CS 466/666 Algorithm Design and Analysis . Spring 2019. Waterloo.

Lecture 10: Probabilistic methods

We will use some probabilistic methods to show the existences of interesting combinatorial objects.

Today we will see the first moment method and the second moment method.

Next time we will see the Lovász local lemma.

Ramsey Graphs

Can you color the edges of a complete graph of n vertices by two colors (say red and blue) so that there are no large monochromatic complete subgraphs?

This is one of the most famous problems in discrete mathematics, and is one of the early problems that the probabilistic method is developed.

Theorem If $\binom{n}{k} 2^{-\binom{k}{2}+1} < 1$, then it is possible to color the edges of K_n so that there are no monochromatic K_k.

Proof Consider a subset of k vertices S.

The probability that it is monochromatic is $2 \cdot 2^{-\binom{k}{2}}$.

There are $\binom{n}{k}$ subsets of size k.

So, by the union bound, if $\binom{n}{k} \cdot 2^{-\binom{k}{2}+1} < 1$, there is at least one coloring with no monochromatic K_k (i.e. the bad events don't cover the sample space) $\square$

Note that the theorem is satisfied if $k = \Omega(\log_2 n)$ (check this).

So, there is a coloring with no monochromatic $K_{\Omega(\log n)}$.

In fact, a random coloring will work with high probability.

Surprisingly, there are no deterministic method known to construct such a coloring.

In fact, it is already very difficult to get $k = \Theta(\sqrt{n})$ for a deterministic construction!

Note that the proof shows that in a random graph, the maximum clique size and the maximum independent set size is $O(\log n)$, as we can think of red edges as edges and blue edges as non-edges.

This is a useful fact to keep in mind.

Magical graphs

Last time we used magical graphs to design a fast algorithm for network coding.

We now show that an efficient randomized construction of magical graphs.

Recall that Hall's theorem for bipartite matching says that given a bipartite graph $G=(U,W;E)$,

a subset $S \subseteq U$ can be perfectly matched to V if and only if $\forall T \subseteq S$, we have $|N(T)| \geq |T|$,

where $N(T)$ is the neighbor set of T.

With Hall's theorem in mind, magical graphs are defined as follows.

Let $G=(L,R;E)$ be a bipartite graph. We say that G is an (n,m,d)-magical graph if

① $|L|=n$, ② $|R|=m$, ③ every left vertex (vertices in L) has d neighbors, and finally

④ $|N(S)| > |S|$ for every $S \subseteq L$ with $|S| \leq |L|/2 = n/2$.

<u>Theorem.</u> For every large enough d and n and $m \geq 3n/4$, there exists an (n,m,d)-magical graph.

<u>Proof</u> Let G be a random bipartite graph with n vertices on the left and m vertices on the right.

Each left vertex connects to a randomly chosen set of d vertices on the right.

We claim that G is a magical graph with high probability.

Let $S \subseteq L$ with $s := |S| \leq n/2$ and $T \subseteq R$ with $t := |T| = |S|$. Note that $|S|=|T|$.

Let $X_{S,T}$ be the indicator variable that all edges from S go to T, and $X = \sum_{\substack{S,T \\ |S|=|T| \leq \frac{n}{2}}} X_{S,T}$

Then $E[X_{S,T}] = \Pr(X_{S,T}=1) = (t/m)^{sd}$.

Then
$$
\begin{aligned}
E[X] = E\Big[\sum_{\substack{S,T \\ |S|=|T|\leq \frac{n}{2}}} X_{S,T}\Big] = \sum_{\substack{S,T \\ |S|=|T|\leq \frac{n}{2}}} E[X_{S,T}] &= \sum_{s \leq n/2} \binom{n}{s}\binom{m}{t}\left(\frac{t}{m}\right)^{sd} \quad (\text{for } t=s) \\
&\leq \sum_{s\leq n/2} \left(\frac{ne}{s}\right)^s \left(\frac{me}{s}\right)^s \left(\frac{s}{m}\right)^{sd} \quad \left(\text{recall } \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k\right) \\
&= \sum_{s \leq n/2} \left[\left(\frac{ne}{s}\right)\left(\frac{me}{s}\right)\left(\frac{s}{m}\right)^d\right]^s \\
&\leq \sum_{s\leq n/2} \left[\frac{4}{3}e^2 \left(\frac{s}{m}\right)^{d-2}\right]^s \quad \left(\text{using } m \geq \frac{3n}{4}\right) \\
&\leq \sum_{s \leq n/2} \left[\frac{4}{3}e^2\left(\frac{2}{3}\right)^{d-2}\right]^s \quad (\text{as } s \leq n/2 \text{ and } m \geq 3n/4) \\
&\leq \sum_{s\leq n/2} \left(\frac{1}{3}\right)^s \leq \frac{1}{2} \quad (\text{for } d \geq 20 \text{ say}).
\end{aligned}
$$

