Percolation on Finite Graphs (0366-4118)

Prof. Michael Krivelevich

Tel Aviv University, Spring 2026

The following notes were written by the participants of the course, compiled by Itay Markbreit and are based on the lectures of a course on percolation on finite graphs, given by Prof. Michael Krivelevich at the School of Mathematical Sciences, Tel Aviv University, Spring 2026. Despite our best efforts, there are probably still some typos and mistakes left. Hence, corrections and other feedback would be greatly appreciated and can be sent to krivelev@tauex.tau.ac.il.

Acknowledgements: We would like to thank Michael Krivelevich for his corrections and invaluable suggestions for improvement throughout the preparation of these notes.

Contents

Percolation on Finite Graphs, Tel Aviv Univ., Spring 2026	April 12, 2026
Lecture 1	
Lecturer: Prof. Michael Krivelevich	*Scribe: Itay Markbreit*

1.1 Standard inequalities

1. For every $x \in \mathbb{R}$, $1 + x \leq e^x$.
2. For every small enough $x > 0$, $1 + x \leq e^{x - \frac{x^2}{3}}$.
3. Recall that: For every $0 \leq k \leq n$, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. For every $1 \leq k \leq n$:
$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \sum_{i=0}^{k} \binom{n}{i} \leq \left(\frac{en}{k}\right)^k.$$
4. Stirling's formula:
$$\lim_{n \to \infty} \frac{n!}{\sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n} = 1.$$
5. Markov's inequality: Let X be a non-negative random variable such that $\mathbb{E}[X]$ exists. Then, for every $t > 0$,
$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}.$$
6. Chebyshev's inequality: Let X be a random variable such that $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ exist. Then, for every $t > 0$,
$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$
7. Chernoff's inequality: Recall the notion of the binomial distribution. Let $\{X_i\}_{i=1}^n$ be i.i.d. Bernoulli(p) random variables, that is, for every $1 \leq i \leq n$, $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = 0) = 1-p$. Set $X = \sum_{i=1}^n X_i$. X is a binomially distributed random variable with parameters n and p, and we denote $X \sim Bin(n,p)$. Notice that $\mathbb{E}[X] = np$ and $\mathrm{Var}(X) = npq$ where $q = 1 - p$. Chernoff's inequalities say that
$$\mathbb{P}(X \leq np - a) \leq e^{-\frac{a^2}{2np}}, \quad \mathbb{P}(X \geq np + a) \leq e^{-\frac{a^2}{2np} + \frac{a^3}{2(np)^2}},$$
for every $a > 0$.

Basic notations and assumption in the course

1. $n \to \infty$ (but finite).

2. Let $\bar{\Omega} = (\Omega_n)_{n=1}^{\infty}$ be a sequence of probability spaces and let $\bar{A} = (A_n)_{n=1}^{\infty}$ such that for every $n \in \mathbb{N}$, $A_n \subseteq \Omega_n$ is an event. We say that $\bar{A}$ occurs *with high probability* (abbv. whp) if

$$\lim_{n \to \infty} \mathbb{P}_{\Omega_n}(A_n) = 1.$$

Example 1.1.1. When we say that $G\left(n, \frac{1}{2}\right)$ is Hamiltonian (=contains a Hamilton cycle), formally, we mean the following: $\Omega_n = G\left(n, \frac{1}{2}\right)$. A_n is the event where $G \sim G\left(n, \frac{1}{2}\right)$ is Hamiltonian and

$$\lim_{n \to \infty} \mathbb{P}\left(G \sim G\left(n, \frac{1}{2}\right) \text{ is Hamiltonian}\right) = 1.$$

1.2 The Gilbert Model $G(n, p)$

In the $G(n, p)$ model, we consider a set of n labeled vertices, denoted as $[n] = \{1, \ldots, n\}$. The total number of possible edges in a complete graph K_n is given by:

$$N \coloneqq \binom{n}{2} = |E(K_n)|.$$

Definition 1.2.1. Let $0 \leq p \coloneqq p(n) \leq 1$. We say that $G \sim G(n, p)$ if $V(G) = [n]$ consists of n labeled vertices, and each possible edge (i, j) exists independently with probability p.

Remark 1.2.2. The fact that the vertices are labeled implies for example that

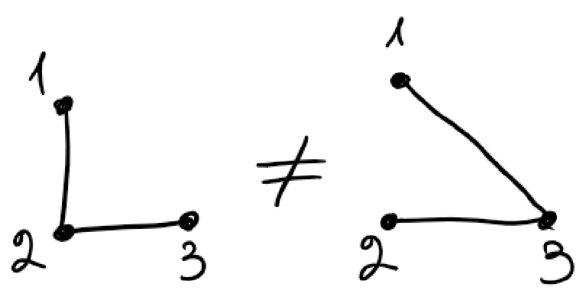


Figure 1.1: The order matters

Equivalently,

Definition 1.2.3 (Equivalent Definition)**.** Let $\Omega_n = \{G = (V, E) \mid V = [n]\}$. For every $G \in \Omega_n$:

$$\mathbb{P}(G) = p^{|E(G)|}(1-p)^{N-|E(G)|}.$$

Example 1.2.4. According to the following picture, $n = 4$, $p = 1/3$, and $|E(G)| = 4$, and hence $\mathbb{P}(G) = \left(\frac{1}{3}\right)^4 \cdot \left(\frac{2}{3}\right)^2$.

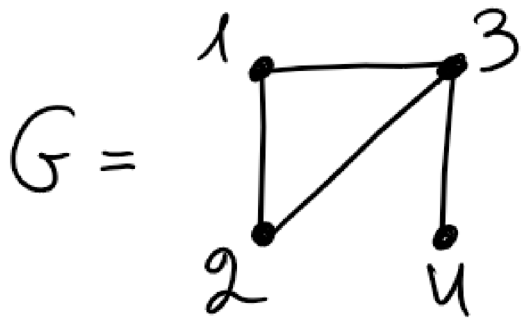


Figure 1.2

Remark 1.2.5. Notice that $G(n,p)$ is a product space.

Remark 1.2.6. Assume that $G = ([n], E)$ satisfies $|E(G)| = m$, then

$$\mathbb{P}(G) = p^m(1-p)^{N-m},$$

and hence the distribution $G(n,p)\big||E(G)| = m$ is a uniform distribution on graphs with vertex set $[n]$ and exactly m edges.

Proposition 1.2.7. *Let $G \sim G(n,p)$ and set $X = |E(G)|$. Then $X \sim Bin(N,p)$.*

Remark 1.2.8. $G(n,p)$ was introduced by Gilbert in 1959 and is often called a *binomial random graph.*

Example 1.2.9. Special case - $p = \frac{1}{2}$: In this case, $\Omega_n = \{G = (V,E) \mid V = [n]\}$, and for every $G \in \Omega_n$:

$$\mathbb{P}(G) = \left(\frac{1}{2}\right)^N.$$

We get a uniform distribution on the space of all graphs with vertex set $[n]$. Hence, we use the following convention: when saying, for example: "Almost every graph on n vertices is connected", we mean that

$$\lim_{n\to\infty} \mathbb{P}\left(G \sim G\left(n, \frac{1}{2}\right) \text{ is connected}\right) = 1.$$

1.3 The Erdős Rényi Model $G(n,m)$

In this model we consider $\Omega = \{G = (V,E) \mid V = [n], |E(G)| = m\}$ and the distribution is uniform: for every $G \in \Omega$,

$$\mathbb{P}(G) = \frac{1}{|\Omega|} = \frac{1}{\binom{N}{m}}.$$

Remark 1.3.1. $G(n,m)$ was introduced and studied by Erdős and Rényi in 1960.

1.4 Random Graph Process

Conceptually, we start with the empty graph $\bar{K_n}$ and finish with the complete graph K_n in a random process. Formally,

Definition 1.4.1. Given a permutation $\sigma \in S_N$ on the edges of K_n, we define the following *graph process* $\tilde{G} := \tilde{G}(\sigma)$: $\tilde{G} = (G_i)_{i=0}^{N}$, where G_i is a graph with $V(G_i) = [n]$ and $E(G_i) = \{e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(i)}\}$ (the first i edges according to σ).

Example 1.4.2. Let $n = 4$. $\tilde{G} = (G_i)_{i=1}^{6}$. According to the following figure,
$G_4 = \big([4], \{(1,2), (1,3), (3,4), (2,3)\}\big)$.

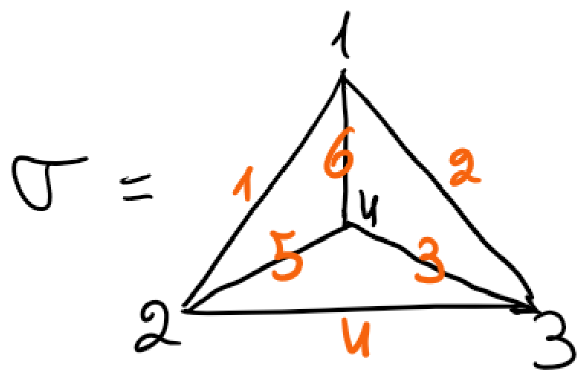


Figure 1.3

Properties:

1. For every $0 \leq i \leq N$, $|E(G_i)| = i$.
2. $\bar{K_n} = G_0 \subsetneq G_1 \subsetneq \dots \subsetneq G_N = K_n$ — a nested sequence.

Definition 1.4.3. When $\sigma \in S_n$ is chosen at random, we say that $\tilde{G} = \tilde{G}(\sigma)$ is a *random graph process*.

Definition 1.4.4 (Equivalent Definition)**.** Start with $G_0 = \bar{K_n}$. For every $1 \leq i \leq N$, set $G_i = G_{i-1} \cup \{e\}$ where e is a uniform chosen edge from the edges of K_n that are not in G_{i-1}.

Remark 1.4.5. This model was introduced and studied by Erdős and Rényi in the 1960s.

Proposition 1.4.6. *Let $\tilde{G}$ be a random graph process. Then, for every $0 \leq m \leq N$, the graph G_m (which is called a snapshot at time m) is distributed as $G(n, m)$.*

Proof. First, recall that $|E(G_m)| = m$, hence G_m is distributed over graphs with vertex set $[n]$ and exactly m edges. Moreover, for every $G = ([n], E)$ with $|E| = m$,

$$\mathbb{P}(G_m = G) = \frac{m!(N-m)!}{N!} = \frac{1}{\binom{N}{m}},$$

where the $m!$ term is for the arrangement of the edges of G at the beginning of the permutation and the $(N - m)!$ term is for the arrangement of the edges that are not in G. Therefore, G_m is uniformly distributed over graphs with vertex set $[n]$ and exactly m edges, as desired. $\square$

Remark 1.4.7. The above proposition demonstrates that random graph process contains all the spaces $G(n,m)$ for $0 \le m \le N$. Therefore, it is more difficult (compared to $G(n,m)$ or $G(n,p)$) to study the random graph process, but results regarding the process can yield consequences on $G(n,m)$ or $G(n,p)$.

1.5 Multiple Exposure

Proposition 1.5.1. *Suppose that $0 \le p, p_1, p_2, \dots, p_k \le 1$ satisfy $1 - p = \prod_{i=1}^{k}(1-p_i)$. Let $G \sim G(n,p)$ and suppose $G_i \sim G(n,p_i)$, $i = 1, \dots, k$, are independent. Define: $G' = \bigcup_{i=1}^{k} G_i$. Then G and G' have the same distribution.*

Proof. First, notice that G, G' are product spaces: the probability of a specific edge appearing in G (as well as in G') is independent of other edges.
Therefore, it is sufficient to show that for any edge $e \in E(K_n)$, $\mathbb{P}(e \in G) = \mathbb{P}(e \in G')$.

- In G: $\mathbb{P}(e \in G) = p$.
- In $G' = G_1 \cup \cdots \cup G_k$:

$$\mathbb{P}(e \in G') = 1 - \mathbb{P}(e \notin G') = 1 - \mathbb{P}\left(\bigwedge_{i=1}^{k} e \notin G_i\right) = 1 - \prod_{i=1}^{k}\mathbb{P}(e \notin G_i) = 1 - \prod_{i=1}^{k}(1-p_i) = p,$$

where the last equality follows from the assumption.
Hence, $\mathbb{P}(e \in G) = \mathbb{P}(e \in G')$, and thus G and G' have the same distribution. $\square$

Remark 1.5.2. It is customary (with some abuse of notation) to use $G(n,p)$ both as a distribution and as a graph drown from this distribution.

Remark 1.5.3 (Typical Scenario — Sprinkling)**.** $0 \le p_1, p_2 \le 1$ and $p_1 \gg p_2$. Suppose that $1 - p = (1-p_1)(1-p_2)$. Let: $G \sim G(n,p)$, $G_1 \sim G(n,p_1)$, $G_2 \sim G(n,p_2)$. Then from the previous proposition: $G \sim G_1 \cup G_2$. Since $p_1 \gg p_2$, most of the edges are typically from G_1, and G_2 "sprinkles" edges onto G_1.

1.6 Monotone Properties and Threshold Function

Definition 1.6.1. Suppose that $\mathcal{A}$ is a property of graphs on vertex set $[n]$, namely $\mathcal{A} \subseteq \Omega_n = \{G = ([n], E)\}$. $\mathcal{A}$ is called *monotone increasing* (or *monotone*) if for every $G \in \mathcal{A}$ and for every $H \in \Omega_n$ such that $G \subseteq H$, it holds that $H \in \mathcal{A}$.

Definition 1.6.2. A property $\mathcal{A}$ is *monotone decreasing* if for every $G \in \mathcal{A}$ and $H \subseteq G$, it holds that $H \in \mathcal{A}$.

Example 1.6.3. Connectivity is a monotone increasing property, whereas planarity is a monotone decreasing property.

Definition 1.6.4. A property $\mathcal{A}$ is called non-trivial if:

1. $\bar{K}_n \notin \mathcal{A}$,
2. $K_n \in \mathcal{A}$.

Definition 1.6.5. Let $\mathcal{A} \subseteq \Omega_n$ be a non-trivial monotone property and let $0 \le p = p(n) \le 1$. We define:

$$f_{\mathcal{A}}(p) = \mathbb{P}(G \sim G(n,p) \in \mathcal{A}) = \sum_{G \in \mathcal{A}} \mathbb{P}(G).$$

Properties:

1. $f_{\mathcal{A}}(0) = 0, f_{\mathcal{A}}(1) = 1$.
2. $f_{\mathcal{A}}(p)$ is a polynomial in p of degree $\le \binom{n}{2}$, therefore it is a continuous function.
3. $f_{\mathcal{A}}(p)$ is a strictly increasing function. (Exercise!)

Question 1.6.6. How fast does $f_{\mathcal{A}}(p)$ jump from nearly 0 to nearly 1?

Proposition 1.6.7. *Let $\epsilon, \delta > 0$ be real numbers. Let $\mathcal{A} \subseteq \Omega_n$ be an increasing monotone property of graphs on a set of vertices $[n]$. Assume there exists an integer $C > 0$ such that $(1-\epsilon)^C \le \delta$. Let $G \sim G(n, p_0)$ satisfy: $\mathbb{P}(G \in \mathcal{A}) \ge \epsilon$. Define $p_1 = C \cdot p_0$.*
Then: $\mathbb{P}(G \sim G(n,p_1) \in \mathcal{A}) \ge 1 - \delta$.

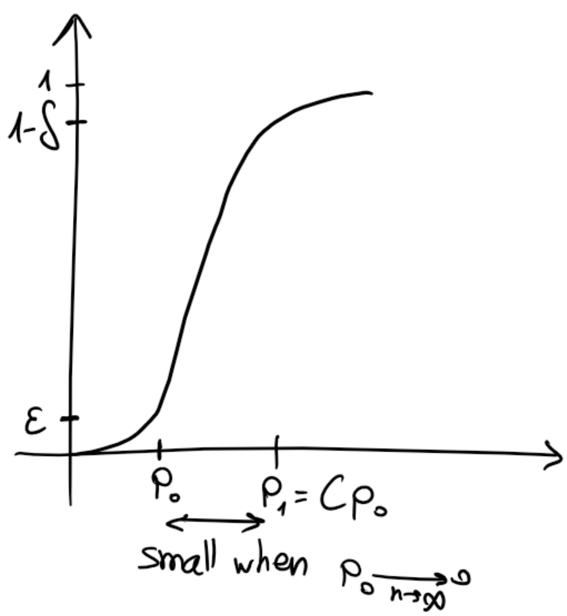


Figure 1.4: Quick jump to high probability

Proof. Suppose that the graphs $G_1, \ldots, G_C$ are independent and satisfy $G_i \sim G(n, p_0)$. Define $G' = \bigcup_{i=1}^{C} G_i$. Then $G' \sim G(n,p')$ where $1-p' = \prod_{i=1}^{C}(1-p_0) = (1-p_0)^C$. Therefore $p' \le Cp_0 = p_1$. Since $\mathcal{A}$ is a monotone property, it suffices to prove that: $\mathbb{P}(G \sim G(n,p') \in \mathcal{A}) \ge 1 - \delta$. Indeed, let us note:

$$\mathbb{P}(G \sim G(n,p') \notin \mathcal{A}) \le \mathbb{P}(G_i \notin \mathcal{A}, 1 \le i \le C) = \prod_{i=1}^{C} \mathbb{P}(G_i \notin \mathcal{A}) \le (1-\epsilon)^C \le \delta.$$

We obtain the required result. □

Conclusion: The jump of $\mathbb{P}(G \sim G(n,p) \in \mathcal{A})$ from nearly 0 to nearly 1 occurs within a range of $\theta(f_{\mathcal{A}}^{-1}(\epsilon))$. Thus if $f_{\mathcal{A}}^{-1}(\epsilon) = o(1)$, then the jump occurs within an interval of negligible width.

Definition 1.6.8. Let $\mathcal{A}$ be an increasing monotone property of graphs on vertex set $[n]$.
The probability $p^* = p^*(n)$ is called a threshold function for the property $\mathcal{A}$ if it satisfies:

$$\lim_{n\to\infty} \mathbb{P}(G \sim G(n,p) \in \mathcal{A}) = \begin{cases} 0 & p \ll p^*, \\ 1 & p \gg p^*. \end{cases}$$

Remark 1.6.9. According to the definition, p^* is not a unique function, and it is defined up to the order of magnitude.

Example 1.6.10. $A = \{G \text{ contains a triangle}\}$.
Let X be a random variable counting the number of triangles in the graph $G(n,p)$. We have:

$$\mathbb{E}X = \binom{n}{3} \cdot p^3.$$

If $\mathbb{E}X \xrightarrow[n\to\infty]{} 0$, then according to Markov's Inequality, $\mathbb{P}(X > 0) = o(1)$.
If $\mathbb{E}X \xrightarrow[n\to\infty]{} \infty$, it can be shown using the Second Moment Method that $\mathbb{P}(X > 0) = 1 - o(1)$.

When $p = o\left(\frac{1}{n}\right)$, then $\binom{n}{3} \cdot p^3 \xrightarrow[n\to\infty]{} 0$.
When $p = \omega\left(\frac{1}{n}\right)$, then $\binom{n}{3} \cdot p^3 \xrightarrow[n\to\infty]{} \infty$.

It can be proven that in this case $p^* = \frac{1}{n}$ is a threshold function.

Theorem 1.6.11 (Bollobás-Thomason 1987)**.** *For every non-trivial monotone property $\mathcal{A}$ of graphs on a set of vertices $[n]$, there exists a threshold function.*

Percolation on Finite Graphs, Tel Aviv Univ., Spring 2026 April 19, 2026

Lecture 2

Lecturer: Prof. Michael Krivelevich *Scribe: Inbal Benbenishty*

2.1 Monotone Properties

Theorem 2.1.1 (Bollobás-Thomason 1987). *For every non-trivial monotone property $\mathcal{A}$ of graphs on a set of vertices $[n]$, there exists a threshold function.*

Proof. We saw that $f_A(p) = \mathbb{P}(G \sim G(n,p) \in A)$ is a continuous, strictly increasing function, and $f_A(0) = 0, f_A(1) = 1$. Therefore, there exists a unique $p^* \in (0,1)$ such that $f_A(p^*) = \frac{1}{2}$.
We will use Proposition 1.6.7 from Lecture 1 in both directions:

1. Let $p_0 = p^*, p_1 = C \cdot p_0, \varepsilon = \frac{1}{2}$. Then by Proposition 1.6.7 from Lecture 1:

$$\mathbb{P}(G \sim G(n,p_1) \in A) \geq 1 - (1-\varepsilon)^C.$$

Since $\lim_{C\to\infty}(\frac{1}{2})^C = 0$, it follows that for any $p \gg p^*$:

$$\mathbb{P}(G \sim G(n,p) \in A) = 1 - o(1).$$

2. Take $p_1 = p^*, p_0 = \frac{p^*}{C}$. Then by Proposition 1.6.7 from Lecture 1:

$$\frac{1}{2} = \mathbb{P}(G \sim G(n,p_1) \notin A) \leq (1 - \mathbb{P}(G \sim G(n,p_0) \in A))^C.$$

For $C \to \infty$ we obtain:

$$\mathbb{P}(G \sim G(n,p_0) \in A) = o(1).$$

□

Remark 2.1.2. In the original proof of the theorem, Bollobás and Thomason used the Kruskal-Katona theorem and obtained a better quantitative estimate.

2.2 Comparison between $G(n,p)$ and $G(n,m)$

One can expect that if $p = \frac{m}{N}$, then $G(n,p)$ and $G(n,m)$ would be "similar."
Indeed, let $X = |E(G)|$.

- In $G(n,p)$: $X \sim \text{Bin}(N,p)$, and $\mathbb{E}[X] = Np = m$.
- In $G(n,m)$: $X = m$ holds with probability 1.

However, this similarity should have some limits.
If $m \to \infty$ and $N - m \to \infty$, then $\mathbb{P}(X = m) = \mathbb{P}(\text{Bin}(N,p) = m) = o_n(1)$ since $\text{Var}(X) = Np(1-p) \to \infty$. However, in $G(n,m)$, the number of edges is always m.

Quantitative comparison between $G(n,p)$ and $G(n,m)$

Proposition 2.2.1. *Let $\mathcal{A}$ be a property of graphs on the set of vertices $[n]$. If $4 \le m \le N$, $m = Np$, then:*

$$\mathbb{P}(G \sim G(n,p) \in \mathcal{A}) \ge \frac{1}{6\sqrt{m}} \mathbb{P}(G \sim G(n,m) \in \mathcal{A}).$$

Proof. Let X be the random variable counting the number of edges in $G \sim G(n,p)$. Notice that $X \sim \mathrm{Bin}(N,p)$, $\mathbb{E}[X] = Np = m$, and $\mathrm{Var}(X) = Np(1-p) < Np = m$.
Therefore, by Chebyshev's inequality, for every $a > 0$, the following holds:

$$\mathbb{P}(|X - m| > a \cdot \sigma) \le \frac{1}{a^2}.$$

Furthermore, for every $0 \le m' \le N$, we have $\mathbb{P}(X = m') \le \mathbb{P}(X = m)$ (check it!), and thus:

$$\mathbb{P}(X = m) \ge \frac{\mathbb{P}(|X - m| \le a \cdot \sigma)}{2a\sigma + 1} \ge \frac{1 - \frac{1}{a^2}}{2a\sigma + 1},$$

since in the interval $[m - a\sigma, m + a\sigma]$ there are at most $2a\sigma + 1$ distinct integer values. By setting $a = \sqrt{3}$, we obtain $\mathbb{P}(X = m) \ge \frac{1}{6\sqrt{m}}$ (we used $m \ge 4$ here).
Recall that the distribution of $G(n,p)$ conditioned on $|E(G)| = m$ is $G(n,m)$. Therefore, according to the Law of Total Probability:

$$\begin{aligned}
\mathbb{P}(G \sim G(n,p) \in \mathcal{A}) &= \sum_{m'=0}^{N} \mathbb{P}(X = m') \cdot \mathbb{P}(G \sim G(n,p) \in \mathcal{A} \mid X = m') \\
&\ge \mathbb{P}(X = m) \cdot \mathbb{P}(G \sim G(n,p) \in \mathcal{A} \mid X = m) \\
&= \mathbb{P}(X = m) \cdot \mathbb{P}(G \sim G(n,m) \in \mathcal{A}) \\
&\ge \frac{1}{6\sqrt{m}} \cdot \mathbb{P}(G \sim G(n,m) \in \mathcal{A}).
\end{aligned}$$

□

We can prove a better estimate for monotone properties.

Proposition 2.2.2. *Let $\mathcal{A}$ be a monotone property of graphs on vertex set $[n]$. Assume N, p, m satisfy $m = N \cdot p$ and $m \to \infty$. Then:*

$$\mathbb{P}(G \sim G(n,p) \in \mathcal{A}) \ge \left(\frac{1}{2} - o(1)\right) \cdot \mathbb{P}(G \sim G(n,m) \in \mathcal{A}).$$

Proof. Note that since $\mathcal{A}$ is a monotone property, for every $0 \le m \le m' \le N$:

$$\mathbb{P}(G \sim G(n,m) \in \mathcal{A}) \le \mathbb{P}(G \sim G(n,m') \in \mathcal{A}).$$

Let $X = |E(G \sim G(n,p))|$. By the Law of Total Probability:

$$\begin{aligned}
\mathbb{P}(G \sim G(n,p) \in \mathcal{A}) &= \sum_{m'=0}^{N} \mathbb{P}(X = m') \cdot \mathbb{P}(G \sim G(n,p) \in \mathcal{A} \mid X = m') \\
&= \sum_{m'=0}^{N} \mathbb{P}(X = m') \cdot \mathbb{P}(G \sim G(n,m') \in \mathcal{A}) \\
&\geq \sum_{m'=m}^{N} \mathbb{P}(X = m') \cdot \mathbb{P}(G \sim G(n,m') \in \mathcal{A}) \\
&\geq \sum_{m'=m}^{N} \mathbb{P}(X = m') \cdot \underbrace{\mathbb{P}(G \sim G(n,m) \in \mathcal{A})}_{\text{monotonicity}} \\
&= \mathbb{P}(G \sim G(n,m) \in \mathcal{A}) \cdot \sum_{m'=m}^{N} \mathbb{P}(X = m') \\
&= \mathbb{P}(G \sim G(n,m) \in \mathcal{A}) \cdot \mathbb{P}(\mathrm{Bin}(N,p) \geq m).
\end{aligned}$$

From the Central Limit Theorem, since $m \to \infty$, $\mathbb{P}(\mathrm{Bin}(N,p) \geq m) \geq \frac{1}{2} + o(1)$. Thus:

$$\mathbb{P}(G \sim G(n,p) \in \mathcal{A}) \geq \left(\frac{1}{2} + o(1)\right) \mathbb{P}(G \sim G(n,m) \in \mathcal{A}).$$

□

2.3 Connected Components in $G(n,p)$ and Phase Transition

Let $G \sim G(n,p)$. Let $L_1, L_2, \ldots$ be the connected components of G ordered by cardinality $|L_1| \geq |L_2| \geq \ldots$.

