College of Computer & Information Science — Fall 2009
Northeastern University — 15 September 2009
CS 7880: Algorithmic Power Tools — Scribe: Dimitrios Kanoulas

Lecture Outline:

- Max-Cut Problem
- Probabilistic method and applications
- "Tail Bounds"
- Chernoff bounds

In this lecture, we complete the introduction to approximation algorithms and by using the max-cut and dominating set problem, we introduce the probabilistic method of solving problems. We also study some techniques for bounding the probability that a random variable takes value far from its expectation, known as tail bounds. For this purpose we introduce Markov's and Chebysevs Inequalities.

1 Max-Cut

In this section we introduce the Max-Cut problem and we analyze a greedy and a randomized algorithm that solve this problem.

Problem 1 (Max-Cut)**.** Given a graph $G = (V, E)$ determine a cut, which is a partition (S, T) of V, such that the number of edges that cross the cut (S, T) is maximized.

It is known that Max-Cut problem is NP-complete [1], that's why we seek good approximation algorithms.

1.1 Local Improvement Algorithm

First consider the following greedy algorithm.

1. Start with an arbitrary partition
2. Find a vertex u with more outgoing edges in its own side than to the other.
3. If such a u exists, switch u to the other side and goto Step 2.
4. Otherwise, end the procedure.

For example, if node u is in partition S and its edges that connect it with nodes into partition S are 4 and the edges that connect it with nodes into partition T are 3, then we will switch node u from partition S to partition T.

Some interesting questions comes from this greedy algorithm:

- Does the greedy Local Improvement Algorithm for Max-Cut terminates?
- If it does, what cut does it produce and what is the size of this cut?

The answers for these questions are:

- The greedy algorithm terminates.

  To prove that, we introduce a measure (such a measure is often referred to as a potential function) that is strictly increasing in each step of the greedy algorithm.
  A strictly increasing potential function for this problem could be the the size of the cut, since if u moves from T to S, the size of cut increases by $d^T(u) - d^S(u) > 0 \Rightarrow d^T(u) - d^S(u) \geq 1$, where $d^T(u)$ is the number of edges from u to T and $d^S(u)$ is the number of edges from u to S.
  Since the number of edges is a fixed number, the algorithm will terminate. In particular, in every step the size of cut increases by 1 and the total number of steps are at most $|E|$, which give a poly-time complexity to the local improvement algorithm.

- The size of the produced cut is:

  $$\frac{\sum_{u\in S} d^T(u) + \sum_{u\in S} d^S(u)}{2} \geq \frac{\sum_{u\in S} d(u) + \sum_{u\in S} d(u)}{2} \geq \frac{\frac{2|E|}{2}}{2} = \frac{|E|}{2}$$

  since for: $u \in S : d^T(u) \geq \frac{d(u)}{2}$ and for $u \in T : d^S(u) \geq \frac{d(u)}{2}$.

Note that the Local Improvement Algorithm achieves a 2-approximation since optimal cut has size at most $|E|$. Also it has been shown that the the weighted version of Local Improvement Algorithm also terminates, but not necessarily in poly-time.

1.2 "Naive" Randomized Algorithm

Consider the following trivial randomized algorithm that does not depend on the kind of the graph:

- Place each vertex u in S or T independently and uniformly at random.

For this randomized method, it makes sense to talk about the expected value of the size of the cut, that we denote by E[size of cut]. To analyze the expected size of the cut we will use the linearity of expectation:

$$E[\alpha x + \beta x] = \alpha E[x] + \beta E[x]$$

Consider the following variable X_e for a fixed edge e:

$$X_e = \begin{cases} 1 & \text{if e is in the cut} \\ 0 & \text{otherwise} \end{cases}$$

Now, the expected value of X_e is:

$$\begin{aligned} E[X_e] &= 1 \cdot \Pr[\text{One node of e is in S and the other in T}] \\ &+ 1 \cdot \Pr[\text{One node of e is in T and the other in S}] \\ &+ 0 \cdot \Pr[\text{Both nodes of e is in S}] \\ &+ 0 \cdot \Pr[\text{Both nodes of e is in T}] \\ &= \frac{1}{4} + \frac{1}{4} = 1/2 \end{aligned}$$

since we placed each vertex u in S or T independently and uniformly at random.

