Random graphs

Emmanuel Abbé

1 Erdős-Rényi model

Notation: for a positive integer n, we define $[n] := \{1, 2, \ldots, n\}$.

Definition 1. *A graph g is defined by a pair $(V(g), E(g))$, where $V(g) = [n]$ for a positive integer n, and $E(g) \subseteq \binom{V(g)}{2}$, where $\binom{V(g)}{2} = \big\{(i,j) : i, j \in V(g), i < j\big\}$.*

An n-graph is a graph on the vertex set $V = [n]$.

Definition 2 (Erdős-Rényi model)**.** *We define the following two variants of the Erdős-Rényi model for random graphs:*

1. *Uniform model: For $m \in \{0, \ldots, \binom{n}{2}\}$, define $\mathbb{G}(n,m)$ as the random graph ensemble such that when $G \sim \mathbb{G}(n,m)$,*

$$\mathbb{P}(G = g) = \begin{cases} 0 & \text{if } |E(g)| \neq m \\ \frac{1}{\binom{\binom{n}{2}}{m}} & \text{otherwise.} \end{cases}$$

   I.e., the graph is drawn uniformly at random among all graphs having exactly m edges.

2. *Binomial model: For $p \in [0,1]$, define $\mathbb{G}(n,p)$ as the random graph ensemble such that when $G \sim \mathbb{G}(n,p)$,*

$$\mathbb{P}(G = g) = p^{|E(g)|}(1-p)^{\binom{n}{2} - |E(g)|}.$$

   I.e., the edges in the graphs are drawn i.i.d. with probability p.

Remark 1. In the Binomial model with $G \sim \mathbb{G}(n,p)$, we have:

$$\mathbb{E}(|E(G)|) = \binom{n}{2} p$$

and

$$\mathbb{V}(|E(G)|) = \binom{n}{2} p(1-p)$$

Remark 2. For the regimes and purpose of this note, the uniform and binomial models are essentially equivalent (when taking $m = \lfloor \binom{n}{2} p \rfloor$), due to standard concentration arguments. We will mainly use the Binomial model.

2 Graph properties

Definition 3. *A graph property is a collection of graphs on the same vertex set.*

For example, the property that "the graph contains a triangle" is the union of all graphs on a given vertex set that contain a triangle.

Definition 4. *For two graphs g and g',*

$$g \subseteq g' \iff V(g) = V(g'), E(g) \subseteq E(g').$$

Definition 5. *For a graph property A:*

- *A is symmetric if it does not depend on the labelling of the vertices, i.e., if a graph has the property and if the vertices are relabelled with a one-to-one map, then the new graph also has the property.*
- *A is increasing if*
$$g \in A, g \subseteq g' \Rightarrow g' \in A.$$
- *A is decreasing if*
$$g' \in A, g \subseteq g' \Rightarrow g \in A.$$
- *A is monotone if it is either increasing or decreasing.*

Example 1.

- *"containing an edge between vertex 1 and 2" $\rightarrow$ not symmetric,*
- *"containing a triangle" $\rightarrow$ symmetric & increasing,*
- *"being connected" $\rightarrow$ symmetric & increasing,*
- *"containing an isolated vertex" $\rightarrow$ symmetric & decreasing.*

Remark 3. From now on, we assume that a graph property is by default symmetric.

Note that a graph g with $|V(g)| = [n]$ can be identified as a boolean vector of length $\binom{n}{2}$, where each component encodes the presence or absence of an edge on a given pair of vertices. In turn, a graph property A can be represented by a boolean indicator function:

$$\mathbb{1}_A : \{0,1\}^{\binom{n}{2}} \to \{0,1\} \tag{1}$$

$$x \mapsto \mathbb{1}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Definition 6. *A Boolean function $f : \{0,1\}^N \to \{0,1\}$ is increasing if*

$$x \leq y \implies f(x) \leq f(y)$$

where $x \leq y$ if $x_i \leq y_i$ for all $i \in [N]$.

Fact 1. *A is an increasing graph property if and only if $\mathbb{1}_A$ is increasing.*

Definition 7. *For $p \in [0,1]$, let μ_p denote the probability measure on the set of n-graphs resulting form the ensemble $\mathbb{G}(n,p)$, i.e.,*

$$\mu_p(g) = p^{|E(g)|}(1-p)^{\binom{n}{2}-|E(g)|}.$$

Lemma 1. *If A is an increasing property, then $[0,1] \ni p \mapsto \mu_p(A)$ is increasing and continuous.*

Proof. To show continuity, let $G \sim \mathbb{G}(n,p)$, and note that

$$\begin{aligned}\mu_p(A) &= \sum_{g \in A} \mu_p(g) \\ &= \sum_{m=0}^{\binom{n}{2}} \sum_{g \in A : |E(g)|=m} p^m (1-p)^{\binom{n}{2}-m} \\ &= \sum_{m=0}^{\binom{n}{2}} |A_m| p^m (1-p)^{\binom{n}{2}-m}\end{aligned}$$

where $|A_m|$ is the number of graphs having property A with $|E(g)| = m$. The above is a polynomial in p which implies continuity.

Now, let $G \sim \mathbb{G}(n,p)$ and $\tilde{G} \sim \mathbb{G}(n,q)$ where $q > p$. Let $G' \sim \mathbb{G}(n,p')$ with $p' = \frac{q-p}{1-p}$ such that

$$G \cup G' \sim \mathbb{G}(n, p + p' - pp') = \mathbb{G}(n,q).$$

Then, by the increasing assumption,

$$\mu_q(A) = \mathbb{P}(\tilde{G} \in A) = \mathbb{P}(G \cup G' \in A) \geq \mathbb{P}(G \in A) = \mu_p(A).$$

□

This technique is an instance of “two round exposure” or “coupling”.