Since the expected value is less than one and X is an integer value random variable, there exist some outcomes in the sample space that $X=0$.

This implies that for all subset S with $|S| \leq \frac{n}{2}$, we have $|N(S)| > |S|$, as otherwise if $|N(S)| \leq |S|$, there exists T with $|T|=|S|$ (possibly a superset of $N(S)$) with all edges from S go to T (i.e. $X_{S,T}=1$ for $|S|=|T| \leq \frac{n}{2}$) contradicting $X=0$. □

In fact, if d is a large enough constant say $d \geq 30$, most graphs in the sample space are magical graphs.

Magical graphs are closely related to expander graphs, which have many applications in theoretical computer science.

Superconcentrator

Definition Let $G=(V,E)$ be a directed graph and let I and O be two subsets of V with n vertices, each called the input and output sets respectively. We say that G is a superconcentrator if for every k and every $S \subseteq I$ and $T \subseteq O$ with $|S|=|T|=k$, there exist k vertex disjoint paths in G from S to T.

We are interested in constructing a superconcentrator with as few edges as possible.

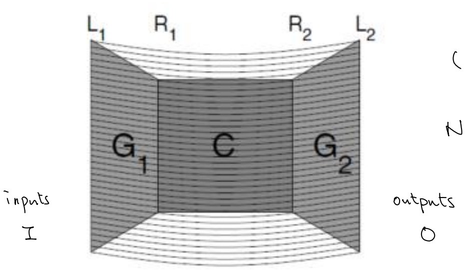
A complete bipartite graph from I to O is a superconcentrator, but it requires n^2 edges.

It was conjectured that a superconcentrator with $O(n)$ edges does not exist, but it turned out that one can use magical graphs to construct such a superconcentrator.

The construction is recursive. We assume the existence of a superconcentrator C with $3n/4$ inputs, $3n/4$ outputs, and $O(3n/4)$ edges. The base case is when n is a constant, for which a complete bipartite graph would do.

We use two $(n, 3n/4, O(1))$-magical graphs, G_1 and G_2.

Now, a superconcentrator with n inputs and n outputs can be constructed by putting C, G_1, G_2 together:



(picture from the survey by Hoory-Linial-Wigderson)

Note that there is a perfect matching between the inputs and the outputs.

First, we show that it is a superconcentrator.

Let $I = O = \{1,2,\ldots,n\}$ and the matching connects vertex j in I to vertex j in O.

Let $S \subseteq I$ and $T \subseteq O$ with $|S|=|T|=k$. We want to show that there are k vertex disjoint paths between S and T. (In the picture, $S \subseteq L_1$ and $T \subseteq L_2$.)

If $S \cap T \neq \phi$ when think of them as subsets of $\{1,2,\ldots,n\}$, then we can use the edges in the matching to connect those pairs in $S \cap T$.

So, we assume that $S \cap T = \phi$. In particular $|S| = |T| = k \le n/2$.

By the property of the magical graph, for any subset $S \subseteq L_1$ (see above picture) with $|S| \le n/2$,

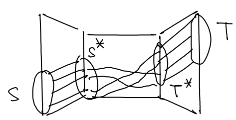
we have $|N(S)| \ge |S|$.

By Hall's theorem, this implies that $S \subseteq L_1$ has a perfect matching to some subset S^* in R_1

By the same argument, $T \subseteq L_2$ has a perfect matching to some subset T^* in R_2.

Since C is a superconcentrator, there are $|S^*| = |T^*|$ vertex disjoint paths between S^* and T^* in C.

Combining with the two matchings, these form $|S| = |T|$ vertex disjoint paths between S and T.