To find the expected size of the cut consider the weighted version of Max-Cut problem, where for each edge we assign a weight $w : E \to Z$. Then the size of the cut equals $\sum_e w(e)X_e$.
The expected size of the cut is:

$$\begin{aligned} E[\text{size of the cut}] &= E[\sum_e w(e)X_e] = \\ &= \sum_e w(e)E[X_e] = \\ &= \frac{1}{2}\sum_e w(e) = \\ &= \frac{1}{2}(\text{The total weights of all edges}) \end{aligned}$$

So the expected size of the cut is:

$$\begin{aligned} E[\text{size of cut}] &= \frac{\text{The total weights of all edges}}{2} \\ &\geq \frac{\text{OPT}}{2} \end{aligned}$$

since we consider OPT to be $\sum_e w(e)$.

Since the randomized algorithm achieves exprected cut size at least $\frac{|E|}{2}$, it follows that for all instances of Max-Cut problem there exists a cut that has at least the half of the total number of edges.

Probabilistic Method

A way to prove the existence of some property or structure is to use the probabilistic method that follows the following rules:

- If $E[X] \geq \alpha$, then there exists an instance for which $X \geq \alpha$.
- If $\Pr[E] > 0$, then there exists an instance in which E holds.

2 Dominating Set Problem

We give one more problem that can be analyzed by using the probabilistic method.

Definition 1 (Dominating Set)**.** In a graph $G = (V, E)$, where $|V| = n$, a dominating set is a subset $S \subseteq V$ such that for every $u \in V$, either $u \in S$ or there exists a node $v \in S$ such that $(u, v) \in E$.

Theorem 1 (Dominating Set Problem)**.** *For every graph G in which the degree of every vertex is at least d, there exists a dominating set of size at most* $\frac{n}{d+1}(1 + ln(d + 1))$*.*

We will try to prove the theorem under two different approaches.

First Approach. Pick a set of vertices of size t, uniformly at random. We need to prove that:

$$\Pr[\text{All vertices are being covered}] > 0$$

If the covered vertices were independent the probability Pr[u is covered] could be calculated easily. Although the way that vertices are covered in dominating sets makes this calculation difficult due to dependences. □

So we have to try a different approach, which whould be more easily analyzed than the first one.

Second Approach. Pick any vertex independently and uniformly at random with probability p. Let us call this picked set X. Then the expected size of X would be $E[|X|] = np$. Since the degree of every vertex is at least d, the probability $\Pr[\text{u is not covered by X}] \leq (1-p)^{d+1}$. Call Y the set of all uncovered vertices. Then the expected size of Y would be $E[|Y|] \leq n(1-p)^{d+1}$. So:

$$\begin{aligned} E[|X \cup Y|] &= E[|X|] + E[|Y|] \\ &\leq n[p + (1-p)^{d+1}] \\ &\leq n[p + e^{-p(d+1)}] \qquad \text{(since } (1-p) \leq e^{-p}) \\ &\leq n[\frac{\ln(d+1)}{(d+1)} + \frac{1}{(d+1)}] \\ &= \frac{n}{(d+1)}[1 + \ln(d+1)] \end{aligned}$$

The last inequality comes from the fact that we choose such a p that minimizes the term $(p + e^{-p(d+1)})$. The derivative of this term is $1 + (d+1)e^{-p(d+1)}$. So we have the minimum value when $1 + (d+1)e^{-p(d+1)} = 0 \Rightarrow p = \frac{\ln(d+1)}{(d+1)}$.

So we proved that there exists a dominating set of size at most $\frac{n}{(d+1)}[1 + \ln(d+1)]$, that completes are proof. □

3 Tail Bounds

In this section we examine ways to bound the probability that a random variable deviates from its expected value. We call these bounds tail bounds. Here we study Markov's and Chebysev's inequalities that can easily apply to many problems when we use probabilistic techniques.

3.1 Markov's Inequality

We first give Markov's inequality which applies to every non-negative real-valued random variable. It only uses the expectation and hence does not yield strong bound in most cases. But it will help us to find better tail bounds.