3 Thresholds

We denote by $\{A_n\}_{n\geq 1}$ a sequence of graph properties on $V = [n]$.

Definition 8. *$\{\hat{p}_n\}_{n\geq 1}$ is a threshold for the sequence of increasing graph properties $\{A_n\}_{n\geq 1}$ if*

- *$\mu_{p_n}(A_n) \to 0$ when $p_n \ll \hat{p}_n$,*
- *$\mu_{p_n}(A_n) \to 1$ when $\hat{p}_n \ll p_n$.*

Flip the limit in the definition for decreasing properties. Recall the notations:

- $a_n \ll b_n \Leftrightarrow a_n = o(b_n)$
- $a_n \asymp b_n \Leftrightarrow a_n = \Theta(b_n)$
- $a_n \sim b_n$ if $a_n/b_n \to 1$.

Remark 4. A threshold, if it exists, is unique up to constants. We will still often talk about "the" threshold. We often talk about a threshold for a graph property A_n without specifying the "sequence", and even drop the subscripts when adequate, e.g., saying that $\hat{p}$ is a threshold.

Definition 9. *Let $\varepsilon \in (0,1)$ and A_n be monotone graph property, then from Lemma 1 there exists a unique $p_n(\varepsilon)$ such that $\mu_{p_n(\varepsilon)}(A_n) = \varepsilon$.*

Theorem 1 (Bollobás-Thomason)**.** *If A_n is a monotone property, then it has a threshold. In particular, $p_n(1/2)$ is a threshold.*

To prove this theorem we need the following lemma.

Lemma 2. *$\hat{p}$ is a threshold for a monotone property if and only if for all $\epsilon \in (0,1)$, $\hat{p} = \Theta(p(\epsilon))$.*

Proof. Assume without loss of generality that the property is increasing.

Suppose $\hat{p}$ is a threshold, so assume that, for some ϵ, we have $\hat{p} \neq \Theta(p(\epsilon))$. Then, after passing possibly to a subsequence, we either have $\hat{p}_{n_k}/p(\epsilon) \to 0$ or $\hat{p}_{n_k}/p(\epsilon) \to \infty$.

In the first case, we have $\hat{p} \ll p(\epsilon)$, hence $\mu_{p_{n_k}(\epsilon)}(A_{n_k}) \to 1$. In the second case, we have $\hat{p} \gg p(\epsilon)$, hence $\mu_{p_{n_k}(\epsilon)}(A_{n_k}) \to 0$. In either case, this contradicts $\mu_{p(\epsilon)}(A_n) = \epsilon$.

Conversely, let $p = p_n$ such that $p \ll \hat{p} = \Theta(p(\epsilon))$. We claim that $\mu_p(A_n) \to 0$. Suppose instead that $\mu_p(A_n) \geq a = \mu_{p(a)}(A_n) > 0$ on a subsequence, hence $p \geq p(a)$ on that subsequence, and by assumption $p(a) \ll p(\epsilon)$. Choosing $\epsilon < a$ yields a contradiction. The other case is done similarly. □

Proof of Theorem 1. Assume without loss of generality that the property is increasing. We show that $p(1/2) = \Theta(p(\epsilon))$ for all $\epsilon \in (0,1)$. Let $\epsilon \in (0,1)$ and $m > \frac{\log \epsilon}{\log(1-\epsilon)}$.

Now, take $\{G_1, G_2, \ldots, G_m\} \overset{iid}{\sim} \mathbb{G}(n, p(\epsilon))$. Define $\tilde{G} = \bigcup_{1 \leq i \leq m} G_i$, and note that $\tilde{G} \sim \mathbb{G}(n, p')$ where $p' = 1 - (1 - p(\epsilon))^m \leq mp(\epsilon)$. Define $G' \sim \mathbb{G}(n, mp(\epsilon))$. We have by the increasing property that

$$\mathbb{P}\left(\cup_{1 \leq i \leq m} G_i \in A_n\right) \leq \mathbb{P}(G' \in A_n) \tag{3}$$

and

$$\mathbb{P}\left(\forall i, G_i \notin A_n\right) = \mathbb{P}(G_1 \notin A_n)^m = (1-\epsilon)^m. \tag{4}$$

Choose m sufficiently large such that $(1-\epsilon)^m < \epsilon$, say $m \geq \frac{\log \epsilon}{\log(1-\epsilon)}$. Thus,

$$1 - \epsilon \leq \mathbb{P}(\cup_{i=1}^m G_i \in A_n). \tag{5}$$

Thus, we have $p(1-\epsilon) \leq m(p(\epsilon))$. Assume $\epsilon < 1/2$, then by monotonicity we have $p(\epsilon) \geq p(1-\epsilon)/m \geq p(1/2)/m$. On the other hand, $p(\epsilon) \leq p(1/2)$ by monotonicity. Since m depends only on ϵ, it holds that $p(\epsilon) = \Theta(p(1/2))$. The case $\epsilon > 1/2$ is similar. □

4 Sharp thresholds

Definition 10 (Threshold width)**.** *For a monotone property A_n and $\epsilon \in (0,1)$, define $\delta_n(\epsilon) = p_n(1-\epsilon) - p_n(\epsilon)$, where p_n is defined in Definition 9.*

Remark 5. For any threshold $\hat{p}$, we have $\delta(\epsilon) = O(\hat{p})$ by Theorem 1. However, its possible that $\delta(\epsilon)/p(\epsilon) \to 0$.

Definition 11.

