University of Washington | Lecturer: Abraham Flaxman/Vahab S. Mirrokni

CSE599m: Algorithms and Economics of Networks

April 13$^{\text{th}}$, 2007 | Scribe: Ben Birnbaum

Lectures 6: Degree Distributions and Concentration Inequalities

1 Summary

In preparation for the upcoming lecture on bias in traceroute sampling, we present some results regarding the degree distributions of standard random graph models. To prove these results we will make use of several theorems bounding the tail distributions of random variables. These theorems, called concentration inequalities, are interesting in their own right and are widely used in the analysis of random structures and algorithms.

2 Topic of lecture

We start the lecture by asking the following question: what is the probability that a fixed vertex v in a random graph $G \sim G_{n,p}$ has degree k? Since each possible edge adjacent to v is present independently with probability p, we have

$$\mathbb{P}\mathrm{r}(\deg(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k} \quad .$$

For $p = d/n$ with d a constant, the expected degree of a vertex is approximately d, and we have

$$\mathbb{P}\mathrm{r}(\deg(v) = k) = (1 + o(1)) \frac{n^k}{k!} p^k e^{-pn} = (1 + o(1)) \frac{d^k e^{-d}}{k!} \quad .$$

Suppose we are interested in bounding the probability that the degree of a vertex is much higher or lower than its expectation. As a first attempt, we might try a union bound:

$$\mathbb{P}\mathrm{r}(\deg(v) \geq k) \leq \binom{n}{k} p^k \leq \frac{(np)^k}{k!} = \frac{d^k}{k!} \quad .$$

It turns out that we can do much better than this using the following theorem, called Chernoff's Bound.

Theorem 1 (Chernoff's Bound)**.** *Let $X_1, X_2, \ldots, X_n$ be independent 0-1 random variables with $\mathbb{P}\mathrm{r}(X_i = 1) \triangleq p_i$ Let $X = \sum_{i=1}^n X_i$, and let $\mu = \mathbb{E}(X) = \sum_{i=1}^n p_i$. Then for all $t > 0$,*

$$\mathbb{P}\mathrm{r}(X \geq \mu + t) \leq e^{\frac{-t^2}{2\mu + 2t/3}} \quad , \tag{1}$$

and for any t, $0 < t < \mu$,

$$\mathbb{P}\mathrm{r}(X \leq \mu - t) \leq e^{\frac{-t^2}{2\mu}} \quad . \tag{2}$$

It is often easier to apply the following bound, which can be easily derived from Theorem 1.

Corollary 2. *In the notation of Theorem 1, for $0 < \varepsilon < 1$,*

$$\mathbb{P}\mathrm{r}(|X - \mu| \geq \varepsilon\mu) \leq 2e^{-\varepsilon^2 \mu/3} \quad . \tag{3}$$

We now return to our task of bounding the tail of the degree distribution of a vertex v in $G \sim G_{n,p}$ for $p = d/n$. Recall that the expectation of the degree of v is $(1+o(1))d$. For vertex v, let X_i be 1 if there is an edge from vertex i to v and 0 otherwise. Then if $X = \sum_{i \in V \setminus \{v\}} X_i$ we have from (2) that

$$\mathbb{P}\mathrm{r}(\deg(v) \leq k) = \mathbb{P}\mathrm{r}(X \leq k) = \mathbb{P}\mathrm{r}(X \leq d - (d-k)) \leq e^{-(d-k)^2/2d} \quad .$$

Now suppose that $p = 8 \ln n / n$. Then $\mathbb{E}(\deg(v)) = 8 \ln n$, and

$$\mathbb{P}\mathrm{r}(\deg(v) \leq 4 \ln n) \leq e^{-\ln n} = \frac{1}{n} \quad .$$

Thus by a union bound, if $p = o(\ln n/n)$, then the probability that there exists a vertex with degree less than $\frac{1}{2}np$ approaches 0 as $n \to \infty$. In fact, it can be shown that $\ln n/n$ is a threshold function for the property of containing a vertex with degree less than $\frac{1}{2}np$.

