Lecture note 5

Sharp thresholds

The Erdős-Rényi random graph model refers to two closely related models for generating random graphs: if $0 \leq p \leq 1$, then $G_n(p)$ is the graph on $n$ vertices whose edges are chosen independently with probability $p$; if $m \in \{0, \ldots, n\}$, then $G_n(m)$ is the random graph chosen uniformly among all graphs with $n$ vertices and $m$ edges.

A graph property $P$ is a family of graphs closed under isomorphism, $P_n$ denotes the elements of $P$ with $n$ vertices. A typical question in the subject of random graphs is that if $P$ is a certain graph, then what is the probability that $G_n(p)$ (or $G_n(m)$) has property $P$? Let $f_{n,P}(p)$ (and $f(n, P)(m)$) denote this probability.

A graph property is monotone, if for every graph $G$ on $n$ vertices, if $G$ contains a member of $P_n$, then $G \in P_n$. One can observe that for many monotone graph properties $P$, the function $p \rightarrow f_{n,P}(p)$ has a sharp threshold. That is, if $f_{n,P}(p) > \epsilon$, then $f_{n,P}(p') > 1 - \epsilon$ for some $p'$ only slightly larger than $p$.

For example, if $P$ is the family of connected graphs, then $f_{n,P}$ has a sharp threshold at $p_0 = \log n/n$. More precisely, for every $\epsilon > 0$ there exists $\delta > 0$ such that

$$f_{n,P}\left(\frac{\log n}{n} - \frac{\delta}{n}\right) \leq \epsilon,$$

while

$$f_{n,P}\left(\frac{\log n}{n} + \frac{\delta}{n}\right) \geq 1 - \epsilon.$$

A similar statement holds for $f_{n,P}(m)$ as well with threshold at $m_0 = n \log n$. Let us provide some explanation of this phenomena.

Monotone families

A family $S \subset 2^{[n]}$ is monotone if there exist no $A \in S$ and $B \in 2^{[n]} \setminus S$ such that $A \subset B$. For $k = 0, \ldots, n$, let $S_k = [n]^{(k)} \cap S$ and let $p_k(S) = |S_k|/(\binom{n}{k})$.

**Theorem 1** (Bollobás, Thomason). Let $\epsilon > 0$, then there exists $c(\epsilon) > 0$ such that the following holds. Let $n$ be a positive integer and $S \subset 2^{[n]}$ be a monotone family. If $k \in \{0, \ldots, n\}$ satisfies that $p_k(S) > \epsilon$, then $p_m(S) \geq 1 - \epsilon$ for every $m \geq c(\epsilon)k$. 

1
Proof. Let \( T = 2^{[n]} \setminus S \). Then \( T \) is an ideal, that is, if \( A \in T \), then for every \( B \subseteq A \) we have \( B \in T \). Therefore, \( \triangle(T_l) \subseteq T_{l-1} \) for \( l = 1, \ldots, n \). Let \( x_l \) be the unique real number satisfying \( |T_l| = \binom{x_l}{l} \). By the theorem of Lovász (Theorem 4 in Lecture note 3), we have

\[
|T_{l-1}| \geq |\triangle(T_l)| \geq \binom{x_l}{l-1}.
\]

This implies that \( x_0 \geq x_1 \geq \ldots \geq x_n \). Consider the functions

\[
f_l(x) = \binom{x}{l} / \binom{n}{l} = \frac{x(x-1)\ldots(x-l+1)}{n(n-1)\ldots(n-l+1)}.
\]

As \( S_l = [n]^{(l)} \setminus T_l \), we have \( p_l(x) = 1 - f_l(x_l) \).

Claim 2. If \( l < m \), then \( f_l(x_l)^{1/l} > f_m(x_m)^{1/m} \).