- *$\hat{p}$ is a coarse threshold if there exists an $\epsilon \in (0,1)$ such that $\delta(\epsilon) = \Theta(\hat{p})$,*
- *$\hat{p}$ is a sharp threshold if for all $\epsilon \in (0,1)$, $\delta(\epsilon) = o(\hat{p})$.*

Remark 6. The following statements are equivalent:

- $\hat{p}$ is sharp
- $p(1-\epsilon)/p(\epsilon) \to 1$ for all $\epsilon \in (0,1)$
- the threshold is unique up to $\sim$
- For all $\delta > 0$, $\mu_p(A_n) \to \begin{cases} 0 & \text{if } p \leq \hat{p}(1-\delta) \\ 1 & \text{if } p \geq \hat{p}(1+\delta) \end{cases}$

What can we say about the sharpness of general monotone property thresholds? Friedgut-Kalai first proved that $\delta(\epsilon) = O(\frac{\log(1/\epsilon)}{\log(n)})$, later improved by Bourgain-Kalai to $\delta(\epsilon) = O(\frac{1}{\log^{2-\nu}(n)})$, $\nu > 0$. An exponent of 2 is conjectured to hold. Note that these bounds are interesting only for large thresholds, i.e., larger than the bounds themselves (since we already know that $\delta(\epsilon) = O(\hat{p})$). Friedgut's work also shows that a monotone property is coarse if it cannot be approximated by a "local property" (in a sense made formal), such as subgraph containment as discussed next.

5 Subgraph containment

We use the following notation throughout the section:

- $X \sim \mathbb{R}_+$ denotes a r.v. X that takes values in $\mathbb{R}_+$.
  Similarly, $X \sim \mathbb{Z}_+$ denotes a r.v. X that takes values in $\mathbb{Z}_+$.
- $\mathrm{Poi}(c)$ denotes a Poisson distribution with mean c.
- $X_n \xrightarrow{(d)} X$ means that X_n tends in distribution to X.

5.1 Single edge

Example 2 (Emptiness)**.** *As an example, consider the emptiness property, which is that the graph has no edges. Let M be the number of edges in a random graph $G \sim \mathbb{G}(n, p)$ and $N = \binom{n}{2}$. We have*

$$M = \sum_{i<j} \mathbb{1}((i,j) \in E(G)) \sim Binom(N, p) \tag{6}$$

We have

$$\mathbb{P}(M = 0) = (1-p)^N \tag{7}$$

If $p_c = c/N$, then, we have $\mathbb{P}(M = 0) \sim e^{-c}$. One can show that $p_c = \Theta(1/n^2)$ is a threshold for all $c > 0$, hence a coarse threshold. Further, $Binom(\binom{n}{2}, p_c)$ tends in distribution to $Poisson(c)$. We will see that this is typical of properties which can be expressed by containing a finite subgraph.

5.2 First and Second Moment Method

This is a typical strategy used in proofs regarding thresholds. A few results:

Theorem 2. (Markov Inequality) *If X takes values in $\mathbb{R}_+$, then for any $a \in \mathbb{R}_+ \setminus \{0\}$,*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Proof.

$$\mathbb{E}[X] \geq \mathbb{E}[X\mathbb{1}(X \geq a)] \geq \mathbb{E}[a\mathbb{1}(X \geq a)] = a\mathbb{P}(X \geq a).$$

$\square$

Lemma 3. *Let $Z \sim \mathbb{R}_+$. Then, the following bounds hold:*

(a) $\mathbb{P}(Z \geq 1) \leq \mathbb{E}(Z)$

(b) $\mathbb{P}(Z = 0) \leq \frac{\mathrm{Var}(Z)}{\mathbb{E}(Z^2)} \leq \frac{\mathrm{Var}(Z)}{[\mathbb{E}(Z)]^2}$

Proof. (a) Follows directly from the Markov Inequality with $a = 1$.

(b) One can proceed with Markov to get the weaker bound: $\mathbb{P}(Z = 0) = \mathbb{P}(Z - \mathbb{E}(Z) = -\mathbb{E}(Z)) \leq \mathbb{P}(|Z - \mathbb{E}(Z)| \geq \mathbb{E}(Z)) \leq \frac{\mathrm{Var}(Z)}{[\mathbb{E}(Z)]^2}$.

One can proceed Cauchy-Schwartz to get the stronger bound: $\mathbb{E}Z = \mathbb{E}[Z\mathbb{1}(Z > 0)] \leq (\mathbb{E}Z^2)^{1/2}(\mathbb{P}(Z > 0))^{1/2}$.

$\square$

Note that both bounds are equivalent and vanish if one shows that $\mathbb{E}Z^2 \sim (\mathbb{E}Z)^2$.

5.3 Triangles

We can write the triangle containment property as:

$$A_n = \{\exists\ i < j < k \in [n] \text{ such that } E_{ij} = E_{jk} = E_{ik} = 1\}$$

A_n is an increasing property, so there exists a threshold.

Let $Z_n = \#$ of triangles in $G \sim \mathbb{G}(n, p)$. Let T denote a triplet of vertices. Denote by $\mathbb{1}_\triangle(T) := \mathbb{1}\{G \text{ contains a triangle with vertices } T\}$. We have:

$$Z_n = \sum_{T \in \binom{[n]}{3}} \mathbb{1}_\triangle(T).$$

Since the expected value function is additive, $\mathbb{E}(Z_n) = \binom{n}{3} \cdot p^3$. Hence, if $p \ll \frac{1}{n}$, we have $\mathbb{E}(Z_n) \to 0$ and by Lemma 3(a), $\mathbb{P}(Z_n \geq 1) \to 0$. So we have a lower bound for our threshold.

