Threshold functions

- The most interesting phenomena in $\mathcal{G}(n,p)$ random graphs emerge when p 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 A if

(i) $p \ll t \Rightarrow G \not\models A$ for a.e. $G \in \mathcal{G}(n,p)$,

(ii) $p \gg t \Rightarrow G \models A$ for a.e. $G \in \mathcal{G}(n,p)$.

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

*) also Thm 4.6

- 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 $n^{-1/\varrho(H)}$.

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 Δ^* (p. 38).

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

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

$$X_S \sim \text{"}S \text{ contains a copy of } H\text{"}$$

and consider the sum $X = \sum_{|S|=k} X_S$.

Now

$$p^{\ell} \leq \Pr(X_S = 1) \leq k!\, p^{\ell}$$

(The upper bound is due to the fact that each ordering of the k vertices induces at most one copy of H.)

Thus, by linearity of expectation:

$$E[X] = \sum_{|S|=k} E[X_S] = \binom{n}{k} \Pr(X_S = 1)$$
$$= \Theta(n^k p^{\ell}).$$

Now if $p \ll n^{-k/\ell}$, 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 $p \gg n^{-k/\ell}$, so that $E[X] \to \infty$ as $n \to \infty$.
By lemma Δ^*, in order to show that $\Pr(X>0) \to 1$ as $n \to \infty$, it suffices to show that $\Delta^* = o(E[X])$, where

$$\Delta^* = \sum_{T \sim S} \Pr(X_T = 1 \mid X_S = 1) \qquad \text{for fixed } S.$$

(Note that all the k-sets of vertices "look the same" except for the numbering of the vertices.)

Here $T \sim S$ iff $T \neq S$ and T & S have common edges, i.e. if $|T \cap S| = r$ for some $r = 2, \ldots, k-1$. Thus:

$$\Delta^* = \sum_{r=2}^{k-1} \sum_{|T \cap S| = r} \Pr(X_T = 1 \mid X_S = 1).$$

Now for a fixed S and given r, there are $\binom{k}{r}\binom{n-k}{k-r}$ $= O(n^{k-r})$ 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

$$\varrho(H) \cdot r = \frac{r\ell}{k}$$

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 $\ell - r\ell/k$ edges one of whose endpoints is <u>not</u> in S, and so

$$\Pr(X_T = 1 \mid X_S = 1) \leq k!\, p^{\ell - \frac{r\ell}{k}} = O(p^{\ell(1-r/k)})$$

Hence

$$\begin{aligned}
\Delta^* &= \sum_{r=2}^{k-1} \binom{k}{r}\binom{n-k}{k-r}\, O(p^{\ell(1-r/k)}) \\
&= \sum_{r=2}^{k-1} O(n^{k-r}\, p^{\ell(1-r/k)}) \\
&= \sum_{r=2}^{k-1} O((n^k p^\ell)^{1-r/k}) \qquad \Big|\ \begin{array}{l} p \gg n^{-k/\ell} \\ \Rightarrow n^k p^\ell \to \infty \end{array} \\
&= O(k \cdot (n^k p^\ell)^{1-r/k}) \\
&= o(n^k p^\ell) \\
&= o(E[X]).
\end{aligned}$$

Lemma Δ^* thus applies, and $\Pr(X > 0) \to 1$ as $n \to \infty$.

$\square$

- <u>Corollary</u> 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.) $\square$

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

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

- Theorem 5.8 can be further generalised as follows: for a graph H, define

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

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

<u>Proof.</u> Omitted. $\square$

Threshold functions for global graph properties

Also known as the "phase transition".

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

0. If $p \prec n^{-2}$, then a.e. G is empty.
1. If $n^{-2} \prec p \prec n^{-1}$, then a.e. G is a forest (a collection of trees).
   - The threshold for the apperarance 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 $n^{-1} \prec p \prec \frac{\ln n}{n}$, then a.e. G is disconnected, consisting of one giant component and trees.
6. If $p \succ \frac{\ln n}{n}$, then a.e. G is connected (in fact Hamiltonian).

