

Chapter 7 Random Graphs (1)

7.1 The notion of a random graph

Let V be a fixed set of n elements

$V = \{0, 1, \dots, n-1\}$. Let \mathcal{G} be the set of graphs with vertex set V .

Goal: Turn \mathcal{G} into a probability space!

Let $p \in [0, 1]$, $p \in \mathbb{R}$. $\forall e = \{i, j\} \subseteq \{0, 1, \dots, n-1\}, i \neq j$
think of some random experiment which decides whether or not e is an edge of G . If these experiments are performed independently and for each of them the probability of success, i.e. the probability of accepting e as an edge of G , is equal to the fixed $p \in [0, 1]$.

Then if G_0 is some fixed graph on V with m edges, the elementary event $\{G_0\}$ has a probability $p^m (1-p)^{\binom{n}{2}-m}$ or with $q := 1-p$ $\mathbb{P}(\{G_0\}) = p^m q^{\binom{n}{2}-m}$.

By defining the prob of every elementary event in \mathcal{G} we define a probability measure in \mathcal{G} .

So it is left to be checked that such a ^{prob} measure on \mathcal{G} , one for which all edges occur independently with probability p , does indeed exist.

Formally we define a probability space $\Omega_e := \{0_e, 1_e\}$ for each $e = \{i, j\} \subseteq \{0, \dots, n-1\}, i \neq j$ by choosing

$$P_e(\{1_e\}) = p \quad P_e(\{0_e\}) = q \text{ as the probabilities}$$

of the elementary events "e exists", "e does not exist".

As our desired probability space \mathcal{G} we take the product space $\Omega := \prod \Omega_e$ where

$$[V]^2 := \{ \{i, j\} : i, j \in \overline{0, n-1}, i \neq j \}$$

Thus, formally, an element of $\sqrt[n]{\Omega}$ ^{the sample space} is a mapping

$$\omega : \{e = (i, j)\} \subseteq \{0, \dots, n-1\} \rightarrow \bigcup_{e = (i, j) \subseteq \{0, \dots, n-1\}} \{0_e, 1_e\}$$

$$\omega(e) \in \{0_e, 1_e\}$$

and the probability measure P on Ω is the product measure of all the measures P_e .

In practice of course we identify ω with the graph G on V whose edge set is

$$E(G) = \{e : \omega(e) = 1_e\}$$

we call G a random graph with on V with edge probability p .

Now we may call any set of graphs on V an event in $G(n, p)$.

In particular, $\forall e \in [V]^2$ the set

$$A_e := \{\omega : \omega(e) = 1_e\}$$

of all graph G on V with $e \in E(G)$ is an event; the event that e is an edge of G .

Now we can formally prove what has been guiding our intuition:

Proposition 7.1 The events A_e are independent and occur with probability p .

Proof: By definition $A_e = \{1_e\} \times \prod_{e' \neq e} \Omega_{e'}$

Since P is the product measure of all $P_{e'}$ we

have
$$\mathbb{P}(A_e) = p \cdot \prod_{e' \neq e} 1 = p$$

Similarly if $\{e_1, \dots, e_k\}$ is any subset of $[V]^2$ (3)

then
$$P(A_{e_1} \cap \dots \cap A_{e_k}) = P(\{e_1\} \times \dots \times \{e_k\} \times \prod_{e \notin \{e_1, \dots, e_k\}} \Omega_e)$$

$= p^k = P(A_{e_1}) \cdot \dots \cdot P(A_{e_k})$, thus proving the independence of $A_e, e \in [V]^2$

\mathbb{P} $P(\cdot)$ \rightarrow set of all 2 element subsets $\{i, j\}$ with $i \neq j, i, j \in \overline{0, n-1}$

Remark. \mathbb{P} is determined uniquely by the value of p and the assumption that the events A_e are independent. In order to calculate probabilities in $G(n, p)$ it is enough to work with two assumptions and the theoretical model of $G(n, p)$ as product of prob. spaces will not be referred to, any more.

Some examples for such computations:

Consider the event that $G \in G(n, p)$ contains some fixed graph H on a subset of V as a subgraph. Let $|H| =: k$ and $|E(H)| =: \ell$. The probability of this event, $H \subseteq G$, is the product of the probab. over all edges $e \in H$ so

$P(H \subseteq G) = p^\ell$

The probability that H is an induced subgraph of G

$P(H \subseteq G) = p^\ell \binom{k}{2} - \ell$

The probability P_H that G has an induced subgraph which is isomorphic to H is not so simple to compute, because the possible instances of H on subsets of V overlap, and the events that they occur in G are not independent.