To show that the upper bound is also $\frac{1}{n}$, we use Lemma 3(b):

$$\mathrm{Var}(Z_n) = \sum_{S,T \in \binom{[n]}{3}} \mathrm{cov}(\mathbb{1}_\triangle(S), \mathbb{1}_\triangle(T)).$$

We must analyze three cases:

- $|S \cap T| \leq 1$, i.e. S and T share at most 1 vertex. Then, $\mathrm{cov}(\mathbb{1}_\triangle(S), \mathbb{1}_\triangle(T)) = 0$.
- $|S \cap T| = 2$. We have two triangles sharing an edge. There are $\leq c\binom{n}{4}$ ways to choose S and T, where c is a constant. Then, $\mathrm{cov}(\mathbb{1}_\triangle(S), \mathbb{1}_\triangle(T)) \leq p^5$.
- $|S \cap T| = 3$. There are $\binom{n}{3}$ such cases. Then, $\mathrm{cov}(\mathbb{1}_\triangle(S), \mathbb{1}_\triangle(T)) \leq p^3$.

Combining these cases, we have $\mathrm{Var}(Z_n) \leq c\binom{n}{4} \cdot p^5 + c'\binom{n}{3} \cdot p^3$. Recall that $\mathbb{E}(Z_n) = \binom{n}{3} \cdot p^3$. Hence, if $p \gg \frac{1}{n}$, we have $\frac{\mathrm{Var}(Z_n)}{[\mathbb{E}(Z_n)]^2} \to 0$. By Lemma 3(b), we have $\mathbb{P}(Z_n = 0) \to 0$.

Theorem 3. *The property* {triangle containment} *has a threshold at* $\frac{1}{n}$.

The message here is that for finite subgraphs, the first and second moment method seems to give the threshold. However, there are additional subtleties that we need to worry about, as discussed in Section 5.5. We first investigate what happens at the threshold.

5.4 Triangles at the threshold

We first introduce the moments method to estimate the limiting distribution of counts.

Recall: $X_n \xrightarrow{(d)} X$ means $\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)\ \forall x$ that are continuity points of the function $u \mapsto \mathbb{P}(X \leq u)$. If $X_n \sim \mathbb{Z}$, then $X_n \xrightarrow{(d)} X$ iff $\mathbb{P}(X_n = k) \to \mathbb{P}(X = k)\ \forall k \in \mathbb{Z}$.

Theorem 4. *Let X be a r.v. whose distribution is determined by its moments. If for all $k \geq 1$, $\mathbb{E}(X_n^k) \to \mathbb{E}(X^k)$ and all moments are finite, then $X_n \xrightarrow{(d)} X$.*

Above, $X^k = X \cdot X \cdot \ldots \cdot X$ (choosing with replacement). An equivalent version of the theorem (see below) concerns $(X)_k = X(X-1)\ldots[X-(k-1)]$ (choosing without replacement).

Theorem 5. *Let X be a r.v. whose distribution is determined by its moments. If for all $k \geq 1$, $\mathbb{E}(X_n)_k \to \mathbb{E}(X)_k$ and all moments are finite, then $X_n \xrightarrow{(d)} X$.*

Thm. 4 and 5 are equivalent since the polynomials X^k and $(X)_k$ generate each other.

Remark 7. If $Z \sim \text{Poi}(\lambda)$, then $\mathbb{E}(Z)_k = \lambda^k$.

Lemma 4. *Let I_α be indicator functions and A be a finite set. If $S = \sum_{\alpha \in A} I_\alpha$, then $S^k = \sum_{\alpha_1,\ldots,\alpha_k} I_{\alpha_1} \cdot \ldots \cdot I_{\alpha_k}$ and $(S)_k = \sum_{\alpha_1,\ldots,\alpha_k \text{ distinct and ordered}} I_{\alpha_1} \cdot \ldots \cdot I_{\alpha_k}$*

Proof. Assume the result to be true for $(S)_k$ (it trivially holds for $k = 0$). We have

$$\begin{aligned}
(S)_k(S-k) &= \sum_{\alpha_1,\ldots\alpha_k \text{ distinct and ordered}} I_{\alpha_1} \cdot \ldots \cdot I_{\alpha_k} \cdot (S-k) \\
&= \sum_{\alpha_1,\ldots\alpha_k \text{ d\&o and } \alpha_{k+1} \text{ free}} I_{\alpha_1} \cdot \ldots \cdot I_{\alpha_{k+1}} - k \sum_{\alpha_1,\ldots\alpha_k \text{ d\&o}} I_{\alpha_1} \cdot \ldots \cdot I_{\alpha_k}.
\end{aligned}$$

Since α_{k+1} is free, it may either be distinct from $\alpha_1, \ldots \alpha_k$, in which case it contributes to a $(k+1)$-tuple of the d&o index set, or α_{k+1} equals one of the $\alpha_1, \ldots \alpha_k$, with k possibilities for that, which are all cancelled out by the term $k \sum_{\alpha_1,\ldots\alpha_k \text{ d\&o}} I_{\alpha_1} \cdot \ldots \cdot I_{\alpha_k}$. □

Theorem 6. *Let $Z_n = \#$ of triangles in $G \sim \mathbb{G}(n, \frac{c}{n})$, then*

$$Z_n \sim \text{Poi}(\frac{c^3}{6}).$$

Proof. (Outline)

$$\begin{aligned}
\mathbb{E}(Z_n)_k &= \sum_{T_1,\ldots,T_k \text{ distinct and ordered triplets}} \mathbb{P}(E(T_1) \in G, \ldots, E(T_k) \in G) \\
&= E'_k + E''_k, \text{ where } E'_k \text{ is over disjoint triplets, and } E''_k \text{ is the rest of the sum} \\
E'_k &= \binom{n}{3} \cdot p^3 \cdot \binom{n-3}{3} \cdot p^3 \cdot \ldots \cdot \binom{n-3(k-1)}{3} \cdot p^3 \\
&\sim \left[\frac{(np)^3}{6}\right]^k = \left(\frac{c^3}{6}\right)^k, \text{ as desired for Poi} \left(\frac{c^3}{6}\right).
\end{aligned}$$