Theorem ~~7.15~~ 5.10 *Let $p_l(n) = \frac{\ln n - \omega(n)}{n}$, $p_u(n) = \frac{\ln n + \omega(n)}{n}$ where $\omega(n) \to \infty$. Then*

(i) *a.e. $G \in \mathcal{G}(n, p_l)$ is disconnected;*

(ii) *a.e. $G \in \mathcal{G}(n, p_u)$ is connected.*

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 $\omega(n) \le \ln\ln n$ and $\omega(n) \ge 10$.

47

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

$$\begin{aligned}\mu = E(X_1) &= n(1-p)^{n-1} = ne^{(n-1)\ln(1-p)} \\ &\le ne^{-np} = ne^{-\ln n + \omega(n)} = e^{\omega(n)} \xrightarrow[n\to\infty]{} \infty.\end{aligned}$$

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,

$$\begin{aligned}\sigma^2 &= \mathrm{Var}(X_1) = E(X_1^2) - \mu^2 \\ &= E(X_1(X_1-1)) + \mu - \mu^2 \\ &= n(n-1)(1-p)^{2n-3} + n(1-p)^{n-1} - n^2(1-p)^{2n-2} \\ &\le n(1-p)^{n-1} + pn^2(1-p)^{2n-3} \\ &\le \mu + (\ln n - \omega(n))ne^{-2\ln n + 2\omega(n)} \underbrace{(1-p)^{-3}}_{\le 2} \\ &\le \mu + \frac{2\ln n}{n} e^{2\omega(n)} \le \mu + 1 \qquad \text{for large } n.\end{aligned}$$

Thus, $\frac{\sigma^2}{\mu^2} \le \frac{\mu+1}{\mu^2} \to 0$ as $n \to \infty$, and by ~~lemma 7.10~~ Corollary 2 to Thm 4.2

$$\Pr(G \text{ is disconnected}) \ge \Pr(X_1(G) > 0) \to 1 \text{ as } n \to \infty.$$

(ii) (Here basic expectation estimation, or "1st moment method" suffices.)
Set $p = p_u = \frac{\ln n + \omega(n)}{n}$ and compute

$$\begin{aligned}\Pr(G \text{ is disconnected}) &= \Pr\left(\sum_{k=1}^{\lfloor n/2 \rfloor} X_k \ge 1\right) \\ &\le E\left(\sum_{k=1}^{\lfloor n/2 \rfloor} X_k\right) = \sum_{k=1}^{\lfloor n/2 \rfloor} E(X_k) \\ &\le \sum_{k=1}^{\lfloor n/2 \rfloor} \binom{n}{k} (1-p)^{k(n-k)} \qquad (5)\end{aligned}$$

Split the sum (5) in two parts:

$$
\begin{aligned}
\text{(a)} \quad & \sum_{1 \le k \le n^{3/4}} \binom{n}{k} (1-p)^{k(n-k)} \\
& \le \sum_{1 \le k \le n^{3/4}} \left(\frac{en}{k}\right)^k e^{k(n-k)(-p)} \\
& = \sum_{1 \le k \le n^{3/4}} \left(\frac{en}{k}\right)^k e^{-knp} e^{k^2 p} \\
& \le \sum_{1 \le k \le n^{3/4}} k^{-k} n^k e^k e^{-k(\ln n + \omega(n))} e^{k^2 \cdot 2 \ln n / n} \\
& = \sum_{1 \le k \le n^{3/4}} k^{-k} e^{(1-\omega(n))k} e^{2k^2 \ln n / n} \\
& \le e^{-\omega(n)} \cdot \underbrace{\sum_{1 \le k \le n^{3/4}} \exp\left(-k \ln k + k + 2k^2 \frac{\ln n}{n}\right)}_{\le 3} \\
& \le 3e^{-\omega(n)}.
\end{aligned}
$$

$$
\begin{aligned}
\text{(b)} \quad & \sum_{n^{3/4} \le k \le n/2} \binom{n}{k} (1-p)^{k(n-k)} \\
& \le \sum_{n^{3/4} \le k \le n/2} \left(\frac{en}{k}\right)^k e^{k(n-k)(-p)} \\
& \le \sum_{n^{3/4} \le k \le n/2} \left(en^{1/4}\right)^k n^{-n/4} \\
& \le \frac{n}{2} e^{n/2} n^{-\frac{1}{4} n^{3/4}} \\
& \le n^{-n^{3/4}/5} \\
& = \exp\left(-\frac{n^{3/4}}{5} \ln n\right) \\
& \le e^{-\omega(n)} \text{ for large } n.
\end{aligned}
$$

