Scribe: Mahshid Montazer

Lecture 2

In this lecture, we study the Max Cut problem in random graphs. We compute the probable value of its optimal solution, we give a greedy algorithm which is nearly optimal on random graphs and we compute a polynomial time upper bound certificate for it using linear algebra methods. We also study the problem of Maximum Independent Set in random graphs and we compute an upper bound to the probable value for its optimal solution.

1 Max Cut

Definition 1 *Max Cut: In an un-weighted graph $G = (V, E)$, a cut is defined as a partition of its vertices into two sets V_1 and V_2. Let $E(V_1, V_2)$ be the size of the cut (V_1, V_2) which is the number of the edges with one endpoint in V_1 and one endpoint in V_2. Max Cut is the the problem of finding a cut of largest size.*

To give a clear example, in every bipartite graph, a bipartition is a maximum cut. It is easy to show that the size of the maximum cut would be at least half of the number of the graph edges. One question that arises here is that how much more than half of the edges can we cut. The answer is: not that much in random graphs. We will show this claim in the following section.

2 Probable Value of Max Cut Optimal Solution

In this section, we compute the probable value of Max Cut optimal solution in random graphs. Our result is for samples of $G_{n,\frac{1}{2}}$, but the analysis will generalize to $G_{n,p}$.

Lemma 2 *For every fixed cut $(S, V - S)$, $\mathbb{E}[E(S, V \setminus S)] \leq \frac{n^2}{8}$.*

Proof: $\mathbb{E}[E(S, V \setminus S)] = |S|\,|V \setminus S|\,\frac{1}{2} = \frac{n^2}{8}$. □

Lemma 3 $\mathbb{P}[E(S, V \setminus S) \geq \frac{n^2}{8} + \epsilon\frac{n^2}{4}] \leq e^{-\Omega(\epsilon^2 n^2)}$ *where* $0 \leq \epsilon \leq \frac{1}{2}$.

Proof: The proof is by applying Chernoff bounds on the result of lemma 2. □

Lemma 4 *There is a constant $c > 0$ such that*

$$\mathbb{P}[\exists(S, V \setminus S) \mid E(S, V \setminus S) \geq \frac{n^2}{8} + \epsilon\frac{n^2}{4}] \leq 2^{-n}$$

where $\epsilon = \frac{c}{\sqrt{n}}$ and the probability is taken over the choice of $G = (V, E)$ from the distribution $G_{n,\frac{1}{2}}$.

Proof:

$$\begin{aligned}\mathbb{P}[\exists(S, V \setminus S) \mid E(S, V \setminus S) \geq \frac{n^2}{8} + \epsilon\frac{n^2}{4}] &\leq 2^n \cdot e^{-\Omega(\epsilon^2 n^2)} \\ &\leq 2^{-n}.\end{aligned}$$

for an appropriate choice of c. □

The above lemma clearly leads us to the following theorem.

Theorem 5 *There is a constant c such that w.h.p. Max Cut in $G_{n,\frac{1}{2}}$ is of size at most $\frac{n^2}{8} + c \cdot n^{1.5}$.*

Thus, we showed that in $G_{n,1/2}$, the probable value of Max Cut is at most $\frac{n^2}{8} + c \cdot n^{1.5}$.

3 Greedy Algorithm for Max Cut

Consider the following greedy algorithm for Max Cut:

```
• $A \leftarrow \emptyset, B \leftarrow \emptyset$
• for $v \in V$
    – if $v$ has more neighbors in $A$ than in $B$, then $B \leftarrow B \cup \{v\}$
    – else $A \leftarrow A \cup \{v\}$
• return $A$ and $B$
```

The above algorithm can be applied to any graph, but we will analyze it on random graphs. A naive analysis of the algorithm guarantees that our greedy algorithm cuts at least half of the edges, giving us an approximation ratio of 2. The reason is that at each step, we add at least half of the processing vertex's incident edges to the cut. However, a more careful analysis of the algorithm shows that it is near-optimal for random graphs. Below, we prove our claim for $G_{n,\frac{1}{2}}$.

Lemma 6 *With high probability over the choice of G from $G_{n,\frac{1}{2}}$, the greedy algorithm finds a cut of size $\frac{n^2}{8} + \Omega(n^{1.5})$.*

PROOF: Let $G(V,E) \sim G_{n,\frac{1}{2}}$ be the given graph and let $v_1, v_2, \cdots, v_n$ be the order in which we process the vertices. Note that at the time of processing v_i $(1 \leq i < n)$, we do not need to know the edges that connect v_i to any vertex v_j $(j > i)$. Let $a_i = |A|$ and $b_i = |B|$ be the size of sets A and B before processing v_i, respectively. Although G is given before we run the algorithm, for the sake of the analysis, we can assume that we are building it on the go and while processing each of the vertices. Remember that each edge of the graph would exists independently with probability $\frac{1}{2}$. For deciding where to put v_i, we generate a_i random bits and call their summation X_i. We also generate b_i random bits and call their summation Y_i. We put v_i in set A (respectively, B) if $X_i \leq Y_i$ (respectively, $Y_i < X_i$). Note that the more balanced A and B get, the worse it gets for the analysis. Also, note that the extra edges that the algorithm cuts other than half of the edges would be:

