Threshold functions

• The most interesting phenomena in \( G(n,p) \) random graphs emerge when \( p = p(n) \) is not a constant but \( p = p(n) \to 0 \) in some controlled way.

• Recall: Function \( t = t(n) \) is a threshold for graph property \( \mathcal{A} \) if
  1. \( p < t \) \( \Rightarrow \) \( G \not\in \mathcal{A} \) for a.e. \( G \in G(n,p) \),
  2. \( p > t \) \( \Rightarrow \) \( G \in \mathcal{A} \) for a.e. \( G \in G(n,p) \).

• As an example, let us review Theorem 4.6: define the density of a graph \( G = (V,E) \) as \( q(G) = |E|/|V|^2 \), and say that \( G \) is balanced if \( q(G') = q(G) \) for all subgraphs \( G' \) of \( G \).

• Theorem 5.8 (Erdős & Rényi 1960). Let \( H \) be a balanced graph. Then the graph property "\( G \) has a subgraph isomorphic to \( H \)" has threshold \( n^{-1/2} q(H) \).

Proof: We apply the first- and second-moment methods as in the special case of Thm 4.3, but now simplify the calculations using Lemma \( A^* \) (p. 38).

For a given balanced graph \( H \), denote \( l = |E|, k = |V| \), so that \( q(H) = l/k \).

1. [Upper threshold / 1st-moment method:] For each vertex set \( S \), \( |S| = k \), define the indicator variable

\[ X_S = \begin{cases} 1 & \text{if } S \text{ contains a copy of } H \\ 0 & \text{otherwise} \end{cases} \]

and consider the sum \( X = \sum_{|S|=k} X_S \).
Now
\[ p^k \leq \Pr(X_s = 1) \leq k! \cdot p^k \]
(The upper bound is due to the fact that each ordering of the \( k \) vertices induces at most one copy of \( H_k \).)

Thus, by linearity of expectation:
\[
E[X] = \sum_{i=1}^{k} E[X_s] = \binom{n}{k} \Pr(X_s = 1) = \Theta(n^k \cdot p^k).
\]

Now if \( p > n^{-k/2} \), then \( E[X] \to 0 \) as \( n \to \infty \), and consequently also \( \Pr(X > 0) \to 0 \) as \( n \to \infty \).

(ii) [Lower threshold / 2nd-moment method:] Now assume that \( p > n^{-k/2} \), so that \( E[X] \to 0 \) as \( n \to \infty \). By lemma \( \Delta^* \), in order to show that \( \Pr(X > 0) \to 1 \) as \( n \to \infty \), it suffices to show that \( \Delta^* = o(E[X]) \), where
\[
\Delta^* = \sum_{T \in S} \Pr(X_T = 1 | X_s = 1) \quad \text{for fixed } S.
\]
(Note that all the \( k \)-sets of vertices "look the same" except for the numbering of the vertices.)

Here \( T \in S \) iff \( T \neq S \) and \( T \cap S \) have common edges, i.e. if \( |T \cap S| = r \) for some \( r = 2, \ldots, k-1 \). Thus:
\[
\Delta^* = \sum_{r=2}^{k-1} \sum_{|T \cap S| = r} \Pr(X_T = 1 | X_s = 1).
\]
Now for a fixed $k$ and given $r$, there are $\binom{k}{r} \binom{n-k}{k-r}$ choices of $T$.

For any choice of $T$, there are at most $k! = O(1)$ copies of $H$ on $T$. Each of these contains at most

\[ g(k) \cdot r = \frac{rk}{k} \]

edges both of whose endpoints are also in $S$. (Consider the induced subgraph of $H$ on $T \cap S$ and note that $H$ is balanced!) Consequently, each copy of $H$ on $T$ contains at least $k - r \frac{k}{k}$ edges one of whose endpoints is not in $S$, and so

\[ \Pr(X_T = 1 | X_e = 1) \leq k! \left( \frac{k}{k} - \frac{r}{k} \right)^k = O\left( p^\frac{k}{k} \right) = O(p^{1 - \frac{r}{k}}) \]

Hence

\[ \Delta^* = \sum_{r=2}^{k-1} \binom{k}{r} \binom{n-k}{k-r} O(p^{2(1 - r/k + r)}) \]

\[ = \sum_{r=2}^{k-1} O(n^{k-r} p^{2(1 - r/k + r)}) \]

\[ = \sum_{r=2}^{k-1} O(n^{k-r} p^{2 - (r/k)}) \]

\[ = O(k \cdot n^{-k} p^2) \]

\[ = o(n^k p^2) \]

\[ = o(\mathbb{E}[X]). \]

Lemma $\Delta^*$ thus applies, and $\Pr(X > 0) \to 1$ as $n \to \infty$. \qed
Corollary 5.9. For $k \geq 3$, the property "$G$ contains a $k$-cycle" has threshold $n^{-1}$. (Note that the threshold is independent of $k$.)