Theorem 2.3.1 (Erdős-Rényi 1960)**.** *Let $p = \frac{c}{n}$ for $c > 0$.*

1. *If $0 < c < 1$ (the sub-critical regime), then whp $|L_1| = O_c(\log n)$.*
2. *If $c > 1$ (the super-critical regime), then whp $|L_1| = \Theta(n)$ and all other components satisfy $|L_i| = O_c(\log n)$ for $i \geq 2$.*

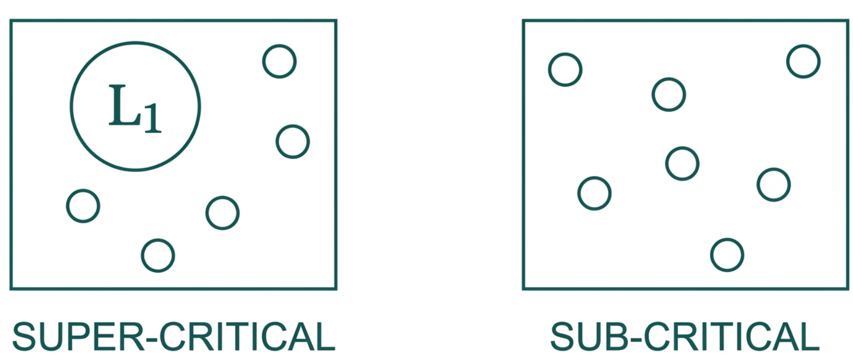


Figure 2.5: Super-critical and Sub-critical states

For the proof of Theorem 2.3.1 we utilize several auxiliary lemmas that are given in Lecture 2 and in Lecture 3.
We begin with the case where $0 < c < 1$.

Lemma 2.3.2. *In $G \sim G(n, \frac{c}{n})$, whp there are no connected components with 2 or more cycles.*

Proof. If a connected graph H has at least 2 cycles, then H contains one of the following 3 configurations:

1. Two disjoint cycles C_1, C_2 connected by a path.
2. Two cycles C_1, C_2 sharing one vertex.
3. Two vertices connected by 3 edge-disjoint paths.

One can easily verify that in any of these cases, H contains a path P and 2 additional edges, each from one end of P inwards ("end edges"). The expected number of such structures in $G(n,p)$ is:

$$\mathbb{E}[\#\text{components with } \geq 2 \text{ cycles}] \leq \sum_{k=4}^{n} \underbrace{\binom{n}{k}}_{\substack{\text{choosing the}\\ \text{vertices}}} \cdot \underbrace{k!}_{\substack{\text{ordering the}\\ \text{vertices}}} \cdot \underbrace{k^2}_{\substack{\text{placing the}\\ \text{"end edges"}}} \cdot \underbrace{p^{k+1}}_{\substack{\text{prob. } p\\ \text{for each edge}}}$$
$$\leq \sum_{k=4}^{n} \frac{n^k}{k!} \cdot k! \cdot k^2 \left(\frac{c}{n}\right)^{k+1} = \frac{c}{n} \sum_{k=4}^{n} k^2 \cdot c^k \xrightarrow[n\to\infty]{} 0.$$

Since the expectation goes to 0, by Markov's inequality, whp no such component exists. $\square$

Lemma 2.3.3. *Whp, the total number of vertices in components containing a (unique) cycle is at most $\omega(n)$ for any $\omega(n)$ such that $\omega(n) \to \infty$ as $n \to \infty$.*

Proof. For $3 \leq k \leq n$, let Z_k be the number of vertices in components of order k containing exactly one cycle. Our task is to show that whp $Z_3 + Z_4 + \ldots + Z_n \leq \omega(n)$. We first estimate the expectation:

$$\mathbb{E}\left[\sum_{k=3}^{n} Z_k\right] = \sum_{k=3}^{n} \mathbb{E}[Z_k].$$

The bound for $\mathbb{E}[Z_k]$ is given by:

$$\mathbb{E}[Z_k] \leq \underbrace{\binom{n}{k}}_{\substack{\text{choosing k}\\ \text{vertices}}} \cdot \underbrace{k^{k-2}}_{\substack{\text{choosing a spanning}\\ \text{tree (Cayley)}}} \cdot \underbrace{\binom{k}{2}}_{\text{forming a cycle}} \cdot \underbrace{p^k}_{\substack{\text{prob. } p\\ \text{for each edge}}} \cdot \underbrace{(1-p)^{k(n-k)+\binom{k}{2}-k}}_{\substack{\text{disconnecting from}\\ \text{other components}\\ \text{and forbidding}\\ \text{other edges inside}}} \cdot \underbrace{k}_{\substack{\text{order of the}\\ \text{component}}}.$$

Note that we have

$$\binom{n}{k} = \frac{n^k}{k!} \cdot \frac{n(n-1)\cdots(n-k+1)}{n^k} = \frac{n^k}{k!}\left(\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n}\right)$$
$$\overset{1-x\leq e^{-x}}{\leq} \frac{n^k}{k!} \cdot e^{-\left(\frac{1}{n}+\frac{2}{n}+\cdots+\frac{k-1}{n}\right)} = \frac{n^k}{k!} \cdot e^{-\frac{k(k-1)}{2n}}.$$

Hence,

$$\begin{aligned}
\mathbb{E}[Z_k] &\leq \frac{n^k}{k!} \cdot e^{-\frac{k(k-1)}{2n}} \cdot k^{k+1} \left(\frac{c}{n}\right)^k \cdot e^{-\frac{c}{n}(kn - \frac{k^2}{2} - \frac{3k}{2})} \\
&\leq \frac{n^k}{(\frac{k}{e})^k} e^{-\frac{k(k-1)}{2n}} \cdot k^{k+1} e^{-ck + \frac{k^2}{2n} + \frac{3k}{2n}} \cdot \left(\frac{c}{n}\right)^k \\
&= k \cdot (ce^{1-c})^k e^{\frac{2k}{n}} \leq e^2 k \cdot (ce^{1-c})^k.
\end{aligned}$$

Notice that for every $c \neq 1$, we have $ce^{1-c} < 1$, hence plugging this into $\mathbb{E}[Z_k]$ we obtain:

$$\mathbb{E}\left[\sum_{k=3}^{n} Z_k\right] \leq \sum_{k=3}^{n} e^2 \cdot k(ce^{1-c})^k = O(1).$$

Therefore, by Markov's inequality, for any $\omega(n) \to \infty$, we have that whp $\sum_{k=3}^{n} Z_k \leq \omega(n)$. $\square$

3.1 Connected Components in $G(n, p)$ and Phase Transition - Continued

Lemma 3.1.1. *For $c > 0$, $c \neq 1$, let $p = \frac{c}{n}$. For $1 \leq k \leq n$, let X_k denote the number of connected components that are trees of order k. The following hold:*

- ***(a)*** *There exists a constant $\beta_1 = \beta_1(c) > 0$ such that whp, for all $k \geq \beta_1 \log n$, it holds that $X_k = 0$. (Meaning: whp, there is no connected component of order $\geq \beta_1 \log n$ in G which is a tree).*
- ***(b)*** *There exists a constant $\beta_0 = \beta_0(c) > 0$ such that whp, for all $1 \leq k \leq \beta_0 \log n$, it holds that:*

$$X_k = (1 + o_n(1))\frac{n}{c} \cdot \frac{k^{k-2}}{k!}(ce^{-c})^k.$$

Proof. **Part (a):** We first calculate the expectation $\mathbb{E}[X_k]$:

$$\mathbb{E}[X_k] = \underbrace{\binom{n}{k}}_{\text{choosing vertices}} \cdot \underbrace{k^{k-2}}_{\text{choosing the tree}} \cdot \underbrace{p^{k-1}}_{\text{tree edges}} \cdot \underbrace{(1-p)^{k(n-k)}}_{\text{no edges to rest of graph}} \cdot \underbrace{(1-p)^{\binom{k}{2}-k+1}}_{\text{no extra edges in tree}}.$$

Notice that for all k, $\binom{n}{k} \leq \frac{n^k}{k!}$, and specifically for $k = O(\log n)$, we have $\binom{n}{k} = (1 + o(1))\frac{n^k}{k!}$. Therefore, we can bound the expectation:

$$\mathbb{E}[X_k] \leq \frac{n^k}{k!} k^{k-2} \left(\frac{c}{n}\right)^{k-1} e^{-ck+O(\frac{k^2}{n})},$$

and for $k = O(\log n)$, it holds that:

$$\mathbb{E}[X_k] = (1 + o(1))\frac{n}{c} \cdot \frac{k^{k-2}}{k!}(ce^{-c})^k.$$

Let us denote $\mu_k = \mathbb{E}[X_k] = \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)+\binom{k}{2}-k+1}$. We can further bound μ_k as follows (assuming $k = O(\log n)$):

$$\mu_k = (1 + o(1))\frac{n}{c} \cdot \frac{k^{k-2}}{k!}(ce^{-c})^k \leq (1 + o(1))\frac{n}{c} e^k (ce^{-c})^k \leq (1 + o(1))\frac{n}{c}(ce^{1-c})^k.$$

Notice that for every $1 \neq c \in \mathbb{R}$ it holds that $ce^{1-c} < 1$. Therefore, if we choose a constant $0 < \beta_1 = \beta_1(c)$ large enough, for $k = \lceil \beta_1 \log n \rceil$ it will hold that $\mu_k = o(n^{-2})$.

Furthermore, analyzing the ratio between consecutive terms yields:

$$\frac{\mu_{k+1}}{\mu_k} = \frac{\binom{n}{k+1}(k+1)^{k-1}p^k(1-p)^{(k+1)(n-k-1)+\binom{k+1}{2}-k}}{\binom{n}{k}k^{k-2}p^{k-1}(1-p)^{k(n-k)+\binom{k}{2}-k+1}} = \frac{n-k}{k+1}\cdot\frac{(k+1)^{k-1}}{k^{k-2}}p(1-p)^{n-k-2}$$
$$= \frac{n-k}{n}\cdot c\cdot\left(\frac{k+1}{k}\right)^{k-2}(1-p)^{n-k-2} \le \frac{n-k}{n}\cdot c\cdot e\cdot e^{-\frac{c(n-k)}{n}}(1-p)^{-2} < (1-p)^{-2}.$$

Hence, summing over the tail gives:

$$\sum_{k=\lceil\beta_1\log n\rceil}^{n}\mu_k \le \mu_{\lceil\beta_1\log n\rceil}\sum_{i=0}^{n}\underbrace{(1-p)^{-2i}}_{=O(1)} = O(n)\mu_{\lceil\beta_1\log n\rceil} = O(n)\cdot o(n^{-2}) = o(n^{-1}).$$

Therefore, according to Markov's inequality, whp, for all $k\ge\beta_1\log n$ it holds that $X_k=0$.

Part (b) (sketch): We will use the Second Moment Method (Chebyshev's inequality) for $k\le\beta_0\log n$, where $\beta_0=\beta_0(c)>0$ is a small enough constant. We already know that:

$$\mathbb{E}[X_k] = (1+o(1))\frac{n}{c}\cdot\frac{k^{k-2}}{k!}(ce^{-c})^k.$$

We will calculate the variance of X_k. Consider the number of ordered pairs of trees T_1, T_2 of order k that are disjoint in vertices:

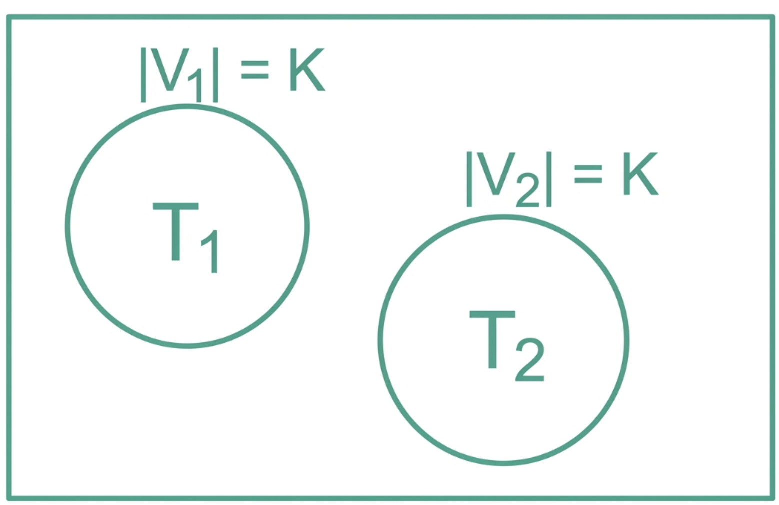


Figure 3.6: T1 and T2

$$\mathbb{E}\begin{bmatrix}\text{Number of ordered pairs of}\\ \text{tree components of order } k\end{bmatrix} = \underbrace{\binom{n}{k}}_{\substack{\text{choosing}\\ V(T_1)}}\cdot\underbrace{\binom{n-k}{k}}_{\substack{\text{choosing}\\ V(T_2)}}\cdot\underbrace{(k^{k-2}p^{k-1})^2}_{\substack{\text{trees of order } k\\ \text{and edges}}}\cdot\underbrace{(1-p)^{2k(n-2k)+\binom{2k}{2}-2k+2}}_{\substack{\text{no edges going out of}\\ T_1\cup T_2 \text{ and between them}}}.$$

We can bound the variance, using the above estimate, as follows:

$$Var(X_k) \leq \mathbb{E}[X_k] + \mathbb{E}^2[X_k] \underbrace{((1-p)^{-k^2} - 1)}_{\approx pk^2} \leq \mathbb{E}[X_k] + \frac{2ck^2}{n}\mathbb{E}^2[X_k].$$

Applying Chebyshev's inequality gives us:

$$\mathbb{P}(|X_k - \mathbb{E}[X_k]| \geq \varepsilon\mathbb{E}[X_k]) \leq \frac{1}{\varepsilon^2\mathbb{E}[X_k]} + \frac{2ck^2}{\varepsilon^2 n}.$$

Recall that:

$$\mathbb{E}[X_k] = (1+o(1))\frac{n}{c} \cdot \frac{k^{k-2}}{k!}(ce^{-c})^k \leq (1+o(1))\frac{n}{c} \underbrace{e^k(ce^{-c})^k}_{(ce^{1-c})^k}.$$

For $k \leq \beta_0 \log n$, we choose $\beta_0 = \beta_0(c) > 0$ small enough such that: $\mathbb{E}[X_k] \geq n^{\frac{1}{2}}$. Hence, by the union bound:

$$\mathbb{P}[\exists 1 \leq k \leq \beta_0 \log n : |X_k - \mathbb{E}[X_k]| > \varepsilon\mathbb{E}[X_k]] = O(\log n) \cdot O\left(\frac{1}{\sqrt{n}}\right) = o(1),$$

which completes the proof. □

Let us conclude for the subcritical regime:

Theorem 3.1.2 (Subcritical)**.** *Let $0 < c < 1$ be a constant, and let $G \sim G(n, \frac{c}{n})$. Then whp, all connected components of G are of order $O_c(\log n)$.*

Proof. According to Lemmas 2.3.2 and 2.3.3 from Lecture 2 and due to Lemma 3.1.1, we get whp that:

1. there are no components with more than one cycle (Lemma 2.3.2);
2. the number of vertices in components with exactly one cycle is at most $\omega(n)$ (where $1 \ll \omega(n) \ll \log n$) (Lemma 2.3.3);
3. there are no connected components that are trees of order $> \beta_1 \log n$ (Lemma 3.1.1).

Combining these facts, it follows that whp, all connected components in G are of order at most $\beta_1 \log n$, as required. □

We now turn to the supercritical regime. We will need the following definition later. Let us define $x = x(c)$ by:

$$x \coloneqq x(c) = \begin{cases} c & c < 1, \\ \text{the unique solution of } xe^{-x} = ce^{-c} & c > 1 \text{ (in the interval } (0,1)). \end{cases}$$

[Note that the function $f(x) = xe^{-x}$ attains its maximum value at $x = 1$, is monotone increasing for $0 < x < 1$, and is monotone decreasing for $x > 1$. Therefore, for every $c > 1$, the solution $x \in (0,1)$ is well-defined and unique.]

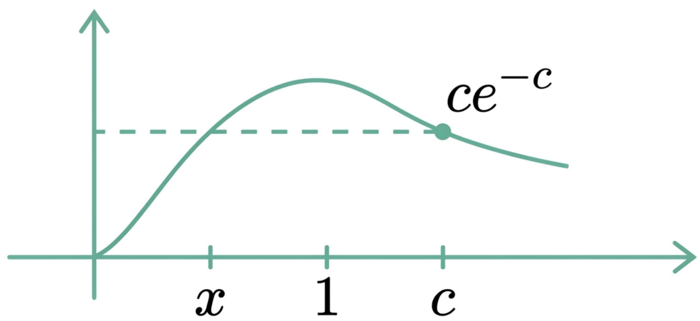


Figure 3.7: x

Lemma 3.1.3. *Let $c > 0$, $c \neq 1$. Then:*

$$x = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(ce^{-c})^k.$$

Proof. We will prove this via the $G(n, p)$ model. Let us define a random variable Z to be the number of vertices in components with a cycle. Additionally, for all $k \geq 1$, we define X_k to be the number of connected components of G that are trees with k vertices.
Summing the vertices of all the components gives the total number of vertices:

$$n = |V(G)| = Z + \sum_{k=1}^{n} k \cdot X_k.$$

Taking the expectation of both sides yields:

$$n = \mathbb{E}[Z] + \sum_{k=1}^{n} k \cdot \mathbb{E}[X_k]. \tag{3.1}$$

Case 1: $0 < c < 1$. From Lemmas 2.3.2 and 2.3.3 from Lecture 2 it follows that $\mathbb{E}[Z] = o(n)$. From Lemma 3.1.1 we know that for $k \leq \beta_0 \log n$, $\mathbb{E}[X_k] = (1 + o(1))\frac{n}{c} \cdot \frac{k^{k-2}}{k!}(ce^{-c})^k$. We have also bounded the tail: $\sum_{k=\beta_1 \log n}^{n} \mathbb{E}[X_k] = o\left(\frac{1}{n}\right)$. This implies that the contribution of the tail is negligible:

$$\sum_{k=\beta_1 \log n}^{n} k \cdot \mathbb{E}[X_k] \leq n \sum_{k=\beta_1 \log n}^{n} \mathbb{E}[X_k] = n \cdot o\left(\frac{1}{n}\right) = o(1).$$

Therefore, substituting these into equation (3.1), we get:

$$n = \underbrace{o(n)}_{\mathbb{E}[Z]} + \underbrace{o(1)}_{\text{tail}} + \sum_{k=1}^{\beta_0 \log n} \frac{n}{c} \cdot \frac{k^{k-1}}{k!}(ce^{-c})^k = o(n) + (1 + o(1))\frac{n}{c}\sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(ce^{-c})^k.$$

Dividing both sides by n and taking the limit as $n \to \infty$, we get:

$$1 = \frac{1}{c}\sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(ce^{-c})^k \quad \Longrightarrow \quad c = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(ce^{-c})^k.$$

Since $x = c$ in this domain, the lemma holds.
Case 2: $c > 1$**.** In this regime, $xe^{-x} = ce^{-c}$, where $0 < x < 1$. Using the result from the first case applied to x, we immediately get:

$$\sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(ce^{-c})^k = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(xe^{-x})^k = x. \qquad \square$$

We are now ready to establish the phenomenon in the supercritical regime.

Theorem 3.1.4 (Supercritical)**.** *Let $c > 1$ be a constant and let $G \sim G(n, \frac{c}{n})$. Then whp:*

1) *In G there exists a unique component L_1 of order $|L_1| = (1 + o(1)) \left(1 - \frac{x}{c}\right) n$, where $x = x(c)$ is defined as in Lemma 3.1.3. This is the **giant component** (of order $\Theta(n)$).*

2) *For every $i \geq 2$ it holds that $|L_i| = O_c(\log n)$.*

Proof Idea. We will prove that whp, the following four facts hold:

a) There are no components with orders in the range $[\beta_3 \log n, \beta_2 n]$ for some constants $\beta_2, \beta_3 > 0$;

b) The total order of components of order at most $\beta_3 \log n$ that are not trees is $o(n)$;

c) The total order of components that are trees of order $\leq \beta_3 \log n$ is $(1 + o(1))\frac{nx}{c}$;

d) In conclusion, the rest of the graph has volume $(1+o(1)) \left(n - \frac{nx}{c}\right)$, forming a single connected component, which is L_1.

Proof. For $1 \leq k \leq n$, let Y_k denote the number of connected components of order k in G. We can bound the expectation $\mathbb{E}[Y_k]$ as follows:

$$\mathbb{E}[Y_k] \leq \binom{n}{k} k^{k-2} p^{k-1} (1-p)^{k(n-k)} \leq \left(\frac{en}{k}\right)^k k^{k-2} \left(\frac{c}{n}\right)^{k-1} e^{-ck+\frac{ck^2}{n}} = \frac{n}{ck^2} \left(ce^{1-c+\frac{ck}{n}}\right)^k.$$

We will choose $\beta_2 = \beta_2(c) > 0$ small enough such that $ce^{1-c+c\beta_2} < 1$ (which is always possible as $ce^{1-c} < 1$). Then, we can choose $\beta_3 = \beta_3(c) > 0$ large enough such that $\left(ce^{1-c+c\beta_2}\right)^{\beta_3 \log n} \ll \frac{1}{n^2}$. Under these choices, $\mathbb{E}[Y_k] = o\left(\frac{1}{n}\right)$ for all $\beta_3 \log n \leq k \leq \beta_2 n$. Therefore, according to Markov's inequality, whp, for all $k \in [\beta_3 \log n, \beta_2 n]$ it holds that $Y_k = 0$.
Now, we estimate the total size of the connected components of G that are small trees ($\leq \beta_3 \log n$). According to Lemma 3.1.1, whp we are guaranteed that for components of order $1 \leq k \leq \beta_0 \log n$, the number of connected components that are trees of order k is $X_k = (1 + o(1))\frac{n}{c} \cdot \frac{k^{k-2}}{k!}(ce^{-c})^k$. Therefore, whp, the total volume of trees of order k (for $1 \leq k \leq \beta_0 \log n$) is $k \cdot X_k = (1 + o(1))\frac{n}{c} \cdot \frac{k^{k-1}}{k!}(ce^{-c})^k$. Summing this up yields:

$$\begin{aligned}\sum_{k=1}^{\beta_0 \log n} k \cdot X_k &= (1 + o(1))\frac{n}{c} \sum_{k=1}^{\beta_0 \log n} \frac{k^{k-1}}{k!}(ce^{-c})^k \\ &= (1 + o(1))\frac{n}{c} \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!}(ce^{-c})^k = (1 + o(1)) \cdot \frac{nx}{c},\end{aligned}$$

where x is from Lemma 3.1.3.

Furthermore, for all $\beta_0 \log n \leq k \leq \beta_3 \log n$, it holds that $\mathbb{E}[X_k] = O\left(\frac{n}{k^2}(ce^{1-c})^k\right)$. If we substitute $k \geq \beta_0 \log n$, this expression becomes $O\left(\frac{n^{1-\delta}}{\log n}\right)$ for some $\delta > 0$. Hence, the expected number of components in this intermediate range is $\mathbb{E}\left[\sum_{\beta_0 \log n}^{\beta_3 \log n} X_k\right] = O(n^{1-\delta})$. According to Markov's inequality, we get whp:

$$\sum_{k=1}^{\beta_3 \log n} k \cdot X_k = (1 + o(1))\frac{nx}{c}.$$

Next, we estimate the typical total volume of components of order $\leq \beta_3 \log n$ with cycles. Let Z_k denote the number of vertices in components of order k with a cycle. We can bound this expectation:

$$\mathbb{E}\left[\sum_{k=1}^{\beta_3 \log n} Z_k\right] \leq \sum_{k=1}^{\beta_3 \log n} \underbrace{\binom{n}{k}}_{\substack{\text{choosing} \\ \text{component}}} \cdot k^{k-2} \cdot \binom{k}{2} \cdot k \cdot \underbrace{p^k}_{\substack{\text{payment} \\ \text{for edges} \\ \text{that close} \\ \text{a cycle}}} \cdot \underbrace{(1-p)^{k(n-k)}}_{\substack{\text{no edges} \\ \text{between the} \\ \text{component} \\ \text{and the rest}}} \leq$$

$$\leq \sum_{k=1}^{\beta_3 \log n} \left(\frac{en}{k}\right)^k k^{k+1} \left(\frac{c}{n}\right)^k e^{-ck + \frac{ck^2}{n}} = O(1).$$

Thus, Markov's inequality implies that whp, $\sum_{k=1}^{\beta_3 \log n} Z_k = o(n)$.

Intermediate Summary: At this point, we know that whp: **(1)** There are no components of order in $[\beta_3 \log n, \beta_2 n]$. **(2)** The total volume of components that are trees of order up to $\beta_3 \log n$ is $(1 + o(1))\frac{nx}{c}$. **(3)** The total volume of components $\leq \beta_3 \log n$ with a cycle is negligible ($o(n)$). Therefore, whp, the remaining volume $(1 + o(1))(1 - \frac{x}{c})n$ must be taken up entirely by connected components of order at least $\beta_2 n$.