Theorem 2 (Markov's Inequality)**.** *Let X be a non-negative real-valued random variable. Then for all $\alpha > 0$:*

$$\Pr[X \geq \alpha] \leq \frac{E[X]}{\alpha}$$

Proof. For $\alpha > 0$, the expected value of X is:

$$\begin{aligned} E[X] &= E[[X : (X < \alpha)] + [X : (X \geq \alpha)]] \\ &\geq E[X : (X \geq \alpha)] \\ &= \sum_{x \geq \alpha} (x \Pr[X = x]) \\ &\geq \alpha \sum_{x \geq \alpha} (\Pr[X = x]) \\ &= \alpha \Pr[X \geq \alpha] \end{aligned}$$

So:

$$\frac{E[X]}{\alpha} \geq \Pr[X \geq \alpha]$$

$\square$

If we consider an non-decreasing function $f : R \to R$, then the following inequality holds:

$$\Pr[x \geq \alpha] \geq \frac{E[X]}{\alpha} \Rightarrow \Pr[f(X) \geq f(\alpha)] \leq \frac{E[f(X)]}{f(\alpha)}$$

For example, if we consider n boolean random variables $x_1, x_2, \ldots, x_n$ such that $\Pr[X_i = 1] = \frac{1}{2}, \forall i$ and $X = \sum_{i=1}^{n} x_i$, then the expected value of X will be $E[X] = \frac{n}{2}$ and by using Markov's inequality $\Pr[X \geq \frac{3n}{4}] = \frac{n/2}{3n/4} = \frac{2}{3}$.

3.2 Chebysev's Inequality

By using the expectation and the variance of random variable X we can get a stronger tail bound known as Chebysev's Inequality.

Theorem 3 (Chebysev's Inequality)**.** *For any $\lambda > 0$:*

$$\Pr[|X - E[X]| \geq \lambda] \leq \frac{Var(X)}{\lambda^2}$$

Proof. By using the fact that $(X - E[X])^2$ is a nonnegative random variable and applying Markov's inequality, we can bound the probability of random variable X to deviate from its expected value $E[X]$. For $\lambda > 0$:

$$\begin{aligned}
\Pr[|X - E[X]| \geq \lambda] &= \Pr[(X - E[X])^2 \geq \lambda^2] \\
&\leq \frac{E[(X - E[X])^2]}{\lambda^2} \\
&= \frac{E[X^2 - 2E[X] + E[X]^2]}{\lambda^2} \\
&= \frac{E[X^2] - E[X]^2}{\lambda^2} \\
&= \frac{Var(X)}{\lambda^2}
\end{aligned}$$

So:

$$\Pr[|X - E[X]| \geq \lambda] \leq \frac{Var(X)}{\lambda^2}$$

□

For the previous example, we have the following tail bound, by using Chebysev's inequality:

$$\Pr[X \geq \frac{3n}{4}] = \frac{n/4}{n^2/16} = \frac{4}{n}$$

We can see that this is a much better bound than Markov's, for n sufficiently large.

3.3 General Example

By generalizing the previous example, consider the following problem:

Problem 2. Assume n independent Bernoulli trials $x_1, x_2, \ldots, x_n$ and let X be the sum of these n trials. Also, suppose that $\Pr[x_i = 1] = p_i$ and $\Pr[x_i = 0] = 1 - p_i$. What is the bound of the following probability:

$$\Pr[X \geq (1 + \delta)E[x]]$$

Let $\mu = E[x]$ be the median. Then $\mu = \sum_{i=1}^{n} p_i$ where $p_i = \frac{1}{2}, \forall i$.

- If we apply Markov's inequality we have that $\Pr[X \geq (1 + \delta)\frac{n}{2}] = \frac{1}{1+\delta}$.
- If we apply Chebysev's inequality we have that $\Pr[X \geq (1 + \delta)\frac{n}{2}] \leq \frac{\frac{n}{4}}{\frac{\delta^2 n^2}{4}} = \frac{1}{\delta^2 n}$.

In the next lecture we will see that we can find even better bounds (Chernoff Bounds) than obtained by using the Markov's and Chebyshev's inequalities.

Note that we can apply Chebysev's inequality even when there are dependences between $x_i' s$, as long as the variables are pairwise dependent.

References

[1] Garey, Michael R.; Johnson, David S. (1979), Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman, ISBN 0-7167-1045-5.