So far, we have proved results about the degree distribution of a fixed vertex, but in the next lecture we will actually be interested in the distribution of the random variable

$$Y_k = \frac{|\{v \ : \ \deg(v) \geq k\}|}{n} \quad .$$

Because it is not clear how Y_k can be written as the sum of independent random variables, it seems that we cannot use Theorem 1 to bound the tail distribution of Y_k. In the hopes of bounding the tail distribution, we present the next theorem, called the Azuma-Hoeffding inequality. First, however, we need the following definition.

Definition 1. *Let $A_1, A_2, \ldots, A_n$ be finite sets. Then a function $f : \prod_{i=1}^n A_i \to \mathbb{R}$ satisfies* the Lipshitz condition *with bound $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ if for all $k \in [n]$ and $\mathbf{x}, \mathbf{x}' \in \prod_{i=1}^n A_i$, $x_i = x'_i$ for all $i \neq k$ implies that $|f(\mathbf{x}) - f(\mathbf{x}')| < c_k$.*

Informally, a function satisfies the Lipshitz condition if changing the value of a single coordinate does not change the value of the function by too much. We are now ready to state the Azuma-Hoeffding Inequality.

Theorem 3 (Azuma-Hoeffding Inequality)**.** *Let $X_1, X_2, \ldots, X_n$ be independent random variables, where X_i can only take on the values in the finite set A_i. Let $A = \prod_{i=1}^n A_i$ and suppose that $f : A \to \mathbb{R}$ satisfies the Lipshitz condition with bound $\mathbf{c} = (c_1, \ldots, c_n)$. Let $Z = f(X_1, \ldots, X_n)$ and let $\mu = \mathbb{E}(Z)$. Then*

$$\mathbb{P}\mathrm{r}(Z \geq \mu + t) \leq e^{\frac{-t^2}{2\sum_{k=1}^n c_k^2}} \tag{4}$$

and

$$\mathbb{P}\mathrm{r}(Z \leq \mu - t) \leq e^{\frac{-t^2}{2\sum_{k=1}^n c_k^2}} \quad . \tag{5}$$

How might Theorem 3 help us bound the tail distribution of Y_k? For all edges e in graph $G \sim G_{n,p}$, let X_e be an indicator random variable that is 1 if and only if edge e is present. Then clearly the X_e's are independent, and since adding or removing an edge can change the degree of at most two vertices, Y_k is a function of the X_e's that satisfies the Lipshitz condition with bound $2/n$. Thus,

$$\mathbb{P}\mathrm{r}(|Y_k - \mathbb{E}(Y_k)| \geq t) \leq 2e^{\frac{-t^2}{2\binom{n}{2}\cdot 4/n^2}} = 2e^{-\Theta(t^2)} \quad .$$

But since $t = O(1)$, this does not provide a very strong bound on the tail distribution of Y_k. The problem is that we needed to write Y_k as a function of $\binom{n}{2}$ independent random variables, which turned out to be too many. Indeed, if we can guarantee that there are $m = o(n^2)$ edges, then we can get more interesting bounds from Theorem 3.

Specifically, consider the following random multi-graph model $\tilde{G}_{n,m}$, in which there are m edges, and each edge is chosen independently and uniformly at random among all possible edges. (Since this is a multigraph, an edge can be chosen more than once, and the events of where each edge is chosen are independent.) Then if X_i is the position of the i^{th} edge, Y_k is again a function of the X_i's that satisfies the Lipshitz condition with bound $2/n$. Thus,

$$\mathbb{P}\mathrm{r}(|Y_k - \mathbb{E}(Y_k)| \geq t) \leq 2e^{\frac{-t^2 n^2}{8m}} \quad .$$

Since in this case, m can be $o(n^2)$, we can get a much stronger bound (if, for example $m = \Theta(n \log n)$ and $t = \Theta(n^{-1/3})$.)