Final Step (Sprinkling): It remains to show that whp, this entire volume of about $(1 - \frac{x}{c})n$ is taken up by a **single component**. We will prove this using the **sprinkling** method.
We have $p = \frac{c}{n}$ for $c > 1$. Let us define $c_1 = c - \frac{\log n}{n}$ and its corresponding $p_1 = \frac{c_1}{n}$; thus p_1 is only slightly smaller than p. We define the "sprinkled" probability p_2 by the relation $1-p = (1-p_1)(1-p_2)$. Let us sample $G_1 \sim G(n, p_1)$ and $G_2 \sim G(n, p_2)$ independently, so that $G = G_1 \cup G_2$. We define $x_1 = x(c_1)$ by $x_1 \in (0, 1)$ and $x_1 e^{-x_1} = c_1 e^{-c_1}$. Notice that $x_1 = x(1 - o(1))$. Applying the previous considerations to G_1, we get that whp the volume $(1 + o(1))\left(1 - \frac{x_1}{c_1}\right)n = (1 + o(1))\left(1 - \frac{x}{c}\right)n$ is taken up by components of order $\geq \beta_2 n$.
The total number of these large components is bounded by a constant: $\ell \leq \frac{(1-\frac{x}{c})n(1+o(1))}{\beta_2 n} = O(1)$. Suppose that $C_1, \ldots, C_\ell$ are these components in G_1, where $|C_i| \geq \beta_2 n$.
When we "sprinkle" the edges of G_2, the probability that two such huge components fail to connect is very small:

$$\mathbb{P}\begin{bmatrix}\text{no edge between } C_i \\ \text{and } C_j \text{ in } G(n, p_2)\end{bmatrix} = (1 - p_2)^{|C_i| \cdot |C_j|} \leq e^{-p_2 \Theta(n^2)} = n^{-\Theta(1)} = o(1),$$

where we used here the following fact: since $1 - p = (1 - p_1)(1 - p_2)$ implies $p_2 \geq p - p_1 = \frac{\log n}{n^2}$. Therefore, after sprinkling, whp, all the large components $C_1, \ldots, C_\ell$ merge into a single giant component L_1. Its total volume is exactly the volume that completes the rest of the graph:

$$|L_1| = (1 + o(1))\left(1 - \frac{x}{c}\right)n.$$

This concludes the proof. $\square$

Percolation on Finite Graphs, Tel Aviv Univ., Spring 2026 May 10, 2026

Lecture 4

Lecturer: Prof. Michael Krivelevich *Scribe: Roy Maulbogat*

4.1 Giant Component in $G \sim G\big(n, \frac{c}{n}\big)$ — Alternative Approach

In the previous lecture we looked at $G \sim G\big(n, \frac{c}{n}\big)$ with $c > 1$. We defined $x \in (0, 1)$ to be the unique solution of $xe^{-x} = ce^{-c}$ in that interval, and proved that whp $|L_1| = (1 + o(1))\big(1 - \frac{x}{c}\big)\, n$, meaning the largest component is a giant component of order linear in n. We also proved that all other components are small: $|L_i| = O_c(\log n)$ for all $i \geq 2$. The proof was a long, tedious calculation. We now want to see an alternative explanation for the order of $|L_1|$ that is more intuitive, and can also be formalized into a rigorous proof.
Let $c > 1$, $G \sim G\big(n, \frac{c}{n}\big)$, and $v \in [n]$ be a vertex. Denote by C_v the connected component of v. We want to evaluate the size of C_v. In fact, we would like to evaluate $y := \Pr |C_v|$ is "big". After that we can argue that most largish components actually merge into one giant component, and by linearity of expectation we expect $|L_1| = (1 + o(1))y \cdot n$.
How do we evaluate $|C_v|$? The idea is to expose the neighbors of v, then expose the neighbors of the neighbors of v, and so on. This algorithm is called *Breadth-First Search (BFS)*, which will be presented formally later.
The number of neighbors of v is distributed as $\mathrm{Bin}(n - 1, p)$. More generally, suppose we have exposed C_v up to a certain stage. The number of neighbors of $u \in C_v$ (outside of C_v) is distributed as $\mathrm{Bin}(n - |C_v|, p)$. As long as $|C_v| = o(n)$, we can estimate $\mathrm{Bin}(n - |C_v|, p) \approx \mathrm{Bin}(n, p)$, meaning the expected value is asymptotically the same.
Now, $\mathrm{Bin}\big(n, \frac{c}{n}\big)$ can be approximated by $\mathrm{Po}(c)$, where $\mathrm{Po}(c)$ is the Poisson distribution given by:

$$\mathbb{P}(X = k) = e^{-c} \cdot \frac{c^k}{k!}, \quad k \geq 0.$$

We model the exposure of $|C_v|$ by a *Galton–Watson branching process* with offspring distribution $\mathrm{Po}(c)$:

Definition 4.1.1. We define a *Galton–Watson branching process* with offspring distribution $\mathrm{Po}(c)$ as follows: denote by X_i the size of generation i. Set $X_0 = 1$, and

$$X_{n+1} = \sum_{j=1}^{X_n} \xi_j^{(n)},$$

where $\xi_j^{(n)}$ are independent $\mathrm{Po}(c)$ random variables. We grow a tree: in generation 0 we start from a single root and assign it a number of children drawn from $\mathrm{Po}(c)$. Each child then independently produces its own children according to $\mathrm{Po}(c)$, and so on.

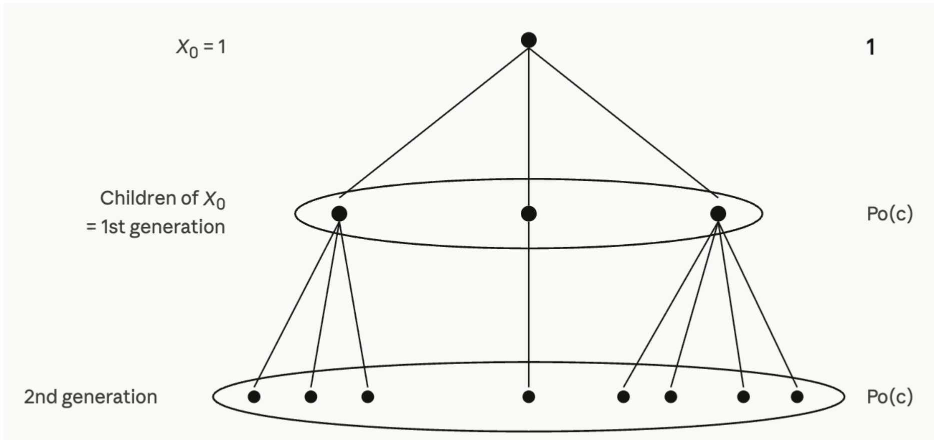


Figure 4.8: Galton–Watson tree.

Let y be the probability that this process is *infinite*, i.e. the survival probability of the dynasty (and $1 - y$ is the probability of extinction). As long as C_v is not too large, this branching process is a good approximation for growing C_v, so y equals the probability that C_v is big.
For the tree rooted at v to *not* survive, every child of the root must also fail to survive. Since the children are independent, we have:

$$1 - y = \sum_{k=0}^{\infty} \mathbb{P}(X = k)\,(1 - y)^k = \sum_{k=0}^{\infty} e^{-c} \cdot \frac{c^k}{k!}(1 - y)^k = e^{-c} \cdot e^{c(1-y)} = e^{-cy}.$$

Hence $1 - y = e^{-cy}$, $y \in (0, 1)$. One can show analytically that this equation has a unique solution for every $c > 1$, from which we can conclude (by further work) that whp $|L_1| = (1 + o(1))y \cdot n$.
In Theorem 2.3.1 we have shown that $|L_1| = (1 + o(1))\left(1 - \frac{x}{c}\right) \cdot n$, hence we have to verify that $y = 1 - \frac{x}{c}$. Recall that $xe^{-x} = ce^{-c}$, thus:

$$\frac{x}{c} = e^{-c+x} = e^{-c(1-x/c)} \implies 1 - \frac{x}{c} = 1 - e^{-c(1-x/c)}.$$

Since y satisfies $y = 1 - e^{-cy}$, the identify $y = 1 - \frac{x}{c}$ follows.

Remark 4.1.2. When $c = 1 + \varepsilon$ for some small $\varepsilon > 0$, one can show that $y = (1 + o_\varepsilon(1))\,2\varepsilon$. Hence, if $G \sim G\big(n, \frac{c}{n}\big)$ with $c = 1 + \varepsilon$ for a small constant $\varepsilon > 0$, then whp $|L_1| = (1 + o_\varepsilon(1))\,2\varepsilon n$.

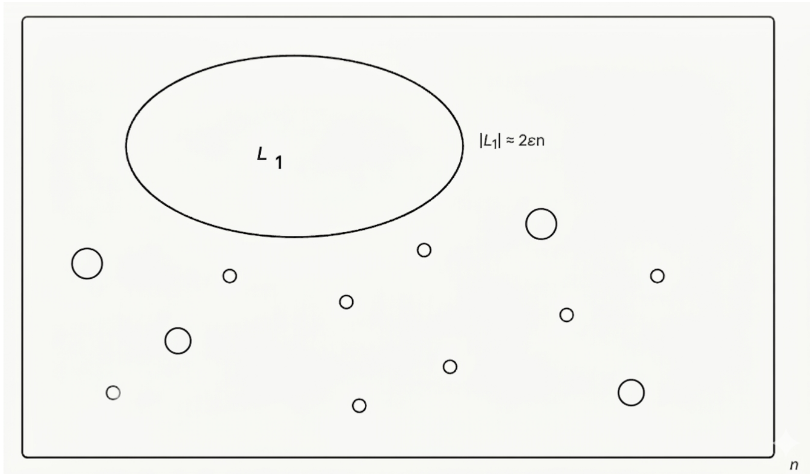


Figure 4.9: Linear approximation $y \approx 2\varepsilon$ near the critical point $c = 1 + \varepsilon$.

4.2 Long Paths and Cycles in Supercritical/Sparse Random Graphs

We will show that for $G \sim G\left(n, \frac{1+\varepsilon}{n}\right)$ (the supercritical case), G contains whp a path of length $\Theta(n)$, and when $G \sim G\left(n, \frac{c}{n}\right)$ for a constant $c > 1$ (sparse random graph), G contains a path of length $n(1-\delta)$, where $\delta \coloneqq \delta(c)$ is a constant satisfying $\lim_{c\to\infty} \delta(c) = 0$.
First, we need some background in graph-search algorithms.

4.2.1 Breadth-First Search (BFS)

Input: A graph $G = (V, E)$ and a permutation σ on V determining the order of vertices. **Output:** The connected components of G.
The algorithm maintains and updates the partition $V = S \cup Q \cup T$, where:

- S = vertices whose processing has been completed;
- Q = vertices currently being processed (**FIFO**[1] queue);
- T = vertices waiting to be processed.

Initialize $S = Q = \emptyset$, $T = V$. The algorithm terminates once $Q = T = \emptyset$, $S = V$.

Algorithm step.

- If $Q \neq \emptyset$: let v be the first vertex in Q. Scan T according to σ and look for a neighbor of v. If a neighbor u is found, move u from T to Q. Otherwise, move v from Q to S.
- If $Q = \emptyset$: take the first vertex v in T (according to σ) and move it to Q.

The algorithm never closes a cycle, and it outputs a spanning forest F with $V(F) = V(G)$ whose components coincide with those of G. The time between Q becoming non-empty and becoming empty again is called an *epoch*, corresponding to discovering one connected component.

[1] First-in-First-out.

4.2.2 Depth-First Search (DFS)

Input: A graph $G = (V, E)$ and a permutation σ on V.
Output: The connected components of G.
The algorithm maintains the partition $V = S \cup U \cup T$, where:

- S = vertices we have completed processing;
- U = vertices currently being processed (**LIFO**[2] stack);
- T = vertices waiting to be processed.

The only difference from BFS is that when $U \neq \emptyset$, we let v be the *last* vertex in U (according to σ) and search for neighbors of v in T.

Example of DFS. Consider the graph in Figure 4.10.

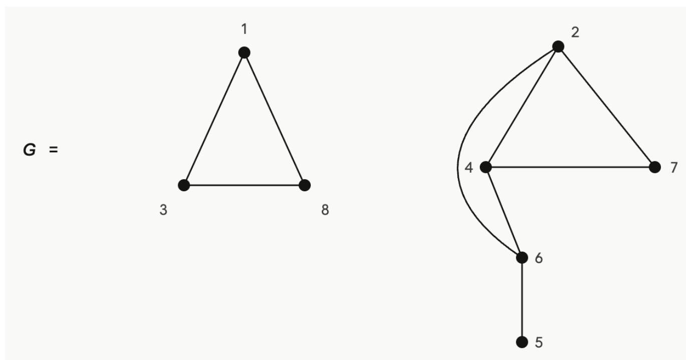


Figure 4.10: Example graph for the DFS walkthrough.

1. Start with $S = U = \emptyset$, $T = V$.
2. Move 1 from T to U.
3. Discover neighbor 3 of 1; move 3 to U.
4. Discover neighbor 8 of 3; move 8 to U.
5. Move 8 to S.
6. Move 3 to S.
7. Move 1 to S.
8. Move 2 from T to U.
9. Discover neighbor 4 of 2; move 4 to U.
10. Discover neighbor 6 of 4; move 6 to U.
11. Discover neighbor 5 of 6; move 5 to U.
12. Move 5 to S.

[2]Last-in-First-out.

13. Move 6 to S.
14. Discover neighbor 7 of 4; move 7 to U.
15. Move 7 to S.
16. Move 4 to S.
17. Move 2 to S.

The resulting spanning forest is shown in Figure 4.11. Note that at every step, U is a path in G.

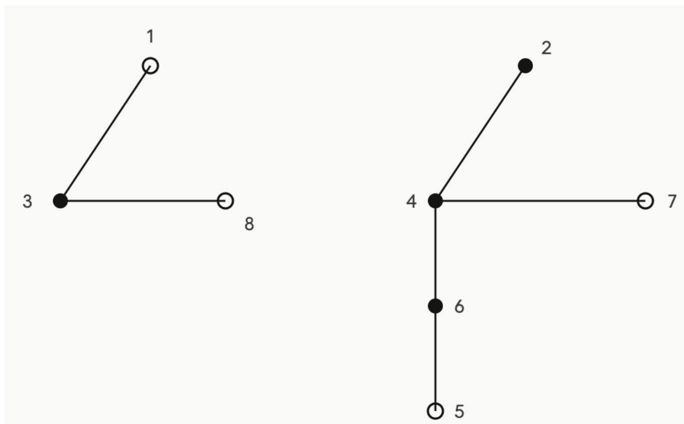


Figure 4.11: Spanning forest produced by DFS on the example graph.

Basic properties of DFS.

1. At each step, exactly one vertex moves: either from T to U, or from U to S.
2. At each step, there are no edges between the current S and the current T.
3. At each step, U spans a *path* in G, ordered by the order in which vertices were added to U (Whenever we add u to U, it is adjacent to the last vertex of U, and thus extends the current path. Removing the last vertex of U shortens the path.).
4. The algorithm generates a spanning forest F with $V(F) = V(G)$ and the same components as G. If $e = (u, v) \in E(G) \setminus E(F)$, then one of u, v is an ancestor of the other in a tree of F (see Figure 4.12).

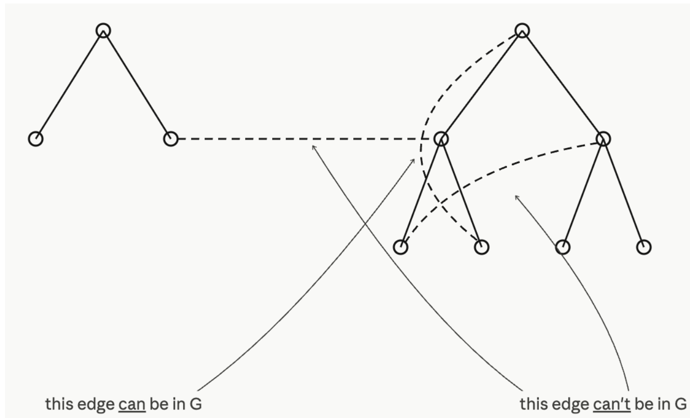


Figure 4.12: A back edge $(u, v) \in E(G) \setminus E(F)$: one endpoint is an ancestor of the other in F.

4.2.3 Using DFS to Find Long Paths and Cycles

Definition 4.2.1. For $G = (V, E)$ and $S \subseteq V$, the *external neighborhood* of S is $N(S) = \{u \in V \setminus S : u \text{ has a neighbor in } S\}$.

Theorem 4.2.2. *Let $k, \ell > 0$ be positive integers and $G = (V, E)$ a graph with $|V| > k$. Suppose that for every $S \subseteq V$ with $|S| = k$ we have $|N(S)| \geq \ell$. Then G contains a path of length at least ℓ.*

Proof. Fix an arbitrary order σ on V and run DFS on (G, σ). Consider the first step for which $|S| = k$ (such a step exists because $|V| > k$ and vertices move one by one). By assumption $|N(S)| \geq \ell$. By Property 2, there are no edges between S and T, hence $N(S) \subseteq U$, and therefore $|U| \geq \ell$. By Property 3, U spans a path, so G contains a path on at least ℓ vertices, i.e. of length at least $\ell - 1$. Moreover, the last operation was moving a vertex v from U to S; just before that move, U contained v plus at least ℓ other vertices, giving a path of length at least ℓ. □

Theorem 4.2.3 (Ben-Eliezer, Krivelevich, Sudakov 2012)**.** *Let $k < n$ be positive integers and $G = (V, E)$ a graph with $|V| = n$. Suppose that for every pair $A, B \subseteq V$ with $|A| = |B| = k$ and $A \cap B = \emptyset$ there is an edge in G between A and B. Then G has a path of length at least $n - 2k + 1$ and a cycle of length at least $n - 4k + 4$ (provided this is positive).*

Proof. Fix an arbitrary order σ on V and run DFS on (G, σ). By Property 1, there exists a step at which $|S| = |T|$. By Property 2, there are no edges between S and T. The theorem's assumption then forces $|S| = |T| < k$, Hence:

$$|U| = n - |S| - |T| \geq n - 2(k - 1) = n - 2k + 2.$$

By Property 3, U spans a path P in G of length at least $n - 2k + 1$.
To obtain a long cycle, let A be the first k vertices of P and B the last k vertices (see Figure 4.13). These sets are disjoint (as $n - 4k + 4 > 0$), and hence by the theorem's assumption, there is an edge between A and B, which closes a cycle of length at least

$$|V(P)| - 2(k - 1) \geq n - 4k + 4.$$

□

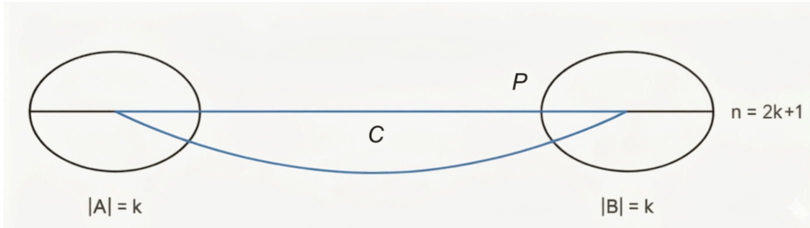


Figure 4.13: Closing a long cycle.

Remark 4.2.4. DFS is also applicable to directed graphs. Essentially, the same proof gives: if $G = (V, E)$ is a directed graph on n vertices, k is a positive integer, and for every ordered pair (A, B) with $|A| = |B| = k$, $A \cap B = \emptyset$, there is an edge from A to B in G, then G contains a directed path of length at least $n - 2k + 1$ and a directed cycle of length at least $n - 4k + 4$.

4.2.4 Using DFS for Random Graphs

Let $G \sim G(n, p)$. We want to find long paths and cycles in G whp by running DFS and exposing G simultaneously.
Let $N = \binom{n}{2}$ and let $\bar{X} = (X_i)_{i=1}^N$ be an i.i.d. sequence of Bernoulli(p) random bits. Fix an ordering on $[n]$ (e.g. the natural order) and run DFS on G using $\bar{X}$ as follows: when DFS makes its i-th query —"is $(u, v) \in E(G)$?"— answer yes if $X_i = 1$, and no if $X_i = 0$. After DFS completes and all components are discovered, answer any remaining unqueried edges from the same sequence $\bar{X}$. The resulting graph is distributed as $G(n, p)$.
Note that a graph is (in a sense) a "2-dimensional" object, while the sequence $\bar{X} = (X_i)_{i=1}^N$ is 1-dimensional. The above approach turn DFS into a technique for "dimension reduction" from 2 to 1.

Percolation on Finite Graphs, Tel Aviv Univ., Spring 2026 May 17, 2026

Lecture 5

Lecturer: Prof. Michael Krivelevich *Scribe: Tom Guy*

5.1 Long Paths and Cycles in $G \sim G(n, p)$

Theorem 5.1.1 (Michael Krivelevich, Benny Sudakov 2013)**.** *For every sufficiently small $\varepsilon > 0$, a random graph $G \sim G\big(n, \frac{1+\varepsilon}{n}\big)$ whp contains a path of size at least $\frac{\varepsilon^2 n}{5}$.*

Proof. Run DFS on G, revealing the edges as the algorithm progresses, based on a vector $\bar{X} = (X_i)_{i=1}^{N}$, where $X_i \sim \text{Bernoulli}\big(\frac{1+\varepsilon}{n}\big)$ i.i.d. ($N = \binom{n}{2}$).
At each step t: if $T \neq \emptyset$, the algorithm checks whether the current top-of-stack vertex in U has a neighbor in T, in which case it moves a vertex from T into U. For each $i \leq t$, every value $X_i = 1$ adds a vertex to U. This vertex might move later to S. Concretely, after t steps, if $T \neq \emptyset$, we have that

$$|S \cup U| \geq \sum_{i=1}^{t} X_i, \qquad |U| \leq 1 + \sum_{i=1}^{t} X_i. \tag{5.2}$$

An explanation for the second inequality in (5.2): the process starts with U as an empty set. The first vertex from T that is added to U is added for free, and then, at step $i \leq t$, a new vertex moves from T to U due to having the corresponding random variable $X_j = 1$ for every $j \leq i \leq t$.
Let $N_0 = \frac{\varepsilon n^2}{2}$ (think of N_0 as an integer number). According to Chebyshev's (or Chernoff's) inequality, we have whp that:

$$\left| \sum_{i=1}^{N_0} X_i - \frac{\varepsilon n^2}{2} \cdot \frac{1+\varepsilon}{n} \right| \leq n^{2/3}.$$

We aim to show that whp at time N_0 we have $|U| > \frac{\varepsilon^2 n}{5}$. First, we will show that whp, at N_0, it must be that $|S| \leq \frac{n}{3}$. Assume by contradiction that $|S| > \frac{n}{3}$. Then, look at the first time t for which $|S| = \frac{n}{3}$. Notice that $t \leq N_0$, and hence

$$|U| \leq 1 + \sum_{i=1}^{t} X_i \leq 1 + \sum_{i=1}^{N_0} X_i \leq \frac{n}{3},$$

where the last inequality holds for sufficiently small $\varepsilon > 0$. Now,

$$|T| = n - |S| - |U| \geq \frac{n}{3}.$$

We know that all queries between S and T (at time t) have been asked (and returned a negative answer), thus:

$$|S| \cdot |T| \geq \frac{n^2}{9} \quad \Longrightarrow \quad \frac{\varepsilon n^2}{2} = N_0 \geq t \geq |S| \cdot |T| \geq \frac{n^2}{9},$$

a contradiction for sufficiently small ε.

Now, $|S| \le \frac{n}{3}$, we assume by contradiction that $|U| \le \frac{\varepsilon^2 n}{5}$ (so in particular $T \neq \emptyset$). Since whp

$$|S \cup U| \ge \sum_{i=1}^{N_0} X_i \ge \frac{\varepsilon(1+\varepsilon)n}{2} - n^{2/3},$$

we have:

$$|S| \ge \frac{\varepsilon(1+\varepsilon)}{2} n - n^{2/3} - \frac{\varepsilon^2 n}{5} = \frac{\varepsilon n}{2} + \frac{3\varepsilon^2 n}{10} - n^{2/3},$$

and as before, all queries between S and T have been asked (and answered negatively). Thus:

$$N_0 \ge |S| \cdot |T| \ge |S| \left(n - |S| - \frac{\varepsilon^2 n}{5} \right).$$

Using $|S| \le \frac{n}{3}$ (and the function $f(x) = x \left(n - x - \frac{\varepsilon^2 n}{5} \right)$ increases in $[0, n/3]$), we derive:

$$\begin{aligned} N_0 &\ge \left(\frac{\varepsilon n}{2} + \frac{3\varepsilon^2 n}{10} - n^{2/3} \right) \left(n - \frac{\varepsilon n}{2} - \frac{\varepsilon^2 n}{5} + n^{2/3} \right) \\ &= \frac{\varepsilon n^2}{2} + \frac{3\varepsilon^2 n^2}{10} - \frac{\varepsilon^2 n}{4} - O(\varepsilon^3) n^2 \\ &= \frac{\varepsilon n^2}{2} + \frac{\varepsilon^2 n^2}{20} - O(\varepsilon^3) n^2 > \varepsilon n^2/2, \end{aligned}$$

giving the desired contradiction. We have shown that typically at time N_0 we have $|U| > \frac{\varepsilon^2 n}{5}$. Therefore, $G \sim G\big(n, \frac{1+\varepsilon}{n}\big)$ has a path of size at least $\frac{\varepsilon^2 n}{5}$. $\square$

Remark 5.1.2.

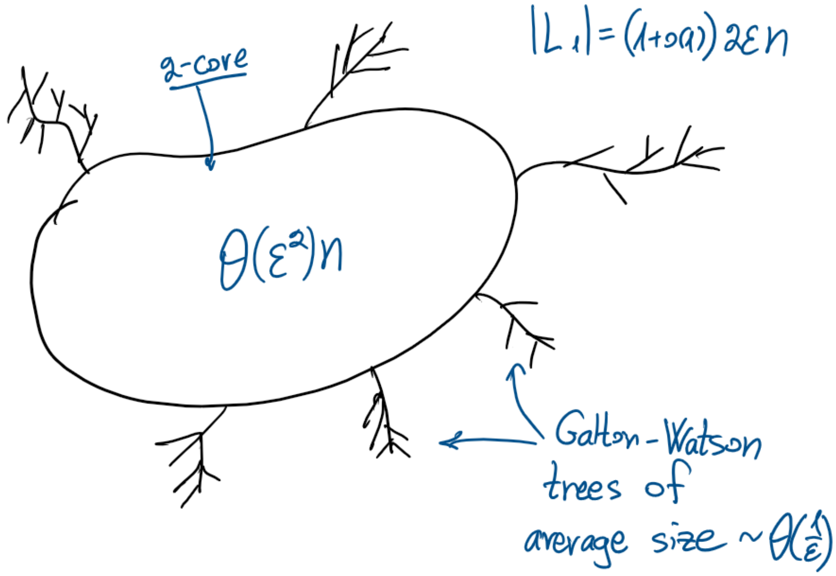
1. Ajtai, Komlós, Szemerédi (1981) showed: In $G \sim G(n, c/n)$ for $c > 1$, whp there is a path with $\Theta(n)$ vertices.

