Topics in Theoretical Computer Science February 16, 2015

Lecture 1

Lecturer: Ola Svensson *Scribes: Igor Malinović, Ashkan Norouzi Fard, Saeid Sahraei*

1 The Probabilistic Method

The probabilistic method was initiated by Paul Erdős and it has been widely used for exploiting randomness in theoretical computer science. Common examples include expander graphs, the Lovász Local Lemma, and the maximum satisfiability problem. The basis of the method is proving the existence of a certain combinatorial object by designing an appropriate probability space and showing that a randomly picked object in the space has the desired structure with positive probability.

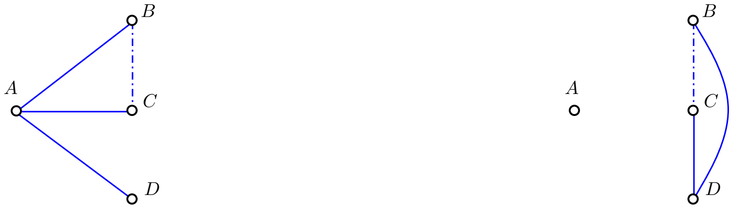
An important aspect of the probabilistic method is that the existential argument can often be turned into an algorithm. We may consider expander graphs as an interesting example. They are highly connected, though fairly sparse graph families. While creating a structure which satisfies either of the properties is easy (a complete graph and a path respectively), designing an expander graph deterministically has been a difficult task so far. Astonishingly, picking a graph at random from a specific simple distribution will return an expander with high probability.

In the first part of the lecture, we prove by induction an illustrative existential statement. It relates the clique size and the independent set size in a graph. Then we introduce the Erdős-Rényi random graph model to investigate the strength of the statement via the probabilistic method. The basic techniques that use expectation and variance are complemented by solving two exercises. We finish the lecture by providing a simple algorithm for finding a clique of size $\log(n)$ in a random graph. Additionally, there is a brief description of the phase transition phenomenon.

2 Cliques and Independent Sets

Our motivation problem called "the theorem on friends and strangers" is coming from the fields of Ramsey theory and extremal graph theory. The questions posed there have the following form: what is a large enough structure such that it contains a substructure with a given property? In graph theory "friends" and "strangers" are formally defined as cliques and independent sets respectively.

Lemma 1 *Suppose you invite six people for a dinner then either three of them are mutual friends or three of them are mutual strangers.*



(a) (b)

Figure 1: Graph interpretation of friends and strangers with the corresponding cases.

Proof Let A, B, C, D, E, F be the invited people. Consider the person A, the following two possible cases come naturally from the problem statement:

- Case 1; see Figure 1(a): A has 3 friends. Without loss of generality, assume them to be B, C and D. If any two of B, C, D are friends, say B and C, then A, B, C are pairwise friends. Otherwise, B, C, D are pairwise strangers.

- Case 2; see Figure 1(b): A has 3 strangers, say B, C and D. If any two of B, C, D are mutual strangers, then there is a group of three people who are all pairwise strangers, otherwise B, C and D are three friends.

■

Now, we introduce formal definitions and show a general statement for an arbitrary number of guests, i.e. on a graph with n vertices. The proof of Lemma 1 uses the case split on a single vertex and analyses a smaller problem of the same kind. It suggests applying induction in the general setting is a good choice.

Definition 2 *Let $G = (V, E)$ be an undirected graph and $V' \subseteq V$ its subset of vertices. Set V' is a clique if for any pair $u, v \in V'$ the edge $\{u, v\} \in E$. If for every $u, v \in V'$ the edge $\{u, v\} \notin E$, V' is called an independent (or stable) set.*

Theorem 3 *In a subgraph of $G = (V, E)$ with $|V| = n$ there is either an independent set of size s or a clique of size t as long as $n \geq 2^{s+t} - 1$.*

Proof By induction on $s + t$.

Basis: Observe that the theorem trivially holds for $s \leq 2$, $t \leq 2$. A single vertex is at is at the same time a clique and an independent set by the definition. If two vertices are connected by an edge they form a clique of size 2. Otherwise, they form an independent set of the same size.

