2MMD30: Graphs and Algorithms	**Lecture 9**	**Date: 11/3/2016**
Instructor: Nikhil Bansal		

1 Probabilistic method with alterations

In this chapter, we will refine the probabilistic approach. Instead of directly constructing an appropriate random object and arguing that it satisfies our properties, we will adopt a more refined approach, by first constructing a random object, but then modifying/tweaking it slightly (and removing the blemishes, which hopefully are not too many). These are called alterations. This often allows us to show much stronger results than the direct probabilistic constructions. We will consider several examples.

1.1 Dominating set

Given a graph G, a subset of vertices S is a called a dominating set if for each vertex v, either $v \in S$ or v has some neighbor in S. We first show the following.

Theorem 1 *Any graph with minimum degree δ has a dominating set of size at least $4n \log n/(\delta+1)$.*

Proof: We assume that $(\delta + 1) \geq 4 \log n$, otherwise the result is trivial.

Let us try to construct a dominating set S by picking each vertex v randomly with probability p, where we will optimize p later. Then, the expected number of vertices picked is np. Moreover, the probability that a particular vertex v is uncovered is at precisely

$$(1 - p)^{d_v+1} \leq (1 - p)^{\delta+1} \leq e^{-p(\delta+1)}$$

and thus the expected number of uncovered vertices is at most $ne^{-p(\delta+1)}$. So if we pick $p = 2 \log n/(\delta + 1)$, then the above is at most $1/n$. Thus, probability that $S$is not a valid dominating set (i.e. at least one vertex uncovered) is at most $1/n \leq 1/4$, assuming $n \geq 8$.

Moreover, by Markov's inequality

$$\Pr[S \geq 2E[S] = 4n \log n/(\delta + 1)] \leq 1/2.$$

So combining these, with probability at least $1/4$, S is both a dominating set and has size at most $4n \log n(\delta + 1)$. Thus such a set exists. $\square$

Note that any dominating set must have size at least $n/(\delta + 1)$, so the above bound is not too far. One might wonder if this can improved. We show below that this is indeed the case and replace $\ln n$ by $\ln \delta$. However, first the reader must convince themselves that the $\log n$ factor cannot be improved based on the above approach.

Theorem 2 *Any graph with minimum degree δ has a dominating set of size at least $n(1+\ln \delta)/(\delta+1)$.*

Proof: Let us pick each vertex with probability p, but add a second step where we add any uncovered vertices to S. Clearly, S is a dominating set by design. Then the expected number of vertices picked is np and as the expected number of uncovered vertices is at most $n(1 - p)^{\delta+1}) \leq ne^{-p(\delta+1)} = n/\delta$, we obtain that the expected size of S is at most $np + n/\delta = n(1 + \ln \delta)/(\delta + 1)$. $\square$

1.2 Independent sets

Theorem 3 *For any graph G with average degree $\bar{d}$, $\alpha(G) \geq n/2\bar{d}$.*

Proof: Let us pick each vertex with probability p to obtain a random set S. But instead of hoping that S be an independent set right away, we will allow it to have a (few) edges and then make it independent by removing (at most) one vertex per edge in the subgraph induced by S. The resulting set is clearly and independent set.

As the number of edges that end up in the random sample is mp^2, we lose at most mp^2 vertices.

So the expected size of the independent set is at most $np - mp^2 = np - n\bar{d}p^2/2$. Here we use that $\bar{d} = 2m/n$. Optimizing for p by taking first derivative and setting it to 0 gives $p = 1/\bar{d}$. The above expression is $n/2\bar{d}$ for this p. $\square$

1.3 Bipartite graphs with a forbidden $K_{r,r}$ (Zarankiewicz problem)

.

This problem (a famous unsolved problem) asks: Given an integer r, what is the largest number of edges that an $n \times n$ bipartite graph can have, if it is not allowed to have a $K_{r,r}$ as a subgraph. Let us call this number $Z(n, r)$.

We first show a lower bound based on alterations. This is essentially the best known for general r (for $r = 2, 3$, better bounds are known). Later we will see an upper bound.

Theorem 4 $Z(n, r) = \Omega(n^{2-2/(r+1)})$.

Proof: Pick each edge with probability p. Whenever there is some $K_{r,r}$ throw away some edge in this $K_{r,r}$. The expected number of edges picked is $n^2 p$. The expected number of $K_{r,r}$ that show up is

$$\binom{n}{r}^2 p^{r^2}.$$

So the expected number of edges after that alteration is at least

$$n^2 p - \frac{n^{2r}}{r^2} p^{r^2} \tag{1}$$

Taking derivatives gives $n^2 - n^{2r} p^{r^2-1}$. Setting this to 0 gives $p = n^{-2/(r+1)}$. Plugging this back in (1) gives that the expected number of edges is $\Omega(n^{2-2/(r+1)})$ which gives the desired result. $\square$

We now show the upper bound. The proof does not use probability, but is very nice. So we discuss it anyway.