2. How to get a cycle from a path? For path P of length $\Theta(n)$, take $o(n)$ vertices from each end of P and connect them bu an edge whp using sprinkling. Thus, the questions on long paths and long cycles are the same asymptotically.

3. We have shown that for $\varepsilon > 0$, in $G \sim G\big(n, \frac{1+\varepsilon}{n}\big)$ there is typically a path/cycle of size $\frac{\varepsilon^2 n}{5}$. The largest a cycle could be is $|L_1|$, which is whp $(1 + o_\varepsilon(1))2\varepsilon n$. It is known that for $p = \frac{1+\varepsilon}{n}$, where $\varepsilon > 0$ small enough, whp the size of the longest cycle is $\Theta(\varepsilon^2)n$.

Open Question 5.1.3. What is the right constant before $\varepsilon^2 n$ for the typical length of a longest cycle in this regime?

Remark 5.1.4. How does L_1 look whp?



5.1.1 Long Paths in Sparse Random Graphs

We can ask ourselves which probability $p(n)$ does it take to have whp a cycle of length n in $G(n,p)$, i.e. a Hamilton cycle.
Answer: The required $p(n)$ is $p = \frac{\ln n + \ln\ln n + \omega(n)}{n}$, where $\lim_{n\to\infty}\omega(n) = \infty$.
If we are willing to settle for a nearly spanning cycle (path), the required probability $p(n)$ is much lower, as given by the following theorem.

Theorem 5.1.5 (Ajtai–Komlós–Szemerédi 1981, Fernandez de la Vega 1979)**.** *For every $\varepsilon > 0$, there exists $C = C(\varepsilon) > 0$ such that whp, for $G \sim G\big(n, \frac{C}{n}\big)$, a longest path has size of at least $(1-\varepsilon)n$.*

Proof. We have shown that for $n > 2k-2$, and for a graph G on n vertices, if for every $A, B \subseteq V(G)$ with $|A| = |B| = k$, $A \cap B = \emptyset$, there is an edge between A and B, then G contains a path of size $n - 2k + 2$.
Let $k = \left\lfloor \frac{\varepsilon n}{2} \right\rfloor$. We will prove that for large enough $C = C(\varepsilon)$, the graph $G \sim G(n, C/n)$ will contain whp at least one edge between every pair of disjoint sets of size k.
The probability that $G \sim G(n, C/n)$ does not satisfy the above property is at most:

$$\underbrace{\binom{n}{k}}_{\text{choosing } A}\underbrace{\binom{n-k}{k}}_{\text{choosing } B}\underbrace{(1-p)^{k^2}}_{\substack{\text{no edges between}\\ A \text{ and } B}} \le \binom{n}{k}^2 (1-p)^{k^2} \le \left(\frac{en}{k}\right)^{2k} e^{-pk^2} \le \left[\left(\frac{en}{k}\right)^2 e^{-pk}\right]^k,$$

where we use $\binom{n}{k} \le \left(\frac{en}{k}\right)^k$ and $1 - p \le e^{-p}$. Now, for our k, we have $(\frac{en}{k})^{2k} = \left(O\big(\frac{1}{\varepsilon^2}\big)\right)^k$, thus by taking $C = \frac{5\ln(1/\varepsilon)}{\varepsilon}$ we get that the above probability is $o_n(1)$ (for small enough ε).
We thus conclude that for $G \sim G(n,p)$ with $p = \frac{5\log(2/\varepsilon)}{\varepsilon n}$, whp G contains a path of size at least $(1-\varepsilon)n$. □

Remark 5.1.6. It is possible to show that for $C = C(\epsilon)$ large enough, a random graph $G \sim G\left(n, \frac{C}{n}\right)$ whp contains also a cycle of length $\ge (1-\epsilon)n$, either through sprinkling or the deterministic argument for the existence of a cycle of length at least $n - 4k + 4$.

5.2 Hitting Time for Connectivity

First, we define the notion of the hitting time for a non-trivial monotone property on graphs on n vertices.

Definition 5.2.1. Let $\tilde{G} = (G_i)_{i=0}^N$ be a graph process, and let $\mathcal{A}$ be a non-trivial monotone property on graphs on n vertices. The *hitting time* $\tau_{\mathcal{A}}(\tilde{G})$ of $\mathcal{A}$ with respect to $\tilde{G}$ is the minimum index i such that $G_i \in \mathcal{A}$, that is,

$$\tau_A(\tilde{G}) = \min\{i : G_i \in \mathcal{A}\}.$$

Remark 5.2.2. When $\tilde{G}$ is a random graph process, then $\tau_{\mathcal{A}}(\tilde{G})$ is a random variable equal to the first time $\tilde{G}$ hits $\mathcal{A}$, and it is possible to study its typical properties.

We study the hitting time of a random graph process $\tilde{G}$ to become connected and denote it by $\tau_c(\tilde{G})$.

Characterizations of Graph Connectivity:

1. G is connected iff G contains a spanning tree.
2. G is connected iff for every partition $V(G) = A \cup B$ such that $\emptyset \neq A, B \subsetneq V(G)$, G has an edge between A and B.

A possible reason for a graph G to be disconnected is the existence of isolated vertices. We denote by $\tau_1(\tilde{G})$ the hitting time of the property $\delta(G) \geq 1$. Notice that $\tau_C(\tilde{G}) \geq \tau_1(\tilde{G})$ deterministically, i.e. in any graph process. It turns out that for a random graph process, the hitting time for connectivity typically *exactly* equals the hitting time for minimum degree ≥ 1. In other words, typically as soon as the last isolated vertex disappears, the graph becomes connected:

Theorem 5.2.3 (Bollobás–Thomeson, 1985)**.** *In a random graph process $\tilde{G}$, whp $\tau_c(\tilde{G}) = \tau_1(\tilde{G})$.*

Before proving the statement we study the typical disappearance of isolated vertices in $G(n, p)$. Our goal is to find the threshold function $p := p(n)$ for this (monotone) property.
For a vertex $v \in [n]$, we have $\mathbb{P}(v \text{ is isolated}) = (1-p)^{n-1}$, and hence:

$$\mathbb{E}[\#\text{isolated vertices in } G(n,p)] = n \cdot (1-p)^{n-1} \sim ne^{-pn}.$$

Notice that for $p = \frac{\ln n}{n}$, one has $ne^{-pn} = 1$. Hence we can guess that typically isolated vertices disappear entirely after $p = \frac{\ln n}{n}$, and we will show the following:

Theorem 5.2.4. *Let $p := p(n) = \frac{\ln(n)+c(n)}{n}$. Then:*

1. *If $\lim_{n\to\infty} c(n) = \infty$, then whp $G \sim G(n,p)$ contains no isolated vertices.*
2. *If $\lim_{n\to\infty} c(n) = -\infty$, then whp $G \sim G(n,p)$ has isolated vertices.*

Remark 5.2.5. This theorem shows that $p = \frac{\ln n}{n}$ is a (very) sharp threshold function for the absence of isolated vertices.

Proof. Case (1): Let X denote the number of isolated vertices in $G \sim G(n,p)$. We have:

$$\mathbb{E}[X] = n(1-p)^{n-1} = \frac{n}{1-p}(1-p)^n \leq \frac{ne^{-pn}}{1-p} = ne^{-\frac{\ln n + c(n)}{n} \cdot n} = \frac{e^{-c(n)}}{1-p} \xrightarrow{n\to\infty} 0.$$

By Markov's inequality whp $X = 0$ (no isolated vertices).
Case (2): $\lim_{n\to\infty} c(n) = -\infty$, $p(n) = \frac{\ln n + c(n)}{n}$.
We need a lower bound for the expectation. Using $1 - x \geq e^{-\frac{x}{1-x}}$ for $0 \leq x < 1$, we get:

$$\begin{aligned}\mu := \mathbb{E}[X] = n(1-p)^{n-1} &\geq ne^{-(n-1)\frac{p}{1-p}} = n \cdot e^{-\frac{np}{1-p}} \cdot e^{\frac{p}{1-p}} \geq ne^{-\frac{np}{1-p}} \\ &= ne^{-\frac{\ln n + c(n)}{1-p}} = \exp\left(\ln n - \frac{\ln n}{1-p} - \frac{c(n)}{1-p}\right) \\ &= \exp\left(-\frac{p\ln n}{1-p} - \frac{c(n)}{1-p}\right) \xrightarrow{n\to\infty} \infty.\end{aligned}$$

Notice that we can assume that $|c(n)| \leq \ln\ln n$ (since the property of having no isolated vertices is monotone (increasing), hence by showing it for small $|c(n)|$ implies to larger $|c(n)|$).

We use the second moment method. For this, we compute $\mathbb{E}[X^2]$. Notice that $X = \sum_{i=1}^{n} X_i$, where X_i is the indicator of vertex i being isolated:

$$X_i = \begin{cases} 1 & d(i) = 0; \\ 0 & d(i) > 0. \end{cases}$$

Then:

$$\begin{aligned}\mathbb{E}[X^2] = \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] &= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + \sum_{i\neq j} \mathbb{E}[X_i X_j] = n\mathbb{E}[X_1^2] + n(n-1)\mathbb{E}[X_1 X_2] \\ &= n(1-p)^{n-1} + n(n-1)(1-p)^{2n-3} \leq \mu + n^2 \cdot \frac{(1-p)^{2n-2}}{1-p} = \mu + \frac{\mu^2}{1-p},\end{aligned}$$

where the penultimate equality holds due to obvious symmetry. Therefore:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mu + \frac{\mu^2}{1-p} - \mu^2 = \frac{\mu^2 p}{1-p} + \mu = o(\mu^2).$$

By Chebyshev's inequality:

$$\mathbb{P}(X = 0) \leq \frac{\mathrm{Var}(X)}{\mu^2} = \frac{o(\mu^2)}{\mu^2} = o(1),$$

hence whp $X > 0$, i.e. G has some isolated vertices. $\square$

6.1 Hitting Time for Connectivity - Continued

Theorem 6.1.1 (Bollobás–Thomason, 1985)**.** *In a random graph process $\tilde{G}$, whp $\tau_c(\tilde{G}) = \tau_1(\tilde{G})$.*

Proof. Define

$$m_1 = \frac{n-1}{2}(\ln n - \ln\ln n); \quad p_1 = \frac{m_1}{N} = \frac{\ln n - \ln\ln n}{n},$$

$$m_2 = \frac{n-1}{2}(\ln n + \ln\ln n); \quad p_2 = \frac{m_2}{N} = \frac{\ln n + \ln\ln n}{n}.$$

We know that for a random graph process $\tilde{G} = (G_i)_{i=0}^N$ we have: $G_i \sim G(n,i)$, for every $0 \le i \le N$. We will prove that for a random graph process, whp:

1. In G_{m_1}: there are isolated vertices, but no more than $\ln^2 n$ such vertices. All other vertices belong to the same connected component L_1;

2. In G_{m_2}: there are no isolated vertices;

3. If we denote the set of isolated vertices in G_{m_1} by V_0, then the process $\tilde{G} = (e_1, \dots, e_N)$ satisfies the following: for every $m_1 < i \le m_2$, $e_i \nsubseteq V_0$.

First, we claim that if a graph process $\tilde{G}$ satisfies these properties, then $\tau_c(\tilde{G}) = \tau_1(\tilde{G})$ (deterministically). Indeed, from property 1, $\tau_1(G) > m_1$. From property 2, $\tau_1(G) \le m_2$. Hence, for every $v \in V_0$, there exists $m_1 < i \le m_2$ such that $v \in e_i$. From property 1: $L_1 = G - V_0$, thus at time m_2, by property 3, all vertices of V_0 have joined L_1. Therefore, if we look at the time $\tau_1(G)$ (when the last isolated vertex receives an edge), then the graph becomes connected at that exact moment, and therefore $\tau_1(G) = \tau_c(G)$.
Now, we prove properties 1-3:

Proof of property 1: Denote by X the number of isolated vertices in G_{m_1}. We prove that whp $0 < X \le \ln^2 n$. We already proved that $X > 0$ whp in $G(n, p_1)$. Note that:

$$\mathbb{E}[X] = n \cdot (1-p_1)^{n-1} \le n \exp(-p_1 \cdot (n-1)) \approx n \exp(-p_1 n) = \ln n.$$

Therefore, by Markov's inequality, we derive that $\mathbb{P}(X \ge \ln^2 n) < \frac{1}{\ln n}$, hence whp $X < \ln^2 n$. The properties $X \le \ln^2 n$ and $X > 0$ are monotone (increasing and decreasing respectively), therefore we have that whp $0 < X \le \ln^2 n$ in $G_{m_1} \sim G(n, m_1)$ as well (by Proposition 2.2.2). It is left to prove that whp $G(n, m_1)$ contains no connected components C such that $2 \le |C| = k \le \frac{n}{2}$. For a fixed vertex set U with $|U| = k$, we have that:

$$\mathbb{P}\left(U \text{ is a connected component}\right) \le k^{k-2} \cdot p_1^{k-1}(1-p_1)^{k\cdot(n-k)}.$$

However, for large values of k, it is better instead of the above estimate to look at the probability that U is disconnected from $[n] \setminus U$:

$$\mathbb{P}\left(U \text{ is a connected component}\right) \le \mathbb{P}\left(U \text{ is disconnected from } G - U\right) \le (1-p_1)^{k\cdot(n-k)}.$$

Overall we get:

$$\mathbb{P}(\exists \text{ connected component } C : 2 \le |C| \le \frac{n}{2}) = \sum_{k=2}^{n/2} \binom{n}{k} \cdot (1-p_1)^{k\cdot(n-k)} \cdot \min\left\{1, k^{k-2} \cdot p_1^{k-1}\right\} = \sum_{k=2}^{n/2} u_k.$$

Whenever $2 \le k \le n^{0.9}$:

$$\begin{aligned} u_k &\le \left(\frac{en}{k}\right)^k k^{k-2} \cdot p_1^{k-1} \cdot \exp(-p_1 k(n-k)) = \frac{1}{k^2 p}(enp_1 \cdot \exp(-(n-k)\cdot p_1))^k \\ &\le n \cdot \left(3\ln n \cdot \exp\left(-0.9n \cdot \frac{\ln n}{n}\right)\right)^k, \end{aligned}$$

where the last inequality holds since $k^2 p \le n^{0.8+o(1)} < n$ and since $n-k \ge 0.9n$. As $k \ge 2$ we get:

$$n \cdot \left(3\ln n \cdot \exp\left(-0.9n \cdot \frac{\ln n}{n}\right)\right)^k \le n \cdot (n^{0.9} \cdot 3\ln n)^k \le n^{-0.3k}.$$

Now we look at the case where $n^{0.9} \le k \le \frac{n}{2}$:

$$u_k \le \left(\frac{en}{k}\right)^k \exp(-k\cdot(n-k)\cdot p_1) \le \left(en^{0.1} \cdot \exp\left(\frac{-np_1}{2}\right)\right)^k \le n^{-0.3k}.$$

Therefore, we get (from convergence of geometric series):

$$\sum_{k=2}^{n/2} u_k \le \sum_{k=2}^{n/2} n^{-0.3k} = O(n^{-0.6}).$$

From the relation between $G(n,m)$ and $G(n,p)$ and from the fact that $O(\sqrt{m_1}) \cdot O(n^{-0.6}) = o(1)$, we have that whp there are no connected components in $G(n, m_1)$ of size $2 \le k \le \frac{n}{2}$. That concludes the proof of property 1.

Proof of property 2: we show that whp $G_{m_2} \sim G(n, m_2)$ has no isolated vertices. Denote by X the number of isolated vertices in $G(n, p_2)$. Notice that:

$$\mathbb{E}[X] = n(1-p_2)^{n-1} \le n(1+o(1)) \cdot \exp(-\ln n - \ln\ln n) = o(1).$$

Therefore, by Markov's inequality whp $X = 0$. From monotonicity we get the same result for $G(n, m_2)$.

Proof of property 3: set V_0 to be the set of isolated vertices in G_{m_1}. We have shown that whp $|V_0| \le \ln^2 n$. We prove that whp no edge $e_i : m_1 < i \le m_2$ is contained in V_0. Indeed,

$$\mathbb{P}(e_i \subseteq V_0 | e_1, \dots e_{i-1}) \le \frac{\binom{|V_0|}{2}}{N-(i-1)} = O\left(\frac{\ln^4 n}{n^2}\right).$$

Therefore, the probability for any e_i to be contained in V_0 for $m_1 < i \le m_2$ is at most (by the union bound):

$$(m_2 - m_1) \cdot O\left(\frac{\ln^4 n}{n^2}\right) = O\left(\frac{\ln^4 n \cdot \ln\ln n}{n}\right) = o(1).$$

□

Corollary 6.1.2 (Erdős-Rényi, 1959)**.** *Assume that* $G \sim G(n,m)$,$m = \frac{n-1}{2} \cdot (\ln n + c(n))$ *then:*

1. *If* $c(n) \to -\infty$, *then whp* G *is not connected.*
2. *If* $c(n) \to +\infty$, *then whp* G *is connected.*

Proof. We showed that in a random graph process $\tilde{G}$ we have whp $\tau_1(\tilde{G}) = \tau_c(\tilde{G})$, hence

1. $m = \binom{n}{2} \cdot \frac{(\ln n + c(n))}{n}$, $c(n) \to -\infty$, then whp G has isolated vertices, therefore not connected.
2. $m = \binom{n}{2} \cdot \frac{(\ln n + c(n))}{n}$, $c(n) \to \infty$, then whp G has no isolated vertices, therefore as $\tau_1(\tilde{G}) = \tau_c(\tilde{G})$ whp we have that G is connected.

□

Corollary 6.1.3. *Let* $G \sim G(n,p)$ *with* $p = \frac{(\ln n + c(n))}{n}$. *Then:*

1. *If* $c(n) \to -\infty$, *then whp* G *is not connected.*
2. *If* $c(n) \to +\infty$, *then whp* G *is connected.*

Proof. From the last theorem and the equivalence between $G(n,m)$ and $G(n,p)$ (connectivity is a monotone property). □

6.2 Perfect Matchings in Random Graphs

Definition 6.2.1. For a graph $G = (V,E)$, a *matching* in G is a set of pairwise-disjoint edges, i.e. $M \subseteq E$, $\forall e_1 e_2 \in M : e_1 \cap e_2 = \emptyset$. A *perfect matching* is a matching of size $\frac{|V|}{2}$. (Can exist only when $|V|$ is even.)

Question 6.2.2. If $G \sim G(n,p)$ (and n is even), what is the minimal $p(n)$ s.t. G contains a perfect matching whp?

Recall that if $p = \frac{\ln n - \omega(n)}{n}$, where $\omega(n) \to -\infty$, then whp G has isolated vertices, and therefore doesn't have a perfect matching. The following theorem shows the positive side of this question:

Theorem 6.2.3. *[Erdős–Rényi, 1966] Suppose* $p = \frac{\ln n + \omega(n)}{n}$, *where* $\omega(n) \to +\infty$ *(and* n *is even). Then whp there exists a perfect matching in* $G \sim G(n,p)$.

There exists a stronger theorem on the hitting time of a perfect matching. Let us denote by $\tau_M(\tilde{G})$ the hitting time of the property "G contains a perfect matching".

Theorem 6.2.4 (Bollobás–Thomason, 1985)**.** *Whp, in a random graph process* $\tilde{G}$, $\tau_M(\tilde{G}) = \tau_1(\tilde{G})$ *(if* n *is even).*

Proof of Theorem 6.2.3. $G \sim G(n,p)$, $p = \frac{\ln n + \omega(n)}{n}$, $\omega(n) \to +\infty$. Notice that (due to monotonicity) we can assume that $\omega(n) \leq \ln n$. Fix a vertex $i \in [n]$, and consider $X = d_G(i) \sim Bin(n-1,p)$. Note that

$$\mathbb{E}[X] = p \cdot (n-1) \geq \ln n(1 - o(1)).$$

By Chernoff's inequality, we get:

$$\mathbb{P}(d_G(i) \leq c \ln n) \leq \exp(-\Theta(\ln n)) = n^{-\Theta(1)}.$$

Assume c is small enough and define:

$$\text{SMALL} = \{i : d_G(i) \leq c \ln n\},\ \text{LARGE} = [n] \setminus \text{SMALL}.$$

Lemma 6.2.5. *Suppose* $p = \frac{\ln n + \omega(n)}{n}$, $\omega(n) \to +\infty$ *and suppose* $G \sim G(n,p)$. *Then:*

1. $\delta(G) \geq 1$;
2. $|SMALL| \leq n^{0.2}$;
3. $\forall u \neq v \in SMALL : dist_G(u,v) \geq 5$ *(this can be any fixed number);*
4. $\forall V_0 \subseteq LARGE : |V_0| \leq \frac{n}{\sqrt{\ln n}} \Rightarrow |N(V_0)| \geq |V_0| \cdot (\ln n)^{1/4}$;
5. $\forall A, B \subseteq [n]$ *disjoint sets s.t.* $|A| = |B| = \frac{n}{2\sqrt{\ln n}}$, *we have:* $E_G(A,B) \neq \emptyset$.

Proof. Proof of property 1: We have already proven this property.
We use (multiple times) the following claim. There exists a small enough $c > 0$ s.t. for every $k > 0$:

$$\mathbb{P}(Bin(n-k,p) \leq c \ln n) \leq n^{-0.9}.$$

Indeed, that probability is:

$$\begin{aligned}
\sum_{i=1}^{c\ln n} \mathbb{P}(Bin(n-k,p) = i) &\leq (1 + c\ln n)\mathbb{P}(Bin(n-k,p) = c\ln n) \\
&= O(\ln n) \cdot \binom{n-k}{c\ln n} \cdot p^{c\ln n} \cdot (1-p)^{n-k-c\ln n} \\
&\leq O(\ln n) \cdot \left(\frac{enp}{c\ln n}\right)^{c\ln n} \cdot \exp(-p(n-k-c\ln n)) \\
&\leq O(\ln n) \cdot \left(\frac{2e}{c}\right)^{c\ln n} e^{-0.95\ln n} = O(\ln n) \cdot \left(\left(\frac{2e}{c}\right)^{c} \cdot e^{-0.95}\right)^{\ln n} \\
&\leq n^{-0.9},
\end{aligned}$$

where the first inequality comes from the monotonicity of the tail in the binomial distribution and the last inequality holds for a sufficiently small c.

Proof of property 2: from symmetry,

$$\mathbb{E}[|\text{SMALL}|] = n \cdot \mathbb{P}(d(1) \leq c\ln n) \leq n \cdot n^{-0.9} = n^{0.1}.$$

Therefore, by Markov's inequality, $\mathbb{P}(|\text{SMALL}| \geq n^{0.2}) \leq n^{-0.1}$, so whp $|\text{SMALL}| \leq n^{0.2}$.

Proof of property 3: we start with distance 1:

$$\begin{aligned}
\mathbb{P}(\exists u,v \in \text{SMALL} : dist_G(u,v) = 1) &= \mathbb{P}(\exists u,v \in \text{SMALL} : (u,v) \in E(G)) \\
&\leq \underbrace{\binom{n}{2}}_{\text{choosing } u,v} \cdot \underbrace{p}_{(u,v)\in E(G)} \cdot \underbrace{\left(\mathbb{P}(Bin(n-2,p) \leq c\ln n)\right)^2}_{\text{pay for } u,v \in \text{SMALL}} \\
&= O(n\ln n) \cdot \left(n^{-0.9}\right)^2 = o(1).
\end{aligned}$$

For distance 2:

$$\mathbb{P}(\exists u,v \in \text{SMALL} : dist_G(u,v) = 2) \leq \binom{n}{3} \cdot p^2 \cdot (\mathbb{P}(Bin(n-3,p) \leq c\ln n))^2 = O(n\ln^2 n) \cdot n^{-1.8} = o(1).$$

Similarly, we can show that whp there are no $u, v \in$ SMALL at distance 3 or 4 from each other.