Although the Azuma-Hoeffding inequality is widely applicable, we still do not have a satisfactory bound on the tail of the distribution of Y_k. We now present a final concentration inequality that will allow us to provide an interesting bound on the tail of the distribution of Y_k.

Theorem 4 (Symmetric Logarithmic Sobolev Inequality)**.** *Suppose that $X_1, X_2, \ldots, X_n$ and $X'_1, X'_2, \ldots X'_n$ are $2n$ i.i.d. random variables, and suppose that X_i and X'_i can only take on values in the finite set A_i. Let $f : \prod_{i=1}^{n} A_i \to \mathbb{R}$ be a function, and define the random variables Z and Z'_i so that $Z = f(X_1, X_2, \ldots, X_n)$ and $Z'_i = f(X_1, \ldots, X_{i-1}, X'_i, X_{i+1}, \ldots, X_n)$. If*

$$\mathbb{E}\left(\sum_{i=1}^{n} (Z - Z'_i)^2 \mathbf{1}_{Z \geq Z'_i} \middle| X_1, \ldots, X_n \right) < c \tag{6}$$

then

$$\mathbb{P}\mathrm{r}(Z \geq \mathbb{E}(Z) + t) \leq e^{\frac{-t^2}{4c}} \quad .$$

We can use this theorem to bound the tail of Y_k in the following way. Let G and G' be two random graphs in $G_{n,p}$. For $1 \leq i \leq \binom{n}{2}$, let X_i (resp. X'_i) be the indicator random variable that is 1 if and only if the i^{th} possible edge in G (resp. G') is present. Let $Z = Y_k$ and let Z'_i be defined as in the theorem. Then the expression on the left-hand side of (6) is a function of the X_i's, and hence a function of the graph G.

Fix a graph G, and suppose that in G, the i^{th} edge is not present. Then regardless of the value of X'_i, $Z \leq Z'_i$. Hence the contribution on the left-hand side of (6) is zero for such edges i. Now suppose that the i^{th} edge is present in G. Then if it is present in G' (which happens with probability p), the value of Z and Z'_i are the same. If it is not present in G', (which happens with probability $1 - p$, then $Z'_i < Z$ if and only if one of the endpoints of edge i has degree k in G. In this case, $(Z - Z'_i)^2 \mathbf{1}_{Z \geq Z'_i} \leq 4/n^2$. Of course, there are at most kn edges that are adjacent to vertices of degree k, so we can take c to be $(1 - p)4k/n \leq 4k/n$. Hence, we have that

$$\mathbb{P}\mathrm{r}(Y_k \geq \mathbb{E}(Y_k) + t) \leq e^{-t^2 n/16k} \quad .$$

Assuming that k is a constant, this probability approaches 0 for $t = o(1/\sqrt{n})$. Thus we finally have a theorem that provides an interesting tail bound on Y_k. This theorem will also be used in the next lecture on the bias of traceroute sampling.

3 Proofs

In this section, we prove the Chernoff and Azuma-Hoeffding bounds. (These proofs were not in the lecture.) We start with the proof of the Chernoff bound.

Proof of Theorem 1. To prove (1), note that for any $\alpha > 0$ (to be fixed later), $\mathbb{P}\mathrm{r}(X \geq \mu + t) = \mathbb{P}\mathrm{r}(e^{\alpha X} \geq e^{\alpha(\mu+t)})$. Hence, by Markov's Inequality, $\mathbb{P}\mathrm{r}(X \geq \mu + t) \leq \mathbb{E}(e^{\alpha X})/e^{\alpha(\mu+t)}$. Note that

$$\mathbb{E}\left(e^{\alpha X}\right) = \mathbb{E}\left(\prod_{i=1}^{n} e^{\alpha X_i}\right) = \prod_{i=1}^{n} \mathbb{E}\left(e^{\alpha X_i}\right) \quad ,$$

