Succinct Graph Structures and Their Applications **Spring 2018**

Lecture 5: May 9

Lecturer: Yinon Nahum

Note: In this course, we make a distinguish between "with high probability" (w.h.p.) which means with probability $\geq 1 - 1/n^c$ for some constant $c \geq 1$, and "with constant probability (w.c.p.) which means with probability $\geq 1 - \epsilon$ for a constant $\epsilon > 0$ which is arbitrarily small.

5.1 Expectation

One of the most important properties of a random variable is its expectation.

Definition 5.1 (Expectation) *The* expectation *of a discrete*[1] *r.v.* X *is given by*

$$\mathbb{E}[X] = \sum_x x \cdot \Pr[X = x] . \tag{5.1}$$

Properties:

1. Linearity: $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, $\mathbb{E}[c \cdot X] = c \cdot \mathbb{E}[X]$
2. For independent r.v's: $\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$.
3. For general r.v.'s (Cauchy Schwartz): $\mathbb{E}[X \cdot Y] \leq \sqrt{\mathbb{E}[X^2]} \cdot \sqrt{\mathbb{E}[Y^2]}$

Use:

Showing that there **exists** an object of at least (respectively, at most) a certain size. This is done by adding probability and calculating the *expected* size of the object.

1. In Ex. 1 you were asked to show the existence of a graph G with $\Omega(n^{1+1/(2k-1)})$ edges and girth at least $2k + 1$. Denoting the number of edges by E and the number of cycles of length at most $2k$ by C, it is enough to show that there exists a graph with $E - C \geq \Omega(n^{1+1/(2k-1)})$ (since then we can remove an edge from each cycle and obtain such G). By picking an appropriate p, we show that in $G(n, p)$, we have $\mathbb{E}[E - C] \geq \Omega(n^{1+1/(2k-1)})$.
2. Note that finding a Maximal IS can be done by traversing the vertices in some order, adding the current vertex to the IS and discarding its neighbors. We can show that in any graph, there exist a Maximal Independent Set of size at least $\sum_v \frac{1}{1+deg(v)}$. This is done by choosing a (uniformly) random order. We then have

$$\begin{aligned}\mathbb{E}[|IS|] &= \mathbb{E}\left[\sum_v 1_{v \in IS}\right] = \sum_v \mathbb{E}[1_{v \in IS}] = \sum_v \Pr[v \in IS] \geq \sum_v Pr[v \text{ appears before its neighbors}] \\ &= \sum_v \frac{1}{1 + deg(v)} .\end{aligned}$$

   Note that an arbitrary permutation can lead to a very small IS. To see this, consider the star graph.

[1] In this course, the R.V.'s will typically correspond to nonnegative integers.

If X is nonnegative, we can also use expectation to show that there **exist many** objects of at most a certain size. This is done by the following inequality.

Lemma 5.2 (Markov's Inequality) *For a nonnegative variable $X \geq 0$ and a real number $c > 0$,*

$$\Pr[X \geq c] \leq \frac{\mathbb{E}[X]}{c} . \tag{5.2}$$

This means that if X is nonnegative, then with constant probability $X = O(\mathbb{E}[X])$ (here w.p. $\geq 1-\epsilon$, we have $X \leq \mathbb{E}[X]/\epsilon$), then the probability tends to 0. This can be used in order to show that w.c.p., the number of edges in $G(n,p)$ is $O(n^2 p)$. Markov's inequality showed that nonnegative r.v.'s cannot exceed their expectation by much, i.e., w.c.p. they are bounded from above by the order of their expectation. But what about bounding them from below? For this we consider sums of independent r.v. and weekly dependent r.v.

5.2 Sum of Independent R.V.'s

A sum of independent r.v.'s is concentrated around its mean. In particular, if X is a *Binomial*(n,p) r.v. (i.e., we can write $X = \sum_{i=1}^n X_i$ where $X_i = 1$ w.p. p and 0 otherwise), then w.c.p. $X = \mathbb{E}[X] \pm O(\sqrt{\mathbb{E}[X]})$, and w.h.p. $X = \mathbb{E}[X] \pm O(\sqrt{\ln n \cdot \mathbb{E}[X]})$. This can be seen by applying the following Chernoff bounds.