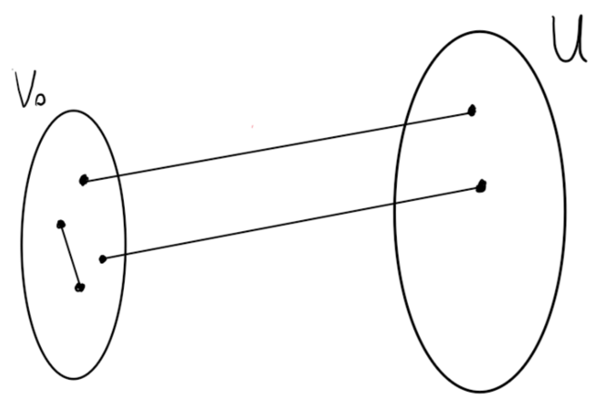
Percolation on Finite Graphs, Tel Aviv Univ., Spring 2026 June 7, 2026

Lecture 7

Lecturer: Prof. Michael Krivelevich *Scribe: Aner Mash*

7.1 Perfect Matchings in Random Graphs - Continued

We continue the proof of Theorem 6.2.3.
Proof of property 4: Assume that $k := |V_0| \leq \frac{n}{\ln^{1/2} n}$. $V_0 \subseteq LARGE$, hence, for every $v \in V_0$, $d(v) \geq c \cdot \ln n$. Thus, V_0 is incident to at least $|V_0| \cdot \frac{c \cdot \ln n}{2} = \frac{k \cdot \ln n}{2}$ edges. These edges lie inside V_0 or between V_0 and $U := N_G(V_0)$ — the external neighborhood of V_0 in G.



$$\mathbb{P}\left(\exists V_0 \subseteq LARGE,\ |V_0| = k,\ 1 \leq k \leq \frac{n}{\ln^{1/2} n},\ |N(V_0)| \leq k \cdot \ln^{1/4} n\right)$$

$$\leq \sum_{k=1}^{n/\ln^{1/2} n} \underbrace{\binom{n}{k}}_{\text{choosing } V_0} \underbrace{\binom{n-k}{k \cdot \ln^{1/4} n}}_{\text{choosing } N_G(V_0)} \underbrace{\binom{\binom{k}{2} + k \cdot k \ln^{1/4} n}{\frac{kc \cdot \ln n}{2}}}_{\text{choosing the edges touching } V_0} \cdot p^{\frac{k \cdot c \ln n}{2}}$$

$$\leq \sum_{k=1}^{n/\ln^{1/2} n} \left(\frac{3n}{k \cdot \ln^{1/4} n}\right)^{k \cdot \ln^{1/4} n} \left(\frac{3k^2 \cdot \ln^{1/4} n}{\frac{k \cdot c \ln n}{2}} \cdot p\right)^{\frac{k \cdot c \ln n}{2}}$$

$$\leq \sum_{k=1}^{n/\ln^{1/2} n} \left[\left(\frac{3n}{k \cdot \ln^{1/4}}\right)^{k \cdot \ln^{1/4} n} \cdot \left(\frac{6k}{c \ln^{3/4} n} \cdot \frac{2 \ln n}{n}\right)^{\frac{c \ln n}{2}}\right]^k$$

$$= \sum_{k=1}^{n/\ln^{1/2} n} \left[\left(\frac{3n}{k \cdot \ln^{1/4} n}\right)^{\ln^{1/4} n} \cdot \left(\frac{12}{c} \cdot \frac{k \ln^{1/4} n}{n}\right)^{\frac{c \ln n}{2}}\right]^k$$

$$\leq \sum_{k=1}^{n/\ln^{1/2} n} \left[\left(\frac{12}{c} \cdot \frac{k \ln^{1/4} n}{n}\right)^{\frac{c \ln n}{4}}\right]^k = o(1).$$

Proof of property 5: It suffices to show that whp, $G \sim G(n,p)$ contains an edge between any two sets A and B satisfying:

$$A, B \subseteq [n], \quad A \cap B = \emptyset, \quad |A| = |B| = \frac{n}{2\ln^{1/2} n}.$$

The probability that this does not happen is at most:

$$\underbrace{\binom{n}{\frac{n}{2\ln^{1/2} n}}^2}_{\text{choosing } A,B} \cdot \underbrace{(1-p)^{\left(\frac{n}{2\ln^{1/2} n}\right)^2}}_{E_G(A,B)=\emptyset} \le \left(2e \cdot \ln^{1/2} n\right)^{\frac{n}{\ln^{1/2} n}} \cdot e^{-\frac{\ln n}{n} \cdot \frac{n^2}{4\ln n}} = \left(2e \cdot \ln^{1/2} n\right)^{\frac{n}{\ln^{1/2} n}} \cdot e^{-\frac{n}{4}}$$

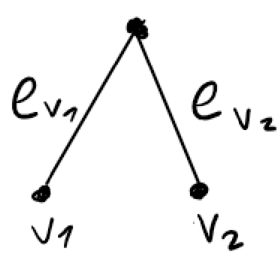
$$\le (\ln n)^{\frac{n}{\ln^{1/2} n}} \cdot e^{-\frac{n}{4}} = \exp\left\{\frac{n \cdot \ln\ln n}{\ln^{1/2} n} - \frac{n}{4}\right\} = o(1).$$

This completes the proof of Lemma 6.2.5. □

We now prove deterministically that if G is a graph on n vertices (n is even and sufficiently large) satisfying properties 1-5, then G contains a perfect matching, proving Theorem 6.2.3. To that end, we use Tutte's Theorem. Before stating Tutte's Theorem, we recall that for a graph $G = (V, E)$, we denote by $o(G)$ the number of odd connected components (i.e., of odd size) in G.

Theorem 7.1.1 (Tutte, 1947)**.** *A graph $G = (V, E)$ contains a perfect matching if and only if for every $S \subseteq V$, we have $o(G - S) \le |S|$.*

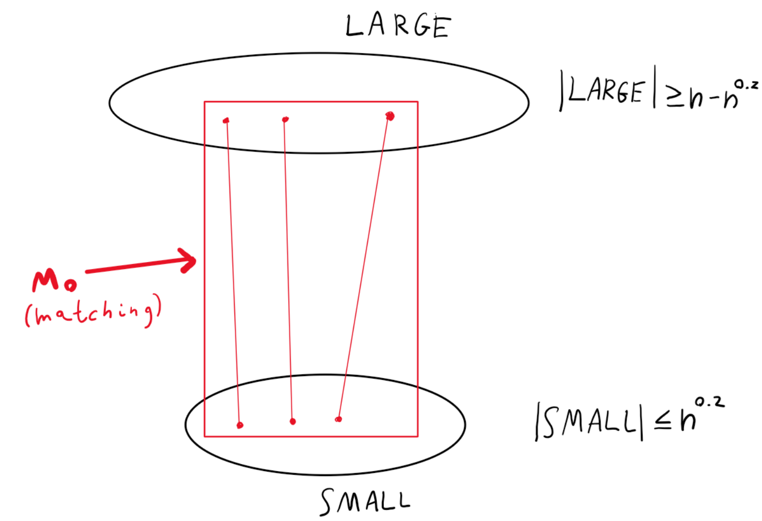
We use this theorem to argue about existence of the desired matching in G. First, for every $v \in$ SMALL, we arbitrarily choose an edge $e_v \in E(G)$ such that $v \in e_v$ (this is possible since property 1 asserts that $\delta(G) \ge 1$).
Note that the chosen edges e_v are disjoint, because if they are not disjoint, then there exist vertices in SMALL at distance of at most 2 from each other, which is a contradiction to property 3.



Therefore, the set of chosen edges

$$M_0 = \{e_v \mid v \in SMALL\} \subseteq E(G)$$

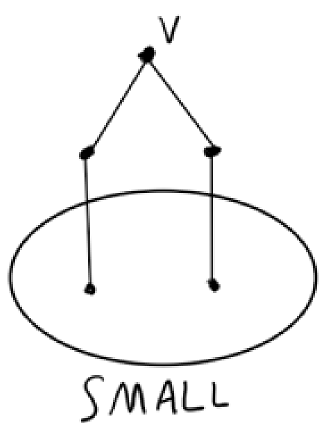
is a matching of size $|M_0| = |SMALL| \le n^{0.2}$ (By property 2).



We construct a perfect matching M in G such that $M_0 \subseteq M$. Let $W \subseteq V$ denote the union of the edges of M_0.

$$|W| = 2 \cdot |M_0| = 2 \cdot |SMALL| \leq 2 \cdot n^{0.2}.$$

For every $v \in V \setminus W$, observe that $v \in LARGE$, hence $d_G(v) \geq c \cdot \ln n$. Moreover, v has at most one neighbor in W, since otherwise there would exist two distinct vertices in SMALL at distance of at most 4 from each other, contradicting property 3).



Denote $U = V \setminus W$. U satisfies $|U| = n - o(n)$. The properties of $G[U]$ are:

(a) $\forall V_0 \subseteq U,\ |V_0| \leq \frac{n}{\ln^{1/2} n}:\ |N(V_0, U)| \geq |N(V_0)| - |V_0| \geq |V_0| \cdot \ln^{1/8} n$ (by property 4).

(b) For every $A, B \subseteq U$ such that $A \cup B = \emptyset$ and $|A|, |B| \geq \frac{n}{2\ln^{1/2} n}$, there is an edge between A and B (by property 5).

We show the existence of a perfect matching in $G_1 := G[U]$ by verifying Tutte's condition for G_1. Let $S \subseteq U$ with $k := |S| \geq 0$. We need to verify that $o(G_1[U \setminus S]) \leq |S|$.
If $k = 0$, we have to show that G_1 is connected. Indeed, if C is a connected component of G_1, then $|C| > \frac{n}{\ln^{1/2} n}$ (every small set expands itself). Otherwise, G_1 has two connected components $C_1 \neq C_2$, $|C_1|, |C_2| \geq \frac{n}{\ln^{1/2} n}$, contradicting property 2 (there is an edge between C_1 and C_2).

So we can assume that $k > 0$. Let $C_1, ..., C_\ell$ denote the connected components (of some parity) in $G_1 - S$. It suffices to show that $\ell \leq k$. WLOG, assume that $|C_1| \geq |C_2| \geq \ldots \geq |C_\ell|$.
First, note that $\alpha(G) \leq \frac{n}{\ln^{1/2} n}$. Otherwise, there exists an independent set I in G with $|I| \geq \frac{n}{\ln^{1/2} n}$, and we can find $A, B \subseteq I$, $A \cap B = \emptyset$, $|A| = |B| \geq \frac{n}{2\ln^{1/2} n}$ such that there is no edge between A and B, in contradiction to property 5.
If $G_1 - S$ has ℓ connected components, then $\alpha(G_1 - S) \geq \ell$ (we choose one representative of each connected component). Hence, $\ell \leq \frac{n}{\ln^{1/2} n}$, so we can assume that $k \leq \frac{n}{\ln^{1/2} n}$.

Case 1: $\sum_{i=2}^{\ell} |C_i| \leq \frac{n}{\ln^{1/2} n}$.
Let $V_0 = \bigcup_{i=2}^{\ell} C_i$. Notice that $\ell - 1 \leq |V_0| \leq \frac{n}{\ln^{1/2} n}$ and $N_{G_1}(V_0) \subseteq S$. From the expansion property 7.1, it follows that:

$$k = |S| \geq |N_G(V_0)| \geq |V_0| \cdot \ln^{1/8} n \geq (\ell - 1) \cdot \ln^{1/8} n,$$

and in particular $k \geq \ell$, as desired.
Case 2: $|V_0| \geq \frac{n}{\ln^{1/2} n}$, $|C_1| \geq \frac{n}{2\ln^{1/2} n}$.
Under these conditions, G contains an edge between C_1 and V_0 — a contradiction.

Case 3: $|\bigcup_{i=2}^{\ell} C_i| = |V_0| \leq \frac{n}{\ln^{1/2} n}$, $|C_1| \leq \frac{n}{2\ln^{1/2} n}$.
Under these conditions, there exists a collection $I \subseteq [\ell]$ of connected components $\{C_i\}_{i \in I}$, such that:

$$\frac{n}{2 \cdot \ln^{1/2} n} \leq |\cup_{i \in I} C_i| \leq \frac{n}{\ln^{1/2} n}$$

(We add components to I one by one until the desired total volume is reached). Hence, the set $\bigcup_{i \in I} C_i$ expands itself:

$$|N_{G_i}(\cup_{i \in I} C_i)| \geq |\cup_{i \in I} C_i| \cdot \ln^{1/8} n = \Theta\left(\frac{n}{\ln^{3/8} n}\right).$$

As before:

$$N_{G_1}(\cup_{i \in I} C_i) \subseteq S,$$

and we assumed that $k = |S| \leq \frac{n}{\ln^{1/2} n}$ — a contradiction.
Overall, We proved that there exists a perfect matching M_1 in $G_1 = G[U] = G[V \setminus W]$, therefore, the union $M := M_0 \cup M_1$ is a perfect matching in the entire graph G, completing the proof of Theorem 6.2.3. $\square$

7.2 A generalization of Cayley's formula for graphs with a given maximum degree

Similar to Cayley's theorem, which determines the number of spanning trees in the complete graph on n vertices, we would like to estimate the number of k-vertex subgraphs of a given graph G that are trees. To that end, we use the following definition:

Definition 7.2.1. Given a graph G, a vertex $v \in V(G)$, and $k \in \mathbb{N}$, let $\mathcal{T}(v, k)$ denote the collection of trees in G of size k rooted at v. Furthermore, we denote $|\mathcal{T}(v, k)| = t(v, k)$.

Theorem 7.2.2 (Cayley's formula)**.** *Let $2 \leq n \in \mathbb{N}$. In K_n, $t(v, n) = n^{n-2}$, for every $v \in V(K_n)$.*

Example 7.2.3. For $k = 1$: $t(v, 1) = 1$. For $k = 2$: $t(v, 2) = d_G(v)$.

Theorem 7.2.4. *Let G be a graph with maximum degree Δ, and let $v \in V(G)$. Then for all $k \geq 2$:*

$$t(v,k) \leq \frac{k^{k-2} \cdot \Delta^{k-1}}{(k-1)!} \leq (e\Delta)^{k-1}.$$

Proof. For a tree $T \in \mathcal{T}(v,k)$, consider a labeling φ of the vertices of T satisfying:

$$\varphi(v) = k, \;\; \varphi : V(T) \to [k], \;\; \varphi \text{ is a bijection.}$$

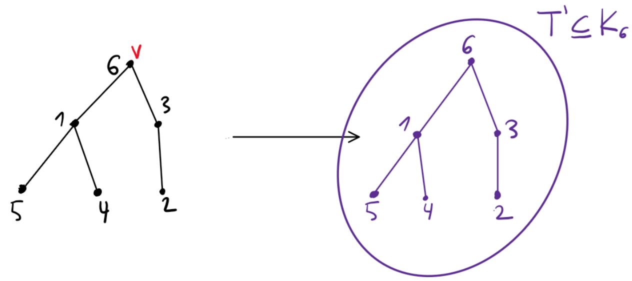
For every $T \in \mathcal{T}(v,k)$ there exist $(k-1)!$ such labelings. Therefore, the total number of pairs (T, φ), where $T \in \mathcal{T}(v,k)$ and φ is a valid labeling, is:

$$|\mathcal{T}(v,k)| \cdot (k-1)! = t(v,k) \cdot (k-1)! \; .$$

Now, given a pair (T, φ), we construct a spanning tree T' in K_k as follows:

$$(i,j) \in E(T') \iff \varphi(x) = i, \;\; \varphi(y) = j \text{ for } x \neq y \in V(T) \text{ and } (x,y) \in E(T).$$

Example



We now bound from above the number of pairs (T, φ) that yield a tree T' in K_k. We run BFS on the tree T' starting from vertex k. When we reach (in the BFS) a vertex $\ell \in [k-1]$, we have to define $\varphi^{-1}(\ell)$: we find the parent $i \in [k]$ of ℓ in T', and choose $\varphi^{-1}(\ell)$ from among the (available) neighbors of $\varphi^{-1}(i)$ — there are at most Δ options. Hence, there are at most Δ^{k-1} ways to choose φ^{-1}. We obtain:

$$t(v,k) \cdot (k-1)! \leq k^{k-2} \cdot \Delta^{k-1} \implies t(v,k) \leq \frac{k^{k-2} \cdot \Delta^{k-1}}{(k-1)!}.$$

Finally, we note that for all $k \geq 2$, one can verify that

$$\frac{k^{k-2}}{(k-1)!} \leq \frac{(k-1)^{k-1}}{(k-1)!} \leq e^{k-1} \implies t(v,k) \leq (e\Delta)^{k-1}.$$

$\square$

Remark 7.2.5. We have proved:

$$\Delta(G) \leq \Delta \implies t(v,k) \leq \frac{k^{k-2} \cdot \Delta^{k-1}}{(k-1)!}.$$

It is easy to see that in fact the essentially same proof yields the following: Let G be a graph with $\delta(G) = \delta \geq k$, then for every $v \in V$ and for every integer $k \geq 2$:

$$t(v,k) \geq \frac{(\delta - k)^{k-1} \cdot k^{k-2}}{(k-1)!}.$$

In particular, if G is a d-regular graph and $d \geq k$, then:

$$\frac{(d-k)^{k-1} \cdot k^{k-2}}{(k-1)!} \leq t(v,k) \leq \frac{d^{k-1} \cdot k^{k-2}}{(k-1)!}.$$

7.3 The Binary (Hyper)cube

Definition 7.3.1. For every integer $d \geq 1$, we define the *binary (hyper)cube* Q^d as follows:

1. $V(Q^d) = \{0,1\}^d$;
2. $\forall \bar{x}, \bar{y} \in V(Q^d) : (\bar{x}, \bar{y}) \in E(Q^d) \iff d_H(x,y) = 1$, i.e., $\bar{x}$ and $\bar{y}$ differ in exactly one coordinate.

Remark 7.3.2. Q^d can be constructed from Q^{d-1} by taking two disjoint copies of Q^{d-1} and connecting the corresponding vertices in both copies by a matching.

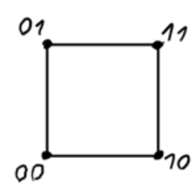
Definition 7.3.3. For $\bar{x} \in V(Q^d)$, we define its *weight*:

$$|\bar{x}| := |\{1 \leq i \leq d \mid x_i = 1\}|.$$

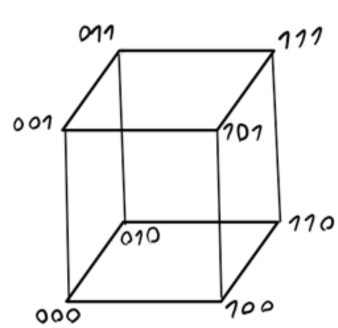
Examples

d=1:

d=2:



d=3:



7.3.1 Basic properties of Q^d

1. $n = |V(Q^d)| = 2^d$.
2. Q^d is a d-regular graph.
3. Q^d is a bipartite graph, with sides:

$$O = \Big\{\bar{x} \in V(Q^d) : |\bar{x}| \text{ is odd}\Big\}, \quad E = \Big\{\bar{x} \in V(Q^d) : |\bar{x}| \text{ is even}\Big\}.$$

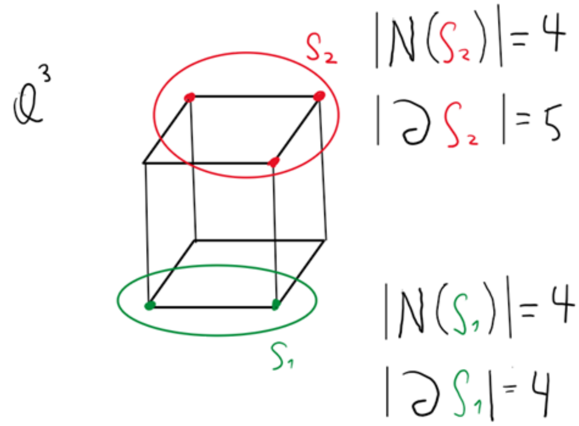
4. Q^d is d-connected.
5. In Q^d, there exist a perfect matching and a Hamilton cycle.

7.3.2 Isoperimetric Inequalities in Q^d

Definition 7.3.4. Let $G = (V, E)$ be a graph and let $\emptyset \neq S \subsetneq V$.

1. The *vertex boundary of* S is $N(S) = \{v \in V \setminus S \mid v \text{ has a neighbor in } S\}$.
2. The *edge boundary of* S is $\partial S = \big\{e \in E \mid |e \cap S| = |e \cap \bar{S}| = 1\big\}$.

Example



The isoperimetric properties of a graph look as follows:
Vertex Isoperimetry:

$$\forall S \subseteq V, 0 < |S| = s < |V| : \ |N(S)| \geq f(s).$$

Edge Isoperimetry:

$$S \subseteq V, 0 < |S| = s < |V(G))| : \ |\partial S| \geq g(s).$$

Edge Isoperimetric properties of Q^d

We begin with the following observation:

Observation 7.3.5. Q^d is a d-regular graph, hence for every $S \subseteq V(Q^d)$:

$$d \cdot |S| = \sum_{v \in S} d(v) = 2 \cdot e(S) + |\partial S| \implies |\partial S| = d \cdot |S| - 2 \cdot e(S).$$

That is, it suffices to bound $e(S)$ from above.

Theorem 7.3.6 (Harper, 1964)**.** *For every $d \geq 1$ and every $S \subseteq V(Q^d)$:*

$$e_{Q^d}(S) \leq \frac{1}{2}|S| \cdot \log_2 |S|.$$

Thus, $|\partial S| \geq |S| \cdot (d - \log_2 |S|)$.

Remark 7.3.7. The above bound is tight for a k-dimensional subcube ($k \leq d$), which is defined as follows:

$$S = \left\{ \bar{x} \in V(Q^d) \mid x_1 = x_2 = ... = x_{d-k} = 0 \right\}$$

(in general, fixing $d - k$ coordinates to arbitrary values).
Indeed, $|S| = 2^k$ and every $v \in S$ has exactly k neighbors in S. Thus,

$$e(S) = \frac{1}{2}|S| \cdot k = \frac{1}{2} \cdot 2^k \cdot k = \frac{1}{2}|S| \cdot \log_2 |S|.$$

8.1 Edge Isoperimetric Inequalities for the Hypercube - Continued

We show a (slightly) weaker version of Harper's result.

Theorem 8.1.1. *Suppose that $S \subseteq V(Q^d)$. Then*

$$e(S) \leq |S| \log_2 |S|.$$

In order to prove this theorem, we show the following two lemmas:

Lemma 8.1.2. *Let $S \subseteq V(Q^d)$. If the minimum degree of the induced subgraph $Q^d[S]$ is δ, then*

$$|S| \geq 2^{\delta}.$$

Proof. Fix a vertex $\bar{v} \in S$. For every $0 \leq i \leq \delta$, define

$$A_i = \{\bar{u} \in S \mid \mathrm{dist}_{Q^d}(\bar{u}, \bar{v}) = i\}.$$

Notice that (for example)

$$A_0 = \{\bar{v}\}, \quad A_1 = N(\bar{v}) \cap S.$$

Observe that the sets A_i are pairwise disjoint, hence

$$|S| \geq \sum_{i=0}^{\delta} |A_i|.$$

For every $i \geq 0$, all vertices in A_i have the same parity, hence A_i is an independent set in Q^d. Therefore, the neighbors of each vertex $\bar{u} \in A_i$ lie in $A_{i-1} \cup A_{i+1}$. Moreover, $\bar{u}$ sends at most i edges into A_{i-1}, thus, at least $\delta - i$ edges into A_{i+1} (since the degree of $Q^d[S]$ is at least δ). Similarly, every $w \in A_{i+1}$ incident to at most $i+1$ edges from A_i. Therefore,

$$|A_i|(\delta - i) \leq |E(A_i, A_{i+1})| \leq |A_{i+1}|(i+1),$$

implying

$$|A_{i+1}| \geq \frac{\delta - i}{i+1} |A_i|.$$

Observe that for every $0 \leq i \leq \delta$,

$$\frac{\binom{\delta}{i+1}}{\binom{\delta}{i}} = \frac{\delta - i}{i+1},$$

hence, it is easy to verify (by induction on i) that for every $0 \leq i \leq \delta$,

$$|A_i| \geq \binom{\delta}{i}.$$

Consequently,

$$|S| \geq \sum_{i=0}^{\delta} |A_i| \geq \sum_{i=0}^{\delta} \binom{\delta}{i} = 2^{\delta}.$$

□

Lemma 8.1.3. *If G is a graph with average degree d, then G contains a subgraph G' whose minimum degree satisfies*

$$\delta(G') \geq \frac{d}{2}.$$

Proof. We construct a sequence

$$G = G_0, G_1, \dots$$

as follows.
As long as $G_i \neq \varnothing$, choose a vertex v whose degree in G_i is smaller than $\frac{d}{2}$ (if exists), and define

$$G_{i+1} = G_i - \{v\}.$$

If the process terminates with a non-empty graph G_t, then every vertex of G_t has degree at least $\frac{d}{2}$, and therefore we are done, by taking $G' = G_t$.
Thus, it remains to show that the process cannot terminate with the empty graph.
Suppose, for contradiction, that eventually $G_t = \varnothing$. Then every deleted vertex had degree smaller than $\frac{d}{2}$ at the moment it was removed.
Therefore the total number of removed edges is less than

$$|V(G)| \cdot \frac{d}{2}.$$

On the other hand, the total number of removed edges is exactly $e(G)$. Hence

$$e(G) < |V(G)|\frac{d}{2},$$

which contradicts the fact that the average degree of G is d. □

Proof of Theorem 8.1.1. Let $S \subseteq V(Q^d)$. Since the average degree of the induced subgraph $Q^d[S]$ is

$$\frac{2e(S)}{|S|},$$

the previous lemma implies that there exists $S_0 \subseteq S$ such that the induced subgraph $G[S_0]$ has minimum degree at least

$$\frac{e(S)}{|S|}.$$

By the first lemma, $|S_0| \geq 2^{e(S)/|S|}$. Since $|S_0| \leq |S|$, we obtain $|S| \geq 2^{\,e(S)/|S|}$. Taking logarithms,

$$\log_2 |S| \geq \frac{e(S)}{|S|},$$

and therefore

$$e(S) \leq |S| \log_2 |S|.$$ □

Remark 8.1.4. If S is a $(d-1)$-dimensional subcube, then

$$e(S) = 2^{d-1} \log_2\left(2^{d-1}\right) = (d-1)2^{d-1}.$$

In particular, Theorem 8.1.1 gives a weak/meaningless bound for large sets $S \subseteq V(Q^d)$.

The following theorem gives another isoperimetric bound that treats large subsets of vertices in Q^d.

Theorem 8.1.5. *For every $S \subseteq V(Q^d)$ satisfying $0 < |S| \leq 2^{d-1}$, we have*

$$|\partial S| \geq |S|.$$

Remark 8.1.6. The theorem is tight when S is a $(d-1)$-dimensional subcube.

Proof. We prove the statement by induction on d. The case $d = 1$ is trivial. Let $d > 1$ and let

$$S \subseteq V(Q^d), \quad 0 < |S| \leq 2^{d-1}.$$

Define

$$S_0 = \{\bar{x} \in S \mid x_1 = 0\}, \quad S_1 = \{\bar{x} \in S \mid x_1 = 1\}.$$

Then

$$S = S_0 \sqcup S_1.$$

Without loss of generality, assume $|S_0| \leq |S_1|$. We further define

$$V_0 = \{\bar{x} \in V \mid x_1 = 0\}, \quad V_1 = \{\bar{x} \in V \mid x_1 = 1\}.$$

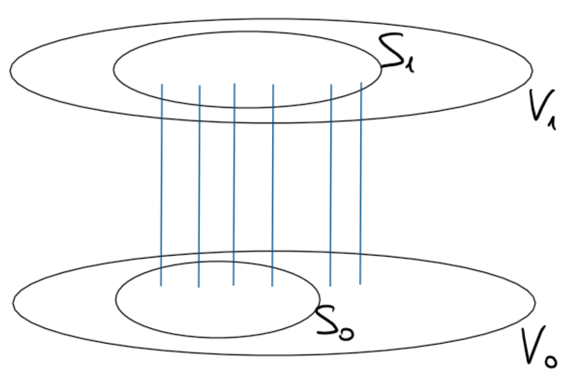
Notice that $S_0 \subseteq V_0, S_1 \subseteq V_1$ and $|V_0| = |V_1| = 2^{d-1}$. By the induction hypothesis, since

$$|S_0| \leq \frac{1}{2}|S|, \quad \text{and} \quad |S_0| \leq \frac{1}{2}|V_0|,$$

we have

$$|E(S_0, V_0 \setminus S_0)| \geq |S_0|, \quad |E(S_1, V_1 \setminus S_1)| \geq \min\{|S_1|,\ |V_1 \setminus S_1|\} \geq |S_0|.$$

Furthermore, Q^d contains a perfect matching between V_0 and V_1. In this matching, exactly $|S_1|$ edges are incident to vertices of S_1, and at least $|S_1| - |S_0|$ of them have their other endpoint in $V_0 \setminus S_0$. Thus, at least $|S_1| - |S_0|$ of such edges belong to ∂S.