where the last equality follows by the independence of the X_i's. Using the inequality $1 + x \leq e^x$, we have that $\mathbb{E}(e^{\alpha X_i}) = 1 + p_i(e^\alpha - 1) \leq e^{p_i(e^\alpha-1)}$. Hence,

$$\frac{\mathbb{E}\left(e^{\alpha X}\right)}{e^{\alpha(\mu+t)}} \leq \frac{\prod_{i=1}^{n} e^{p_i(e^\alpha-1)}}{e^{\alpha(\mu+t)}} = \frac{e^{\mu(e^\alpha-1)}}{e^{\alpha(\mu+t)}} \quad .$$

Substituting $\alpha = \ln(1 + t/\mu) > 0$, we get

$$\mathbb{P}\mathrm{r}(X \geq \mu + t) \leq \frac{e^t}{\left(1 + \frac{t}{\mu}\right)^{\mu+t}} \quad . \tag{7}$$

To finish the proof of (1), we now show that for all $t > 0$ the right-hand side of (7) is no greater than the right-hand side of (1). Taking natural logarithms, we have the equivalent task of showing that the function

$$f(t) = t - (\mu + t)\ln\left(1 + \frac{t}{\mu}\right) + \frac{t^2}{2\mu + 2t/3}$$

is no greater than zero for all $t > 0$. Taking the first and second derivatives, we have

$$f'(t) = \frac{3t(t + 6\mu)}{2(t + 3\mu)^2} - \ln\left(1 + \frac{t}{\mu}\right)$$

and

$$f''(t) = \frac{-t^2(t + 9\mu)}{(t + \mu)(t + 3\mu)^2} \quad .$$

Since $f(0) = f'(0) = 0$, and since $f''(t) < 0$ for all $t > 0$, we have that $f(t) \leq 0$ for all $t > 0$. This finishes the proof of (1).

The proof of (2) is similar. For any $\beta < 0$, we have by Markov's Inequality that $\mathbb{P}\mathrm{r}(X \leq \mu - t) = \mathbb{P}\mathrm{r}(e^{\beta X} \geq e^{\beta(\mu-t)}) \leq \mathbb{E}(e^{\beta X})/e^{\beta(\mu-t)}$. Proceeding as we did in the proof of (1), we see that

$$\mathbb{P}\mathrm{r}(X \leq \mu - t) \leq \frac{e^{\mu(e^\beta-1)}}{e^{\beta(\mu-t)}} \quad ,$$

and substituting $\beta = \ln(1 - t/\mu) < 0$ yields

$$\mathbb{P}\mathrm{r}(X \leq \mu - t) \leq \frac{e^{-t}}{\left(1 - \frac{t}{\mu}\right)^{\mu-t}} \quad . \tag{8}$$

As in the proof of (1), it is straightforward to use elementary calculus to show that the right-hand side of (8) is no greater than the right-hand side of (2) for $0 < t < \mu$, which concludes the proof. $\square$

Next, we prove the Azuma-Hoeffding inequality.

Proof of Theorem 3. For $0 \le i \le n$, define the random variable $Y_i = \mathbb{E}(Z|X_1, X_2, \ldots, X_i)$. Note that $Y_0 = \mathbb{E}(Z)$ and $Y_n = Z$. We first prove the following claim.

Claim 5. *For all* $\mathbf{x} = (x_1, \ldots, x_n) \in A$ *and for all* $0 \le k < n$, $|Y_{k+1}(\mathbf{x}) - Y_k(\mathbf{x})| \le c_{k+1}$.