Corollary 5.10. For $k \geq 2$, the property "$G$ contains a $k$-clique" has threshold $n^{-2/(k-1)}$.

Corollary 5.11. For $k \geq 2$, the property of $G$ containing a specific tree structure on $k$ nodes has threshold $n^{-k/(k-1)}$.

Theorem 5.9. Theorem 5.8 can be further generalised as follows: for a graph $H$, define

$$g^*(H) = \max \{ g(H') \mid H' \text{ is a subgraph of } H \}.$$ 

Theorem 5.9. For any given graph $H$, the graph property "$G$ has a subgraph isomorphic to $H$" has threshold $n^{-1/g^*(H)}$.

Proof. Omitted.
7. Random Graphs

Threshold functions for global graph properties

Also known as the "phase transition".

The "epochs of evolution": Consider the structure of random graphs \( G \in G(n, p) \) as \( p = p(n) \) increases. The following results can be shown (note that \( np = \) average node degree):

0. If \( p < n^{-2} \), then a.e. \( G \) is empty.

1. If \( n^{-2} < p < n^{-1} \), then a.e. \( G \) is a forest (a collection of trees).
   - The threshold for the appearance of any \( k \)-node tree structure is \( p = n^{-k/(k-1)} \).
   - The threshold for the appearance of cycles (of all constant sizes) is \( p = n^{-1} \).

2. If \( p \sim cn^{-1} \) for any \( c < 1 \) (i.e. \( np \to c < 1 \) as \( n \to \infty \)), then a.e. \( G \) consists of components with at most one cycle and \( \Theta(\log n) \) nodes.

3. "Phase transition" or "emergence of the giant component" at \( p \sim n^{-1} \) (i.e. \( np \to 1 \)).

4. If \( p \sim cn^{-1} \) for any \( c > 1 \) (i.e. \( np \to c > 1 \)), then a.e. \( G \) consists of a unique "giant" component with \( \Theta(n) \) nodes and small components with at most one cycle.

5. If \( n^{-1} < p < \frac{\ln n}{n} \), then a.e. \( G \) is disconnected, consisting of one giant component and trees.

6. If \( p \sim \frac{\ln n}{n} \), then a.e. \( G \) is connected (in fact Hamiltonian).

\[ \text{Theorem 6.10} \]

Let \( p_t(n) = \frac{\ln n - \omega(n)}{n} \), \( p_u(n) = \frac{\ln n + \omega(n)}{n} \) where \( \omega(n) \to \infty \). Then

(i) a.e. \( G \in G(n, p_t) \) is disconnected;

(ii) a.e. \( G \in G(n, p_u) \) is connected.

\[ \text{Proof.} \] We shall use the second moment method on random variables \( X_k = X_k(G) = \) number of components on \( G \) with exactly \( k \) nodes.

Assume without loss of generality that \( \omega(n) \leq \ln \ln n \) and \( \omega(n) \geq 10 \).
Part II. Combinatorial Models

(i) Set \( p = p_t \) and compute \( \mu = E(X_1), \sigma^2 = \text{Var}(X_1) \). By linearity of expectation,

\[
\mu = E(X_1) = n(1 - p)^{n-1} = ne^{(n-1)\ln(1-p)} \\
\leq ne^{-np} = ne^{-\ln n + o(n)} = e^{o(n)} \xrightarrow{n \to \infty} 0.
\]

Furthermore, the expected number of ordered pairs of isolated nodes is

\[
E(X_1(X_1 - 1)) = n(n-1)(1-p)^{2n-3}.
\]

Hence,

\[
\sigma^2 = \text{Var}(X_1) = E(X_1^2) - \mu^2 \\
= E(X_1(X_1 - 1)) + \mu - \mu^2 \\
= n(n-1)(1-p)^{2n-3} + n(1-p)^{n-1} - n^2(1-p)^{2n-2} \\
\leq n(1-p)^{n-1} + pn^2(1-p)^{2n-3} \\
\leq \mu + (\ln n - o(n))ne^{-2\ln n + 2o(n)}(1-p)^{-3} \\
\leq \mu + \frac{2\ln n}{n}e^{o(n)} \leq \mu + 1 \quad \text{for large } n.
\]

Thus, \( \frac{\sigma^2}{\mu^2} \xrightarrow{n \to \infty} 0 \) and by Lemma 7.16.

\[
\Pr(\text{G is disconnected}) \geq \Pr(X_1(G) > 0) \to 1 \quad \text{as } n \to \infty.
\]

(ii) (Here basic expectation estimation, or "1st moment method" suffices.)