To conclude,

$$|\partial S| \geq |E(S_0, V_0 \setminus S_0)| + |E(S_1, V_1 \setminus S_1)| + (|S_1| - |S_0|) \geq |S_0| + |S_0| + (|S_1| - |S_0|) = |S_0| + |S_1| = |S|. \quad \square$$

8.2 Vertex Isoperimetric Inequalities for the Hypercube

For general interest, we also present a vertex isoperimetric inequality. We begin by exhibiting a subset of vertices in the hypercube which serves as an example attaining equality in the isoperimetric inequality that we will present later.

Example 8.2.1. For $0 \le k \le d$, define

$$B_k = \{\bar{x} \in V(Q^d) \mid |\bar{x}| \le k\},$$

a ball of radius k around $\bar{0}$. The number of vertices of weight at most k is

$$|B_k| = \sum_{i=0}^{k} \binom{d}{i}.$$

Moreover,

$$N(B_k) = \{\bar{x} \in V(Q^d) \mid |\bar{x}| = k+1\},$$

and therefore

$$|N(B_k)| = \binom{d}{k+1}.$$

Definition 8.2.2. We define the *lexicographic order* on the vertices of Q^d as follows:
For $\bar{x} \neq \bar{y} \in V(Q^d)$,

$$\bar{x} <_{\text{lex}} \bar{y} \iff x_{i_0} = 0,\ y_{i_0} = 1, \quad \text{where} \quad i_0 = \min\{1 \le i \le d \mid x_i \neq y_i\}.$$

Definition 8.2.3. We define the *simplicial order* on the vertices of Q^d as follows:

$$\bar{x} <_{\text{sim}} \bar{y} \iff |\bar{x}| < |\bar{y}|, \quad \text{or} \quad (|\bar{x}| = |\bar{y}| \quad \text{and} \quad \bar{x} <_{\text{lex}} \bar{y}).$$

Example 8.2.4. For $d = 3$, the simplicial order on $V(Q^3)$ is

$$\sigma = (000, 001, 010, 100, 011, 101, 110, 111).$$

Theorem 8.2.5 (Harper, 1966)**.** *Let $A \subseteq V(Q^d)$. Let B be the set of the first $|A|$ vertices in the simplicial order on $V(Q^d)$. Then*

$$|N(A)| \ge |N(B)|.$$

In particular, if

$$|A| = \sum_{i=0}^{k} \binom{d}{i}, \quad 0 \le k < d,$$

then

$$|N(A)| \ge \binom{d}{k+1}.$$

8.3 The Model G_p

Definition 8.3.1. Let $G = (V, E)$ be a finite graph, and let $0 \leq p \leq 1$. We define a probability distribution on the subgraphs $H \subseteq G$ as follows: for every edge $e \in E(G)$,

$$\mathbb{P}(e \in E(H)) = p.$$

That is, each edge is chosen independently with probability p.
Equivalently, for every $H \subseteq G$, we can define:

$$\mathbb{P}(G_p = H) = p^{|E(H)|}(1-p)^{|E(G)|-|E(H)|}.$$

Remark 8.3.2. The value of p may depend on the parameters of the graph G.

Remark 8.3.3. If $G = K_n$ (the complete graph on n vertices), then G_p is exactly $G(n, p)$.

8.3.1 Typical Questions about G_p

1. What is the typical size of the connected components of G_p? For which values of p is the existence of a giant component guaranteed whp?

2. What are the typical combinatorial properties of G_p? For example, perfect matching, connectivity, Hamilton cycle, etc.

3. What is the critical probability for which a fixed graph H appears in G_p whp?

We focus on the following question:

Question 8.3.4. We consider a d-regular graph G, where d can be a constant satisfying $d \geq 3$ or $d \to \infty$. What is the critical probability p of the existence of the giant component in G_p?

Guess/Explanation: Choose a vertex v of G and expose the connected component of G_p containing v, denoted by C_v. To that end, we perform a search process from v (BFS or DFS), while revealing the edges of G_p. If we get (during the process) to a vertex u in C_v (due to an edge entering u), we expect to have $d-1$ edges leaving u and going outside the currently explored connected component C_v. Intuitively, by the GW process, if the expected number of edges leaving u is $(d-1)p =: c < 1$, then one may expect that no large component will emerge from v. On the other hand, if $(d-1)p = c > 1$, then the situation changes dramatically, and one expects that C_v will be large, with probability that bounded away from 0.
We therefore might guess that the critical probability for the emergence of a giant component in G_p is

$$p^* = \frac{1}{d-1}.$$

Remark 8.3.5. Notice that for large d we have

$$\frac{1}{d-1} \approx \frac{1}{d},$$

whereas for constant values of d, there is a substential difference between $\frac{1}{d-1}$ and $\frac{1}{d}$.

8.3.2 Sub-Critical Percolation

Theorem 8.3.6. *Let G be a d-regular graph on n vertices, and let $\epsilon > 0$ be a fixed constant. Set $p = \frac{1-\epsilon}{d-1}$. Then whp (as $n \to \infty$), every connected component L_i of G_p satisfies*

$$|L_i| \leq \frac{9 \ln n}{\epsilon^2}.$$

Proof. We run a search process on G (BFS or DFS) and "feed" it with random bits. Let $m = |E(G)|$ and notice that $m = \frac{nd}{2} \leq n^2$. Let $\bar{X} = (X_i)_{i=1}^m$ be a sequence of independent random variables satisfying

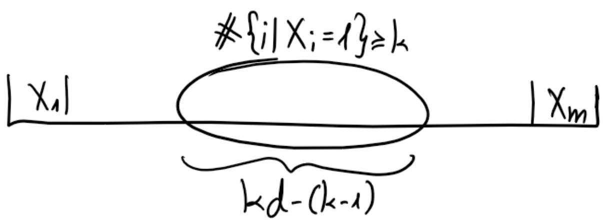
$$\mathbb{P}(X_i = 1) = p, \quad \mathbb{P}(X_i = 0) = 1 - p.$$

When the algorithm asks whether the i-th edge (the order is determined by the execution of the algorithm) of G falls into G_p, if $X_i = 1$, then $e_i \in E(G_p)$, and if $X_i = 0$, then $e_i \in E(G_p)$. At the end of the algorithm, we obtain a random subgraph G_p, together with its connected components.
Fix $k = k(n, \epsilon)$, whose value will be chosen later. Suppose there exists a connected component K with $|K| > k$. Let us consider the precise moment during the execution of the algorithm at which the $(k+1)$-st vertex of K was discovered. Since, in each connected component, the first vertex comes "for free", by that moment we have obtained exactly k positive answers. In addition, note that

1. we queried only edges incident to the first k vertices (according to the order in which the algorithm discovered them) of K.

2. K contains a tree that spans the first k vertices in K.

Hence, we have queried at most $kd - (k-1)$ edges (where $k-1$ is a lower bound on the number of edges inside these k vertices).
Therefore, $\bar{X}$ must contain a segment of length $kd - (k-1)$, with at least k positive answers, i.e. there are k indices in this segment for which $X_i = 1$.



Since $p = \frac{1-\epsilon}{d-1}$, the number of $X_i = 1$ in this segment is distributed as

$$Y \sim Bin\big(kd - (k-1), p\big).$$

Hence

$$\mathbb{E}[Y] = \big(kd - k + 1\big) \cdot \frac{1-\epsilon}{d-1} = (k(d-1) + 1) \cdot \frac{1-\epsilon}{d-1} = (1-\epsilon)k + \frac{1-\epsilon}{d-1}.$$

By Chernoff's inequality, we have

$$\mathbb{P}(Y \geq k) \leq e^{-\frac{\epsilon^2 k}{4}}.$$

Now, the probability that there exists an interval of length $kd - (k-1)$ with at least k successes in $[m]$ is at most

$$m \cdot e^{-\frac{\epsilon^2 k}{4}} < n^2 \cdot e^{-\frac{\epsilon^2 k}{4}}.$$

By choosing

$$k = \left\lfloor \frac{9 \ln n}{\epsilon^2} \right\rfloor,$$

this probability is $o(1)$.
Therefore, whp, G_p has no component of size $> k$, namely for all $i \geq 1$,

$$|L_i| \leq \frac{9 \ln n}{\epsilon^2}.$$

□

Remark 8.3.7. A few comments are in place.

1. The theorem remains true even if the assumption that G is d-regular is replaced by the assumption that $\Delta(G) \leq d$, with the same proof.

2. Applying the theorem to the case where $G = K_n$, we obtain that in $G(n,p)$, for $p = \frac{1-\epsilon}{n}$, whp $|L_i| \leq \frac{9 \ln n}{\epsilon^2}$, for every $i \geq 1$.

3. In fact, in $G(n,p)$ with $p = \frac{1-\epsilon}{n}$ ($\epsilon > 0$ is a small constant), whp $|L_1| = \Theta\left(\frac{\ln n}{\epsilon^2}\right)$.

Later, we look at Q_p^d with $p = \frac{c}{d}$, where $c > 1$ is a constant, and show that whp, there exists a giant component L_1 in Q_p^d such that $|L_1| = (1 + o(1))yn$ where $y \in (0,1)$ satisfies $y = 1 - e^{-cy}$. Furthermore, for every $i \geq 2$, we have $|L_i| = O(d)$.

Percolation on Finite Graphs, Tel Aviv Univ., Spring 2026 June 21, 2026

Lecture 9

Lecturer: Prof. Michael Krivelevich *Scribe: Gali Maman*

9.1 The Super-Critical Case — Q_p^d

Theorem 9.1.1 (Ajtai-Komlós-Szemerédi 1982; Bollobás, Kohayakawa, Łuczak 1992)**.** *Assume $c > 1$ is a constant, and consider Q_p^d with $p = \frac{c}{d}$. Then whp:*

1. $|L_1| = (1 + o(1))yn$, *where $n := 2^d$ and $y := y(c)$ is the (unique) solution of*

$$1 - y = e^{-cy} \tag{$*$}$$

   in $(0, 1)$.

2. *Every other component L_i, $i \geq 2$, satisfies*

$$|L_i| \leq \frac{d}{c - 1 - \ln c}.$$

Remark 9.1.2.

1. The order of L_1 in Q_p^d is asymptotically equal to the order of the giant component of $G(n, p)$ for $p = \frac{c}{n}$.

2. If $c = 1 + \epsilon$ for small $\epsilon > 0$, then it is easy to see that

$$c - 1 - \ln c > 0 \text{ and } c - 1 - \ln c = \Theta(\epsilon^2),$$

   and therefore whp

$$|L_i| \leq O\left(\frac{d}{\epsilon^2}\right) = O\left(\frac{\ln n}{\epsilon^2}\right)$$

   (this can be shown to be tight up in n and in ϵ).

Proof. The proof is based on a paper by Michael Krivelevich and consists of a sequence of lemmas. We will not prove the following two lemmas here; proofs may be found in a variety of standard sources.

Lemma 9.1.3. *Consider a Galton–Watson (GW) process with offspring distribution* $\mathrm{Bin}\left(d, \frac{c}{d}\right)$*. When $c > 1$, as $d \to \infty$ the process survives (i.e. does not become extinct at any finite stage) with probability $(1 + o(1))y$, where $y \in (0, 1)$ is defined as in $(*)$.*

Lemma 9.1.4 (Measure concentration / edge-exposure martingale / McDiarmid's inequality)**.** *Let $G = (V, E)$ be a finite graph with $|E| = m$ edges, and let $f : 2^G \to \mathbb{R}$ be a function on the subgraphs of G (i.e. $f(G')$ is a function of the subgraph $G' \subseteq G$). Suppose f is Lipschitz with parameter C: for every $G', G'' \subseteq G$ differing in exactly one edge,*

$$|E(G') \triangle E(G'')| = 1 \implies |f(G') - f(G'')| \leq C.$$

Let $X = f(G')$ for $G' \sim G_p$ (the random subgraph of G obtained by including each edge independently with probability p) with $p \in [0,1]$. Then for all $\beta > 0$,

$$\mathbb{P}\big[\, |X - \mathbb{E}[X]| \geq \beta \,\big] \leq 2\exp\left\{-\frac{\beta^2}{C^2 m}\right\}.$$

The following two lemmas were proved in previous lectures.

Lemma 9.1.5 (Counting trees in Q^d given a root)**.** *Let $v \in V(Q^d)$ and let $k \geq 1$ be an integer. The number of trees on k vertices in Q^d that contain v is at most*

$$(ed)^{k-1}.$$

Lemma 9.1.6 (Harper's isoperimetric inequality, weak version)**.**

1. *For every $S \subseteq V(Q^d)$,*
$$e(S) \leq |S| \cdot \log_2 |S|,$$
and consequently
$$|\partial S| \geq |S|\,(d - 2\log_2 |S|).$$

2. *For every $S \subseteq V(Q^d)$ with $|S| \leq \frac{n}{2}$,*
$$|\partial S| \geq |S|.$$

From now on, assume $p = \frac{c}{d}$ for a constant $c > 1$, and we study Q^d_p. We also fix an integer $t > 0$ (to be chosen later). Lastly, we consider y as defined in $(*)$.

Lemma 9.1.7. *In Q^d_p, whp there is no connected component of size in the range $\left[\frac{d}{c-1-\ln c},\, d^t\right]$.*

Proof. Let $v \in V(Q^d)$ and denote by C_v the connected component of v in Q^d_p. By the union bound, it suffices to show

$$\mathbb{P}\left(|C_v| \in \left[\frac{d}{c-1-\ln c},\, d^t\right]\right) = o\left(\frac{1}{n}\right).$$

Suppose $|C_v| = k$ for some $k \in \left[\frac{d}{c-1-\ln c},\, d^t\right]$. Then:

1. There is a tree $T \subseteq Q^d_p$ containing v, with $|V(T)| = k$.
2. There is no edge of Q^d_p between $V(T)$ and $V(Q^d) \setminus V(T)$.

Therefore,

$$\mathbb{P}(|C_v| = k) \leq \underbrace{(ed)^{k-1}}_{\text{choice of tree } T} \cdot \underbrace{p^{k-1}}_{\text{pay for } E(T)} \cdot \underbrace{(1-p)^{k(d-2\log_2 k)}}_{\text{no edges between } V(T) \text{ and } V(Q^d)\setminus V(T)}.$$

Summing over k,

$$\begin{aligned}
\mathbb{P}\left(|C_v| \in \left[\frac{d}{c-1-\ln c},\, d^t\right]\right) &\le \sum_{k=d/(c-1-\ln c)}^{d^t} (ed)^{k-1} p^{k-1} \cdot (1-p)^{k(d-2\log_2 k)} \\
&\le \sum_{k=d/(c-1-\ln c)}^{d^t} (ec)^k \cdot e^{-\frac{c}{d}k(d-2\log_2 k)} \\
&= \sum_{k=d/(c-1-\ln c)}^{d^t} \left[ec \cdot e^{-c+\frac{2c\log_2 k}{d}}\right]^k \\
&= \sum_{k=d/(c-1-\ln c)}^{d^t} \left[e^{1+\ln c - c + o(1)}\right]^k \\
&= (1+o(1))\, e^{(1+\ln c - c + o(1))\,\frac{d}{c-1-\ln c}} \\
&= (1+o(1))\, e^{-d+o(d)} = o\left(\frac{1}{n}\right) \qquad \text{(since } n = 2^d\text{)}.
\end{aligned}$$

□

Lemma 9.1.8. *For every $v \in V(Q^d)$, we have*

$$\mathbb{P}\left(|C_v| \ge d^t\right) = (1+o(1))y.$$

Proof. First, we estimate $\mathbb{P}\left(|C_v| \ge d^{1/2}\right)$. We run a search algorithm (BFS or DFS) on Q^d_p to grow C_v. As long as $|C_v| \le d^{1/2}$, for every vertex $u \in C_v$ whose neighbors we are about to expose, at most $d^{1/2}$ of its edges have already been exposed. Hence we can lower-bound (couple) the growth of C_v by a GW process with offspring distribution $\mathrm{Bin}(d - d^{1/2}, p)$, and therefore $|C_v| \ge d^{1/2}$ with probability $y(c')$, which equals to $(1-o(1))y$, since

$$c' = (d - d^{1/2})\, p = (1-o(1))c.$$

Next, we estimate $\mathbb{P}\left(d^{1/2} \le |C_v| \le d^t\right)$. This event is handled very similarly to before:

$$\mathbb{P}\left(d^{1/2} \le |C_v| \le d^t\right) \le \sum_{k=d^{1/2}}^{d^t} (ed)^{k-1} p^{k-1} (1-p)^{k(d-2\log_2 k)} \le \sum_{k=d^{1/2}}^{d^t} \left(ec \cdot e^{-c+\frac{2c\log_2 k}{d}}\right)^k = o(1),$$

because $ec \cdot e^{-c} < 1$ for all $c > 1$. Therefore

$$\mathbb{P}\left(|C_v| \ge d^t\right) = 1 - \mathbb{P}\left(|C_v| \le d^{1/2}\right) - \mathbb{P}\left(|C_v| \in \left(d^{1/2}, d^t\right)\right) = y + o(1) = (1+o(1))y.$$

□

Lemma 9.1.9. *Denote*

$$W = \{v \in V(Q^d) \mid |C_v| \ge d^t\}$$

(the collection of "largish" components). Then, whp

$$|W| = (1+o(1))yn.$$

Proof. Notice that we can write

$$|W| = \sum_{v \in V(Q^d)} \mathbb{1}_v, \qquad \mathbb{1}_v = \begin{cases} 1 & |C_v| \geq d^t, \\ 0 & \text{otherwise.} \end{cases}$$

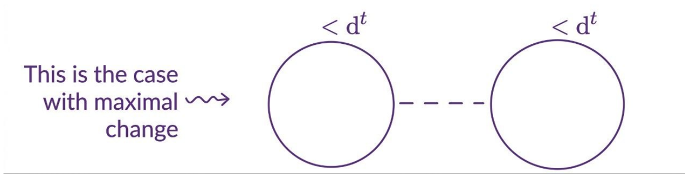
By linearity of expectation,

$$\begin{aligned} \mathbb{E}[|W|] &= \sum_{v \in V(Q^d)} \mathbb{E}[\mathbb{1}_v] = \sum_{v \in V(Q^d)} \mathbb{P}\left(|C_v| \geq d^t\right) \\ &= (1 + o(1))yn \qquad \text{(by Lemma 9.1.8).} \end{aligned}$$

We now apply the concentration result (Lemma 9.1.4). For $G' \subseteq Q^d$, define $f(G') = |W|$, and set a random variable $X = f(G')$. Then:

(1) $\mathbb{E}[X] = (1 + o(1))yn$.

(2) Toggling the state of a single edge of $E(Q^d)$ (present in Q_p^d or not) changes the value of X by at most $2d^t$ (this is maximal precisely when the toggled edge is the unique connection between two components, each of size $< d^t$, that would merge into a component of size $\geq d^t$).



By McDiarmid's inequality (Lemma 9.1.4),

$$\mathbb{P}\left[|X - \mathbb{E}[X]| \geq n^{2/3}\right] \leq 2 \exp\left\{-\frac{n^{4/3}}{\frac{nd}{2} \cdot (2d^t)^2}\right\} = o(1).$$

Hence, whp

$$X = (1 + o(1))\,\mathbb{E}[X] = (1 + o(1))yn.$$

$\square$

Informally, in the next lemma we show that W is spread "nicely" throughout $V(Q^d)$.

Lemma 9.1.10. *Let W be as in Lemma 9.1.9. Whp, in Q_p^d, every vertex $v \in V(Q^d)$ is at distance at most two from W.*

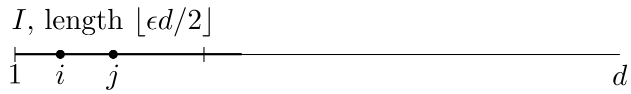
Proof. By symmetry, it suffices to show that

$$\mathbb{P}\Big[\bar{0} \text{ is not within distance 2 of } W\Big] = o\left(\frac{1}{n}\right).$$

Define $\epsilon = \frac{c-1}{c} > 0$ and $I = \left[\left\lfloor \frac{\epsilon d}{2} \right\rfloor\right]$. For every $i \neq j \in I$, define

$$H_{ij} = \left\{\bar{x} \in Q^d \mid x_i = 1,\ x_j = 1,\ \forall k \in I \setminus \{i, j\},\ x_k = 0\right\},$$

and let $u_{ij} \in H_{ij}$ be the point with $(u_{ij})_i = (u_{ij})_j = 1$ and $(u_{ij})_k = 0$ for all $k \neq i, j$.



Then:

1. H_{ij} is a subcube of Q^d of dimension $d - \left\lfloor \frac{\epsilon d}{2} \right\rfloor$.
2. $u_{ij} \in H_{ij}$, and u_{ij} is at distance 2 from $\bar{0}$.
3. For $(i,j) \neq (i',j')$, observe that $V(H_{ij}) \cap V(H_{i'j'}) = \emptyset$.

Since percolation restricted to each H_{ij} is distributed as $Q_p^{d-\lfloor \epsilon d/2 \rfloor}$, we note

$$\left(d - \left\lfloor \frac{\epsilon d}{2} \right\rfloor\right) p \geq \left(d - \frac{\epsilon d}{2}\right) \frac{c}{d} = \left(1 - \frac{\epsilon}{2}\right) c = \left(1 - \frac{c-1}{2c}\right) c = \frac{c+1}{2} > 1.$$

That is, percolation on H_{ij} is supercritical. Hence, for every $i \neq j \in I$,

$$\mathbb{P}\big[u_{ij} \text{ belongs to a component of size } \geq d^t \text{ of } H_{ij}\big] \geq \delta$$

for some $\delta \coloneqq \delta(c) > 0$. Therefore,

$$\begin{aligned} &\mathbb{P}(\text{for all } i \neq j \in I,\ u_{ij} \text{ does not belong to a component of size } \geq d^t \text{ of } H_{ij}) \\ &\qquad \leq (1-\delta)^{\binom{|I|}{2}} \leq e^{-\delta\binom{|I|}{2}} = e^{-\Theta(d^2)} = o\left(\frac{1}{n}\right), \end{aligned}$$

as required.
Therefore, with probability $1 - o(1/n)$, there exist $i \neq j \in I$ such that the component of u_{ij} in H_{ij} has at least d^t vertices. Since the component of u_{ij} in Q_p^d contains the one in H_{ij}, it follows that whp, every vertex $v \in V(Q^d)$ is at distance at most 2 from W. $\square$

Let

$$t = 31,$$

and define p_1, p_2 by

$$p_2 = \frac{1}{d^5}, \qquad 1 - p = (1-p_1)(1-p_2),$$

so in particular $p_2 \geq p - p_1$. Let

$$G_1 \sim Q_{p_1}^d, \qquad G_2 \sim Q_{p_2}^d$$

be independent, and set $G = G_1 \cup G_2$. Notice that $G \sim Q_p^d$. Denote

$$W_1 = \{v \in V(Q^d) : |C_{G_1,v}| \geq d^t\}.$$

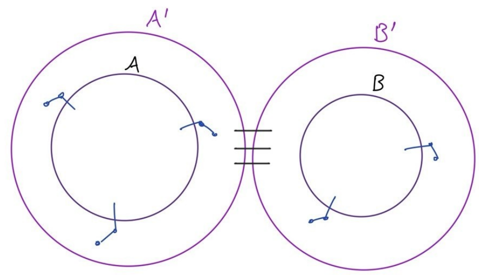
Lemma 9.1.11. *Whp, all components of W_1 merge into a single component after adding the edges of G_2.*

Proof. By Lemma 9.1.10 (applied to G_1), every vertex $v \in V(Q^d)$ is at distance ≤ 2 from W_1. Suppose toward contradiction that exposing the edges of G_2 does not unite the components of W_1. Then, there is a partition

$$W_1 = A \cup B, \qquad A, B \neq \emptyset,$$

respecting the connected components of W_1, such that there is no path in G_2 between A and B. Suppose W_1 has $s \leq n$ components, and (WLOG) one of A, B has $\ell \leq s/2$ of them. Define

$$A' := \{v \in V(Q^d) \setminus B : \mathrm{dist}_{Q^d}(v, A) \leq 2\}, \qquad B' := V(Q^d) \setminus A'.$$



Then $A \subseteq A'$, $B \subseteq B'$, and $A' \cup B' = V(Q^d)$. By the isoperimetric inequality in Q^d, there are at least $\min\{|A'|, |B'|\}$ edges between A' and B'. Since

$$|A'| \geq |A| \geq \ell\, d^{31}, \qquad |B'| \geq |B| \geq \ell\, d^{31},$$

there are at least $\ell\, d^{31}$ edges of Q^d between A' and B'. By construction of A, A', B, B', each such edge lies on a path of length ≤ 5 between A and B (at most two vertices inside each of A and B, plus the crossing edge).
Hence there are in total at least $\ell\, d^{31}$ (not necessarily disjoint) paths of length ≤ 5 between A and B in Q^d. Since every edge of Q^d lies on at most $5\, d^4$ such paths, a greedy argument yields at least

$$\frac{\ell\, d^{31}}{5 \cdot 5\, d^4 + 1} \geq \frac{\ell\, d^{27}}{30}$$

pairwise edge-disjoint paths of length ≤ 5 between A and B. Each such path is entirely contained in G_2 with probability $\geq p_2^5$, independently across the paths. Hence

$$\mathbb{P}\big[A, B \text{ not connected in } G_2\big] \leq \big(1 - p_2^5\big)^{\ell d^{27}/30} \leq e^{-p_2^5 \cdot \ell d^{27}/30} = e^{-\Theta(\ell d^2)}.$$

