More on random graphs and discrepancy for directed graphs

Prepared by: Steven Butler

November 9, 2005

1 More on random graphs

In an earlier lecture we introduced some random graph models. We now return briefly to look at random graphs. There are two main applications that random graphs are used for. The first is to show the existence of graphs which satisfy certain properties. The second is to describe what a "typical" graph looks like. We will show examples of both applications.

1.1 Graphs with large girth and chromatic number

Consider the following problem: Does there exist a graph with large girth and large chromatic number? The girth of a graph is the length of the shortest cycle and the chromatic number is the minimum number of colors needed to color the vertices of the graph so no two adjacent vertices have the same color.

These two conditions seem contradictory, intuitively to have high chromatic number we need something dense, i.e., clusters of vertices all close to one another. On the other hand to have large girth means that locally everything looks like a tree (i.e., no local cycles) and hence locally the graph is 2-colorable.

An amazing result of Erdős is that such graphs exist with girth at least g and chromatic number at least k for any positive g and k. We now give a proof using random graphs (this proof is from [4]).

Proof. Let n be large and consider a graph in $G(n,p)$. A set t is independent in G (i.e., no edges between the t vertices) with probability $(1-p)^{\binom{t}{2}} \approx e^{-pt^2/2}$. There are $\binom{n}{t} < n^t$ such sets and so the probability that $\alpha(G) \geq t$ (where $\alpha(G)$ is the size of the largest independent set in G) is at most $n^t e^{-pt^2/2}$. When $pt/2 > (1+o(1))\ln n$ this quantity is $o(1)$ so that $\alpha(G) < t$ almost always. Note that the chromatic number,

$\chi(G)$, satisfies $\chi(G) \geq n/\alpha(G)$, since in any coloring each color class is an independent set. Thus $\chi(G) > n/t$ almost always.

We might be tempted to now select $t = n/k$ and $p = (2 + o(1))k(\ln n)/n$ so that $\chi(G) > k$ almost always. This is not enough because such a G is going to have many small cycles. We just need to tweak it a little. Let $p = (4 + o(1))k(\ln n)/n$ so that $\alpha(G) < n/2k$ almost always. Let X be the number of cycles of length at most g. For $3 \leq i \leq g$ there are less than n^i potential i-cycles, each of which is in G with probability p^i. By linearity of expectation we have

$$E(X) = \sum_{i=3}^{g} (np)^i < c(\ln n)^g,$$

where c is some constant depending only on g and k. With g fixed, we have $E(X) = o(n)$ so that $X < n/2$ almost always. Almost always $\alpha(G) < n/2k$ and G has fewer than $n/2$ cycles of size at most g. Fix a G with these properties. Delete from G one vertex from each such cycle. The remaining graph has no small cycles, at least $n/2$ vertices, and independence number $n/2k$ so that the chromatic number is at least k. □

Note that the proof shows that a large number of such graphs exist, but actually giving a construction of such a graph is a difficult problem. One of the characteristics of probabilistic combinatorics is its ability to show that certain structures exist without being able to describe one. [For instance recall an earlier example with calculating a lower bound for the Ramsey number, this was done by randomly coloring the edges; giving an actual construction for such a graph is much more difficult.]

1.2 Describing a typical graph

What can you say about a typical graph? Usually this question is interpreted as what properties are almost surely satisfied by graphs in our $G(n, p)$ model. Usually these properties will vary with p (more generally with the underlying graph model). For our first example of this type of question we will show that when $np \to 0$ as $n \to \infty$ a typical graph is a forest. In other words when our edge density gets sufficiently small then typically we would expect to have a forest. We now give short proof of this result (taken from [3]).

Proof. Let X_k count the number of cycles in a graph in $G(n, p)$. Then

$$P(G \text{ has a cycle}) \leq \sum_{k=3}^{n} P(G \text{ has a cycle of length } k) \leq \sum_{k=3}^{n} E(X_k).$$

But

$$E(X_k) = \binom{n}{k}\frac{1}{2}(k-1)!p^k < (np)^k,$$

so substituting and simplifying shows $P(G \text{ has a cycle}) = o(1)$. □

For our second example of what we can say about a typical random graph we consider a graph in the $G(\mathbf{w})$ model. Recall that in this model that we have an expected degree sequence $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ and that the probability of including an edge between vertices i and j is $w_i w_j / \sum_k w_k$. One nice property that we showed about such graphs is that the expected degree at vertex i is w_i.