Pictorically,  . So, it is a superconcentrator.

Finally, let $E(n)$ be the number of edges in this graph.

Then $E(n) = \underbrace{2 \cdot d \cdot n}_{\text{two magical graphs}} + \underbrace{n}_{\text{matching}} + \underbrace{E(3n/4)}_{\text{small super concentrator}}$, and solving the recurrence gives $E(n) = O(n)$.

Superconcentrators are efficient communication networks (switching networks).

Expander graphs can also be used to construct optimal sorting networks [Ajtai, Kolmós, Szemerédi].

High Girth High Chromatic Number Graphs

Graph coloring: the objective is to use the minimum number of colors to color all vertices so that every pair of adjacent vertices receive different colors.

In general we would like to use as few colors as possible, and would like to understand that what graphs require large chromatic number.

One straightforward condition for a graph to have large chromatic number is to have a large clique, but there exist graphs with no triangles (clique of size 3) and yet with large chromatic number.

The following theorem is even more surprising.

Theorem For all k, ℓ, there exist graphs of chromatic number $> k$ and with no cycles of length $\le \ell$.

Proof Consider a random graph where each edge is chosen with probability p.

Let $\chi(G)$ be the chromatic number of G and $\alpha(G)$ be the size of a maximum independent set in G.

Note that $n/\chi(G) \le \alpha(G)$, as the graph is partitioned into $\chi(G)$ independent sets.

Thus $\chi(G) \geq n/\alpha(G)$. To lower bound $\chi(G)$, we will upper bound $\alpha(G)$ as in above.

$\Pr(\alpha(G) \geq t) \leq \binom{n}{t}(1-p)^{\binom{t}{2}} \leq n^t e^{-p\binom{t}{2}} = (n e^{-p(t-1)/2})^t$

Set $t = \lceil 3\ln n / p \rceil$, this probability is at most $1/2$ (just a loose bound).

So we know that there are no independent sets of size $\Omega(\ln n/p)$ with good probability.

Next, we bound the number of cycles of length at most l. Call this number X.

Then $E[X] = \sum_{i=3}^{l} \binom{n}{i} \frac{i!}{2i} p^i$ ($\binom{n}{i}$ subsets. for each subset, $i!$ permutations, each cycle of length i counted $2i$ times)

$\leq \sum_{i=3}^{l} \frac{n^i}{2i} p^i$ (recall $\binom{n}{i} \leq \frac{n^i}{i!}$).

Ideally, we would like to choose p so that $E[X] < \frac{1}{2}$ say, and conclude that there are no such cycles.

However, to do so, we need $p < \frac{1}{n}$.

But then the bound on $\alpha(G)$ would become $O(n \ln n)$, and we could not say anything about $\chi(G)$.

The new idea is to set p larger and do some <u>deterministic modifications</u> later.

Set $p = n^{\varepsilon - 1}$ where $\varepsilon < 1/l$.

Then $E[X] \leq \sum_{i=3}^{l} \frac{n^{\varepsilon i}}{2i} = o(n)$ as $\varepsilon < 1/l$.

In particular, $\Pr(X \geq n/2) < \frac{1}{2}$ by Markov's inequality.

So, with positive probability, the graph has less than $n/2$ "short" (length $\leq l < 1/\varepsilon$) cycles

and $\alpha(G) < 3n^{1-\varepsilon} \ln n$.

Now, the idea is to delete one vertex in each short cycle to obtain G^*.

Since we need to delete at most $n/2$ vertices, G^* has at least $n/2$ vertices.

Furthermore, it has no short cycles and no independent sets of size $3n^{1-\varepsilon} \ln n$.

Therefore, $\chi(G^*) \geq \frac{|V(G^*)|}{\alpha(G^*)} \geq \frac{n/2}{3n^{1-\varepsilon}\ln n} = \frac{n^{\varepsilon}}{6 \ln n}$.

It is larger than any k for sufficiently large n. $\square$

This "deletion method" (more generally "deterministic modification) is a useful technique in probabilistic methods.

Second Moment Method

The first moment method is to compute the expected value of a random variable, and conclude that there is an outcome with value at least $E[X]$ or at most $E[X]$.