Inductive step: We assume that the claim holds for all s and t satisfying $s + t \leq k - 1$ and prove it for $s + t = k$. Consider a vertex $A \in V$:

- Case 1: A has at least $\frac{n-1}{2} = \frac{2^{s+t}-2}{2} = 2^{s+t-1} - 1$ neighbours. By the induction hypothesis, the graph consisting of neighbours of A has either an independent set of size s or a clique of size $t-1$; thus G has either an independent set of size s or a clique of size t.

- Case 2: A has $\frac{n-1}{2} = 2^{s-1+t} - 1$ non-neighbours. Using the induction hypothesis on the set of non-neighbours with $s - 1$ and t proves the theorem statement analogously to Case 1.

■

3 Random Graphs

One might be interested to know whether the bounds of the Theorem 3 are tight. It is hard to explicitly construct a graph which has neither a clique nor an independent set of size $2\log(n)$. The random graph model introduced by Erdős and Rényi in 1950s brings us closer to the answer.

Definition 4 *For given n and $p \in [0, 1]$, a graph G sampled from $G(n, p)$ is a graph with labelled set of vertices $V(G) = \{1, 2, \ldots, n\}$, obtained by taking each edge $e \in \binom{[n]}{2}$ with probability p, independently from any other edge; thus forming the edge set $E(G)$.*

The above model is one of the most widely used random graph models. It is an adequate way for modelling ”logical” networks such as peer-to-peer networks and social networks. For such graphs it is necessary to let p depend on n. The reason is that the number of edges per vertex grows linearly in n for any fixed p. However, the real-world networks are usually much sparser.

After this small digression we go back to our initial problem with sampling a random graph G from $G(n, \frac{1}{2})$. Observe that G and its complement $\bar{G}$ are equiprobable in this setting. A clique in G is an independent set in $\bar{G}$ and vice versa; therefore it is equally probable to have a clique of size t in G and the independent set of the same size.

Theorem 5 *A graph $G \sim G(n, \frac{1}{2})$ has no independent set or clique of size $2\log_2(n) + 1$ with high probability.*

Proof For any subset $S \subseteq V(G)$ of size $t = 2\log_2(n) + 1$, let X_S be an indicator variable such that $X_S = 1$ if S is a clique and $X_S = 0$ otherwise. For X_S to be equal to 1 all $\binom{t}{2}$ edges between vertices of S have to be present in the set of edges $E(G)$. This gives $\mathbb{E}[X_S] = \Pr[S \text{ is a clique}] = (\frac{1}{2})^{\binom{t}{2}}$. Let X be the number of cliques of size t in G then its expectation is:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{|S|=t} X_S\right] = \sum_{|S|=t} \mathbb{E}[X_S] = \sum_{|S|=t} 2^{-\binom{t}{2}} = \binom{n}{t} \cdot 2^{-\binom{t}{2}}$$
$$\leq \frac{n^t}{t!} \cdot \frac{1}{2^{\frac{t(t-1)}{2}}} = \frac{1}{t!} \cdot \left(\frac{n}{2^{\frac{t-1}{2}}}\right)^t \leq \frac{1}{t!} \cdot \left(\frac{n}{n}\right)^t = \frac{1}{t!} = o(1).$$

Now by using Markov's inequality:

$$\Pr[X \geq 1] \leq \frac{\mathbb{E}[X]}{1} = o(1).$$

As we discussed earlier, the probability of having an independent set of size t is the same as the probability of having a clique of size t. Applying the union bound gives the statement of the theorem. ■

The proof above uses linearity of expectation, the fact that $\binom{n}{t} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-t+1)}{t!} \leq \frac{n^t}{t!}$, and the following basic probabilistic tool:

Theorem 6 *(Markov's inequality). If X is a nonnegative random variable and $a > 0$, then*

$$Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

4 Exercises

We continue by solving two exercises in order to better our understanding of the probabilistic method.

Exercise 1

For a given constant positive integer l, what is the largest value of p so that $G \sim G(n, p)$ has no cycle of length l with high probability?

