Luca Trevisan
August 31, 2017

Summary of Lecture 3

In which we complete the study of Independent Set and Max Cut in $G_{n,p}$ random graphs.

1 Maximum Independent Set

Last time we proved an upper bound of $O\left(\frac{1}{p}\log np\right)$ to the probable value of the maximum independent set in a $G_{n,p}$ random graph, a bound that holds also for $p = p(n)$ being a function of n.

Consider the greedy algorithm

- $S := \emptyset$
- for each $v \in V$
  - if v has no neighbors in S then $S := S \cup \{v\}$
- return S

To analyze the algorithm, consider the following random variables: let t_i be the number of for-loop iterations between the time the i-th element is added to S and the time the $(i+1)$-th element is added to S (t_i is undefined if the algorithm terminates with a set S of size less than $i+1$). Thus the size of the independent set found by the algorithm is the largest i such that t_{i-1} is defined.

Consider now the following slightly different probabilistic process: in addition to our graph over n vertices $\{1,\ldots,n\}$, we also consider a countable infinite number of other vertices $n+1, n+2, \ldots$, we sample an infinite super-graph of our graph so that each possible edge has probability p of being generated, we continue to run the greedy algorithm for every vertex of this infinite graph, and we call t_i the (now, always defined) number of for-loop iterations between the i-th and the $(i+1)$-th time that we add a node to S. In this revised definition, the size of the independent set found by algorithm in our actual graph is the largest i such that $t_0 + t_1 + \ldots + t_k \leq n$.

We show that t_i has a geometric distribution with success probability $(1-p)^i$, and so $\mathbb{E}\, t_i = \frac{1}{(1-p)^i}$ and $\mathbf{Var} t_i = \frac{1-(1-p)^i}{(1-p)^{2i}}$, meaning that

$$\mathbb{E}\, t_0 + t_1 + \cdots t_k = \frac{\frac{1}{(1-p)^k} - 1}{\frac{1}{1-p} - 1} \leq \frac{1}{(1-p)^k}\left(\frac{1}{1-p} - 1\right)^{-1} = \frac{1}{p \cdot (1-p)^{k-1}}$$

$$\textbf{Var} t_0 + t_1 + \cdots t_k \leq \sum_{i=0}^{k} \frac{1}{(1-p)^{2i}} = \frac{\frac{1}{(1-p)^{2k}} - 1}{\frac{1}{(1-p)^2} - 1} \leq \frac{1}{(1-(1-p)^2) \cdot (1-p)^{2k-2}}$$
$$\leq \frac{1}{p \cdot (1-p)^{2k-2}} = p\,(\mathbb{E}\, t_0 + \cdots t_k)^2$$

If we choose a k such that $\mathbb{E}\, t_0 + \cdots t_k \leq \frac{n}{2}$, which is true if we choose

$$k = \log_{\frac{1}{p-1}} \frac{pn}{2} \approx \frac{1}{p} \ln pn$$

then we are also getting that the standard deviation of $t_0 + \cdots t_k$ is at most $pn/2$ and, if $p(n) \to 0$ we have a $1 - o(1)$ probability that $t_0 + \cdots t_k \leq n$, meaning that $|S| \geq k$.

Thus, if $p(n) \to 0$, the greedy algorithm has a $1 - o(1)$ probability of finding an independent set of size $\Omega(p^{-1} \log pn) = \Omega\left(\frac{n}{d} \log d\right)$.

In terms of certifiable upper bounds, the key bound is

Lemma 1 *If $p = p(n) > \frac{\log n}{n}$, G is sampled from $G_{n,p}$ and A is the adjacency matrix of G, then there is a $1 - o(1)$ probability that*

$$||A - pJ|| \leq O(\sqrt{pn})$$

If S is an independent set of size k, then $\mathbf{1}_S^T A \mathbf{1}_S = 0$, $\mathbf{1}_S^T J \mathbf{1}_S = k^2$, and $||\mathbf{1}_S||^2 = k$, so that

$$||A - pJ|| \geq pk$$

so we have that, if we denote by $\alpha(G)$ the size of the largest independent set in G,

$$\alpha(G) \leq \frac{1}{p} ||A - pJ||$$

In $G_{n,p}$ random graph, the above upper bound is, with high probability, $O(\sqrt{n/p}) = O(n/\sqrt{d})$.

In conclusion, in $G_{n,p}$ random graphs, the probable value of the largest independent set is $O\left(\frac{n}{d} \log d\right)$, the independent set found by the greedy algorithm has size $\Omega\left(\frac{n}{d} \log d\right)$ with high probability, and spectral methods provide with a high probability an $O(n/\sqrt{d})$ upper bound certificate, where $d = pn$.

2 Max Cut

The probability that a $G_{n,p}$ random graph, $d := pn$, has a cut cutting more than $\frac{dn}{4} + \epsilon dn$ is at most $e^{-\Omega(\epsilon^2 dn)}$, there are 2^n possible cuts, so with $2^{-\Omega(n)}$ probability the size of the maximum cut is at most $O(dn/4 + \sqrt{d}n)$.

Consider the greedy algorithm

- $A := \emptyset$
- $B := \emptyset$
- for each $v \in V$
  - if v has more neighbors in B than in A then $A := A \cup \{v\}$
  - else $B := B \cup \{v\}$
- return (A, B)

Let $V = \{1, \ldots, n\}$, A_i and B_i be the sets A, B when vertex i is considered in the for-loop and let a_i and b_i be their cardinality. Then the absolute value of the difference between the number of neighbors of i in A_i versus B_i has expectation $\Omega(\sqrt{pi})$ and variance $O(pi)$. Adding over all i, the sum of the differences (which is the gain over cutting half the edges), has mean $\Omega(n\sqrt{pn})$ and variance $O(pn^2)$, so the gain is $\Omega(n\sqrt{pn}) = \Omega(n\sqrt{d})$ with $1 - o(1)$ probability.

In terms of certifiable upper bounds, we have that if $S, V - S$ is a cut of cost $\frac{dn}{4} + C$, then we have

$$\mathbf{1}_S^T A \mathbf{1}_{V-S} = \frac{dn}{4} + C$$

$$\mathbf{1}_S^T pJ \mathbf{1}_{V-S} = p \cdot |S| \cdot |V - S| \leq p \cdot \frac{n^2}{4} = \frac{dn}{4}$$

$$||\mathbf{1}_S|| \cdot ||\mathbf{1}_{V-S}|| = \sqrt{|S| \cdot |V - S|} \leq \sqrt{\frac{n^2}{4}}$$

so

$$C \leq 2n \cdot ||\mathbf{1}_S|| \cdot ||\mathbf{1}_{V-S}||$$

This means that, in every graph, the maximum cut is upper bounded by

$$\frac{dn}{4} + \frac{n}{2} \left\| A - \frac{d}{n} J \right\|$$

which, in $G_{n,p}$ random graphs with $d = pn$ is with high probability $\frac{dn}{4} + O(n\sqrt{d})$.

So we have that the probable optimum is at most $\frac{dn}{4} + O(n\sqrt{d})$, the greedy algorithm finds a cut that, with high probability, has cost at least $\frac{dn}{4} + \Omega(n\sqrt{d})$ and, for $p > \frac{\log n}{n}$, spectral algorithms give upper bound certificates that the optimum is at most $\frac{dn}{4} + O(n\sqrt{d})$