Therefore

$$\mathbb{P}\big[\exists \text{ components of } W_1 \text{ disconnected in } G_2\big] \leq \sum_{\ell \leq s/2} \binom{s}{\ell} e^{-\Theta(\ell d^2)} \leq \sum_{\ell \leq s/2} n^\ell\, e^{-\Theta(\ell d^2)} = o(1),$$

using $s \leq n$ and $n = 2^d$. $\square$

Proof of the theorem. Set p_1, p_2, G_1, G_2, W_1 as above. Since $p = (1 + o(1))p_1$, Lemma 9.1.9 gives that whp

$$|W_1| = (1 + o(1))yn.$$

Moreover, every connected component of G_1 outside of W_1 has size $\leq \frac{d}{c-1-\ln c}$. By Lemma 9.1.11, all components of W_1 merge into a single component whp, and therefore

$$|L_1| \geq |W_1| = (1 + o(1))yn.$$

On the other hand, recalling

$$W = \{v \in V(Q^d) : |C_{G,v}| \geq d^t\},$$

Lemma 9.1.9 gives $|L_1| \le |W|$, and whp

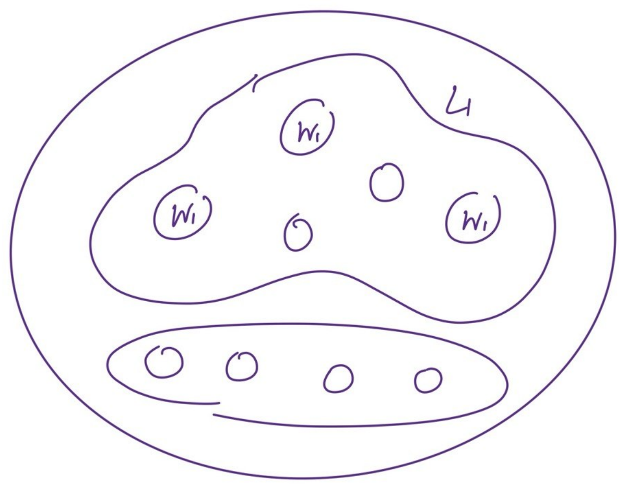
$$|W| = (1 + o(1))yn,$$

so that

$$|L_1| \le (1 + o(1))yn.$$

Combining the two bounds,

$$|L_1| = (1 + o(1))yn.$$



It remains to show that all components of $G = Q_p^d$ outside W_1 have size whp at most $\frac{d}{c-1-\ln c}$ in G. Define an auxiliary random graph Γ whose vertices are the connected components of G_1 outside W_1, where two such components are joined by an edge in Γ iff there is a G_2-edge between them. Each vertex of Γ corresponds to a component of size $O(d)$ with maximum degree d, hence connected to $O(d^2)$ components; between any two components there are at most d^2 possible connecting edges. Therefore, any two vertices of Γ are joined by an edge with probability

$$O(d^2) \cdot p_2 = O(d^{-3}).$$

Thus in Γ the degrees are $O(d^2)$ and the edge probability is $O(d^{-3})$ — so we are in the sub-critical regime. By the sub-critical analysis, whp every connected component of Γ has size

$$O(\log_2 n) = O(d).$$

Expanding each vertex of Γ back into its underlying component of G_1 (size $O(d)$), a component of G outside W_1 has size whp

$$O(d) \cdot O(d) = O(d^2).$$

But by Lemma 9.1.7, $G = Q_p^d$ has no component of size in $\left[\frac{d}{c-1-\ln c}, d^2\right]$, so this bound forces the actual size to fall below the lower end of that interval. Hence, for all $i \ge 2$, we have

$$|L_i| \le \frac{d}{c - 1 - \ln c},$$

which completes the proof. $\square$

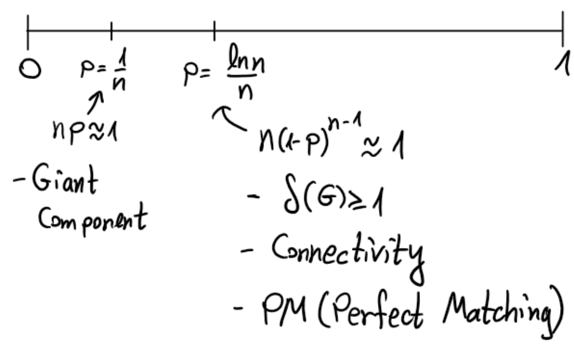
Percolation on Finite Graphs, Tel Aviv Univ., Spring 2026 June 28, 2026

Lecture 10

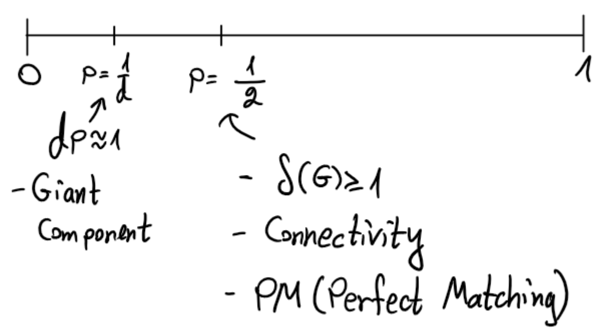
Lecturer: Dr. Sahar Diskin *Scribe: Aner Mash*

10.1 A comparison between $G(n, p)$ and Q_p^d

Consider the random graph $G(n, p)$:



Consider the random hypercube Q_p^d:



Intuitively, $p = \frac{1}{2}$ comes from the calculation of the first moment:

$$\mathbb{E}[\# \text{ isolated vertices}] = 2^d \cdot (1-p)^d = 1 \iff p = \frac{1}{2}.$$

10.2 Perfect Matching in Q_p^d

Theorem 10.2.1 (Bollobás, 1990)**.**

$$\lim_{d \to \infty} \mathbb{P}(Q_p^d \textit{ contains a perfect matching}) = \begin{cases} 0 & p \leq \frac{1}{2} - \frac{\omega(d)}{d} \\ 1 & p \geq \frac{1}{2} + \frac{\omega(d)}{d} \end{cases}$$

The proof is due to Sahar Diskin and Anna Geisler.

10.2.1 Preparation

Lemma 10.2.2. *Let* $\frac{2^d}{d^{\ln^2 d}} \le m \le 2^d$ *be an integer. Then:*

$$\#\{S \subseteq V(Q^d) : |S| = m,\ |\partial S| < m \cdot \ln^4 d\} \le \exp\left\{\frac{2m}{\ln^2 d}\right\}.$$

Remark 10.2.3. For such m,

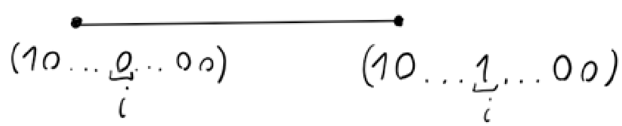
$$\#\{\text{subsets of size } m\} = \binom{2^d}{m} \ge \left(\frac{2^d}{m}\right)^m.$$

This lemma shows that relatively few subsets of size m have weak expansion in Q^d.

Proof. Fix $\frac{2^d}{d^{\ln^2 d}} \le m \le 2^d$. Define the family:

$$\mathcal{F} := \{S \subseteq V(Q^d) \mid |S| = m,\ |\partial S| < m \cdot \ln^4 d\}.$$

Given $i \in [d]$ and $S \subseteq V$, let $E_i(S, S^c)$ denote the set of edges in $E(S, S^c) = \partial S$ that are oriented along the i-th dimension.



Denote $e_i(S, S^c) := |E_i(S, S^c)|$. Given a subset $I \subseteq [d]$, we denote:

$$e_I(S, S^c) := \sum_{i \in I} e_i(S, S^c).$$

We say that S is a *"bad" set w.r.t. I* if

$$e_I(S, S^c) < m \cdot \ln^4 d \cdot \frac{|I|}{d}.$$

Averaging for every $1 \le k \le d$, if $S \in \mathcal{F}$, then there exists a subset I with $|I| = k$, such that S is "bad" with respect to I. Let us denote this family by $\mathcal{F}_I$.
We wish to bound $|\mathcal{F}|$. Observe that

$$\mathcal{F} \subseteq \bigcup_{\substack{I \subseteq [d] \\ |I| = k}} \mathcal{F}_I,$$

hence

$$\forall k \quad |\mathcal{F}| \le \binom{d}{k} \cdot \max_{\substack{I \subseteq [d] \\ |I| = k}} |\mathcal{F}_I|.$$

Set

$$k = \log_2(\ln^5 d).$$

Our goal is to estimate $|\mathcal{F}_I|$ from above, for $I \subseteq [d]$ with $|I| = k$. Let $\mathcal{Q}_I$ be the set of all subcubes of Q^d with all fixed coordinates outside I. Observe that every $Q \in \mathcal{Q}_I$ is a k-dimensional subcube of Q^d, and in particular $|V(Q)| = 2^k$. Moreover, notice that $|\mathcal{Q}_I| = \frac{2^d}{2^k}$.

Claim 10.2.4. *Given a k-dimensional cube Q and a subset $A \subseteq V(Q)$ such that $A \neq \emptyset$ and $A \neq V(Q)$, we have $|\partial A| \geq k$. (Exercise - use Harper)*

Let $S \in \mathcal{F}_I$. The number of distinct subcubes from $\mathcal{Q}_I$ such that the intersection with S is neither $\emptyset$ nor S is at most $\frac{|S| \cdot \ln^4 d}{d}$. Otherwise, the number of edges in the directions of I would be at least $k \cdot \frac{|S| \cdot \ln^4 d}{d}$, which implies $S \notin \mathcal{F}_I$ — a contradiction.
Therefore, S contains at least $\frac{|S|}{2^k} - \frac{|S| \cdot \ln^4 d}{d}$ subcubes from $\mathcal{Q}_I$ and at most $\frac{|S| \cdot \ln^4 d}{d} \cdot 2^k$ additional vertices. We obtain:

$$|\mathcal{F}_I| \leq \binom{\frac{2^d}{2^k}}{\frac{m}{2^k}} \cdot \binom{2^d}{\frac{m \cdot \ln^4 d}{d} \cdot 2^k} \leq \binom{\frac{2^d}{\ln^5 d}}{\frac{m}{\ln^5 d}} \cdot \binom{2^d}{\frac{m \cdot \ln^9 d}{d}} \overset{\text{exercise}}{\leq} \exp\left(\frac{m}{\ln^2 d}\right),$$

where we used $m \geq \frac{2^d}{d^{\ln^2 d}}$. Thus,

$$|\mathcal{F}| \leq \binom{d}{\log_2(\ln^5 d)} \cdot \exp\left(\frac{m}{\ln^2 d}\right) < \exp\left(\frac{2m}{\ln^2 d}\right).$$

$\square$

Lastly, recall the well-known Hall's condition for having a perfect matching in a bipartite graph:

Claim 10.2.5. *Let $G = (A \cup B, E)$ be a bipartite graph such that $|A| = |B|$. The graph G contains a perfect matching if and only if for every $X \subseteq A$, we have $|N_G(X)| \geq |X|$.*

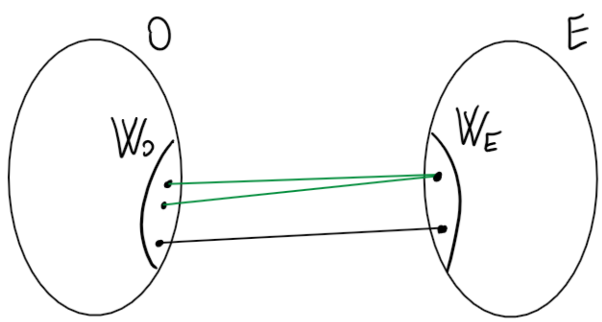
It easily follows from Hall's theorem that if a bipartite graph $G = (A \cup B, E)$ has no perfect matching, then there are sets $A_0 \subseteq A$ and $B_0 \subseteq B$ such that $G[A_0 \cup B_0]$ is connected and

1. $|A_0| = |B_0| + 1$ and $N_G(A_0) = B_0$ or
2. $|B_0| = |A_0| + 1$ and $N_G(B_0) = A_0$.

In both cases $|A_0 \cup B_0|$ is odd.

10.2.2 Proof of Theorem 10.2.1

Let $p \geq \frac{1}{2} + \frac{\omega(d)}{d}$. Whp, Q_p^d contains no isolated vertices (the expected number of isolated vertices in Q_p^d tends to 0, hence the assertion is implied by Markov's inequality). Denote by $\mathcal{O}$ and $\mathcal{E}$ the sides of the hypercube (which is a bipartite graph). For every $W \subseteq V(Q^d)$, denote by $W_{\mathcal{O}} = W \cap \mathcal{O}$ and $W_{\mathcal{E}} = W \cap \mathcal{E}$.



Hall's condition fails because of the green edges

Let $m \in \mathbb{N}_{odd}$. A set $W \subseteq V(Q^d)$ is called an *m-obstacle* if $|W| = m$, W is connected in Q^d, $|W_{\mathcal{O}}| < |W_{\mathcal{E}}|$ and $N_{Q^d_p}(W_{\mathcal{O}}) = W_{\mathcal{E}}$. One can define an analogous notion of an m-obstacle by exchanging the roles of $\mathcal{O}$ and $\mathcal{E}$. Hence, by the remark after Claim 10.2.5, it suffices to show that whp, there are no m-obstacles on either side for any odd $1 \leq m \leq 2^{d-1}$. We show it for our definition of an m-obstacle, where the proof for the analog definition is very similar. First observe that $m = 1$ is an easy to check case. Hence, assume that $m \geq 3$. We split the proof into cases:

Case 1: $3 \leq m \leq 2^{\frac{d}{100}}$.
By Theorem 7.3.6, we have $e(W) \leq \frac{m \cdot \log_2 m}{2}$. Also, $|W_{\mathcal{O}}| \geq \frac{m}{2}$. Therefore,

$$e_{Q^d}(W_{\mathcal{O}}, V \setminus W_{\mathcal{E}}) \geq \frac{m}{2} \cdot d - e(W) \geq \frac{m}{2} \cdot (d - \log_2 m) \geq \frac{49}{100} md.$$

Now,

$$\mathbb{E}[\#m\text{-obstacles}] \leq 2^d \cdot (ed)^m \cdot (1-p)^{\frac{49}{100} md} \leq 2^d \cdot (ed)^m \cdot \left(\frac{1}{2}\right)^{\frac{49}{100} md}$$
$$\leq 2^{-\frac{49}{100} md + d + 2m \cdot \log_2 d} \leq 2^{-\frac{45}{100} md + d}.$$

Since $m \geq 3$, we have

$$\sum_{m=3}^{2^{\frac{d}{100}}} \mathbb{E}[\#m\text{-obstacles}] \to 0,$$

and finish by Markov's inequality.

Case 2: $2^{\frac{d}{100}} \leq m \leq \frac{2^d}{d^{\ln^2 d}}$.

$$\mathbb{E}[\#m\text{-obstacles}] \leq 2^d \cdot (ed)^m \cdot \left(\frac{1}{2}\right)^{\frac{m}{2} \cdot (d - \log_2 m)} \leq 2^{d + 2m \cdot \log_2 d - \frac{m \cdot d}{2} + \frac{m \cdot \log_2 m}{2}} \leq 2^{-\frac{m \ln^3 d}{3}}.$$

Again, summing over all values of m, we finish by Markov's inequality.

Case 3: $\frac{2^d}{d^{\ln^2 d}} \leq m \leq 2^{d-1}$.
Let W be an m-obstacle. We split this case into two subcases according to the size of ∂W.

(1) Suppose $|\partial W| \geq m \cdot \ln^4 d$.
Note that $dm = d|W| = 2e(W) + |\partial W|$. Hence, $e(W) \leq \frac{m}{2}(d - \ln^4 d)$. We have $|W_{\mathcal{O}}| \geq \frac{m}{2}$ and thus $e(W_{\mathcal{O}}, V \setminus W_{\mathcal{E}}) \geq \frac{m}{2} \cdot d - e(W) \geq \frac{m \ln^4 d}{2}$. Therefore, the probability that such m-obstacle W exists in Q^d_p is at most:

$$\binom{2^d}{m} \cdot \left(\frac{1}{2}\right)^{\frac{m \ln^4 d}{2}} \leq \exp\left(m\left(2 \ln\left(\frac{2^d}{m}\right) - \frac{\ln^4 d}{2}\right)\right) \leq \exp\left(m\left(2 \ln^3 d - \frac{\ln^4 d}{2}\right)\right)$$
$$\leq \exp\left(-\frac{m \ln^4 d}{3}\right).$$

Summing over $\leq 2^d$ values of m and using Markov's inequality completes this case.

(2) Suppose $|\partial W| < m \cdot \ln^4 d$.
In this case, by Lemma 10.2.2 we have that the probability we have such an m-obstacle in Q^d_p is at most

$$\exp\left(\frac{2m}{\ln^2 d}\right) \cdot \left(\frac{1}{2}\right)^{\frac{m}{2} \cdot (d - \log_2 m)} \leq \exp\left(\frac{2m}{\ln^2 d}\right) \cdot \left(\frac{1}{2}\right)^{\frac{m}{2}}.$$

Lastly, as before, we finish by Markov's inequality.

Percolation on Finite Graphs, Tel Aviv Univ., Spring 2026	July 5, 2026
Lecture 11	
Lecturer: Prof. Michael Krivelevich	*Scribe: Nati Pupko*

11.1 Preliminaries

Definition 11.1.1. Given a graph $G(V, E)$, for any two vertices $v, u \in V(G)$ we define the *distance* between u and v to be the (edge) length of a shortest u, v path in G, and denote it by $\text{dist}_G(u, v)$. In case no path exists between u and v, we define $\text{dist}_G(u, v) = \infty$.

Definition 11.1.2. The *diameter* of a graph $G(V, E)$ is defined to be the largest distance between a pair of vertices in G. We denote it by $\text{diam}(G)$. Note that G is connected iff $\text{diam}(G) < \infty$.

Definition 11.1.3. Given two vertices in the hypercube, $u, v \in Q^d$, we define the *Hamming distance* between u and v to be the number of coordinates in which u and v differ. Formally, define

$$\mathcal{D}(u, v) = \{i \in [d] \mid u_i \neq v_i\},$$

and the Hamming distance is defined as

$$d(u, v) = |\mathcal{D}(u, v)|.$$

We now prove that for the hypercube, these two notions of distance are equivalent.

Claim 11.1.4. *Given two vertices in the hypercube, $u, v \in Q^d$, we have:*

$$d(u, v) = dist_{Q^d}(u, v).$$

Proof. In order to establish equality, we first prove that $\text{dist}_{Q^d}(u, v) \geq d(u, v)$ by proving that any $u - v$ path in Q^d must include at least $d(u, v)$ edges. After proving this, we show that $\text{dist}_{Q^d}(u, v) \leq d(u, v)$ by constructing an explicit u, v path with $d(u, v)$ edges, proving equality.
Indeed, every edge of Q^d connects between two vertices of Hamming distance 1, therefore, a path of length k can change at most k coordinates. In particular, in order to connect two vertices with $d(u, v)$ different coordinates we need at least $d(u, v)$ edges.
For the other direction, we can define $P = (v = v_0, ..., v_{d(u,v)} = u)$ by defining v_j to be the vector obtained by altering the first j coordinates of v to match u. $\square$

Corollary 11.1.5. $\text{diam}(Q^d) = d$.

11.2 Diameter of the giant in Q_p^d

We now turn to analyzing the typical diameter of the giant component in Q_p^d, where $p = \frac{1+\epsilon}{d}$ and $\epsilon > 0$ is a small constant. We can imagine theoretical cases where the diameter is asymptotically greater than d (where many edges are deleted, and the shortest path between two vertices u, v is highly convoluted). However, we wish to understand the typical behavior.
The question of what is the right order of $\text{diam}(Q_p^d)$ has been open for some time, until recently.

In 2023, Erde, Kang and Krivelevich proved that whp $\mathrm{diam}(Q_p^d) = O(d^3)$. This was the first result to establish that the correct order is polynomial in d. Later, in 2026, Anastos, Diskin, Lichev and Zhukovskii established that whp $\mathrm{diam}(Q_p^d) = O(d)$. This is also the best possible bound, in the sense that whp $\mathrm{diam}(Q_d^p) = \Theta(d)$, (but not exactly d).
We prove a weaker statement, that the diameter is typically polynomial in d.

Theorem 11.2.1. *Let $\epsilon > 0$ be a sufficiently small constant. Let L_1 be the largest component of Q_p^d, where $p = \frac{1+\epsilon}{d}$. Then, whp*

$$diam(L_1) = O(d^{12}).$$

11.2.1 Proof of Theorem 11.2.1

Throughout the proof we use the following notation: $n = 2^d$, $p = \frac{1+\epsilon}{d}$, where $\epsilon > 0$ is a sufficiently small constant. We also denote $G \sim Q_p^d$, and let L_1 be the unique giant component of size $|L_1| = \Theta_\epsilon(n)$ (which existence is asserted by Theorem 9.1.1).
The proof relies on a series of lemmas:

Lemma 11.2.2. *For any $\epsilon > 0$ small enough, there exist $C_1 := C_1(\epsilon) > 0$ large enough, and $\alpha := \alpha(\epsilon) > 0$ small enough, such that the following holds whp. Let $S \subseteq V(G)$ such that $G[S]$ connected, and of size $C_1 d \leq |S| \leq n^\alpha$. Then $|N_G(S)| \geq \alpha|S|$.*

Proof. Let $C_1 d \leq k \leq n^\alpha$ (where C_1 and α are constants that will be chosen later). We define

$$\mathcal{A}_k = \{S \subseteq V \mid |S| = k \text{ and } G[S] \text{ is connected}, |N(S)| \leq \alpha k\}.$$

It is enough to show that $\mathbb{P}(\mathcal{A}_k) = o(\frac{1}{n})$, then apply the union bound to obtain:

$$\mathbb{P}\left(\bigcup_{k=C_1 d}^{n^\alpha} \mathcal{A}_k\right) \leq \sum_{k=C_1 d}^{n^\alpha} \mathbb{P}(\mathcal{A}_k) \leq \sum_{k=C_1 d}^{n^\alpha} o\left(\frac{1}{n}\right) = o(1).$$

We now proceed to proving $\mathbb{P}(\mathcal{A}_k) = o(\frac{1}{n})$. If $\mathcal{A}_k$ occurs (i.e. the set is not empty), then there exists a tree T in Q^d on k vertices such that:

(1) All edges of T are present in G.

(2) Denote $S = V(T)$. The S has at most αk neighbors in G outside of S.

Given a tree T, conditions (1) and (2) address disjoint sets of edges, and therefore are independent events. We will bound the probability of each of them separately, than multiply, and apply a union bound over all possible trees T.
Bounding the probability of (1): A tree of size k has $k-1$ edges, therefore the probability of it being present in G is p^{k-1}.
Bounding the probability of (2): Condition (2) implies that in the graph $G[S, V \setminus S]$, the maximal size of a matching is i, for some $0 \leq i \leq \alpha k$. Let M be a maximal such matching. Then:

1. All i edges of M in G.

2. All edges of Q^d between S and $V \setminus S$ (i.e., ∂S) that are not incident to vertices of M are not present in G (as otherwise M could be made larger).

Hence,

$$\mathbb{P}((2)) \leq \sum_{i=0}^{\alpha k} \binom{|\partial S|}{i} p^i (1-p)^{|\partial S|-2id} \leq \sum_{i=0}^{\alpha k} \binom{kd}{i} p^i (1-p)^{k(d-2\log_2 k)-2id},$$

where the last inequality holds by (weak) Harper's inequality. We note that $k \leq n^{\alpha}$, and that for α small enough we have:

$$k(d - 2\log_2 k) \geq (1-\epsilon^3)dk.$$

Thus, overall, by the union bound:

$$\begin{aligned}
\mathbb{P}(\mathcal{A}_k) &\leq n \cdot (ed)^{k-1} p^{k-1} \sum_{i=0}^{\alpha k} \binom{kd}{i} p^i (1-p)^{(1-\epsilon^3)kd-2id} \\
&\leq n\left(e(1+\epsilon)\right)^k (1-p)^{(1-\epsilon^3)kd} \sum_{i=0}^{\alpha k} \binom{kd}{i} p^i (1-p)^{-2id}.
\end{aligned}$$

Note that:

$$\sum_{i=0}^{\alpha k} \binom{kd}{i} p^i (1-p)^{-2id} \leq \sum_{i=0}^{\alpha k} \left(\frac{ekd}{i}\right)^i \left(\frac{1+\epsilon}{d}\right)^i \cdot 10^i = \sum_{i=0}^{\alpha k} \left(\frac{ek(1+\epsilon)}{i} \cdot 10\right)^i.$$

The terms in this sum grow exponentially, and the maximum is achieved at the boundary $i = \alpha k$:

$$\sum_{i=0}^{\alpha k} \left(\frac{ek(1+\epsilon)}{i} \cdot 10\right)^i \leq O(k) \cdot \left(\frac{30}{\alpha}\right)^{\alpha k} \leq e^{\epsilon^3 k},$$

for sufficiently small $\alpha \coloneqq \alpha(\epsilon) > 0$, where we used the fact that $\lim_{\alpha \to 0^+} \left(\frac{30}{\alpha}\right)^{\alpha} = 1$.
Plugging this estimate back into $\mathbb{P}(\mathcal{A}_k)$, we obtain:

$$\begin{aligned}
\mathbb{P}(\mathcal{A}_k) &\leq n\left(e(1+\epsilon)\right)^k (1-p)^{(1-\epsilon^3)kd} e^{\epsilon^3 k} \\
&\leq n\left(e(1+\epsilon) e^{-(1+\epsilon)(1-\epsilon^3)d \cdot \frac{1}{d}} e^{\epsilon^3}\right)^k \\
&\leq n\left(e(1+\epsilon)e^{-1-\epsilon+3\epsilon^3}\right)^k = n\left((1+\epsilon)e^{-\epsilon+3\epsilon^3}\right)^k.
\end{aligned}$$

Using $1+\epsilon \leq e^{\epsilon - \frac{\epsilon^2}{3}}$ for small $\epsilon > 0$, we obtain:

$$\mathbb{P}(\mathcal{A}_k) \leq n\left(e^{-\frac{\epsilon^2}{3}+3\epsilon^3}\right)^k \leq ne^{-\frac{\epsilon^2}{4}k}.$$

Choosing $C_1 \coloneqq C_1(\epsilon)$ sufficiently large in terms of ϵ, we obtain:

$$\mathbb{P}(\mathcal{A}_k) = o\left(\frac{1}{n}\right).$$

This completes the proof of Lemma 11.2.2. $\square$

We will apply Lemma 11.2.2 on balls centered at v, i.e., $B(v,k) = \{u \in V : \mathrm{dist}_G(u,v) \leq k\}$. This is clearly a connected set, therefore, by Lemma 11.2.2 if we have $C_1 d \leq |B(v,k)| \leq n^{\alpha}$, then observing that $B(v,k+1) = B(v,k) \cup N_G\left(B(v,k)\right)$, we obtain:

$$|B(v,k+1)| \geq (1+\alpha)|B(v,k)|.$$

In order to ensure that $|B(v,k)| \geq C_1 d$, we let $k_0 = C_1 d$, thus, $|B(v,k_0)| \geq k_0 = C_1 d$. Now, by induction, for every $i \geq 0$,

$$|B(v, k_0 + i)| \geq \min\{(1+\alpha)^i |B(v,k_0)|,\ n^\alpha\}. \tag{11.3}$$

We use this observation for the next lemma:

Lemma 11.2.3. *For every sufficiently small constant $\epsilon > 0$, there exist $\alpha = \alpha(\epsilon) > 0$ and $C_2 = C_2(\epsilon) > 0$ such that whp there exists a partition $V(L_1) = W_1 \cup \ldots \cup W_t$ satisfying:*

1. *$W_i \cap W_j = \emptyset$ for all $1 \leq i \neq j \leq t$;*
2. *$n^\alpha \leq |W_i| \leq 2n^\alpha$;*
3. *for all $u, v \in W_j$, $dist_{L_1}(u,v) \leq 4C_2 d$.*

Proof. By (11.3), we conclude that there exists a constant $C_2 = C_2(\epsilon) > 0$, such that for every $v \in V(L_1)$ we have $|B(v, C_2 d)| \geq n^\alpha$. Choose $X \subseteq V(L_1)$ to be a maximal set of vertices such that for any $u \neq v \in X$:

$$B_{L_1}(u, C_2 d) \cap B_{L_1}(v, C_2 d) = \emptyset.$$

Write $X = \{v_1, \ldots, v_{t_0}\}$, for some $t_0 \in \mathbb{N}$. By maximality, for every $u \in V(L_1)$, there exists some $v_i \in X$ such that:

$$B(u, C_2 d) \cap B(v_i, C_2 d) \neq \emptyset \implies \mathrm{dist}_{L_1}(u, v_i) \leq 2C_2 d.$$

We define a family of disjoint sets $U_1, \ldots, U_{t_0} \subseteq V(L_1)$ as follows:

1. Initialize $U_i = B_{L_1}(v_i, C_2 d)$.
2. For any vertex $u \notin \bigcup_{i=1}^t U_i$, we assign u to exactly one U_i for which $\mathrm{dist}(u, v_i) \leq 2C_2 d$.