Solution to Exercise 1:

We proceed in a similar manner as the proof of Theorem 5. The main difference is that for each subset $S \subseteq V(G)$ of size l there is $\frac{l!}{2l}$ distinct labelled cycles up to automorphism, while there is only a single labelled clique (or independent set). The number $\frac{l!}{2l}$ is obtained by dividing the number of permutations with l (since cycle can be rotated) and 2 (for two possible directions). Denote with X_C an indicator variable for a specific labelled cycle of size l, then $\mathbb{E}[X_C] = \Pr[E(C) \subseteq E(G)] = p^l$. Let X be the number of cycles of size l. We will calculate its expectation:

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{|S|=l} \sum_{\substack{C \ s.t. \\ V(C)=S}} X_C\right] = \sum_{|S|=l} \sum_{\substack{C \ s.t. \\ V(C)=S}} \mathbb{E}[X_C] = \sum_{|S|=l} \sum_{\substack{C \ s.t. \\ V(C)=S}} p^l$$
$$= \binom{n}{l} \cdot \frac{l!}{2l} \cdot p^l \leq \frac{n^l}{2l} \cdot p^l = \frac{1}{2l} \cdot (p \cdot n)^l.$$

Again by using Markov's inequality:

$$\Pr[X \geq 1] \leq \frac{\mathbb{E}[X]}{1} \leq \frac{1}{2l} \cdot (p \cdot n)^l .$$

For $p = \frac{1}{n^{(1+\epsilon)}}$ and an arbitrary small constant $\epsilon > 0$, the probability goes to 0 as $n \to \infty$.

We can use the expected number of cycles from Exercise 1 to create a dense graph with no cycles of length l. The algorithm is following:

1. Sample $G \sim G(n, p)$;
2. For each cycle of length l remove an edge.

The expected number of edges in the resulting graph $\tilde{G}$ is:

$$\mathbb{E}\left[|E(\tilde{G})|\right] \geq \mathbb{E}\left[|E(G)|\right] - \mathbb{E}[X] \geq \binom{n}{2} \cdot p - \frac{n^l}{2l} \cdot p^l$$

We are looking for a function $p = p(n,l)$ that is going to maximize the above expression. In order to simplify the calculation we make several simple assumptions. Since we are mainly interested in asymptotic behaviour we consider only the leading terms without coefficients, i.e. $n^2 \cdot p - n^l \cdot p^l$. We assume $l \geq 3$ and write p as $n^{-f(l)}$, where $\forall l : f(l) \geq 0$. The restated expression is:

$$\begin{aligned}\max_{f(l)} \; n^2 \cdot n^{-f(l)} - n^l \cdot n^{-lf(l)} &= \max_{f(l)} \; n^{2-f(l)}(1 - n^{(l-2)-(l-1)f(l)}) \\ &= \max_{f(l)} \; n^{2-f(l)}(1 - n^{1-\frac{(l-1)}{(l-2)}f(l)}).\end{aligned}$$

For $f(l) < \frac{l-2}{l-1}$ the the whole expression becomes negative for large enough n. On the other hand, for $f(l) \geq \frac{l-2}{l-1}$ the second product term $(1 - n^{1-\frac{(l-1)}{(l-2)}f(l)})$ goes to 1 as $n \to \infty$, while the first term $n^{2-f(l)}$ is maximized for $f(l) = \frac{l-2}{l-1}$. The expected number of edges $\mathbb{E}\left[|E(\tilde{G})|\right] = \Omega(n^{1+\frac{1}{l-1}})$ for $p = n^{-\frac{l-2}{l-1}}$.

Exercise 2

What is the largest value of t so that $G \sim G(n, \frac{1}{2})$ has in expectation at least one clique of size t?

Solution to Exercise 2:

We would like to prove the statement for $t = 2(1-\epsilon)\log_2(n) + 1$ and an arbitrary small constant $\epsilon > 0$. Let X the number of cliques of size t. We use a part of the calculation from the proof of Theorem 5 with the difference that we want to lower-bound the expectation:

$$\mathbb{E}[X] = \binom{n}{t} \cdot \left(\frac{1}{2}\right)^{\binom{t}{2}} \geq \left(\frac{n}{t}\right)^t \cdot \frac{1}{2^{\frac{t(t-1)}{2}}} = \left(\frac{n}{tn^{1-\epsilon}}\right)^t = \left(\frac{n^\epsilon}{t}\right)^t ,$$

where we used $\binom{n}{t} = \frac{n}{t} \cdot \ldots \cdot \frac{n-t+1}{1} \geq \left(\frac{n}{t}\right)^t$ since $\frac{n-i}{t-i} \geq \frac{n}{t}$ for $0 \leq i < t$ and $n \geq t$. We obtain that:

$$\lim_{n \to +\infty} \mathbb{E}[X] = +\infty.$$

This still does not imply that $Pr[X \geq 1]$ also grows arbitrarily close to 1 as n grows large. Consider for example a random variable X with the following PDF:

$$f(x) = 0.99\delta(x) + 0.01\delta(x - 10^{12}),$$

where $\delta(x)$ indicates the Dirac delta function.[1]