For some applications (and for some group of people) it is sufficient to know the expected value. But to really understand behavior and give meaningful interpretations we need to bound the range of the actual degrees. [Failure to bound behavior has led many astray into false notions and bad mathematics.]

To bound our behavior we will use a concentration inequality. (A comprehensive survey on concentration inequalities and their applications can be found in [2].)

Lemma 1. *Let $X_1, \ldots, X_n$ be independent random variables with $P(X_i = 1) = p_i$ and $P(X_i = 0) = 1 - p_i$. For $X = \sum_{i=1}^n X_i$, we have $E(X) = \sum_{i=1}^n p_i$. Then we have*

$$P(|X - E(X)| \geq \lambda) \leq 2e^{-\lambda^2/[2(E(X)+\lambda/3)]}.$$

We can apply this to bound the behavior of the degree. In our case the variable X represents the actual degree of vertex i, which we denote d_i, and each X_j represents the random variable of whether we have included an edge from i to j. So in our case we have that $P(X_j = 1) = w_i w_j / \sum_k w_k$ and $E(X) = w_i$. We then have the following.

Fact 1. Almost surely every vertex in the random graph in $G(\mathbf{w})$ satisfies

$$|d_i - w_i| < \ln n + 2\sqrt{w_i \ln n}. \qquad (*)$$

The proof of this fact follows from the concentration inequality and noting that when $w_i < \ln n$ we can use the error term $\lambda = 2\sqrt{w_i \ln n}$ and when $w_i > \ln n$ we can use the error term $\lambda = \ln n$. At any vertex it follows that $(*)$ fails to hold with probability at most $1/n$. Summing the probabilities it follows that with probability $o(1)$ that $(*)$ fails to hold at some vertex giving the result.

What this demonstrates is that if $w_i \gg \ln n$ then the expected degree and the actual degree are relatively close in size and we can control the behavior. On the other hand if $w_i \ll \ln n$ then the expected degree and the actual degree can vary widely and we do not have good control on behavior. Any applications to graphs in these models should make sure that this has been accounted for.

1.3 Quasirandom graphs

One drawback of the approach in the previous section is that we are arguing about a "typical graph". But when we come to apply these results we are usually given a specific graph (the internet, telephone, or Hollywood graph for example). How can we tell if such a graph is typical?

This is one of the advantages of the quasirandom approach to graphs. A quasirandom property is a property which is measurable and is such that if we have good control on that property then we know we have a "typical quasirandom graph" and so have good control on other quasirandom properties. Examples of these quasirandom properties for $G(n, 1/2)$ are that there are at least $(1 + o(1))n^2/4$ edges and at most $(1+o(1))n^4/16$ cycles of length 4, another is that there is at least $(1+o(1))n^2/4$ edges, $\lambda_1 = (1 + o(1))n/2$ and the remaining $\lambda_i = o(n)$ where the λ_i are the eigenvalues of the adjacency matrix, and so on.

In the next section we will look at another quasirandom property, namely that of discrepancy of edges.

2 Discrepancy for graphs

The idea of discrepancy is to measure how "randomly" the edges have been placed in a graph. More precisely we want to look at the difference between the actual edges and the expected edges. The original form of discrepancy was given as the minimal α so that for all subsets X, Y of the vertices

$$\big|e(X,Y) - \rho|X|\,|Y|\big| \leq \alpha\sqrt{|X|\,|Y|},$$

where $e(X, Y)$ is the number of edges with a vertex in X and a vertex in Y (by convention any edge in $X \cap Y$ is counted twice), ρ is the edge density, and $|X|$ denotes the number of vertices in X and should be thought of as a measurement of the size of X.

In the above definition the $e(X, Y)$ is the number of actual edges between the subsets X and Y. The term $\rho|X|\,|Y|$ is the number of expected edges. To see this, one way to think about ρ is that it is the probability that any particular edge is in the graph, then there are $|X|\,|Y|$ edges between X and Y so that (thinking as a random graph) we would expect $\rho|X|\,|Y|$ edges. Again thinking as including edges randomly then the $\sqrt{|X|\,|Y|}$ term is present to normalize the error term.

With this version of discrepancy the quasirandom property is that α is small, $\alpha = o(d)$ where d is the average degree of the graph.

2.1 A better version of discrepancy

The original form of discrepancy works very well when we deal with random graphs in the $G(n,p)$ model, or more generally any regular graphs. However, for graphs in the $G(\mathbf{w})$ model, or any graph whose degree sequence is highly imbalanced this definition does not work well. The problem lies in that the expected number of edges was using the assumption that the graph is regular.