Proof. Let H be the set of $\mathbf{x}'$ that agree with $\mathbf{x}$ on the first $k+1$ coordinates. More precisely,

$$H = \left\{\mathbf{x}' = (x'_1, \ldots, x'_n) \in A \ : \ x'_i = x_i \text{ for } i \le k+1\right\} \ .$$

By the definition of Y_{k+1} and the fact that the X_i's are independent, we have

$$Y_{k+1}(\mathbf{x}) = \sum_{\mathbf{x}' \in H} f(\mathbf{x}') \prod_{i=k+2}^{n} \Pr(X_i = x'_i) \ .$$

For each $\mathbf{x}' \in H$, let

$$H[\mathbf{x}'] = \left\{\mathbf{x}^* = (x^*_1, \ldots, x^*_n) \in A \ : \ x^*_i = x'_i \text{ for } i \ne k+1\right\} \ .$$

Note that the $H[\mathbf{x}']$ partition the set of $\mathbf{x}^*$ that agree with $\mathbf{x}$ on the first k coordinates. Therefore,

$$\begin{aligned} Y_k(\mathbf{x}) &= \sum_{\mathbf{x}' \in H} \sum_{\mathbf{x}^* \in H[\mathbf{x}']} f(\mathbf{x}^*) \prod_{i=k+1}^{n} \Pr(X_i = x^*_i) \\ &= \sum_{\mathbf{x}' \in H} \sum_{\mathbf{x}^* \in H[\mathbf{x}']} f(\mathbf{x}^*) \Pr(X_{k+1} = x^*_{k+1}) \prod_{i=k+2}^{n} \Pr(X_i = x'_i) \ , \end{aligned}$$

and

$$\begin{aligned} |Y_{k+1}(\mathbf{x}) - Y_k(\mathbf{x})| &= \left| \sum_{\mathbf{x}' \in H} \prod_{i=k+2}^{n} \Pr(X_i = x'_i) \left(f(\mathbf{x}') - \sum_{\mathbf{x}^* \in H[\mathbf{x}']} f(\mathbf{x}^*) \Pr(X_{k+1} = x^*_{k+1}) \right) \right| \\ &\le \sum_{\mathbf{x}' \in H} \prod_{i=k+2}^{n} \Pr(X_i = x'_i) \sum_{\mathbf{x}^* \in H[\mathbf{x}']} \Pr(X_{k+1} = x^*_{k+1}) \left| f(\mathbf{x}') - f(\mathbf{x}^*) \right| \ , \end{aligned}$$

where to obtain the inequality we have rewritten $f(\mathbf{x}')$ as $\sum_{\mathbf{x}^* \in H[\mathbf{x}']} f(\mathbf{x}') \Pr(X_{k+1} = x^*_{k+1})$ and applied the triangle inequality. The Lipshitz condition gives $|f(\mathbf{x}') - f(\mathbf{x}^*)| \le c_{k+1}$, and so we conclude that $|Y_{k+1}(\mathbf{x}) - Y_k(\mathbf{x})| \le c_{k+1}$. □

We will also need the following.

Claim 6. *For* $0 \le k < n$, $\mathbb{E}(Y_{k+1}|Y_1, \ldots, Y_k) = Y_k$.

Proof. Using the fact that $\mathbb{E}(\mathbb{E}(V|U, W)|W) = \mathbb{E}(V|W)$ for all random variables U, V, and W, we have

$$\begin{aligned} \mathbb{E}(Y_{k+1}|Y_1, \ldots, Y_k) &= \mathbb{E}(Y_{k+1}|X_1, \ldots, X_k) \\ &= \mathbb{E}(\mathbb{E}(Z|X_1, \ldots, X_{k+1})|X_1, \ldots, X_k) \\ &= \mathbb{E}(Z|X_1, \ldots, X_k) \\ &= Y_k \ . \end{aligned}$$

□

We now prove (4). For $1 \le k \le n$, let $\Delta Y_k = Y_k - Y_{k-1}$. Then, by Claim 5, $\Delta Y_k \le c_k$, and by Claim 6,

$$\mathbb{E}(\Delta Y_k | Y_0, \ldots, Y_{k-1}) = \mathbb{E}(Y_k | Y_0, \ldots, Y_{k-1}) - \mathbb{E}(Y_{k-1} | Y_0, \ldots, Y_{k-1}) = 0 \ .$$