Note two triplets are distinct if they are not the exact same triplet, and they disjoint if they do not intersect. It remains to show that $E''_k = o(1)$, which takes place since the probability of overlaps vanishes fast enough. (Note the distinction between distinct triplets and disjoint triangles on these: one can have two distinct triplets that share one vertex, and one can have to disjoint triangles on these despite the common vertex). □

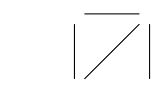
5.5 Balanced subgraphs

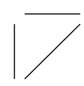
Consider a finite graph H. We attempt to find the threshold for the property of containing H. Let v_H, e_H be the number of vertices and the number of edges in H, respectively. Let $Z_n = \#$ copies of H in $\mathbb{G}(n,p)$. Below c_H is a constant corresponding to the number of automorphism of H (see below).

$$\mathbb{E}(Z_n) = \binom{n}{v_H} \cdot p^{e_H} \cdot c_H \asymp n^{v_H} \cdot p^{e_H}$$

We can see above that indeed, if $p \ll (\frac{1}{n})^{v_H/e_H}$, then $\mathbb{E}(Z) \to 0$. Hence $p = (\frac{1}{n})^{v_H/e_H}$ is a lower bound for the threshold.

Consider the following example, with graph H to the left and $G \subseteq H$ to the right.





In this example, if $p \ll n^{-\frac{5}{6}}$, we do not contain H as a subgraph with high probability.
Question: If $p \gg n^{-\frac{5}{6}}$, do we contain H with high probability?
Answer: No. Choose $n^{-\frac{5}{6}} \ll p \ll n^{-\frac{4}{5}}$. (Ex: $p = n^{-\frac{9}{11}}$), then $\mathbb{E}(Z_n) \to \infty$. However, consider $G \subset H$ pictured to the right of H. Let $\tilde{Z_n} = \#$ of copies of G. Then, $\mathbb{E}(\tilde{Z_n}) \asymp n^4 p^5 \to 0$ by our choice of p. There is no G w.h.p., so we cannot have H w.h.p.!
The catch: $G \subset H$, but G is *denser*.

Definition 12. *For a graph H, define $m(H) = \max\{\frac{|E(H')|}{|V(H')|}\}$ over all $H' \neq \emptyset$ contained in H (ignoring isolated vertices). A graph H is balanced if $m(H) = \frac{|E(H)|}{|V(H)|}$ and strictly balanced if only H achieves this ratio. Let $\mathrm{Aut}(H)$ be the number of automorphisms of H, i.e., the number of ways to relabel the vertices of the graph without changing the edge set.*

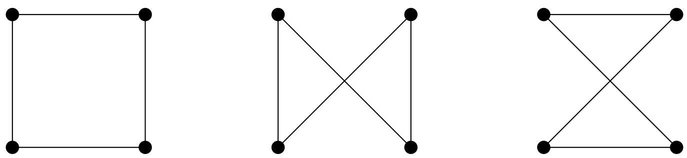
Theorem 7. *The property of containing a finite graph H in $G(n,p)$ has a threshold at $p = n^{-1/m(H)}$. Further, if $p = cn^{-1/m(H)}$ for some $c > 0$ and H is strictly balanced and connected, then $Z_H \to Poi(c^{E(H)}/\mathrm{Aut}(H))$ in distribution.*

The proof of this theorem is similar to the triangle case, taking into account that

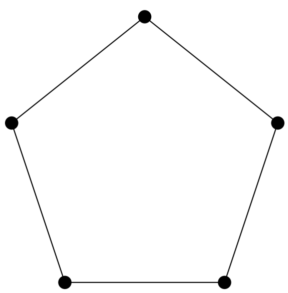
1. $m(H)$ matters because in order to contain a copy of H, the graph must contain a copy of its densest subgraph. Also, if H contains a subgraph that is denser than it, there is an elevated chance of two copies of H overlapping in order to share that subgraph.

2. Assuming that H is connected and contains no subgraph denser than itself, we have that
$$E_k \sim \left(C(H)\binom{n}{V(H)}p^{E(H)}\right)^k \sim \left(\frac{n^{V(H)}p^{E(H)}}{Aut(H)}\right)^k$$
allowing us to establish the Poisson law.

Above $C(H) = V(H)!/Aut(H)$ is the number of different ways to place the graph H on the $V(H)$ selected vertices. Recall that $A : V(H) \to V(H)$ is an automorphism if $(i, j) \in E(H)$ implies that $(A(i), A(j)) \in E(H)$. For example, for a square $H = C_4$, we have $C(H) = 3$ different ways to draw the square (that are not isomorphic), and for each way, there are $Aut(H) = 8$ equivalent labelings of the vertices (starting with the label 1 in each of the 4 possible vertices and assigning the other sequence of labels on the right or on left side of 1).



More generally, for any $k > 2$, $m(C_k) = 1$ and $Aut(C_k) = 2k$.



Containing a finite subgraph has thus always a coarse threshold that is of the form n^{-d} where $d \in \mathbb{Q}$. Friedgut 1999 shows that a type of "converse" statement also holds; if the threshold is coarse, then it must be local in some sense and it must be a rational power of n.

6 Connectivity

A graph is connected if there is a path between any pair of vertices.

