

Refined Probabilistic Methods: Dependent Random Choice

Graph Theory (Fall 2011)
Rutgers University
Swastik Kopparty

Two trivial but profound facts:

- If there are two queues of people, most people are in the longer line (even though they probably all tried to join the shorter one!).
- Suppose you have a graph where the vertices have widely differing degrees. Consider the following 2 experiments: (1) you pick a vertex uniformly at random, and (2) you pick an edge uniformly at random, and then pick one of its two vertices. The second experiment is more likely to produce a vertex of a higher degree.

We will see some real concrete graph theoretic situations where such facts play an important role.

We start with a warmup lemma. For a set S of vertices, let $N(S)$ denote the intersection of the neighborhoods of all the vertices in S . (as opposed to $\Gamma(S)$, which denotes the union).

Lemma 1. *Let $G = (L, R, E)$ be a bipartite graph with $|L| = |R| = n$. Suppose $|E| \geq \alpha n^2$. Then there is a set $T \subseteq L$, $|T| \geq \frac{\alpha}{2}n$, such that at least 0.99 fraction of the pairs of vertices (w_1, w_2) in T have $|N(w_1, w_2)| \geq \frac{\alpha^2}{200}n$.*

Proof. Pick a vertex $v \in R$ uniformly at random, and let $T = N(v)$.

By design, every pair of vertices in T has a common neighbor (namely v). As we will see, this magically biases T to have the desired property (in expectation).

Let $\beta = \frac{\alpha^2}{200}$. We call a pair of vertices $(w_1, w_2) \in L^2$ *bad* if $|N(w_1, w_2)| < \beta n$.

Let B be the number of bad pairs (w_1, w_2) such that w_1, w_2 both lie in T . Then $\mathbb{E}[B]$ equals:

$$\sum_{(w_1, w_2) \in L^2, |N(w_1, w_2)| < \beta n} \Pr[w_1, w_2 \in T] = \sum_{(w_1, w_2) \in L^2, |N(w_1, w_2)| < \beta n} \Pr[v \in N(w_1, w_2)] \leq \beta n^2.$$

Let P be the total number of pairs (w_1, w_2) such that w_1, w_2 both lie in T . Then $\mathbb{E}[P]$ equals:

$$\frac{1}{n} \sum_v d_v^2 \geq (\alpha n)^2.$$

Therefore

$$\mathbb{E}[P - 100B] \geq (\alpha^2 - 100\beta)n^2 \geq \frac{\alpha^2}{2}n^2.$$

Thus there is a choice of v for which $P - 100B \geq \frac{\alpha^2}{2}n^2$. In particular, $P \geq \frac{\alpha^2}{2}n^2$ (so that $|T| = \sqrt{P} \geq \frac{\alpha}{2}n$) and $B \leq \frac{P}{100}$. This T has the desired property. \square

1 Application to Turan numbers of some bipartite graphs

We now see an application of this style of argument to Turan numbers. Here we want to say that a dense enough graph contains a certain graph as a subgraph; we will use dependent random choice to find that subgraph.

Theorem 2 (Alon-Krivilevich-Sudakov). *Let H be a bipartite graph such that every left vertex has degree at most r . Then $\text{ex}(n, H) \leq O(n^{2-\frac{1}{r}})$.*

Proof. Let L and R be the two parts of the bipartite graph H , and let a, b be their cardinalities.

Let $G = (V_G, E_G)$ be a graph with n vertices and $m = \Omega(n^{2-\frac{1}{r}})$ edges. We want to show that G contains a copy of H . We first make a simple observation:

Observation: If there exists a set $S \subseteq V_G$ such that:

- $|S| = b$,
- for every $T \subseteq S$ with $|T| = r$, we have $|N(T)| \geq a + b$,

then G contains a copy of H .

The proof of the observation is very simple; we just map the vertices of R to S (one-to-one arbitrarily). Then for each v vertex of L , the second condition guarantees that we will be able to choose a corresponding vertex u of V_G so that the set of neighbors of v in R corresponds to the set of neighbors of u in S .

We now show that such an S exists. This S we are looking for should have the property that every small subset of it should have many common neighbors; this suggests that we should look for S which is the common neighborhood of some c vertices.

Let $T = (v_1, \dots, v_c)$ be c uniformly random vertices of G . Let $S' = N(T)$. Then (using Holder's inequality), we get:

$$\mathbb{E}[|S'|] \geq \frac{(2m)^c}{n^{2c-1}}.$$

Let B be the random variable which counts the number of bad r -subsets in S' , namely the number of $T \subseteq S'$ with $|T| = r$ such that $|N(T)| < a + b$. Then

$$\mathbb{E}[B] \leq \binom{n}{r} \cdot \left(\frac{a + b - 1}{n} \right)^c.$$

Then if we remove from S' one element from each of the B bad r -subsets, we will get a set S with size at least $|S'| - B$ which has no bad r -subsets, and this S will have the desired property.

It remains to note that

$$\mathbb{E}[|S'| - B] \geq \frac{(2m)^c}{n^{2c-1}} - n^{r-c} \frac{(a + b - 1)^c}{r!},$$

which for $r = c$ and $m = \Omega(n^{2-\frac{1}{r}})$ is larger than a . □

Thus it turns out that we are choosing r uniformly random vertices in T and making $S = N(T)$ (after some deletions). This seems to be significant, and that there should be a way of motivating this choice (but it is not clear to me).