For a non-negative random variable X, we can use the Markov inequality to prove that $\Pr(X \geq 1) \leq E[X]$, and thus conclude that $X=0$ with high probability if $E[X] \ll 1$ for an integral X.

Often we also want to prove that $\Pr(X \geq 1)$ is large. It is not enough to just show that $E[X]$ is large, as it could be the case that X is very large for a small fraction of the outputs, while $X=0$ for a large fraction of the outputs.

To exclude this case, we need some kind of concentration inequalities (e.g. variance of X is small), and the second moment method provides one way to establish this.

<u>Theorem</u> If X is an integral-valued random variable, then $\Pr(X=0) \leq \frac{\mathrm{Var}[X]}{(E[X])^2}$

<u>Proof</u> By Chebyshev's inequality, $\Pr(X=0) \leq \Pr(|X - E[X]| \geq E[X]) \leq \mathrm{Var}[X]/(E[X])^2$. □

<u>Corollary</u> If $\mathrm{Var}[X] = o((E[X])^2)$ or $E[X^2] = (1+o(1))(E[X])^2$, then $X > 0$ almost always.

<u>Threshold Behavior in Random Graphs</u>

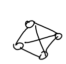
Let $G_{n,p}$ be a graph with n vertices where each edge appears with probability p.

A property has a <u>threshold behavior</u> if there is a function f such that:

- when $\lim_{n \to \infty} \frac{g(n)}{f(n)} = 0$, then almost surely $G(n, g(n))$ does not have this property.
- when $\lim_{n \to \infty} \frac{h(n)}{f(n)} = \infty$, then almost surely $G(n, h(n))$ has this property.

A property has a <u>sharp threshold behavior</u> if there is a function f such that for any $\varepsilon > 0$,

- $G(n, (1-\varepsilon) f(n))$ almost surely does not have the property.
- $G(n, (1+\varepsilon) f(n))$ almost surely has the property.

For example, consider the property of having a clique of size 4, i.e. .

Let X be the number of 4-cliques in $G_{n,p}$. Then $E[X] = \binom{n}{4} p^6$.

When $p = o(n^{-\frac{2}{3}})$, then $E[X] \to 0$, and we can conclude that $\Pr(X=0) \to 1$ by the first moment method.

On the other hand, when $p = \omega(n^{-\frac{2}{3}})$, then $E[X] \to \infty$, and we can use the second moment method to conclude that $\Pr(X \geq 1) \to 1$, by showing that $\mathrm{Var}[X] = o((E[X])^2)$. See Mitzenmacher-Upfal chapter 6.5.1.

Theorem The property of having a clique of size 4 has a threshold function $f(n) = n^{-2/3}$.

Let's see a property with a sharp threshold.

Consider the property that a random graph has diameter less then or equal to two, i.e. for every pair of vertices there is a path of length at most two connecting them.

Theorem The property that $G(n,p)$ has diameter two has a sharp threshold at $p = \sqrt{\frac{2\ln n}{n}}$.

Proof We say a pair of vertices i and j is a bad pair, if there is no edge between i and j, and no other vertex in G is adjacent to both i and j.

Note that a graph has diameter at most two if and only if there is no bad pair.

For every pair of vertices i and j with $i<j$, let X_{ij} be an indicator variable of (i,j) being a bad pair.

Let $X = \sum_{i<j} X_{ij}$ be the number of bad pairs.

Then $E[X_{ij}] = (1-p)(1-p^2)^{n-2}$, as each of the other $n-2$ vertices is not adjacent to both i and j.

So, $E[X] = \binom{n}{2}(1-p)(1-p^2)^{n-2} \sim \frac{n^2}{2}(1-p^2)^n \sim \frac{n^2}{2}e^{-p^2 n}$.

Set $p = \sqrt{\frac{c\ln n}{n}}$. Then $E[X] \sim \frac{1}{2}n^2 e^{-c\ln n} = \frac{1}{2}n^{2-c}$.

Therefore, if $c>2$, then $E[X] \to 0$, and so diameter is two almost surely.

Now, consider the case when $c<2$, then $E[X] \to \infty$.

We use the second moment method to show that $X \geq 1$ almost surely.

$$E[X^2] = E\Big[\Big(\sum_{i<j} X_{ij}\Big)^2\Big] = E\Big(\sum_{\substack{i<j\\k<l}} X_{ij}X_{kl}\Big) = \sum_{\substack{i<j\\k<l}} E[X_{ij}X_{kl}].$$

We split the sum into three cases: ① all i,j,k,l are different.