Theorem 5 $Z(n, r) = O(n^{2-1/r})$.

Proof: We want to upper bound the number of edges. Let G be some graph with the maximum number of edges. Let d_v be the degree of vertex v on the left hand side. Consider the following auxiliary graph H.

H is bipartite with n vertices on the left, each corresponding to a left vertex of G. The right hand side of H has $\binom{n}{r}$ vertices, corresponding to r-tuples of vertices on the right hand side of G.

There is an edge in H from v to a r-tuple $(w_1, \ldots, w_r)$ if and only if v is adjacent to each of $w_1, \ldots, w_r$ in G.

The following two observations are crucial:

(1) The degree of a vertex v on left of H is exactly $deg_H(v) = \binom{d_v}{r}$, where d_v is the degree of v is in G.

(2) The degree of any vertex on the right of H is at most $r-1$. Otherwise, if it is r or more, then this r-tuple would be connected to some r vertices on the left of H. But this would mean that this r-tuple and those r vertices form a $K_{r,r}$ in G, which is not allowed.

So, the total number of edges in H is at most $(r-1)\binom{n}{r}$. But the number of edges is exactly $\sum_{v \in L(G)} \binom{d_v}{r}$ which implies that

$$\sum_{v \in L(G)} \binom{d_v}{r} \leq (r-1)\binom{n}{r}.$$

We want to use this relation to bound the total number of edges in G (or equivalently the average left degree of G). Let $\bar{d}$ denote the average degree of a left vertex in G. We can now use convexity of the function $\binom{d_v}{r}$ to help us out. In particular,

$$n\binom{\bar{d}}{r} \leq \sum_{v \in L(G)} \binom{d_v}{r} \leq (r-1)\binom{n}{r}$$

Ignoring lower order terms (it is not hard to formally argue that they have negligible effect), this implies that $n\bar{d}^r \leq rn^r$ and hence that $\bar{d} = O(n^{1-1/r})$ which implies the result. $\square$

1.4 Graphs with large girth and chromatic number

Our final example of alterations is a beautiful result of Erdös about graphs that have a high girth (the girth of a graph G is the length of the smallest cycle in G), yet have high chromatic number. Note that if the girth is g, this means that for every vertex v, its distance $(g-1)/2$ neighborhood looks like a tree.

If a graph contains a k clique, then its chromatic number $\chi(G) \geq k$. It is natural to wonder if some approximate converse of the this is true. The result of Erdös disproves this in a strong sense. It also means that 'local considerations' are not useful for determining $\chi(G)$.

Theorem 6 *For any k, there exists graph with girth more than k and $\chi(G) \geq n^{1/4k}$.*

Proof: We will try to construct such a graph randomly, and then apply alterations. First, let us think how can we hope to argue that $\chi(G) \geq n^{1/4k}$?

Luckily, we realize that $\chi(G) \geq n/\alpha(G)$, so we can try to show that $\alpha(G)$ is small. In fact, in general this is the only strategy we know to lower bound $\chi(G)$.

We create a random graph on n vertices by including every possible edge with probability p, where we fix $p = \frac{n^{1/2k}}{n}$. We can then try to remove cycles of length $\leq k$, by removing one vertex arbitrarily for each such cycle. The hope would be that we do not remove essentially the whole graph. Second, note that removing a vertex can never increase the size of the maximum independent set in the resulting graph (can you see why?).

So let us calculate how many expected cycles of length at most k this random graph contains. As we only need an upper bound, we can be a bit sloppy. A particular, cycle of length l appears with probabiity p^l. Moreover, we can obtain a cycle by choosing l vertices and picking some order, so there are at most $n^l l!$ such possibilities. So,

$$E(\text{number of cycles of length} \le k) \le \sum_{l=3}^{k} \binom{n}{l} l! p^l \le \sum_{l=1}^{k} \frac{n^l}{l!} l! p^l$$
$$= \sum_{l=1}^{k} (np)^l = \sum_{l=1}^{k} (n^{1/2k})^l \le 2\sqrt{n} \qquad \text{(assuming } n \text{ is large enough s.t. } n^{1/(2k)} \le 1/2)$$

So we can easily eliminate such cycles and still retain $n - \Omega(\sqrt{n})$ vertices.

We now try to bound $\alpha(G)$. Let $r = 2n^{1-1/4k}$. Fix some subset S with r vertices. The probability that S is an independent set is $(1-p)^{\binom{r}{2}}$. There are $\binom{n}{r}$ subsets of size r, so the expected number of independent sets of size r is

$$\binom{n}{r}(1-p)^{\binom{r}{2}} \le 2^n e^{-pr^2/4} \le 2^n e^{-n}$$

which is exponentially small in n.

Now, by Markov's inequality, at least $9/10$ of the graphs have no more than $20\sqrt{n}$ cycles smaller than k and at least $9/10$ of the graphs have no independent sets bigger than r. So, there must exist (plenty of) graphs that satisfy both these properties simultaneously. This proves the result. $\square$