Theorem 5.3 (Chernoff's Inequality) *Let X be a Binomial(n,p) r.v. Then for $0 \leq \delta \leq 1$,*

$$\Pr[X \geq (1+\delta)] \leq e^{-\frac{\delta^2 \cdot \mathbb{E}[X]}{3}} , \tag{5.3}$$

$$\Pr[X \leq (1-\delta)] \leq e^{-\frac{\delta^2 \cdot \mathbb{E}[X]}{2}} . \tag{5.4}$$

and for $\delta \geq 1$

$$\Pr[X \geq (1+\delta)] \leq e^{-\frac{\delta \cdot \mathbb{E}[X]}{3}} . \tag{5.5}$$

For bounded X_i's which are not necessarily bernoulli, we can show that w.c.p. $X = \mathbb{E}[X] \pm O(\sqrt{n})$, and w.h.p. $X = \mathbb{E}[X] \pm O(\sqrt{n \cdot \ln n})$. This follows from Hoeffding's Inequality.

Theorem 5.4 (Hoeffding's Inequality) *Letting $X_1, \ldots, X_n$ be independent random variables such that $a_i \leq X_i \leq b_i$. Then for the sum $X = \sum_{i=1}^n X_i$ we have for every real $c \geq 0$,*

$$\Pr[X - \mathbb{E}[X] \geq c] \leq \exp\left(\frac{-2c^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \tag{5.6}$$

and

$$\Pr[X - \mathbb{E}[X] \leq -c] \leq \exp\left(\frac{-2c^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \tag{5.7}$$

This can be used to show that a random walk on $\mathbb{Z}$ that starts at 0, will be after n steps, w.c.p. at $O(\sqrt{n})$ and w.h.p at $O(\sqrt{n \ln n})$. Another example is that the number of times a 6 was rolled in n attempts is w.c.p. $n/6 \pm O(\sqrt{\mathbb{E}[X]})$, and w.h.p. $n/6 \pm O(\sqrt{\ln n \cdot \mathbb{E}[X]})$.

5.3 Sum of Weekly Dependent R.V.'s - Variance and Second Moment

For this we consider the second moment of a random variable X, defined as $\mathbb{E}\left[X^2\right]$. We also define the variance.

Definition 5.5 *The Variance of a r.v. X is given by*

$$VAR[X] \;=\; \mathbb{E}\left[(X-\mathbb{E}[X])^2\right] \;=\; \mathbb{E}\left[X^2\right]-(\mathbb{E}[X])^2\ . \tag{5.8}$$

Note that the variance is nonnegative, thus the second moment is at least as large as the expectation squared.

Use:

If $X=\sum_{i=1}^{n} X_i$ is a sum of r.v's which are not necessarily independent, and $\mathbb{E}\left[X^2\right]/(\mathbb{E}[X])^2 \to 1$ (**happens when the X_i's have small dependencies!!**), then w.c.p. $X=\mathbb{E}[X]\pm o(\mathbb{E}[X])$ and w.h.p., $X=\mathbb{E}[X]\pm o(\sqrt{n\cdot VAR[X]})$ This follows from Chebyshev's Inequality.

Theorem 5.6 (Chebyshev's Inequality) *For a r.v. X with finite expectation and variance, and a real number $c \geq 0$,*

$$\Pr[|X-\mathbb{E}[X]| \geq c\cdot\sqrt{VAR[X]}] \leq \frac{1}{c^2}. \tag{5.9}$$

1. We can prove that w.c.p. the graph $G(n, p = c\ln n/n)$ contains isolated vertices if $c<1$. Denote the number of isolated vertices by $X=\sum_v 1_{v \text{ isolated}}$. We have

$$\mathbb{E}[X]=\sum_v \Pr[v \text{ isolated}] = n\cdot(1-p)^{n-1} \approx n^{1-c}\ , \tag{5.10}$$

   where the last step follows since $1-p\approx e^{-p}$ for small p's. We also have

$$\mathbb{E}\left[X^2\right]=\sum_{v,u}\Pr[v \text{ isolated}\wedge[u \text{ isolated}] = n\cdot(1-p)^{n-1}+n(n-1)\cdot(1-p)^{2n-3}\approx n^{1-c}+n^{2-2c}\ . \tag{5.11}$$

   Therefore, the ratio tends to 1 and w.c.p. $X=\mathbb{E}[X]\pm o(\mathbb{E}[X])$. In particular, w.c.p $X>0$. Note that for $c>1$, we have $\mathbb{E}[X]\to 0$, and by Markov $\Pr[X\geq 1]\to 0$, thus w.c.p. there are no isolated vertivces.

