CS 466/666 Algorithm Design and Analysis . Spring 2022

Lecture 9 : 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 random experiment where we color each edge of K_n uniformly independently.

Consider a subset S of vertices with $|S|=k$.

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

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

So, by the union bound, $\Pr(\text{some subset of size } k \text{ is monochromatic}) \le \binom{n}{k} 2^{-\binom{k}{2}+1}$.

By assumption, this probability is strictly smaller than one.

So, there is an edge coloring (i.e. an outcome) so that there are no monochromatic K_k. □

Check that the theorem is satisfied when $k \ge 2\log_2 n$.

So, there is an edge coloring with no monochromatic $K_{2\log_2 n}$.

In fact, if say $k > 2\log_2 n$, then a random coloring will work with high probability.

Surprisingly, there are no known deterministic algorithm to construct an edge coloring with no monochromatic $K_{O(\log n)}$.

It is already very difficult to construct an edge coloring with no monochromatic $K_{O(\sqrt{n})}$.

Note that the proof shows that if we generate a graph where each pair of vertices have an edge with probability $1/2$ independently, then there are no complete subgraph of size $\ge 2\log_2 n$ and no independent set of size $\ge 2\log_2 n$ with high probability (edges = red, non-edges = blue).

This is an useful fact to keep in mind.

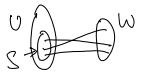
Magical graphs

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

We will show how to construct a superconcentrator using probabilistic method.

For this, we will first construct a useful gadget called "magical graphs".

A magical graph is a sparse bipartite graph $G=(U,W,E)$ with the (magical) property that every subset $S \subseteq U$ with $|S| \le |U|/2$ has a perfect matching to W.



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 W if and only if $\forall T \subseteq S$, we have $|N(T)| \ge |T|$, where $N(T)$ is the neighbor set of T, i.e. $N(T) = \{ w \in W \mid uw \in E \text{ for some } u \in T \}$.

With Hall's theorem, magical graphs can be defined as follows.

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

① $|U|=n$ and $|W|=m$.

② every vertex in U has d neighbors.

③ $|N(S)| > |S|$ for every $S \subseteq U$ with $|S| \le |U|/2 = n/2$.

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

Proof Let G be a random bipartite graph with n vertices on the left and m vertices on the right, where each left vertex is connected to a random set of d vertices on the right independently.

We will show that G is an (n,m,d)-magical graph with high probability.

Let $S \subseteq U$ with $s := |S| \le n/2$ and $T \subseteq W$ 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 $X = \sum_{\substack{S,T \\ |S|=|T| \le \frac{n}{2}}} X_{S,T}$

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

Then $E[X] = E\left[\sum_{\substack{S,T \\ |S|=|T| \le \frac{n}{2}}} X_{S,T}\right] = \sum_{\substack{S,T \\ |S|=|T| \le \frac{n}{2}}} E[X_{S,T}] = \sum_{s \le n/2} \binom{n}{s}\binom{m}{t}\left(\frac{t}{m}\right)^{sd}$ (for $t=s$)

$\le \sum_{s \le n/2} \left(\frac{ne}{s}\right)^s \left(\frac{me}{s}\right)^s \left(\frac{s}{m}\right)^{sd}$ (recall $\binom{n}{k} \le \left(\frac{ne}{k}\right)^k$)

$= \sum_{s \le n/2} \left[\left(\frac{ne}{s}\right)\left(\frac{me}{s}\right)\left(\frac{s}{m}\right)^d \right]^s$

$\le \sum_{s \le n/2} \left[\frac{4}{3}e^2 \left(\frac{s}{m}\right)^{d-2} \right]^s$ (using $m \ge \frac{3n}{4}$)

$\le \sum_{s \le n/2} \left[\frac{4}{3}e^2 \left(\frac{2}{3}\right)^{d-2} \right]^s$ (as $s \le n/2$ and $m \ge 3n/4$)

$\le \sum_{s \le n/2} \left(\frac{1}{2}\right)^s \le \frac{1}{2}$ (for $d \ge 8$ say).

