CS 591-06, Lecture 10

Jared Saia
University of New Mexico

Class Outline

A reasonable probability is the only certainty. - E.W. Howe

- Majority-Preserving Property
- Chernoff Bounds
- In-Class Exercise
- Proof for Spam Resistant P2P network

Majority-Preserving Property

A bipartite graph $G = (L, R)$ where $|L| = |R| = n$ is said to have the *majority preserving property* iff for some $\epsilon > 0$,

- For all sets $L' \subset L$ such that $|L'| \geq 2n/3$, the number of nodes in R which have a majority of their neighbors in the set L' is at least $n - \epsilon n / \log n$.

Claim:

- A bipartite graph where every node in L has edges to $C \log n$ random nodes in R has the *majority preserving property*. Further the constant C depends only on ϵ.

Chernoff Bounds

- Let $X = X_1, \ldots, X_m$ where the X_i are random variables which are: independent and all take on values between 0 and 1.

then for any $0 < \delta \leq 1$,

- $P(|X - E(X)| \geq \delta E(X)) \leq 2e^{\frac{-E(X)\delta^2}{4}}$

Review: Linearity of Expectation

Linearity of Expecation Fact: If there are random variable $X, X_1, \ldots, X_n$, such that $X = \sum_{i=1}^{n} X_i$, then

- $E(X) = \sum_{i=1}^{n} E(X_i)$

Example:

Let the variable X_i be 1 if the i-th flip of a coin is heads and 0 otherwise and say that $E(X_i) = 1/2$. Let $X = \sum_{i=1}^{n} X_i$. Then:

$E(X) = n/2$

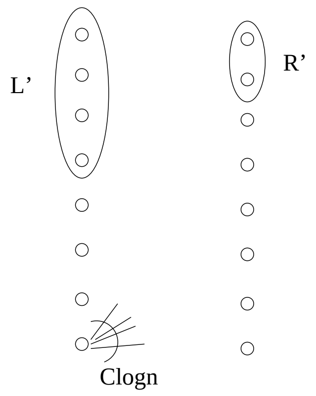
Chernoff Example

- Let $X = X_1, \ldots, X_m$, where each X_i is 1 if the i-th flip of a coin flip is heads tails otherwise. Let the coin turn up heads with probability 1/2.
- Q: What is a bound on the probability that either the number of heads exceeds $(3/4)m$ or the number of heads is less than $(1/4)m$?

Majority Preserving Property

- Assume we've got a biparitite graph $G = (L, R)$, where each vertex in L has $C \log n$ edges to random nodes in R. (The edges are chosen uniformly and independently at random),
- Consider some fixed sets $L' \subset L, R' \subset R$ such that $|L'| = 2n/3$ and $|R'| = \epsilon n / \log n$, for some fixed $\epsilon > 0$
- **Question:** What is the probability that no node in R' has a majority of its edges to nodes in L'?

Expander Picture



Probability of Failure

- **Question:** What is the probability that no node in R' has a majority of its edges to nodes in L'?
- This bad event occurs only if the number of edges between L' and R' is no more than the number of edges between $L-L'$ and R'
- Q1: What is the expected number of edges between L' and R'?
- Q2: What is the expected number of edges between $L-L'$ and R'?

Probability of Failure

- **Question:** What is the probability that no node in R' has a majority of its edges to nodes in L'?
- This bad event occurs only if the number of edges between L' and R' is no more than the number of edges between $L-L'$ and R'
- Q1: What is the expected number of edges between L' and R'? A: $(2/3)\epsilon nC$
- Q2: What is the expected number of edges between $L-L'$ and R'? A: $(1/3)\epsilon nC$

Probability of Failure

- Bad Event 1: The number of edges between L' and R' is less than $9/10 * ((2/3)\epsilon nC)$
- Bad Event 2: The number of edges between $L-L'$ and R' is greater than $10/9 * ((1/3)\epsilon nC)$
- We can use a union bound to bound the probability that either bad event occurs.