This yields a partition of $V(L_1)$ to sets $U_1, \ldots, U_{t_0}$ satisfying:

- $U_i \supseteq B(v_i, C_2 d) \implies |U_i| \geq n^\alpha$.
- For any $u \in U_i$, we have $\mathrm{dist}_{L_1}(u, v_i) \leq 2C_2 d$.

Remark 11.2.4. The distance between any $u, v \in U_i$ is at most $4C_2 d$ inside L_1.

Finally, we divide each U_i into smaller disjoint connected subsets $W_j \subseteq U_i$ of sizes $n^\alpha \leq |W_j| \leq 2n^\alpha$. This gives a partition:

$$V(L_1) = W_1 \cup \cdots \cup W_t,$$

where for each $j \in [t]$:

1. $W_i \cap W_j = \emptyset$ for $i \neq j$;
2. $n^\alpha \leq |W_i| \leq 2n^\alpha$;
3. For all $u, v \in W_j$, $\mathrm{dist}_{L_1}(u, v) \leq 4C_2 d$.

$\square$

Our strategy now becomes proving that the sets from Lemma 11.2.3 are close to one another, which would allow us to construct a short path between any pair of vertices of G in the following way: If $u \in W_i$ and $v \in W_j$, if we prove that W_i and W_j are close, we can take representatives $u' \in W_i$ and $v' \in W_j$ so that u' and v' are close, which allows us to construct a short $u - v$ path in the following fashion:

$$u \to u' \to v' \to v,$$

where Lemma 11.2.3 tells us that u and u', as well as v and v' are at distance at most $4C_2 d$ from one another.

In order to show something of this sort, we apply sprinkling. Given our $p = \frac{1+\epsilon}{d}$, we choose $p_2 = \frac{\delta}{d}$, for some $\delta := \delta(\epsilon)$ small enough, and p_1 such that $1 - p = (1 - p_1)(1 - p_2)$, which gives $p_1 \geq \frac{1+\epsilon-\delta}{d} \geq \frac{1+\epsilon/2}{d}$.

Let $G_1 \sim Q^d_{p_1}$ and $G_2 \sim Q^d_{p_2}$, so $G = G_1 \cup G_2 \sim Q^d_p$. Let $L_1' \subseteq L_1$ be the giant component in G_1 (which exists, whp, by Theorem 9.1.1).

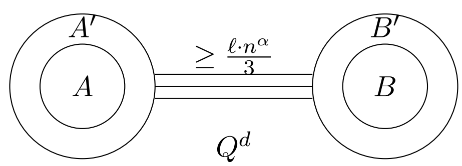
Lemma 11.2.5. *Whp, any partition of $L_1' = A \cup B$ respecting the sets $W_1, .., W_t$ as in Lemma 11.2.3 (applied in G_1) so that A contains ℓ of these sets, $1 \leq \ell \leq \frac{3t}{4}$, in G_2 there are $\Omega\big(\frac{\ell \cdot n^\alpha}{d^9}\big)$ edge-disjoint paths of length at most 5 between A and B. Therefore, whp there exist at least $\Omega\big(\frac{\ell \cdot n^\alpha}{d^{10}}\big)$ different endpoints of paths of length at least 5 between A and B.*

Proof. Let:

$$A' = \{v \in V(Q^d) \setminus B \mid \mathrm{dist}(v, A) \leq 2\}, \quad B' = V(Q^d) \setminus A'.$$

Notice that $A \subseteq A'$ and $B \subseteq B'$, hence $\min\{|A'|, |B'|\} \geq \min\{|A|, |B|\}$. We use this inequality later. Observe that since A contains ℓ of these sets, with $1 \leq \ell \leq \frac{3t}{4}$, then $|A| \geq \ell \cdot n^\alpha$, and also $|B| \geq (t - \ell)n^\alpha \geq \frac{\ell}{3} \cdot n^\alpha$. Together, by Theorem 8.1.5,

$$|E_{Q^d}(A', B')| \geq \min\{|A'|, |B'|\} \geq \min\{|A|, |B|\} \geq \frac{\ell \cdot n^\alpha}{3}.$$



Also, as we have proven that whp in G_1 every vertex of Q^d is at distance at most 2 from L_1' (see Theorem 9.1.10), we thus have that every $v \in B'$ is at distance at most 2 from B, hence, every edge $e \in E_{Q^d}(A', B')$ is part of a path of length at most 5 in Q^d between A and B. Hence, the number of edge-disjoint paths in Q^d between A and B is at least:

$$\frac{|E(A', B')|}{5 \cdot 5d^4 + 1} \geq \frac{\ell \cdot n^\alpha}{80d^4}.$$

Each such path is present in $G_1 \cup G_2$ with probability at least $p_2^5 = \big(\frac{\delta}{d}\big)^5$. Thus, the expected number of open paths connecting A and B in $G = G_1 \cup G_2$ is:

$$\frac{\ell \cdot n^\alpha}{80d^4} \cdot \left(\frac{\delta}{d}\right)^5 = \frac{\delta^5 \cdot \ell \cdot n^\alpha}{80d^9}.$$

Using the standard concentration of the binomial distribution and the union bound over all partitions $V(L_1') = A \cup B$, we yield whp at least $\Omega\big(\frac{\ell \cdot n^\alpha}{d^9}\big)$ edge-disjoint paths of length at most 5 between A

and B in G. For the second part of the lemma, we note that for any $v \in B$ has degree d in Q^d, hence it is an endpoint of at most d such paths. $\square$

We are now close to completing the proof. To finish, we first prove that the vertices of L_1' are close to one another (in L_1):

Lemma 11.2.6. *Whp, the distance in L_1 between any pair of vertices $u, v \in L_1'$ is $O(d^{12})$.*

Proof. For each partition of L_1' as in Lemma 11.2.5, whp there are at least $\Omega\left(\frac{\ell \cdot n^\alpha}{d^{10}}\right)$ distinct endpoints in B that correspond to at least $\Omega\left(\frac{\ell \cdot n^\alpha}{d^{10}}\right)$ edge-disjoint paths between A and B. Now, since $|W_j| \le 2n^\alpha$, for every $1 \le j \le t$, these endpoints must belong to at least:

$$\frac{\Omega\left(\frac{\ell \cdot n^\alpha}{d^{10}}\right)}{2n^\alpha} = \Omega\left(\frac{\ell}{d^{10}}\right)$$

distinct components $\{W_j\}$ in B. Look at $W_j \subseteq B$. Recall that the distance (in L_1) between any $u, v \in W_j$ is at most $4C_2 d$. Hence, there are

$$\Omega\left(\frac{\ell}{d^{10}}\right) \cdot n^\alpha = \Omega\left(\frac{\ell \cdot n^\alpha}{d^{10}}\right)$$

distinct vertices at distance at most $5 + 4C_2 d$ from A. Therefore.

$$|B(A, 5 + 4C_2 d)| \ge |A| \left(1 + \Omega\left(\frac{1}{d^{10}}\right)\right).$$

Now, let $v \in V(L_1')$. There is $1 \le j \le t$ such that $v \in W_j$. Denote $A_0 = W_j$ and $B_0 = V(L_1') \setminus A_0$. Apply the above argument on the partition $V(L_1') = A_0 \cup B_0$. Define $A_1 \subseteq V(L_1')$ as the union of all $\{W_j\}$ at distance at most $5 + 4C_2 d$ from A_0. By the above,

$$A_1 \supseteq B(A, 5 + 4C_2 d) \implies |A_1| \ge |A_0| \left(1 + \Omega\left(\frac{1}{d^{10}}\right)\right).$$

Iteratively define A_i in such a way that $|A_{i+1}| \ge |A_i| \left(1 + \Omega\left(\frac{1}{d^{10}}\right)\right)$ for every $i \in \mathbb{N}$.
After $O(d^{10} \cdot d) = O(d^{11})$ steps of the procedure we double the volume of A_0, and thus, since $|L_1'| = \Theta(n)$, we have to move $O(d^{11} \cdot \log_2 n) = O(d^{12})$ steps to cover more than half of L_1'. Therefore, the distance (in L_1) between any two vertices from L_1' is whp $O(d^{12})$. $\square$

Now, all that remains is to show that typically we do not have a vertex in $L_1 \setminus L_1'$ that is too far from L_1'. If we prove this, we are done, because for any pair of vertices $u, v \in L_1$ we can take vertices $u', v' \in L_1'$, such that u is close to u', and v is close to v', providing a short path from u to v by:

$$u \to u' \to v' \to v.$$

Definition 11.2.7. Let $v \in V(L_1')$. We denote by C_v the set of connected components of $L_1 \setminus L_1'$ that connect to v in G_2 (the set of components C_i such that there exists a $v_i \in C_i$ with $(v_i, v) \in E(G_2)$).

We have already proven that in G_1, whp, every connected component outside of L_1' is of size at most $K_1 d$ for some $K_1 := K_1(\epsilon) > 0$. We now prove the following lemma:

Lemma 11.2.8. *There exists a constant $K_2 := K_2(\epsilon) > 0$, such that whp, for every $v \in V(L_1')$ the set C_v is of size at most $|C_v| \le K_2 d$.*

Observe that

$$L_1 = L_1' \cup \bigcup_{v \in V(L_1)} C_v.$$

Note that the sets C_v are not necessarily disjoint. Assuming Lemma 11.2.8, we have that whp for every $u \in V(L_1)$, there exists $v \in V(L_1')$ such that $u \in C_i \in C_v$ for some component C_i in $L_1 \setminus L_1'$ that is connected to v by an edge from G_2. Thus, the distance between u and L_1' is at most $|C_i| \leq |C_v|$ steps, which the lemma argues is $O(d)$. To conclude, the distance (in L_1) between any two vertices $u, v \in V(L_1)$ is $O(d) + O(d^{12}) + O(d) = O(d^{12})$, completing the proof of Theorem 11.2.1.
We now prove Lemma 11.2.8.

Proof. Assume in contradiction that there exists a $v \in V(L_1)$ such that $|C_v| \geq K_2 d$. Since all connected components in $L_1 \setminus L_1'$ are of size at most $K_1 d$, we can find a subset $\tilde{C} = \cup C_{i_j}$, of components from C_v ($\tilde{C}$ is a set of vertices that form the union of some subset of components of C_v), such that:

$$K_2 d \leq |\tilde{C}| \leq (K_1 + K_2) d.$$

Denote the size of $\tilde{C}$ by k. We note that $\tilde{C}$ satisfies the following properties:

1. $\tilde{C} \cup \{v\}$ is a connected graph in G with $k+1$ vertices;

2. All edges between $\tilde{C}$ and $V(Q^d) \setminus \tilde{C}$ are not present in G_1.

We aim to prove that whp no such set exists. We note that these two conditions concern disjoint sets of edges and are thus independent. We bound the probability by:

1. Applying the union bound over all possible sizes k of $\tilde{C}$: $K_2 d \leq k \leq (K_1 + K_2) d$.

2. Choosing a tree T on $k+1$ vertices ($\leq (k+1)(ed)^k$ possible spanning trees) and requiring its edges to be presented in G (p^k, k edges in G).

3. Choosing a vertex v in T (in $k+1$ ways) and requiring all edges at the boundary of $T \setminus \{v\}$ to be closed in G_1.

We obtain that the probability that such a set $\tilde{C}$ exists is bounded from above by:

$$\begin{aligned}
\sum_{k=K_2 d}^{(K_1+K_2)d} n(k+1)(ed)^k p^k (1-p_1)^{k(d-2\log_2 k)} &\leq n^2 \sum_{k=K_2 d}^{(K_1+K_2)d} \big(e(1+\epsilon)\big)^k e^{-\frac{1+\epsilon-\delta}{d} k(d-2\log_2 k)} \\
&\leq n^2 \sum_{k=K_2 d}^{(K_1+K_2)d} \exp\big(1+\epsilon-\frac{\epsilon^2}{3}-1-\epsilon+\delta+o(1)\big)^k \\
&\leq n^2 \sum_{k=K_2 d}^{(K_1+K_2)d} \exp\big(-\frac{\epsilon^2}{3}+\delta+o(1)\big)^k.
\end{aligned}$$

Taking K_2 to be sufficiently large, and $\delta = \delta(\epsilon) > 0$ sufficiently small, this probability tends to 0, completing the proof. $\square$

Lecture 12

Lecturer: Prof. Michael Krivelevich *Scribe: Binyamin kobzantsev*

12.1 Hitting Time for Connectivity in a Random Cube

12.1.1 Background / Definitions

Notation: $N = |E(Q^d)| = \frac{nd}{2}$, where $n = 2^d$.

Definition 12.1.1. Let $\sigma \in S_N$ be a permutation of $E(Q^d)$. We define a process $\tilde{Q} = \{Q_i\}_{i=0}^N$ on Q^d by:

$$V(Q_i) = V(Q^d) = \{0,1\}^d;$$
$$E(Q_i) = \{e_{\sigma(1)}, \dots, e_{\sigma(i)}\}.$$

Remark 12.1.2. Note that:

$$\overline{K_{2^d}} = Q_0 \subsetneq Q_1 \subsetneq \cdots \subsetneq Q_N = Q^d$$

are graphs on $V(Q^d)$, and $|E(Q_i)| = i$ for all $0 \le i \le N$.

Definition 12.1.3. If, when defining a graph process on Q^d, the permutation $\sigma \in S_N$ is chosen uniformly at random, we say that the process is a *random graph process* on Q^d.

Definition 12.1.4. We say that $\mathcal{A}$ is an *increasing (non-trivial) monotone property* of subgraphs of Q^d if:

1. $\mathcal{A} \subseteq \{G \subseteq Q^d\}$
2. If $G_1 \in \mathcal{A}$ and $G_1 \subseteq G_2 \subseteq Q^d$, then $G_2 \in \mathcal{A}$.
3. $Q_0 \notin \mathcal{A}$ and $Q_N = Q^d \in \mathcal{A}$.

Definition 12.1.5. Given a graph process $\tilde{Q} = \{Q_i\}_{i=0}^N$ and a monotone non-trivial property $\mathcal{A}$, we define the *hitting time* of $\mathcal{A}$ with respect to $\tilde{Q}$ by:

$$\tau_{\mathcal{A}}(\tilde{Q}) = \min\{0 \le i \le N \mid Q_i \in \mathcal{A}\}.$$

Thus, $\tau_{\mathcal{A}}(\tilde{Q})$ is the first moment in the process where $\mathcal{A}$ holds.

Remark 12.1.6. If $\sigma \in S_N$ is chosen randomly, then for any property $\mathcal{A}$, the hitting time $\tau_{\mathcal{A}}(\tilde{Q})$ is a random variable. One can therefore study its distribution, find its expectation, variance, etc.

Theorem 12.1.7 (Bollobás, 1990)**.** *In a random graph process $\tilde{Q}$, whp:*

$$\tau_C(\tilde{Q}) = \tau_1(\tilde{Q}),$$

where $\tau_C(\tilde{Q})$ is the hitting time for connectivity, and $\tau_1(\tilde{Q})$ is the hitting time for the disappearance of isolated vertices.

Remark 12.1.8. Meaning in words: In a random graph process of the cube, whp, the graph becomes connected exactly at the time when the last isolated vertex disappears.

Remark 12.1.9. For every graph process on Q^d, deterministically:

$$\tau_C(\tilde{Q}) \geq \tau_1(\tilde{Q}).$$

Proof (Diskin, Krivelevich). It is enough to prove that there exists $t := t(d)$ such that whp in a random graph process $\tilde{Q}$, the graph Q_t satisfies:

1. $\delta(Q_t) = 0$.
2. The connected components of Q_t are either isolated vertices or a giant component L_1 of size $|L_1| = 2^d(1 - o(1))$.
3. For any two isolated vertices $u \neq v$ in Q_t, we have that u and v are not connected by an edge of Q^d.

Indeed, at time t, there are isolated vertices and therefore $\tau_C(\tilde{Q}) \geq \tau_1(\tilde{Q}) > t$. Also, every edge touching an isolated vertex has its other endpoint in L_1. Therefore, adding any edge touching an isolated vertex v in Q_t connects v to L_1. Hence, exactly at the moment when the last isolated vertex of Q_t disappears, then this vertex and all other isolated vertices are connected to L_1, thus we get a connected graph.
Due to monotonicity (similar to $G(n, p)$ and a random graph process of K_n), it is enough to prove the following claim:

Claim 12.1.10. *Assume $p = \frac{1}{2} - \varepsilon$, where $\varepsilon > 0$ is a sufficiently small constant. Consider a random cube Q_p^d. Then, whp:*

(1) There are isolated vertices in Q_p^d.

(2) $|L_1(Q_p^d)| = 2^d(1 - o(1))$, and any other component besides L_1 is an isolated vertex.

(3) No two isolated vertices in Q_p^d are adjacent in Q^d.

Proof of (1). Recall that Q^d is a bipartite graph with parts O, E of size $|O| = |E| = 2^{d-1}$. The random variable X counts the number of isolated vertices in part O of Q_p^d and is distributed as $\text{Bin}(2^{d-1}, (1-p)^d)$. Hence,

$$\mathbb{E}[X] = 2^{d-1} \cdot (1-p)^d = 2^{d-1}\left(\frac{1}{2} + \varepsilon\right)^d = 2^{d-1}\frac{(1+2\varepsilon)^d}{2^d} = \frac{(1+2\varepsilon)^d}{2} \xrightarrow[d\to\infty]{} \infty.$$

Therefore, whp $X > 0$, which implies that whp there are isolated vertices in Q_p^d. $\square$

Proof of (3). We want to show that whp, for every edge $(u, v) \in E(Q^d)$, it is not the case that both u and v are isolated in Q_p^d, i.e., $d_{Q_p^d}(u) = d_{Q_p^d}(v) = 0$ is false.
For any edge $e = (u, v) \in E(Q^d)$:

$$\mathbb{P}\left(d_{Q_p^d}(u) = d_{Q_p^d}(v) = 0\right) = (1-p)^{2d-1} = \left(\frac{1}{2} + \varepsilon\right)^{2d-1} = \frac{(1+2\varepsilon)^{2d-1}}{2^{2d-1}}.$$

Thus, the expected number of such pairs of adjacent (in Q^d) isolated vertices (in Q_p^d) is

$$|E(Q^d)| \cdot \frac{(1+2\varepsilon)^{2d-1}}{2^{2d-1}} = \frac{2^d \cdot d}{2} \cdot \frac{(1+2\varepsilon)^{2d-1}}{2^{2d-1}} \xrightarrow[d\to\infty]{} 0,$$

for a sufficiently small constant $\varepsilon > 0$. By Markov's inequality, whp there are no such pairs. $\square$

Proof of (2). We use two-round exposure (sprinkling). Let us define $p_2 = \varepsilon$, and define p_1 such that:

$$1 - p = (1 - p_1)(1 - p_2).$$

This implies $p_1 \geq \frac{1}{2} - 2\varepsilon$. Recall that $G = G_1 \cup G_2 \sim Q_p^d$ where $G_i \sim Q_{p_i}^d$ for $i = 1, 2$. Let L_1' denote the giant component of G_1, and let L_1 denote the giant component of G. We will show that whp in G_1 there is no component of size:

$$k \in \left[2, n^{\frac{1}{3}}\right].$$

Indeed, if there exists a component in G_1 of size k, then there exists a tree T of size k in Q^d whose edges are contained in G_1, and there are no edges in G_1 between $V(T)$ and $V(Q^d) \setminus V(T)$.
Therefore, using (weak) Harper's isoperimetric inequality, the probability that there exists a component of size $k \in [2, n^{1/3}]$ is at most:

$$\begin{aligned}
\sum_{k=2}^{n^{1/3}} n(ed)^{k-1}(1-p_1)^{k(d-2\log_2 k)} &\leq \sum_{k=2}^{n^{1/3}} n(ed)^{k-1}\left(\frac{1}{2}+2\varepsilon\right)^{k(d-2\log_2 k)} \\
&\leq \sum_{k=2}^{d} n\left(ed\left(\frac{1}{2}+2\varepsilon\right)^{d(1-o(1))}\right)^k + \sum_{k=d+1}^{n^{1/3}} n\left(ed\left(\frac{1}{2}+2\varepsilon\right)^{\frac{d}{3}}\right)^k \\
&\leq d \cdot n^{-\frac{1}{3}} + n^{\frac{1}{3}} \cdot n^{-\frac{d}{4}} = o(1),
\end{aligned}$$

where the last inequality holds since $\epsilon > 0$ is sufficiently small. Therefore, whp there are no components of size $k \in [2, n^{1/3}]$ in G_1. Furthermore,

$$\mathbb{P}(d_{G_1}(v) = 0) = (1-p_1)^d \leq \left(\frac{1}{2}+2\varepsilon\right)^d \leq n^{-1+9\varepsilon},$$

which implies:

$$\mathbb{E}[\text{number of isolated vertices in } G_1] \leq n^{9\varepsilon}.$$

By Markov's inequality, whp there are at most $n^{10\varepsilon}$ isolated vertices in G_1. Now, we will use the edges of G_2 to connect the large components of G_1. Let us define:

$$W = \{v \in V(Q^d) \mid v \text{ belongs to a component of size } \geq n^{\frac{1}{3}} \text{ in } G_1\}.$$

Then, whp:

$$|W| \geq n - n^{10\varepsilon}.$$

If W does not merge into one component in G, then there exists a partition $W = A \cup B$ where $A, B \neq \emptyset$, with no connecting edge between A and B in G_2. We can assume $a = |A| \leq |B|$. Notice that

$$|E_{Q^d}(A, A^c)| \geq |A| = a.$$

Some of these edges leaving A in Q^d can lead to $A^c \setminus B$. Since there are at most $n^{10\varepsilon}$ vertices outside of W, at most $n^{10\varepsilon}d$ such edges do not touch B. Hence, since $a \geq n^{1/3}$, the number of edges in Q^d between A and B is at least:

$$|E_{Q^d}(A, B)| \geq |A| - n^{10\varepsilon}d \geq \frac{a}{2}.$$

Therefore:

$$\mathbb{P}(E_{G_2}(A, B) = \emptyset) \leq (1-p_2)^{\frac{a}{2}} = (1-\varepsilon)^{\frac{a}{2}}.$$

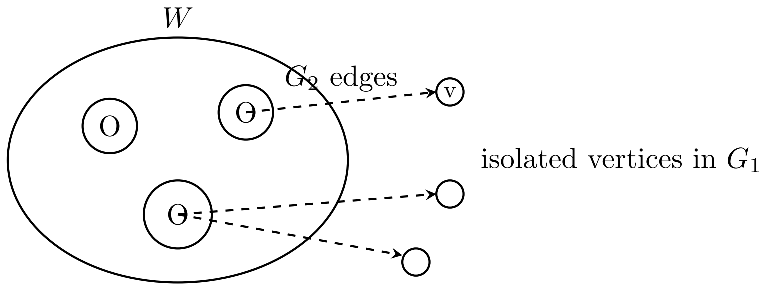


Figure 12.14: The set W contains components of size $\geq n^{1/3}$ in G_1. Some edges from G_2 connect these components to the isolated vertices outside W.

Denote by s the number of components of size at least $n^{1/3}$. Obviously, $s \leq n$. Thus, the probability that W does not connect into a single component in $G = G_1 \cup G_2$ is whp at most

$$\sum_{a=n^{1/3}}^{n/2} \left(\sum_{i=1}^{a/n^{1/3}} \binom{s}{i} \right) \cdot (1-\varepsilon)^{\frac{a}{2}} \leq \sum_{a=n^{1/3}}^{n/2} s^{\frac{a}{n^{1/3}}} \cdot e^{-\frac{\varepsilon a}{2}} \leq \sum_{a=n^{1/3}}^{n/2} \left(n^{\frac{1}{n^{1/3}}} \cdot e^{-\frac{\varepsilon}{2}} \right)^a = o(1).$$

Therefore, whp in $G = G_1 \cup G_2$, the vertices of W merge into a single component. Recall that whp, for every edge in Q^d incident to a vertex isolated in G_1, its other endpoint is in W (apply (3) to G_1). Thus, after exposing G_2, we obtain whp a single giant component L_1 in G, and, outside of it, only isolated vertices. □

□