The way to generalize is change the underlying graph model about how edges are placed. Now instead of thinking of placing edges as in $G(n,p)$ we place edges as if they were coming from the $G(\mathbf{w})$ model. Then the expected number of edges between X and Y will be

$$\sum_{x\in X}\sum_{y\in Y}\frac{d_x d_y}{\operatorname{vol} G}=\frac{\sum_{x\in X} d_x \sum_{y\in Y} d_y}{\operatorname{vol} G}=\frac{\operatorname{vol} X \operatorname{vol} Y}{\operatorname{vol} G}.$$

The correct error normalizing term now becomes $\sqrt{\operatorname{vol} X \operatorname{vol} Y}$. Putting this into our definition we have that

$$\left|e(X,Y)-\frac{\operatorname{vol} X \operatorname{vol} Y}{\operatorname{vol} G}\right|\leq \alpha\sqrt{\operatorname{vol} X \operatorname{vol} Y},$$

where now our quasirandom property is that $\alpha = o(1)$. It is known that α is closely tied to $\max_{i\neq 0}|1-\lambda_i|$ in that if one of these terms is small then the other term is also small.

Note that intuitively this says that when $\operatorname{vol} X$ and $\operatorname{vol} Y$ are large we should expect more edges connecting them and when these terms are small we should expect fewer edges. So this definition is more reasonable for general graphs.

2.2 An even better version of discrepancy

We can improve on this even more. Discrepancy as given above works for undirected graphs. A similar definition can be given that works for directed graphs. Now we will let

$$e(X\to Y)=\sum_{\substack{x\in X\\ y\in Y}}[x\to y \text{ an edge}],$$

where $[x\to y$ an edge$]$ will be 1 if there is an edge and 0 otherwise.

Volume is the sum of the degrees, when we move to directed graphs we have two sets of degrees. The in-degree is the number of edges directed into the vertex, and the out-degree is the number edges directed out of the vertex. We will denote these quantities as $d_{in}(u)$ and $d_{out}(u)$ respectively. We now also have an in-volume

and an out-volume defined in the obvious way, namely $\mathrm{vol}_{in}(X) = \sum_{x\in X} d_{in}(X)$ and $\mathrm{vol}_{out}(X) = \sum_{x\in X} d_{out}(X)$. We note that $\mathrm{vol}_{in}(G) = \mathrm{vol}_{out}(G) := \mathrm{vol}\, G$.

If in our underlying graph model we now assume that an edge $x\to y$ is included with probability $d_{out}(x)d_{in}(y)/\mathrm{vol}\, G$ then our definition of discrepancy now generalizes to finding the minimal β so that for all X, Y subsets of the vertices

$$\left| e(X\to Y) - \frac{\mathrm{vol}_{out} X \ \mathrm{vol}_{in} Y}{\mathrm{vol}\, G} \right| \leq \beta \sqrt{\mathrm{vol}_{out} X \ \mathrm{vol}_{in} Y},$$

with the quasirandom property being that $\beta = o(1)$. The minimum such β we will refer to as the discrepancy of G and denote it $\mathrm{disc}\, G$.

As with the undirected case there is a close connection between discrepancy and the (normalized) adjacency matrix. But with the directed case instead of using eigenvalues we will use singular values. [When the graph is undirected, so that our matrix is symmetric, using eigenvalues is equivalent to using singular values, now we are being more clear about what tool we are using.]

2.3 Connecting discrepancy to the normalized adjacency matrix

We will let A denote our adjacency matrix where $A_{uv} = [u\to v \text{ an edge}]$. Note that the row sums of A are the out-degrees and the column sums of A are the in-degrees.

As we did in the undirected case we normalize, so we will consider $D_{out}^{-1/2} A D_{in}^{-1/2}$ where D_{out} and D_{in} are the diagonal out-degree and in-degree matrices. The first thing to note is that as before 1 is the largest singular value of the normalized adjacency matrix. To see this we note that

$$\big(D_{out}^{-1/2} A D_{in}^{-1/2}\big)^* \big(D_{out}^{-1/2} A D_{in}^{-1/2}\big) D_{in}^{1/2}\mathbf{1} = \big(D_{in}^{-1/2} A^* D_{out}^{-1/2}\big) D_{out}^{1/2}\mathbf{1} = D_{in}^{1/2}\mathbf{1},$$

where the * indicates that we are taking the conjugate-transpose and $\mathbf{1}$ is the all 1s vector. By application of the Perron-Frobenius this is the largest eigenvalue of $\big(D_{out}^{-1/2} A D_{in}^{-1/2}\big)^* \big(D_{out}^{-1/2} A D_{in}^{-1/2}\big)$ and hence the largest singular value of $D_{out}^{-1/2} A D_{in}^{-1/2}$. It is also easy to see from this calculation that the right and left singular vectors are $D_{in}^{1/2}\mathbf{1}$ and $D_{out}^{1/2}\mathbf{1}$.