Proof. Let \( s_i = (x - i + 1)/(n - i + 1) \). Then \( s_1 \geq s_2 \geq \ldots \) and \( f_l(x) = s_1 \ldots s_l \). Hence,

\[
f_l(x_l)^{1/l} \geq s_{l+1} \geq s_{l+2} \geq \ldots.
\]

But \( f_m(x) = f_l(x)s_{l+1} \ldots s_m \), so we get

\[
f_m(x) \leq f_l(x)f_l(x)^{(m-l)/l} = f_l(x)^{m/l}.
\]

The previous claim implies that

\[
(1 - p_l(S))^{1/l} \geq (1 - p_m(S))^{1/m}
\]

holds for every \( l < m \). Setting \( c(\epsilon) = \log \epsilon / \log(1 + \epsilon) \), we get that for every \( m \geq c(\epsilon)k \),

\[
1 - p_m(S) \leq (1 - p_k(S))^{m/k} \leq (1 - \epsilon)^{c(\epsilon)} = \epsilon,
\]

so \( p_m(S) \geq 1 - \epsilon \).

\[ \square \]

Symmetric monotone families

A family \( S \subseteq 2^{[n]} \) is symmetric if there exists a transitive permutation group \( \Gamma \) on \([n]\) that is invariant on \( S \). In other words, for every \( A \in S \), the set \( \Gamma(A) = \{ \Gamma(i) : i \in A \} \) is also an element of \( S \).

Note that graph properties correspond to symmetric families. Indeed, if \( N = \binom{n}{2} \) and we identify the edges of an \( n \) vertex graph with \([N]\), then the group of all permutations acting on the vertices of the graph induce a transitive permutation group on \([N]\), which is invariant on any family of \( 2^{[N]} \) corresponding to a graph property.

In case \( S \) is monotone and symmetric, we can improve Theorem 1 in certain ranges, as shown by the following theorem of Friedgut and Kalai (1996).
\textbf{Theorem 3.} Let $S \subset 2^n$ be a symmetric monotone family. If $p \in (0, 1)$ satisfies $\mu_p(S) > \epsilon$, then $\mu_q(S) > 1 - \epsilon$ for every

$$q > p + c_1 \frac{\log(1/2\epsilon)}{\log n},$$

where $c_1$ is an absolute constant.

Let us extend the definition of influence for the $p$-biased measure on $2^n$. For $f : 2^n \to \{0, 1\}$ and $i \in [n]$, let

$$\text{Inf}_{i,p}(f) = \sum_{f(x) \neq f(x \Delta \{i\})} \mu_p(x) = P(f(x) \neq f(x \Delta \{i\})),$$

where the probability is meant with respect to the $p$-biased measure. Also, if $S \subset 2^n$, then $\text{Inf}_{i,p}(S) = \text{Inf}_{i,p}(\chi(S))$, where $\chi(S)$ is the characteristic function of $S$.

The following extension of the KKL theorem, called the BKKKL theorem after Bourgain, Kahn, Kalai, Katznelson and Linial (1992), can be proved similarly as KKL theorem. Therefore, we omit its proof.

\textbf{Theorem 4 (BKKKL theorem).} For every $S \subset 2^n$ and $p \in (0, 1)$, if $t = \mu_p(S)$, then

$$\max_{i \in [n]} \text{Inf}_{i,p}(S) \geq c_2 t \frac{\log n}{n},$$

where $t' \in \min\{t, 1 - t\}$ and $c_2$ is an absolute constant.

Also, let us define $p$-biased analogue of the edge boundary. Let $S \subset 2^n$. For $A \in S$, let $h_S(A) = |\{B \in 2^n \setminus S : |A \Delta B| = 1\}|$. Define

$$\psi_p(S) = \sum_{A \in S} h_S(A) \mu_p(A).$$

Note that $\psi_p(S)/p = \sum_{i \in [n]} \text{Inf}_{i,p}(S)$. We use the following lemma of Margulis (1974) and Russo (1978) to relate the $p$-biased measure and edge boundary of $S$.

\textbf{Lemma 5.} Let $S \subset 2^n$ be a monotone family. Then

$$\frac{d\mu_p(S)}{dp} = \psi_p(S) / p.$$

\textit{Proof.} We proceed by induction on $n$. To indicate that our family is in $2^n$, let us write $\psi_p^{(n)}, \mu_p^{(n)}, h_S^{(n)}$ instead of $\psi_p, \mu_p, h_S$.

If $n = 1$, there are three cases: $S = \emptyset$, $S = \{[1]\}$ and $S = 2^1$. In each of these cases, it is easy to check that the statement holds.