2. Letting X the number of triangles in $G(n, p=\omega(1/n))$, we can show that w.c.p $X=\mathbb{E}[X]\pm o(\mathbb{E}[X])$. This is done by calculating $\mathbb{E}\left[X^2\right]$ and $(\mathbb{E}[X])^2$ and showing that their ratio tends to 1. Letting $Y_{i,j,k}=1$ if the nodes i,j,k form a triangle and 0 otherwise. Then $X=\sum_{(i,j,k)} Y_{i,j,k}$. Thus $\mathbb{E}[X]=\binom{n}{3}p^3$. We then have

$$X^2=\sum_{(i,j,k),(i',j',k')} Y_{i,j,k}\cdot Y_{i',j',k'}\ . \tag{5.12}$$

   Note that if the tuples (i,j,k) and (i',j',k') share at most one node, then $Y_{i,j,k}$ and $Y_{i',j',k'}$ are independent. Let S be the set of all such tuple pairs. If they share exactly 2 nodes (there are $\binom{n}{4}\cdot\binom{4}{2}$ such summands), then the probability that the summand is 1 is p^5. If they share all 3 nodes (there are $\binom{n}{3}$ such summands), then the probability that the summand is 1 is p^3. Therefore, we have

$$\begin{aligned}
\mathbb{E}\left[X^2\right] &= O(n^4p^5) + O(n^3p^3) + \sum_S \mathbb{E}\left[Y_{i,j,k} \cdot Y_{i',j',k'}\right] \\
&= O(n^4p^5) + O(n^3p^3) + \sum_S \mathbb{E}\left[Y_{i,j,k}\right] \cdot \mathbb{E}\left[Y_{i',j',k'}\right] \\
&\leq O(n^4p^5) + O(n^3p^3) + \sum_{(i,j,k),(i',j',k')} \mathbb{E}\left[Y_{i,j,k}\right] \cdot \mathbb{E}\left[Y_{i',j',k'}\right] \\
&= O(n^4p^5) + O(n^3p^3) + (\mathbb{E}\left[X\right])^2 \ .
\end{aligned}$$

Noting that the ratio between $\mathbb{E}\left[X^2\right]$ and $(\mathbb{E}\left[X\right])^2$ tends to 1, we apply Chebyshev and obtain that w.c.p. $X = \binom{n}{3} \cdot p^3 \pm O(\max\{n^2p^{2.5}, n^{1.5}p^{1.5}\})$

5.4 The Geometric Distribution and Memorylessness.

Definition 5.7 *A random variable X is* memoryless *if for all $m, n \geq 0$*

$$\Pr[X > m + n \mid X \geq m] = \Pr[X > n] \ . \tag{5.13}$$

Intuitively, if someone tells us that X is at least a certain value m, then $X - m$ has the same distribution as X. I.e., X does not "remember" that it is larger than m, but simply behaves as usual in the interval $[m, \infty)$. An example of a memoryless r.v. is the geometric distribution.

Definition 5.8 (Geometric Distribution) *A r.v. X has a geometric distribution with parameter p if $\Pr[X = k] = (1-p)^{k-1}p$.*

Example: If I buy a lottery ticket every week, and I have a winning probability of p every time independently from the other times, then my first winning week has a geometric distribution, since in order to win on the k-th week for the first time, I'll need to not win for $k-1$ weeks (happens w.p. $(1-p)^{k-1}$), and then win on the k'th week. (happens with probability p). Note that $\Pr[X > k] = (1-p)^k$ (can easily be seen from the lottery example), and thus the geometric distribution is memoryless. Formally, for every $m, n \geq 0$

$$\Pr[X > m + n \mid X > m] = \frac{\Pr[X > m + n]}{\Pr[X > m]} = \frac{(1-p)^{m+n}}{(1-p)^m} = (1-p)^n = \Pr[X > n] \ . \tag{5.14}$$

5.5 Summary

When approaching a question in probability, Consult the following list.

1. If asked to show that **there exists** an object of at least (respectively, at most) a certain size a, then denote the size of the object by X, add probability (if no hint was given, try uniform distribution), and show that $\mathbb{E}\left[X\right] \geq a$ (resp., $\mathbb{E}\left[X \leq a\right]$).
2. If X has a binomial distribution, use Chernoff's Inequalities.
3. If X is a sum of independent r.v.'s which are not bernoulli, use Hoeffding's Inequality.
4. If X is a sum of r.v.'s which are not independent, but are weakly dependent, i.e., $\mathbb{E}\left[X^2\right]/(\mathbb{E}\left[X\right])^2 \to 1$, use Chebyshev's Inequality.

5. If non of the above holds, try Markov's Inequality.

Memorylessness is the property of seeing the same distribution looking forward in time, no matter where you stand (thus ”forgetting” where you are). Geometric r.v.'s are memoryless.