Theorem 8. *Let $G \sim \mathbb{G}(n, p)$ with $p = \frac{c \log(n)}{n}$,*

- *If $c < 1$ then G is not connected w.h.p.*
- *If $c > 1$ then G is connected w.h.p.*

i.e., $\frac{\log(n)}{n}$ is a sharp threshold for connectivity

We will first prove a weaker statement which will imply the first part of the above theorem.

Lemma 5. *Let $G \sim \mathbb{G}(n, p)$ with $p = \frac{c \log(n)}{n}$,*

- *If $c < 1$ then G has an isolated vertex w.h.p.,*

- *If $c > 1$ then G does not have an isolated vertex w.h.p.,*

i.e., $\log(n)/n$ *is a sharp threshold for containing an isolated vertex.*

Proof. We use the first and second moment method. Let

$$Z_{iso} = \sum_{v \in [n]} \mathbb{1}(v \text{ is iso})$$

i.e., Z_{iso} counts the number of isolated vertices. Note that the probability that a given vertex is isolated is

$$E[\mathbb{1}(v \text{ is iso})] = (1-p)^{n-1}$$

hence

$$E[Z_{iso}] = n(1-p)^{n-1}$$

and

$$\begin{aligned} \mathbb{P}(Z_{iso} \geq 1) &\leq E[Z_{iso}] \\ &\leq n(1-p)^{n-1} \\ &\sim ne^{-c\log(n)} \\ &= n \cdot n^{-c} \end{aligned}$$

If $c > 1$ we then have that $\mathbb{P}(Z_{iso} \geq 1) \to 0$ proving the second part of the lemma. For the first part,

$$\mathbb{P}(Z_{iso} = 0) \leq \frac{\mathrm{Var}(Z_{iso})}{E[Z_{iso}]^2}$$

Note that

$$\begin{aligned} \mathrm{Var}(Z_{iso}) &= \mathrm{Var}\sum_v \mathbb{1}(v \text{ is iso})) \\ &= \sum_{v,w \in [n]} \mathrm{Covar}(\mathbb{1}(v \text{ is iso}), \mathbb{1}(w \text{ is iso})) \\ &= \sum_v \mathrm{Var}\mathbb{1}(v \text{ is iso})) + \sum_{v \neq w} \mathrm{Covar}(\mathbb{1}(v \text{ is iso}), \mathbb{1}(w \text{ is iso})) \end{aligned}$$

The following two equalities follow from elementary calculations

$$\begin{aligned} \sum_v \mathrm{Var}\mathbb{1}(v \text{ is iso})) = n((1-p)^{n-1} - (1-p)^{2n-2}) &= n(1-p)^{n-1}(1-(1-p)^{n-1}) \\ &= n(1-p)^{n-1}(1-o(1)) = o(E[Z_{iso}]^2) \end{aligned}$$

$$\begin{aligned} \sum_{v \neq w \in [n]} \mathrm{Covar}(\mathbb{1}(v \text{ is iso}), \mathbb{1}(w \text{ is iso})) &= n(n-1)((1-p)^{n-1}(1-p)^{n-2} - (1-p)^{2n-2}) \\ &= n(n-1)(1-p)^{2n-3}(1-(1-p)) \\ &\sim (n(1-p)^{n-1})^2 p \\ &= o(E[Z_{iso}]^2) \end{aligned}$$

Hence
$$\mathbb{P}(Z_{iso} = 0) = o(1)$$
if $c < 1$. □

Note that Lemma 5 directly implies the first part of Theorem 8. To prove the second part we bound the probability of any subset of size up to $n/2$ to be disconnected. Note that for a fixed subset of size k, the probability that it is disconnected is $(1-p)^{k(n-k)}$. Therefore a union bound gives

$$\mathbb{P}(\exists S : |S| = k, S \text{ is isolated}) \leq \binom{n}{k}(1-p)^{k(n-k)}$$

and again by a union bound

$$\mathbb{P}(\exists S : |S| \leq n/2, S \text{ is isolated}) \leq \sum_{k=1}^{n/2} \binom{n}{k}(1-p)^{k(n-k)}.$$

It is not hard to verify that the above summation goes to 0 if $c > 1$. The first term at $k = 1$ is
$$n(1-p)^n = n \cdot n^{-c} \to 0$$
and the last term is

$$\leq 2^n(1-p)^{n^2/4} = 2^n(1 - \frac{c\log(n)}{n})^{n^2/4} \sim 2^n e^{-\frac{cn\log(n)}{4}}$$

which falls exponentially fast.

7 Giant Component

A giant component in a graph drawn from $G(n,p)$ is a connected component of linear size (in n).

Definition 13. *For a given n and p, let $LCC(G(n,p))$ denote the size of the largest connected component of a graph drawn from $G(n,p)$ and $2LCC(G(n,p))$ denote the size of its second largest connected component.*

The above does not rule out having multiple largest components.

Theorem 9. *Let $G \sim G(n,p)$ with $p = \frac{c}{n}$,*

1. *If $c < 1$ then $LCC(G) \leq \frac{2.1}{(1-c)^2}\log n$ w.h.p.*
2. *If $c > 1$, then*
   (a) *$LCC(G) \geq (1+o(1))\beta n$ w.h.p., where β is the unique solution in $(0,1)$ to $e^{-\beta c} = 1 - \beta$,*
   (b) *$2LCC(G) \leq \frac{16c}{(1-c)^2}\log n$ w.h.p..*

In particular, $1/n$ is a sharp threshold for containing a giant.

Remark 8. The 2.1 can be made $2+\epsilon$ for any $\varepsilon > 0$. In 2(a), the largest component is indeed unique.