Thus, altogether

$$
\Pr(G \text{ is disconnected}) \le 4e^{-\omega(n)} \xrightarrow[n \to \infty]{} 0. \quad \square
$$

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" $\widetilde{G} = (G_t)_{t=0}^{N}$, 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 $\widetilde{G} \in \mathcal{G}(n,M)$, where $M = t$.)

Theorem ~~7.16~~ 5.11 *Let $c > 0$ be a constant and $\omega(n) \to \infty$. Denote $\beta = (c - 1 - \ln c)^{-1}$ and $t = t(n) = \lfloor cn/2 \rfloor$. Then*

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

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

(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

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

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

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

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 $t \sim N_p \sim n/2$.)

Theorem ~~7.17~~ 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 $1 - n^{-a}$, a random $G \in \mathcal{G}(n,p)$ has no component C that satisfies*

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

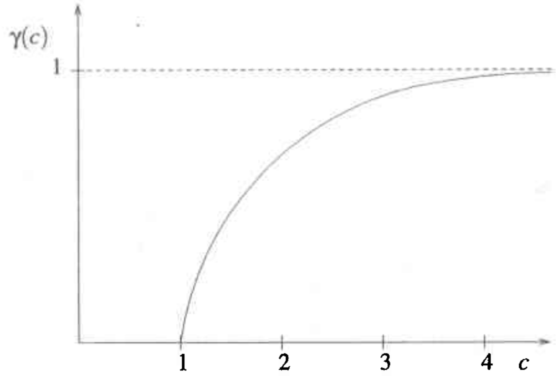

Figure 8: Fraction of nodes in the giant component.

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

1. (Stage 0:) Set $A_0 = \varnothing, B_0 = \{u\}$.
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 $A_i = A_i \cup \{v\}$, $B_{i+1} = B_i \cup \{\text{neighbours of } v \text{ in } G\}$.

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

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

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

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

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 $k_0 = \lceil 8a\varepsilon^{-2}\ln n \rceil$, $k_1 = \lceil \varepsilon^2 n/12 \rceil$.
Since presumably $k_0 \le k_1$, we may assume $\varepsilon^4 \ge \frac{96a\ln n}{n} \ge \frac{1}{n}$.
We may now estimate

$$p_k \le \Pr(|B_i| = k) \le \frac{n^k}{k!} e^{-\frac{k^2}{2n}} (kp)^k (1-p)^{k(n-k-1)}, \tag{6}$$

because

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

$$(1-p)^k \ge 1 - kp.$$

Applying Stirling's formula

$$\sqrt{2\pi k}\left(\frac{k}{e}\right)^k \le k! \le e^{\frac{1}{12k}} \sqrt{2\pi k}\left(\frac{k}{e}\right)^k$$

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

$$\begin{aligned} p_k &\le \exp\left(\frac{-k^2}{2n} - \frac{\varepsilon^3 k}{3} + \frac{k^2(1+\varepsilon)}{n}\right) \\ &\le \exp\left(\frac{-\varepsilon^2 k}{3} + \frac{k^2}{n}\right) \\ &\le \exp\left(\frac{-\varepsilon^2 k}{4}\right), \end{aligned}$$

and consequently

$$\begin{aligned} \sum_{k=k_0}^{k_1} p_k &\le \sum_{k=k_0}^{k_1} e^{-\varepsilon^2 k/4} \le e^{-\varepsilon^2 k_0/4} \cdot (1 - e^{-\varepsilon^2/4})^{-1} \\ &\le \frac{5}{\varepsilon^2} \cdot e^{-\varepsilon^2 k_0/4} \le 5\sqrt{n} \cdot n^{-2a} \\ &= 5n^{-2a+1/2} < n^{-a-1}. \end{aligned}$$

for large n. □