Fix an $\alpha > 0$, to be determined later. Writing ΔY_k as $-\frac{c_k}{2}(1 - \Delta Y_k / c_k) + \frac{c_k}{2}(1 + \Delta Y_k / c_k)$, we have by the convexity of the exponential function that

$$e^{\alpha \Delta Y_k} \le \frac{1 - \Delta Y_k / c_k}{2} e^{-\alpha c_k} + \frac{1 + \Delta Y_k / c_k}{2} e^{\alpha c_k} = \frac{e^{\alpha c_k} + e^{-\alpha c_k}}{2} + \frac{\Delta Y_k}{2 c_k}(e^{\alpha c_k} - e^{-\alpha c_k}) \ .$$

Hence,

$$\begin{aligned} \mathbb{E}(e^{\alpha \Delta Y_k} | Y_0, \ldots, Y_{k-1}) &\le \mathbb{E}\left(\frac{e^{\alpha c_k} + e^{-\alpha c_k}}{2} + \frac{\Delta Y_k}{2 c_k}(e^{\alpha c_k} - e^{-\alpha c_k}) \ \middle| \ Y_0, \ldots, Y_{k-1} \right) \\ &= \frac{e^{\alpha c_k} + e^{-\alpha c_k}}{2} \\ &\le e^{(\alpha c_k)^2 / 2} \ , \end{aligned}$$

where the last inequality follows because

$$e^{x^2/2} = \sum_{i \ge 0} \frac{(x^2/2)^i}{i!} \ge \sum_{i \ge 0} \frac{x^{2i}}{(2i)!} = \frac{1}{2} \sum_{i \ge 0} \frac{x^i}{i!} + \frac{1}{2} \sum_{i \ge 0} \frac{x^i (-1)^i}{i!} = \frac{e^x + e^{-x}}{2} \ .$$

We have

$$\begin{aligned} \mathbb{E}\left(e^{\alpha(Y_n - Y_0)}\right) = \mathbb{E}\left(\prod_{k=1}^{n} e^{\alpha \Delta Y_k} \right) &= \mathbb{E}\left(\left(\prod_{k=1}^{n-1} e^{\alpha \Delta Y_k} \right) \mathbb{E}(e^{\alpha \Delta Y_n} | Y_0, \ldots, Y_{n-1}) \right) \\ &\le \mathbb{E}\left(\prod_{k=1}^{n-1} e^{\alpha \Delta Y_k} \right) e^{(\alpha c_n)^2/2} \\ &\le e^{\alpha^2 \sum_{k=1}^{n} c_k^2 / 2} \ . \end{aligned}$$

Setting $\alpha = t / \sum_{k=1}^{n} c_k^2$, we use Markov's inequality to conclude that

$$\begin{aligned} \Pr(Z \ge \mu + t) = \Pr(Y_n - Y_0 \ge t) &= \Pr(e^{\alpha(Y_n - Y_0)} \ge e^{\alpha t}) \\ &\le \frac{\mathbb{E}\left(e^{\alpha(Y_n - Y_0)}\right)}{e^{\alpha t}} \\ &\le e^{\alpha^2 \sum_{k=1}^{n} c_k^2 / 2 - \alpha t} \\ &= e^{\frac{-t^2}{2 \sum_{k=1}^{n} c_k^2}} \ . \end{aligned}$$

A similar argument can be used to prove (5). □

4 Further reading

Concentration inequalities are treated in a number of references. The proofs of the Chernoff and Azuma-Hoeffding inequalities were adapted from [1] and [3]. The proof of Theorem 4 can be found in [2].

References

[1] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley, 2000.

[2] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 31:1583–1614, 2003.

[3] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.