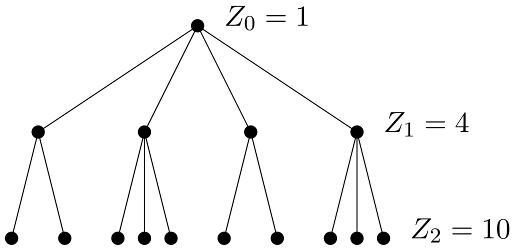
A quick intuition for previous theorem is that the number of neighbors of a vertex has distribution $\mathrm{Bin}(n-1, \frac{c}{n})$, whose expectation tends to c. Thus depending on whether c is lesser or greater than 1, neighborhoods tend to either expand or die out. This can be described more precisely with Galton-Watson trees.

7.1 Galton-Watson process

Let μ be a probability distribution on $\mathbb{Z}_+$.

Definition 14. *$GW(\mu)$ is the random process $\{Z_k\}_{k\geq 0}$ defined by:*

- $Z_0 = 1$,
- $Z_k = \sum_{i=1}^{Z_{k-1}} X_i^{(k)}$, *where $X_i^{(k)}$ are i.i.d. under μ, $k \geq 1$.*



Remark 9. For any k, $E[Z_k] = E_{Z_{k-1}}E[Z_k|Z_{k-1}] = E(E(\mu)Z_{k-1}) = E(\mu)^k$. Thus the expected number of descendants at generation k is vanishing if $\mathbb{E}(\mu) < 1$ and diverging if $\mathbb{E}(\mu) > 1$.

Definition 15. *Let $Z = \sum_{i\geq 0} Z_i$ (total number of descendants) and $z_\mu = \mathbb{P}(Z < \infty)$ (the extinction probability).*

Note that for any k, if $Z_k = 0$, then $Z_n = 0$ for all $n \geq k$. Thus the events $\{Z_k = 0\}$ are increasing and

$$z_\mu = \mathbb{P}(Z < \infty) = \mathbb{P}(\exists k \geq 1 : Z_k = 0) = \lim_{k\to\infty} \mathbb{P}(Z_k = 0). \tag{8}$$

Since $\mathbb{P}(Z_k = 0)$ is increasing in k, the limit exists in $(0, 1)$.

Note that if $\mu(0) = 0$, then $z_\mu = 0$.

Theorem 10. *Assume $\mu(0) > 0$ and $\mathbb{E}[\mu] < \infty$, $\mathrm{Var}(\mu) < \infty$.*

1. *If $\mathbb{E}(\mu) \leq 1$, then $z_\mu = 1$.*

2. *If* $\mathbb{E}(\mu) > 1$*, then* z_μ *is the unique solution in* $(0,1)$ *to* $\phi_\mu(z) = z$*, where* $\phi_\mu(z) = \mathbb{E}[z^X]$*,* $X \sim \mu$ *(probability generating function).*

Remark 10 (Properties of the Generating Function). $\phi_\mu(z)$ has the following properties:

(A) $\phi_\mu(z)$ is strictly increasing and convex for $0 \le z \le 1$;
(B) $\phi_\mu(z)$ is continuous for $0 \le z \le 1$;
(C) $\phi_\mu(1) = 1$;
(D) $\mathbb{E}(\mu) = \phi'_\mu(1)$.

Proof. 1. We have

$$\mathbb{P}(Z_k \ge 1) \le \mathbb{E}[Z_k] = \mathbb{E}(\mu)^k. \tag{9}$$

Therefore, if $\mathbb{E}(\mu) < 1$, $\lim_{k\to\infty} \mathbb{P}(Z_k \ge 1) = 0$ and $\lim_{k\to\infty} \mathbb{P}(Z_k = 0) = z_\mu = 1$. The case $\mathbb{E}(\mu) = 1$ is covered by the next part.

2. Denote $\phi_{Z_k}(z) = \mathbb{E}[z^{Z_k}]$ and note that $\mathbb{P}(Z_k = 0) = \phi_{Z_k}(0)$. We have

$$\mathbb{E}[z^{Z_k}] = \mathbb{E}\mathbb{E}[z^{Z_k}|Z_{k-1}] \tag{10}$$
$$\phi_{Z_k}(z) = \phi_{Z_{k-1}}(\phi_\mu(z)) = (\phi_\mu \circ \ldots \circ \phi_\mu)(z) = \phi_\mu(\phi_{Z_{k-1}}(z)). \tag{11}$$

Thus, using (8) and the continuity of ϕ at 0, we obtain

$$\phi_\mu(z_\mu) = \phi_\mu(\lim_{k\to\infty} \phi_{Z_k}(0)) \tag{12}$$
$$= \lim_{k\to\infty} \phi_\mu(\phi_{Z_k}(0)) \tag{13}$$
$$= \lim_{k\to\infty} \phi_{Z_{k+1}}(0) = z_\mu. \tag{14}$$

Notice that 1 is always a fixed point of ϕ_μ. From the properties of ϕ_μ it follows that if $\mathbb{E}(\mu) > 1$, there exists a unique fixed point in $(0,1)$ (see Figure 1).

Let $z_0 > 0$ be a fixed point of ϕ_μ. Since ϕ_{Z_k} is increasing, we have

$$\phi_{Z_k}(0) \le \phi_{Z_k}(z_0) = z_0. \tag{15}$$

Taking the limit for $k \to \infty$, we obtain $z_\mu \le z_0$, hence z_μ is the smallest fixed point of ϕ_μ.

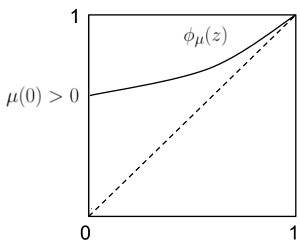
If $\mathbb{E}(\mu) = 1$, there exists a unique fixed point at $z = 1$, which clears the case left open in part 1.