$$\sum_{1 \leq i \leq n} |X_i - Y_i| = E(A,B) - \frac{|E|}{2}.$$

We know that

$$X_i - Y_i = \frac{a_i - b_i}{2}.$$

Note that

$$\mathbb{E}[|X_i - Y_i|] = \Omega(\sqrt{i})$$

and

$$\mathbf{Var}(|X_i - Y_i|) = O(i).$$

Thus, we have that $\sum_{1 \leq i \leq n} |X_i - Y_i|$ has mean $\Omega(n^{1.5})$ and standard deviation $O(n)$. Thus, with $1 - O(1)$ probability we have:

$$\sum_{1 \leq i \leq n} |X_i - Y_i| = \sum_{1 \leq i \leq n} \Omega(\sqrt{i}) \geq \Omega(n^{1.5}).$$

$$\Rightarrow E(A,B) \geq \frac{n^2}{8} + \Omega(n^{1.5}).$$

□

4 Polynomial Time Upper Bound for Max Cut

In this section, we find polynomial time upper bound certificates for Max Cut in random graphs using linear algebra techniques.

Lemma 7 *Let $G = (V,E)$ be a graph, A be its adjacency matrix, J be the matrix all whose entries are 1 and $(S, V \setminus S)$ be the Max Cut of G. Then*

$$E(S, V \setminus S) \leq \frac{n^2}{8} + \frac{n}{2}||A - J/2||$$

Proof: we have:

$$\begin{aligned} E(S, V\setminus S) - \frac{n^2}{8} &\leq \mathbf{1}_S^T \cdot (A - J/2)\cdot \mathbf{1}_{V\setminus S} \\ &\leq ||A - J/2|| \cdot ||\mathbf{1}_S|| \cdot ||\mathbf{1}_{V\setminus S}|| \\ &\leq ||A - J/2|| \cdot \sqrt{|S|} \cdot \sqrt{|V \setminus S|} \\ &\leq ||A - J/2|| \cdot \frac{n}{2} \end{aligned} \tag{1}$$

□

Recall that, with high probability over the choice of a graph G from $G_{n,\frac{1}{2}}$, if A is the adjacency matrix of G then we have $||A - J/2|| \leq O(\sqrt{n})$ with high probability.

We conclude that, with high probability over the choice of G from $G_{n,\frac{1}{2}}$ we can find in polynomial time a certificate the max cut optimum of G is at most $\frac{n^2}{8} + O(n^{1.5})$.

5 Maximum Independent Set

In this section, we discuss the Maximum Independent Set problem for $G_{n,p}$ (especially $G_{n,\frac{1}{2}}$) and we show its close connection with Max Clique problem. Finally, we compute its optimal solution's probable value.

Definition 8 *Maximum Independent Set: In a graph $G(V,E)$, an independent set is a set of vertices that are mutually disconnected. A Maximum Independent Set in G is an independent set of largest possible size. The Maximum Independent Set problem is the problem of finding such a set.*

Note that the Maximum Independent Set in $G_{n,p}$ corresponds to the Maximum Clique in $G_{n,1-p}$. Thus, for $p = \frac{1}{2}$, everything that we argued for Max Clique is usable for Maximum Independent Set as well.

In this section, we compute an upper bound to the probable value of Maximum Independent Set's optimal solution in $G_{n,p}$.

Fix a set $S \subset V$ of size k. We have

$$\mathbb{P}[S \text{ is an independent set in } G] = (1-p)^{\binom{k}{2}}$$

where the probability is over the choice of $G \sim G_{n,p}$.

The following lemma holds.

Lemma 9 $\mathbb{P}[\exists$ *Independent Set of size* $k] \leq e^{-\frac{k}{2}\left(((k-1)\cdot \ln \frac{1}{1-p} - 2\ln\frac{n}{k}\right)}$

Proof:

$$\begin{aligned}
\mathbb{P}[\exists \text{ Independent Set of size } k] &\leq \mathbb{E}[\#\text{Independent Sets of size k}] \\
&= \binom{n}{k} \cdot (1-p)^{\binom{k}{2}} \\
&\leq \left(\frac{n}{k}\right)^k \cdot (1-p)^{\frac{k^2}{2}-\frac{k}{2}} \\
&= e^{k \cdot \ln \frac{n}{k} - \left(\frac{k^2}{2}-\frac{k}{2}\right)\cdot \ln \frac{1}{1-p}} \\
&= e^{-\frac{k}{2}((k-1)\cdot \ln \frac{1}{1-p} - 2\ln \frac{n}{k})}. \qquad (2)
\end{aligned}$$

□

Now, what would be the maximum value of k such that with high probability we can still make sure that there exists an independent set of size k? Note that the value of (2) goes to 0 when $k \geq 2\log_{\frac{1}{1-p}} \frac{n}{k} + 2$.

A sufficient condition for $k \geq 2\log_{\frac{1}{1-p}} \frac{n}{k} + 2$ is to have $k = 2\log_{\frac{1}{1-p}} n + 2$, showing us that there is a high probability that maximum independent set in $G_{n,p}$ is at most $O\left(\log_{\frac{1}{1-p}} n\right) = O\left(\frac{1}{p}\log n\right)$. A more careful bound is that we can have $k \geq 2\log_{\frac{1}{1-p}} \frac{n}{k} + 2$ provided, say, $k \geq 3\log_{\frac{1}{1-p}} np + 100$, and so with high probability the maximum independent set in $G_{n,p}$ is at most $O\left(\frac{1}{p}\log pn\right)$. If we call $d = pn$, then the bound is $O\left(\frac{n}{d}\log d\right)$