed graphs).

This tool is very useful, especially for directed graph since it doesn't require the matrix A to be symmetric.

Claim 1. If G' is a subgraph of G then $\alpha_1(G') \leqslant \alpha_1(G)$.
Proof. By the Perron-Frobenius Theorem the vector which gives the largest eigenvalue of A can be assumed to have all non-negative entries. So

$$\begin{aligned} \alpha_1(G') &= \sup_{\|x\|=1} x^* A' x = \sup_{\|x\|=1, x \geqslant 0} x^* A' x \\ &\leqslant \sup_{\|x\|=1, x \geqslant 0} x^* A x = \alpha_1(A) \end{aligned}$$

The inequality follows by noting that since G' is a subgraph of G then $A' \leqslant A$ entrywise. So $X^*A'Z \leqslant X^*AX$ because all of the terms are non-negavite and at most we are adding additional terms.

Proof of Proposition. Take x to be a vertex with maximum degree in G. Then G contain a star having x as the center vertex as a subgraph. This star has largest eigenvalue $\sqrt{d_{\max}}$. Then the proposition is proved by Claim 1.

It is important that we used the adjacency matrix in this example. The Laplacian does not have a similar property because it is not a non-negative matrix. When stating results we need to be clear which spectrum we are using and careful not to generalize results of one kind of spectra to another kind.

References

[1] J. Spencer, N. Alon, The probabilistic methods, 2nd Ed., Wiley & Sons Inc., 2000.

[2] F.C. Graham, R. L. Graham and R. M. Wilson, Quasi-random graphs, A short version appeared in Proc. Natl. Acad. Sci. USA, 85 (1988), 969-970. A long version appeared in Combinatorica 9 (1989), 345-362.

[3] F.C. Graham, Lincoln Lu and Van Vu, The spectra of random graphs with given expected degrees, short version, Proceedings of National Academy of Sciences, 100, no. 11, (2003), 6313-6318. (long version with full proofs will appear in Internet Mathematics.)

[4] R.Horn and C.Johnson, Matrix Analysis, Cambridge University Press, Cambridge, UK, 1999.