② three distinct vertices out of i,j,k,l.

③ two distinct vertices out of i,j,k,l (i.e. the same pair).

For ①, the two variables are independent, and thus $\sum_{i<j,k<l} E[X_{ij}X_{kl}] = \sum_{i<j,k<l} E[X_{ij}]E[X_{kl}]$

$$\leq \sum_{i<j} E[X_{ij}] \sum_{k<l} E[X_{kl}] = (E[X])^2.$$

For ③, the two variables are the same, and thus $\sum_{i<j,k<l} E[X_{ij}X_{kl}] = \sum_{i<j} E[X_{ij}^2]$.

Since X_{ij} is an indicator variable, $X_{ij} = X_{ij}^2$, and thus this is just $\sum_{i<j} E[X_{ij}] = E[X]$.

For ②, let (i,j) and (i,k) be the two bad pairs. For any other vertex u, either it is not adjacent to i or it is not adjacent to both j and k. This happens with probability $(1-p) + p(1-p)^2 = 1-2p^2+p^3 \approx 1-2p^2$.

So, $E[X_{ij} X_{ik}] \approx (1-p)^2 \cdot (1-2p^2)^{n-3} \approx (1-2p^2)^n \approx e^{-2p^2 n}$.

Since there are at most $3\binom{n}{3}$ such triples, the sum in 2 is at most $\frac{n^3}{2} e^{-2p^2 n}$.

Now, recall that $p = \sqrt{\frac{c \ln n}{n}}$, this is at most $\frac{1}{2} n^3 e^{-2c\ln n} = \frac{1}{2} n^{3-2c} = o(n^{4-2c}) = o((E[X])^2)$.

Therefore, $E[X^2] \leq (E[X])^2 + o((E[X])^2) + E[X] = (1+o(1))(E[X])^2$.

Hence, by the corollary of the second moment method, we conclude that $X \geq 1$ almost surely.

This implies that the graph has diameter at least three when $p = \sqrt{\frac{c \ln n}{n}}$ for $c<2$. □

Random graphs and random structures

Threshold phenomenons are common in random graphs and it is a subject of intense study.

An important result is the emergence of "giant component": when we sample a random graph with edge probability $p = \frac{1}{n}$, the largest components are of size $O(n^{\frac{2}{3}})$ and they are almost surely trees. But when $p \geq (1+\varepsilon)n$, then there is a unique giant component of size $\Omega(n)$, while all other components are of size $O(\log n)$.

Other random structures also have this phenomenon of "phase transition".

For random 3-SAT formula where each clause has three random variables (or their negations).

It is conjectured and generally believed that when the clause-to-variable ratio is less then 4.2, then the formula is almost surely satisfiable, and when this ratio is greater than 4.2, then the formula is almost surely unsatisfiable.

The same is conjectured for k-SAT, and the conjecture is very recently settled by Ding, Sly and Sun in the paper "Proof of the satisfiability conjecture for large k".

Algorithmic Issues

The second moment method can be used to show that $G_{n,1/2}$ has a clique of size $2\log_2 n$ almost surely.

While it is easy to find a clique of size $\log_2 n$ (a simple greedy algorithm would work), it is not known how to find a clique of size $(1+\varepsilon)\log_2 n$ in polynomial time, and in fact there is some evidence suggesting it may be computationally hard.

Similarly, there is no known polynomial time algorithms to determine whether a random 3-SAT formula with clause-to-variable ratio 4.2 is satisfiable or not. In fact, these are some hardest instances for 3-SAT that we know how to generate efficiently.

References and pointers

- Magical graphs and superconcentrators are from the survey "Expander graphs and their applications" by Hoory, Linial and Wigderson, in which you can also read about deterministic constructions of expander graphs.
- The book "Probabilistic methods" by Alon and Spencer is a great resource for various applications in combinatorics and also computing. The high girth high chromatic number example is from there.
- One important example of the first moment method is the work of Shannon who showed the existence of good error correcting codes.
- The diameter two example is taken from a new book "Foundations of data science" by Hopcroft and Kannan.