Set \( p = p_u = \frac{\ln n + o(n)}{n} \) and compute

\[
\Pr(\text{G is disconnected}) = \Pr \left( \sum_{k=1}^{\lfloor n/2 \rfloor} X_k \geq 1 \right) \\
\leq E \left( \sum_{k=1}^{\lfloor n/2 \rfloor} X_k \right) = \sum_{k=1}^{\lfloor n/2 \rfloor} E(X_k) \\
\leq \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} (1-p)^k(1-p)^{n-k} \\
= \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} (1-p)^n (1-p)^{-k} \\
= \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} (1-p)^{n-k} \\
\]

(5)
Split the sum (5) in two parts:

\[(a) \quad \sum_{1 \leq k \leq n^{3/4}} \binom{n}{k} (1 - p)^{k(n-k)} \leq \sum_{1 \leq k \leq n^{3/4}} \left( \frac{en}{k} \right)^k e^{k(1-\omega(n))} e^{k^2 \ln n/n} \]
\[= \sum_{1 \leq k \leq n^{3/4}} k^{-1} e^{(1-\omega(n))k} e^{2k \ln n/n} \leq e^{-\omega(n)} \sum_{1 \leq k \leq n^{3/4}} \exp \left( -k \ln k + k + \frac{2k^2 \ln n}{n} \right) \leq 3e^{-\omega(n)}.\]

\[(b) \quad \sum_{n^{3/4} \leq k \leq n/2} \binom{n}{k} (1 - p)^{k(n-k)} \leq \sum_{n^{3/4} \leq k \leq n/2} \left( \frac{en}{k} \right)^k e^{k(1-\omega(n))} \leq \frac{n}{2} e^{n/2} n^{-1/4} n^{3/4} \leq n^{-3/4/5} \leq e^{-\omega(n)} \quad \text{for large } n.\]

Thus, altogether

\[\Pr(G \text{ is disconnected}) \leq 4e^{-\omega(n)} \xrightarrow{n \to \infty} 0. \qed\]
Part II. Combinatorial Models

What happens at the “phase transition” $p \sim n^{-1}$? For fixed values of $n$ and $N = \binom{n}{2}$, consider the space of “graph processes” $\mathcal{G} = (G_t)_{t=0}^N$, where at each “time instant” $t$ a new edge is selected uniformly at random for insertion into an $n$-node graph. (Thus, picking graph $G_t$ from a randomly chosen process $G \in \mathcal{G}(n,M)$, where $M = t$.)

5.11

Theorem 5.11 Let $c > 0$ be a constant and $\omega(n) \to \infty$. Denote $\beta = (c - 1 - \ln c)^{-1}$ and $t = t(n) = \lceil cn/2 \rceil$. Then

(i) At $c < 1$, every component $C$ of a.e. $G_t$ satisfies

$$|C| - \beta \left( \ln n - \frac{5}{2} \ln \ln n \right) \leq \omega(n).$$

(ii) At $c = 1$, for any fixed $h \geq 1$ the $h$ largest components $C$ of a.e. $G_t$ satisfy

$$|C| = \Theta(n^{2/3}).$$

(iii) At $c > 1$, the largest component $C_0$ of a.e. $G_t$ satisfies

$$|C_0| - \gamma n \leq \omega(n) \cdot n^{1/2},$$

where $0 < \gamma = \gamma(c) < 1$ is the unique root of

$$e^{-\gamma} = 1 - \gamma.$$

The other components $C$ of a.e. $G_t$ satisfy also in this case

$$|C| - \beta \left( \ln n - \frac{5}{2} \ln \ln n \right) \leq \omega(n).$$

Thus, the fraction of nodes in the “giant” component of a.e. $G_t$ for $t = cn/2$ behaves as illustrated in Figure 8.

Let us prove one part of this result, the emergence of a gap in the component sizes of $G \in \mathcal{G}(n,p)$ at $p \sim n^{-1}$. (This corresponds to $t \sim N_p \sim n/2$.)