□

Remark 11. In the case of a Poisson offspring, we have

$$\phi_{Poi(c)}(z) = e^{c(z-1)} \tag{16}$$

Galton-Watson trees play an important role for the ER model with bounded degree, since the close neighborhood of a vertex in the graph can be approximated by a Galton-Watson tree of offspring $Poi(c)$. This in fact gives some insight behind the giant component threshold: if $c < 1$, the neighborhoods tend to die out, whereas if $c > 1$, the neighborhoods



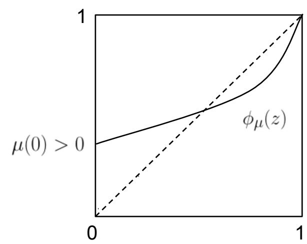


Figure 1: On the left: in the subcritical ($\mathbb{E}(\mu) < 1$) and critical ($\mathbb{E}(\mu) = 1$) regime the smallest fixed point is at $z = 1$. On the right: in the supercritical ($\mathbb{E}(\mu) > 1$) regime there is a unique fixed point in $(0, 1)$.

expand with probability $z_{Poi(c)}$. Thus a handwaving argument suggests that each vertex reaches in expectation roughly $(1 - z_{Poi(c)})n + o(n)$ vertices, which gives the size of the giant component since $\phi_{Poi(c)}(z) = e^{c(z-1)}$. Showing that the giant appears when $c > 1$ based on this intuition requires a bit more work, and we refer to the book of Janson et al. on Random Graphs. We next prove part 1 of the Theorem, revising first the Chernoff bound.

7.2 Chernoff bound

Recall that Markov's inequality for a positive random variable X

$$\mathbb{P}(X \geq a) \leq \frac{E[X]}{a}$$

can be used for $t \geq 0$ as

$$\begin{aligned}
\mathbb{P}(X \geq E[X] + \Delta) &= \mathbb{P}(e^{tX} \geq e^{t(E[X]+\Delta)}) \\
&\leq \frac{E[e^{tX}]}{e^{t(E[X]+\Delta)}}
\end{aligned}$$

In the case when X is drawn from $\mathrm{Bin}(n, p)$, we have that

$$E[e^{tX}] = (1 - p + pe^t)^n$$

therefore

$$\mathbb{P}(X \geq E[X] + \Delta) \leq \frac{(1 - p + pe^t)^n}{e^{t(np+\Delta)}}$$

Since the above expression is for any t we can optimize over t, choosing

$$e^t = \frac{(\Delta + np)(1 - p)}{p(n - np - \Delta)}$$

and thus obtaining

$$\mathbb{P}(X \geq E[X] + \Delta) \leq \exp(-\frac{\Delta^2}{2(E[X] + \frac{\Delta}{3})}) \leq \exp(-\frac{\Delta^2}{2E[X] + \Delta}). \tag{17}$$

7.3 Proof of Theorem 9 part 1

Take a vertex v and let C_v be its connected component. We want to upper bound the probability of $|C_v| \geq k$. To bound this we consider the process in which we grow a tree depth by depth from v and at each intermediate leaf i we assume that the number of descendants is an independent Binomial variable $X_i \sim \text{Bin}(n, c/n)$, i.e., a Galton-Watson tree with offspring distribution $\text{Bin}(n, c/n)$. Note that this model over counts the vertices in C_v because there may be common descendants (i.e., loops) and because the number of possible descendants should normally reduce with the depth. However we are over-counting and obtain the following upperbound

$$\mathbb{P}(|C_v| \geq k) = \mathbb{P}(\text{number of descendent of } v \geq k-1) \leq \mathbb{P}(\sum_{i=1}^{k} X_i \geq k-1)$$

where the 'descendants' are simply the vertices connected to v, which can be seen as descendants in a depth first search from v. Note that $X = \sum_{i=1}^{k} X_i$ is distributed as $X \sim \text{Bin}(nk, c/n)$. Therefore we want to bound

$$\mathbb{P}(X \geq k-1) = \mathbb{P}(X \geq kc + (k-1-kc))$$

and applying the Chernoff bound with $\Delta = k - 1 - kc$ and $E(X) = kc$, we obtain (after simplifications using the stronger bound in (17))

$$\mathbb{P}(|C_v| \geq k) \leq \exp(-\frac{(1-c)^2 k}{2})$$

Now if $k = \frac{2.1}{(1-c)^2}\log(n)$ we have that

$$\mathbb{P}(\exists v \in [n] : |C_v| \geq k) \leq n \cdot n^{-1.05} \to 0$$

7.4 Further intuition on the giant size

We now provide an intuitive explanation on why the giant component should have size βn where β is the unique solution to

$$1 - \beta = \psi_{poi(c)}(1-\beta)$$

Let us say that the giant component has size βn. Then the probability that a vertex does not belong to the giant component is $1 - \beta$. We can also write this probability as

$$\mathbb{P}(v \notin Giant) = \sum_k \mathbb{P}(N(v) \cap Giant = \emptyset \,|\, |N(v)| = k)\mathbb{P}(|N(v)| = k)$$

where $N(v)$ stands for the set of immediate neighbours of vertex v. We have that $|N(v)|$ is roughly distributed as $\mathrm{Bin}(n, c/n) \to Poi(c)$, hence

$$\sum_k \mathbb{P}(N(v) \cap Giant = \emptyset \,|\, |N(v)| = k)\mathbb{P}(|N(v)| = k) \approx \sum_k (1-\beta)^k \mathbb{P}(Poi(c) = k) = \psi_{poi(c)}(1-\beta)$$

which gives the desired fixed point.