Now suppose that $n > 1$. Let $S^- = \{A \in S : n \notin S\}$ and $S^+ = \{A \setminus \{n\} : n \in A \in S\}$. Then $S^-, S^+ \subset 2^{[n-1]}$ and

$$\mu_p^{(n)}(S) = (1 - p)\mu_p^{(n-1)}(S^-) + p\mu_p^{(n-1)}(S^-),$$

which gives

$$\frac{d\mu_p^{(n)}(S)}{dp} = (1 - p)\frac{d\mu_p^{(n-1)}(S^-)}{dp} - \mu_p^{(n-1)}(S^-) + p\frac{d\mu_p^{(n-1)}(S^+)}{dp} + \mu_p^{(n-1)}(S^+). \quad (1)$$
Also, as $S$ is monotone, $S^- \subset S^+$. But then if $A \in S^-$, then $h_S^{(n)}(A) = h_S^{(n-1)}(A)$, if $A \in S^+ \cap S^-$ then $h_S^{(n)}(A \cup \{n\}) = h_S^{(n-1)}(A)$, and if $A \in S^+ \setminus S^-$, then $h_S^{(n)}(A \cup \{n\}) = h_S^{(n-1)}(A) + 1$. This implies that
\[
\frac{\psi_p^{(n)}(S)}{p} = (1 - p) \frac{\psi_p^{(n-1)}(S^-)}{p} + p \frac{\psi_p^{(n-1)}(S^+)}{p} + \mu_p^{(n-1)}(S^+ \setminus S^-).
\] (2)

By our induction hypothesis (1) and (2) are equal, finishing the proof.

We are now ready to prove our main theorem.

**Proof of Theorem 3.** Let $r \in (0, 1)$. As $S$ is symmetric, we have Inf$_{1,r}(S) = \ldots$Inf$_{n,r}(S)$. But then such that $\mu_r(S) \leq 1/2$. If $\mu_r(S) < 1/2$, then by the BKKKL theorem, we have Inf$_{i,r}(S) \geq c_2 \mu_r(S) \log n/n$ for $i \in [n]$, which gives

\[
\psi_r(S) = \sum_{i \in [n]} \text{Inf}_{i,r}(S) \geq c_2 \mu_r(S) \log n.
\]

But then Lemma 5 implies that

\[
\frac{d \mu_r(S)}{dr} \geq c_2 \mu_r(S) \log n,
\]

which gives

\[
\frac{d \log \mu_r(S)}{dr} \geq c_2 \log n.
\]

Hence, for $r = p + \log(1/2\epsilon)/c_2 \log n$, either $\mu_r(S) > 1/2$, or we get

\[
\log \mu_r(S) \geq \log \mu_p(S) + \int_p^r c_2 \log n \, dx = \log \mu_p(S) + \log(1/2\epsilon) \geq \log(1/2),
\]

which also gives $\mu_r(S) \geq 1/2$. Similarly, if $q > r + \log(1/2\epsilon)/c_2 \log n$, then $\mu_q(S) > 1 - \epsilon$, so we can take $c_1 = 2/c_2$ to finish our proof.

**Research problems**

We have seen that for graph properties, $\mu_p(S)$ changes from $\epsilon$ to $1 - \epsilon$ in an interval of size $O(1/\log n)$. However, it is believed that this is not sharp. There are examples showing that for certain graph properties, this interval needs to have size $\Omega(1/(\log n)^2)$ (see Exercise problem 1), and it is conjectured that this should be sharp, see [1].

**Conjecture 6.** Let $P$ be a monotone graph property. If $p \in (0, 1)$ satisfies $f_{n,p}(p) \geq \epsilon$, then $f_{n,p}(q) \geq 1 - \epsilon$ for every $q \geq p + \Omega(1/(\log n)^2)$.

**References**

Exercise problems

1. Let $k$ be a positive integer. Pick each element of $[n]$ independently with probability $p$ and let $A$ be the resulting set. Prove that if $p = o(n^{-2/k})$, then

$$\mathbb{P}(A \text{ contains a } k\text{-term arithmetic progression}) = o(1).$$

2. Let $P$ be the graph property that a graph $G$ on $n$ vertices contains a clique of size $2\log_2 n$. Prove that there exists a constant $c > 0$ such that

$$f_{n,P}(1/2 - c/(\log n)^2) > 0.1 \text{ and } f_{n,P}(1/2 + c/(\log n)^2) < 0.9.$$ 

3. * Let $0 < p < 1/2$ and $\epsilon > 0$. Prove that if $n$ is sufficiently large, then for every $P \subset 2^n$ symmetric intersecting family, we have $\mu_p(S) < \epsilon$. 