5.12

Theorem 5.12 Let $a \geq 2$ be fixed. Then for large $n$, $\varepsilon = \varepsilon(n) < 1/3$ and $p = p(n) = (1 + \varepsilon)n^{-1}$, with probability at least $1 - n^{-\alpha}$, a random $G \in \mathcal{G}(n,p)$ has no component $C$ that satisfies

$$\frac{8a}{\varepsilon^2 \ln n} \leq |C| \leq \frac{\varepsilon^2}{12} n.$$
7. Random Graphs

![Image of a graph with nodes and edges]

Figure 8: Fraction of nodes in the giant component.

**Proof.** Let us consider "growing" the component $C(u)$ of an arbitrary node $u$ in $G$ incrementally as follows:

1. (Stage 0:) Set $A_0 = \emptyset, B_0 = \{u\}.$

2. (Stage $i + 1$:) If $B_i = A_i,$ then stop with $C(u) = B_i.$ Otherwise pick an arbitrary $v \in B_i \setminus A_i$; set $A_{i+1} = A_i \cup \{v\},$ $B_{i+1} = B_i \cup \{\text{neighbours of } v \text{ in } G\}.$

Now what is the probability distribution of $|B_i|$ (=size of set $B_i$)?

Consider any node $v \in G \setminus \{u\}.$ It participates in $i$ independent Bernoulli trials for being included in $B_i,$ each with success probability equal to $p.$ Thus the inclusion probability for any fixed $v \neq u$ is $1 - (1 - p)^i,$ independently of each other.

Consequently, the size of each $B_i$ obeys a simple binomial distribution

$$\Pr(|B_i| = k) = \binom{n-1}{k} (1 - (1 - p)^i)^k (1 - p)^{(n-k-1)}.$$ 

This gives also for each $k$ an upper bound on the probability

$$\Pr(|C(u)| = k) = \Pr(|B_i| = k \land \text{ process stops at stage } i).$$

Denoting $p_k = \Pr(|C(u)| = k)$ for any fixed $u \in G,$ it is clear that

$$\Pr(G \text{ contains a component of size } k) \leq np_k,$$

and to prove the theorem it suffices to show that

$$\sum_{k=k_0}^{k_1} p_k \leq n^{-a-1},$$
where \( k_0 = [8\sigma e^{-2}\ln n] \), \( k_1 = [\varepsilon^2 n/12] \).

Since presumably \( k_0 \leq k_1 \), we may assume \( \varepsilon^4 \geq \frac{9k_1 \ln n}{n} \geq \frac{1}{n} \).

We may now estimate

\[
P_k \leq \Pr(|B_1| = k) \leq \frac{n^k}{k!} e^{-\frac{k^2}{2n}} (kp)^k (1 - p)^{k(n-k-1)},
\]

because

\[
\binom{n-1}{k} = \frac{n^k}{k!} \prod_{j=1}^{k} \left(1 - \frac{j}{n}\right) \leq \frac{n^k}{k!} e^{-\frac{k^2}{2n}}, \text{ and}
\]

\[(1 - p)^k \geq 1 - kp.
\]

Applying Stirling's formula

\[
\sqrt{2\pi k} \left(\frac{k}{e}\right)^k \leq k! \leq e^{12\pi} \sqrt{2\pi k} \left(\frac{k}{e}\right)^k
\]

and the bounds \( k_0 \leq k \leq k_1 \) to (6) we obtain

\[
P_k \leq \exp \left(\frac{-k^2}{2n} - \frac{\varepsilon^3 k}{3} + \frac{k^2(1 + \varepsilon)}{n}\right)
\]

\[
\leq \exp \left(\frac{-\varepsilon^2 k}{3} + \frac{k^2}{n}\right)
\]

\[
\leq \exp \left(\frac{-\varepsilon^2 k}{4}\right),
\]

and consequently

\[
\sum_{k=k_0}^{k_1} p_k \leq \sum_{k=k_0}^{k_1} e^{-\varepsilon^2 k/4} \leq e^{-\varepsilon^2 k_0/4} \cdot (1 - e^{-\varepsilon^2/4})^{-1}
\]

\[
\leq \frac{5}{\varepsilon^2} e^{-\varepsilon^2 k_0/4} \leq 5\sqrt{n} \cdot n^{-2\alpha}
\]

\[
= 5n^{-2\alpha+1/2} < n^{-a-1},
\]

for large \( n \). \( \Box \)