If we subtract out the largest singular value then we have the matrix

$$D_{out}^{-1/2} A D_{in}^{-1/2} - \frac{D_{out}^{1/2}\mathbf{1}(D_{in}^{1/2}\mathbf{1})^*}{\|D_{out}^{1/2}\mathbf{1}\|\,\|D_{in}^{1/2}\mathbf{1}\|} = D_{out}^{-1/2}\big(A - \frac{1}{\mathrm{vol}\, G} D_{out} J D_{in}\big) D_{in}^{-1/2},$$

where $J = \mathbf{1}\,\mathbf{1}^*$ is the all 1s matrix.

The glue connecting discrepancy and the normalized adjacency matrix is

$$e(X{\to}Y) - \frac{\operatorname{vol}_{out} X \ \operatorname{vol}_{in} Y}{\operatorname{vol} G} = \big\langle \psi_X, \big(A - \frac{1}{\operatorname{vol} G} D_{out} J D_{in}\big)\psi_Y \big\rangle$$

where ψ_X is the indicator function of X, i.e., $\psi_X(u) = [u \in X]$, and $\langle a, b\rangle$ denotes the usual inner product.

2.4 Bounding behavior

We have the following explicit connection between discrepancy and singular values. This will show that these two quantities are equivalent and that if we can bound one of these then we can bound both of them. Note in the theorem below σ_2 refers to the second largest singular value of the matrix.

Theorem 1. $\operatorname{disc} G \le \sigma_2\big(D_{out}^{-1/2} A D_{in}^{-1/2}\big) \le 150 \operatorname{disc} G\big(1 - 8 \ln \operatorname{disc} G\big)$.

We will prove one of these inequalities (the proof of the other can be found in [1], we will not leave it as an exercise as the details are not trivial).

Proof. The key to our proof will lie in the fact that for any matrix M and vectors a, b we have $|\langle a, Mb\rangle| \le \sigma_1(M)\|a\| \, \|b\|$. Applying this to our case we have for any X, Y

$$\begin{aligned}
\left| e(X{\to}Y) - \frac{\operatorname{vol}_{out} X \ \operatorname{vol}_{in} Y}{\operatorname{vol} G} \right| &= \left| \big\langle \psi_X, \big(A - \frac{1}{\operatorname{vol} G} D_{out} J D_{in}\big)\psi_Y \big\rangle \right| \\
&= \left| \big\langle D_{out}^{1/2}\psi_X, D_{out}^{-1/2}\big(A - \frac{1}{\operatorname{vol} G} D_{out} J D_{in}\big) D_{in}^{-1/2} \, D_{in}^{1/2}\psi_Y \big\rangle \right| \\
&\le \sigma_1\big(D_{out}^{-1/2}\big(A - \frac{1}{\operatorname{vol} G} D_{out} J D_{in}\big) D_{in}^{-1/2}\big) \|D_{out}^{1/2}\psi_X\| \, \|D_{in}^{1/2}\psi_Y\| \\
&= \sigma_2\big(D_{out}^{-1/2} A D_{in}^{-1/2}\big) \sqrt{\operatorname{vol}_{out} X \ \operatorname{vol}_{in} Y}.
\end{aligned}$$

This shows that $\operatorname{disc} G \le \sigma_2\big(D_{out}^{-1/2} A D_{in}^{-1/2}\big)$. □

The proof in the other direction is similar. But now the idea is we have a group of vectors for which we can control the inner product (namely the $D_{out}^{1/2}\psi_X, D_{in}^{1/2}\psi_Y$) and so we take an arbitrary vector and approximate using these "step vectors". The crux of the proof being to show that we can approximate and still keep good control on the behavior.

References

[1] S. Butler, "Relating singular values and discrepancy of weighted directed graphs", to appear in the SODA'06 proceedings.

[2] F. Chung and L. Lu, "Concentration inequalities and martingale inequalities—a survey", preprint.

[3] M. Karoński, "Random graphs" in *Handbook of Combinatorics*, MIT Press, 1995, 351-380.

[4] J. Spencer, "Probabilistic methods" in *Handbook of Combinatorics*, MIT Press, 1995, 1785-1817.