This is clearly a PDF, as integral of f equals to 1. Furthermore, the expected value of x is quite large:

$$\mathbb{E}[X] = 0.99 \cdot 0 + 0.01 \cdot 10^{12} = 10^{10}.$$

Nonetheless the probability that $x > 1$ is only 0.01.
What is noticeable in this example is that the standard deviation of the random variable x is also very large:

$$\mathsf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 10^{22} - 10^{20} \approx 10^{22} \Rightarrow \sigma[X] \approx 10^{11}$$

where $\mathsf{Var}[X] = \sigma^2[X]$.

One can empirically conclude that a random variable can be arbitrarily far from its expected value with a high probability, given that its standard deviation is sufficiently large. This gives rise to the following question: does having a small standard deviation guarantee that a random variable remains close to its expected value with high probability? Chebyshev's inequality gives an affirmative answer:

Theorem 7 *(Chebyshev's inequality). If X is a random variable with finite expected value μ and finite non-zero standard deviation σ, then for any positive λ:*

$$\Pr[|X - \mu| \geq \lambda \cdot \sigma] \leq \frac{1}{\lambda^2}.$$

Proof

$$\sigma^2 = \mathsf{Var}[X] = \mathbb{E}[(X-\mu)^2] \geq Pr[|X - \mu| \geq \lambda \cdot \sigma] \cdot \lambda^2\sigma^2,$$

where the last inequality follows from Markov's inequality for the random variable $|X - \mu|^2$ and the parameter $\lambda^2\sigma^2$. Dividing both sides by $\lambda^2\sigma^2$ gives Chebyshev's inequality. ■

Back to our problem. We would like to prove that $\Pr[X > 0] \to 1$ as $n \to \infty$, having in mind that X is integer valued and thus $\Pr[X > 0] = \Pr[X \geq 1]$. We apply Chebyshev's inequality to bound the standard deviation of X, by setting $\lambda = \frac{\mu}{\sigma}$ we obtain:

$$\Pr[X = 0] \leq \Pr[|X - \mu| \geq \mu] \leq \frac{\sigma^2}{\mu^2} = \frac{\mathsf{Var}[X]}{\mathbb{E}[X]^2}$$

Now, we use the definition of variance to bound the last expression:

$$\mathsf{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[(\sum_{|X_S|=t} X_S)^2] - (\mathbb{E}[\sum_{|X_S|=t} X_S])^2 = \mathbb{E}[\sum_{\substack{|A|=t,\\ |B|=t}} X_A X_B] - \sum_{\substack{|A|=t,\\ |B|=t}} \mathbb{E}[X_A]\mathbb{E}[X_B]$$

If A and B are disjoint, then X_A and X_B are independent. Thus $\mathbb{E}[X_A X_B] - \mathbb{E}[X_A]\mathbb{E}[X_B] = 0$. The same argument holds when A and B intersect in only one vertex, as in such case they do not have any edges in common, thus again X_A and X_B are independent. So, we assume that A and B intersect in at least two vertices.

$$\mathsf{Var}[X] \leq \sum_{|X_A|=t} \sum_{\substack{|X_B|=t,\\ |A\cap B|\geq 2}} \mathbb{E}[X_A X_B] = \binom{n}{t} 2^{-\binom{t}{2}} \sum_{i=2}^{t} \binom{t}{i}\binom{n-t}{t-i} 2^{-\binom{t}{2}+\binom{i}{2}}$$

Here the term $\binom{n}{t}$ stands for the number of different ways one can choose X_A. The probability that X_A is a clique is $2^{-\binom{t}{2}}$. The term $\binom{t}{i}$ accounts for all possible ways for X_B and X_A to share i vertices. X_B

[1] To be precise, δ is a generalized function that ranges over the reals, such that $\delta(x) = 0$ if $x \neq 0$, and $\int_{-\infty}^{\infty} \delta(t)dt = 1$.