Bad Event 1

- Let X be the number of edges between L' and R'
- Note that $X = X_1, \ldots, X_m$ where X_i is 1 if the i-th edge from L' lands in R' and is 0 otherwise and $m = (2/3)nC \log n$
- From before, $E(X) = ((2/3)\epsilon nC)$
- By Chernoff bounds:

$$\begin{aligned}
P(|X - E(X)| > (1/10)E(X)) &\leq 2e^{\frac{-E(X)(1/10)^2}{4}} && (1)\\
&\leq 2e^{\frac{-(2/3)\epsilon nC}{400}} && (2)\\
&= 2e^{\frac{-\epsilon nC}{600}} && (3)\\
& && (4)
\end{aligned}$$

Bad Event 2

- Let Y be the number of edges between $L - L'$ and R'
- Note that $Y = Y_1, \ldots, Y_l$ where Y_i is 1 if the i-th edge from $L - L'$ lands in R' and is 0 otherwise, and $l = (1/3)nC \log n$
- From before, $E(Y) = (1/3)\epsilon nC$
- By Chernoff bounds:

$$
\begin{aligned}
P(|Y - E(Y)| > (1/10)E(Y)) &\leq 2e^{\frac{-E(Y)(1/10)^2}{4}} && (5)\\
&\leq 2e^{\frac{-(1/3)\epsilon nC}{400}} && (6)\\
&= 2e^{\frac{-\epsilon nC}{1200}} && (7)
\end{aligned}
$$

Probability of failure

- Let γ_1 be bad event 1 and γ_2 be bad event 2
- Then:

$$
\begin{aligned}
P(\gamma_1 \bigcup \gamma_2) &\leq P(\gamma_1) + P(\gamma_2) && (8)\\
&\leq 2e^{\frac{-\epsilon nC}{600}} + 2e^{\frac{-\epsilon nC}{1200}} && (9)\\
&\leq 4e^{\frac{-\epsilon nC}{1200}} && (10)
\end{aligned}
$$

- This is a bound on the probability that the majority preserving property fails for a fixed set L' and a fixed set R'.
- It's very small, but is it small enough???

Prob of Any Failure

- We know that the prob. of failure for a fixed set L' and fixed set R' is no more than $4e^{\frac{-\epsilon nC}{1200}}$
- There are $\binom{n}{n/2}$ ways to choose a set L'
- There are $\binom{n}{\epsilon n/\log n}$ ways to choose a set L'
- $\binom{n}{n/2}$ and $\binom{n}{\epsilon n/\log n}$ are both no more than e^n

Prob of Any Failure

- For any set L' and R' of appropriate size, let $\gamma_{L',R'}$ be the probability that the bad event occurs for the sets L' and R'
- We want $P(\bigcup_{L' \subset L, R' \subset R \text{ s.t. } |L'|=n/2, |R'|=\epsilon n/\log n} \gamma_{L',R'})$
- Use the union bound:

$$
\begin{aligned}
P(\bigcup_{L',R'} \gamma_{L',R'}) &\leq \sum_{L',R'} P(\gamma_{L',R'}) && (11)\\
&= \binom{n}{n/2} * \binom{n}{\epsilon n/\log n} * (4e^{\frac{-\epsilon nC}{1200}}) && (12)\\
&\leq e^n * e^n * 4e^{\frac{-\epsilon nC}{1200}} && (13)\\
&= e^{\frac{-\epsilon nC + 2400n + 2400}{1200}} && (14)
\end{aligned}
$$

Analysis

- For C chosen dependent only on ϵ, this probability can be to go to 0 exponentially quickly as n grows large.
- This means that with high probability, for appropriate constant C, a *random* bipartite graph will have the majority preserving property.
- Note: Our proof made use of the union bound and Chernoff bounds.

Some Useful Bounds

- Stirlings Formula: $n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n\left(1+\frac{1}{12n}+O(\frac{1}{n^2})\right)$
- $n! \geq \left(\frac{n}{e}\right)^n$
- $\binom{n}{k} \leq \frac{n^k}{k!}$
- $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$

In Class Exercise

- Consider a directed random graph with possible loops on n vertices where each of the n^2 possible edges is included with probability p.
- Q1: What is the probability that some fixed set of $n/2$ vertices share more than $n^2p/4$ edges?
- Q2: What is an upperbound on the probability that the graph contains *some* set of $n/2$ vertices which share $n^2p/4$ edges?