Since the expected value is less than one and X is an integer value random variable, there exists some outcome in the sample space (i.e. some (n,m,d)-graphs) with $X=0$.

This implies that for all subset S with $|S| \le \frac{n}{2}$, we have $|N(S)| > |S|$, as otherwise if $|N(S)| \le |S|$, 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 $|S| = |T| \le \frac{n}{2}$) contradicting $X = 0$. □

In fact, if d is slightly larger. say $d \ge 15$, then most graphs in the sample space are magical graphs. Magical graphs are closely related to expander graphs. which are very useful in theoretical computer science.

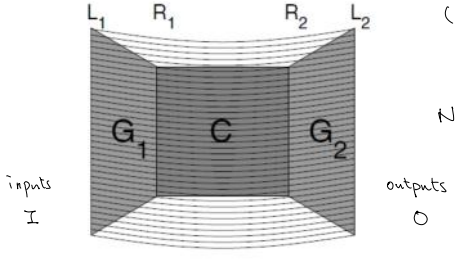
Superconcentrator

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 by Valiant that a superconcentrator with $O(n)$ edges does not exist. but later he disproved himself by constructing a superconcentrator with $O(n)$ edges using magical graphs.

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

For the induction step. 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 prove 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 need to show that there are k vertex disjoint paths between S and T.
If $S \cap T \neq \phi$ when we think of them as subsets of $\{1, 2, \ldots, n\}$, then we can use the edges

in the perfect matching to connect those pairs in S and T.

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

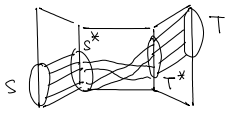
By the property of the magical graph G_1, for any subset $S \subseteq L_1$ with $|S| \le n/2$, we have $|N(S)| \ge |S|$.

By <u>Hall's theorem</u>, 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 a superconcentrator with n inputs and n outputs.

Then, $E(n) = 2 \cdot d \cdot n + n + E(3n/4)$, and solving the recurrence gives $E(n) = O(n)$.

($2 \cdot d \cdot n$: two magical graphs; n: perfect matching; $E(3n/4)$: smaller superconcentrator)

Superconcentrators are efficient communication networks (switching networks).

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

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 $Pr(X \ge 1) \le E[X]$, 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 $Pr(X \ge 1)$ 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 $Pr(X = 0) \le \frac{Var[X]}{(E[X])^2}$

<u>Proof</u> By Chebyshev's inequality. $Pr(X = 0) \le Pr(|X - E[X]| \ge E[X]) \le Var[X] / (E[X])^2$. □

<u>Corollary</u> If $Var[X] = o((E[X])^2)$ or $E[X^2] = (1 + o(1))(E[X])^2$, then $X > 0$ with high probability.

Threshold Behavior in Random Graphs

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 $Var[X] = o((E[X])^2)$. See [MU] Chapter 6.5.1.

<u>Theorem</u> 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 than or equal to two, i.e. for every pair of vertices there is a path of length at most two connecting them.

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

<u>Proof</u> 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 at most two with high probability.

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.

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

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 $\sum_{i<j,\, k<l} E[X_{ij} X_{kl}] = \sum_{i<j,\, k<l} E[X_{ij}]\, E[X_{kl}]$

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

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 $\frac{1}{2} n^3 e^{-2c \ln n} = \frac{1}{2} n^{3-2c} = o(n^{4-2c}) = o\left(\left(E[X]\right)^2\right)$.

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

Hence, by the corollary of the second moment method, we conclude that $X \geq 1$ with high probability.

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.

A classical 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 than 4.2,

then the formula is almost surely satisfiable, and when this ratio is greater than 4.2, then the 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 algorithm 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.

Reference 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 expanders.
- The book "Probabilistic methods" by Alon and Spencer is a great resource.
- One important example of the first moment method is the work of Shannon who showed the existence of good error correcting codes. See [MU].
- The diameter two example is from a new book "Foundations of data science" by Hopcroft and Kannan.