may choose the remaining vertices in $\binom{n-t}{t-i}$ different ways. Finally, $2^{-\binom{t}{2}+\binom{i}{2}}$ is the probability that the remaining edges of X_B are also connected and thus X_B is a clique as well.

$$\frac{\mathsf{Var}[X]}{\mathbb{E}[X]^2} \leq \frac{\binom{n}{t}2^{-2\binom{t}{2}} \sum_{i=2}^{t} \binom{t}{i}\binom{n-t}{t-i}2^{\binom{i}{2}}}{\binom{n}{t}^2 2^{-2\binom{t}{2}}} = \sum_{i=2}^{t} \frac{\binom{t}{i}\binom{n-t}{t-i}2^{\binom{i}{2}}}{\binom{n}{t}}.$$

The remaining part of the proof is inspired by the original paper on cliques in random graphs by Erdős and Bollobás dated in 1975. Denote the i-th summation term with T_i for $2 \leq i \leq t$. For a sufficiently large n and $3 \leq i \leq t-1$ it is not hard to prove the following statement:

$$T_i < T_3 + T_{t-1}.$$

The inequality trivially holds for $i = 3$ and $i = t-1$, while for other terms we use monotonicity $\frac{T_{i+1}}{T_i} = \frac{(t-i)^2 \cdot 2^i}{(t-i+1)(i+1)} = \frac{(t^2-2i+i^2)\cdot 2^i}{ti-i^2+t+1} > 1$. We proceed by giving an upper-bound for our main expression:

$$\begin{aligned}\frac{\mathsf{Var}[X]}{\mathbb{E}[X]^2} &< T_2 + T_t + t \cdot (T_3 + T_{t-1}) < \frac{t^4}{2n^2}2 + \frac{1}{\mathbb{E}[X]} + t \cdot \left(\frac{t^6}{6n^3}2^3 + \frac{tn2^{-(t-1)}}{\mathbb{E}[X]}\right) \\ &= \frac{t^4}{n^2} + \frac{1}{\mathbb{E}[X]} + \frac{4t^7}{3n^3} + \frac{1}{\mathbb{E}[X]} \cdot \frac{t^2}{n^{1-2\epsilon}}.\end{aligned}$$

We have already shown that $\mathbb{E}[X] \to \infty$ as $n \to \infty$, thus its inverse goes to 0. All the other terms go to 0 so it does $\Pr[X = 0] \leq \frac{\mathsf{Var}[X]}{\mathbb{E}[X]^2}$ too.

We have proven that almost all graphs from $G(n, \frac{1}{2})$ have a clique of size $2(1-\epsilon)\log_2(n) + 1$ for any small and positive constant ϵ. But this does not mean that it is easy to detect such a clique. In fact, it is still an open problem to find a clique of expected size $c\log(n)$ where $c > 0$ in polynomial time in a graph from $G(n, \frac{1}{2})$. As a result, we have a double threshold chart; see Figure 2. It is interesting that up to size $2\log(n)$ there is a huge number of cliques and than suddenly this number drops to zero. This phenomenon is called phase transition. The same behaviour has been recently proven for k-SAT.

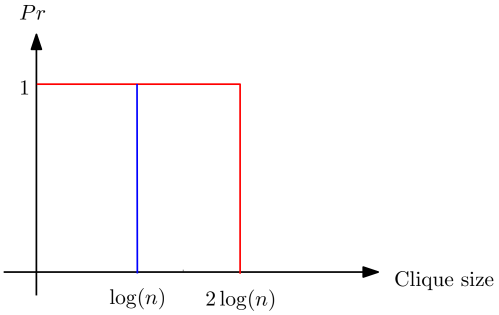


Figure 2: In a graph $G \sim G(n, \frac{1}{2})$ there is a known polynomial algorithm to find a clique of the given size (blue line). A clique of a given size exists with high probability (red line).

Here we describe an algorithm which finds a clique in the case $c = 1$.

1. Set $S = \emptyset$;
2. Pick any vertex v at random and set $S = S \cup \{v\}$;
3. Discard v and all vertices not connected to v from the graph;
4. If the graph is not empty, go back to step 2.

Since in $G(n, \frac{1}{2})$ every edge exists with probability $\frac{1}{2}$, in every step of the algorithm half of the nodes are discarded on average. Thus the expected size of the returned clique is $\log(n)$.