However we get an upper bound: $P_H \leq \sum_{\substack{U \subseteq V \\ |U|=k}} P[H \simeq G[U]]$ (4)

$E = \bigcup_{\substack{U \subseteq V \\ |U|=k}} [G[U] \simeq H]$

This holds because P_H is the measure of the union $\bigcup_{\substack{U \subseteq V \\ |U|=k}} [H \simeq G[U]]$ where $[H \simeq G[U]]$ denotes the set of all graphs $G \in \mathcal{G}(n, p)$ for which $G[U] \simeq H$ holds.

If $H = \overline{K_k}$ (the complement of K_k) we get

Lemma 7.2 $\forall n, k \in \mathbb{N}$ with $n \geq k \geq 2$ the probability that $G \in \mathcal{G}(n, p)$ has a set of k independent vertices is at most $P(\alpha(G) \geq k) \leq \binom{n}{k} p^{\binom{k}{2}}$.

Proof: The probability that a fixed k -set $U \subseteq V$ is independent in G is $p^{\binom{k}{2}}$. And there are $\binom{n}{k}$ such sets. \square

Analogously the probability that $G \in \mathcal{G}(n, p)$ contains a K_k is at most $P(\omega(G) \geq k) \leq \binom{n}{k} p^{\binom{k}{2}}$

Now notice that if k is fixed and n is small enough the sum of these bounds is < 1 (convince yourself or a homework). $\Rightarrow P(\alpha \geq k \vee \omega \geq k) < 1 \Rightarrow P(\alpha < k \wedge \omega < k)$
 This means: there are graphs which contain neither K_k nor $\overline{K_k}$ as an induced subgraph.

Then any such n is a lower bound for the Ramsey number of K_k to be defined in the following section.

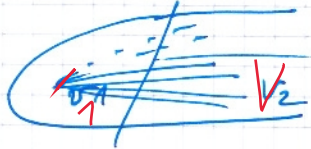
Theorem 7.3 (Ramsey 1930) On Ramsey numbers

$\forall r \in \mathbb{N}$ there exists an $n \in \mathbb{N}$ such that every graph of order at least n contains either K_r or $\overline{K_r}$ as an induced subgraph.

Now how would we proof this theorem?

We could try to build a K_r or a \bar{K}_r in G inductively, starting with an arbitrary vertex $v_1 \in V_1 := V(G)$.

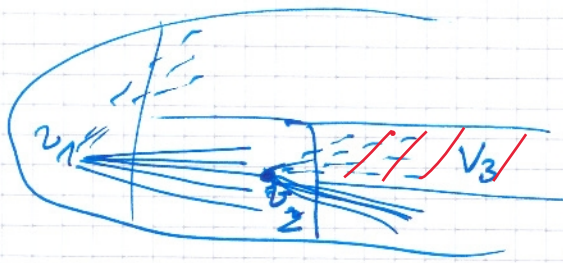
If $|G|$ is large, then there will be a large set of vertices $V_2 \subseteq V_1 \setminus \{v_1\}$ that are either all adjacent to v_1 or all non-adjacent to v_1 . We may think of v_1 as



the first vertex of K_r or \bar{K}_r all other vertices of it lying in V_2

Then let us choose another vertex v_2 in V_2 for K_r or \bar{K}_r . Since V_2 is large we will have a set of vertices in V_2 , call it V_3 , which is large enough and is such that all of its vertices have the same status with respect to v_2 : either all of them are adjacent to v_2 or none of them is.

We then continue our search inside V_3 and so on



How long can we go in this way? This depends of course on the size of the initial

set V_1 ; each set V_i has at least half the size of its predecessor V_{i-1} , so we shall be able to complete s construction steps if $|G| \geq 2^s$.

It turns out that $s = 2r - 3$ vertices suffices to find among them either a K_r or a \bar{K}_r .

Proof The statement is trivial for $r=1$. So let us assume $r \geq 2$. let $n := 2^{2r-1}$ and let G be a graph of order at least n , $|G| \geq n$.

We define a sequence V_1, \dots, V_{2r-2} of sets of chosen vertices $v_i \in V_i$ with the following properties (inductively)

(i) $|V_i| = 2^{2r-2-i}$ ($i = 1, 2, \dots, 2r-2$)

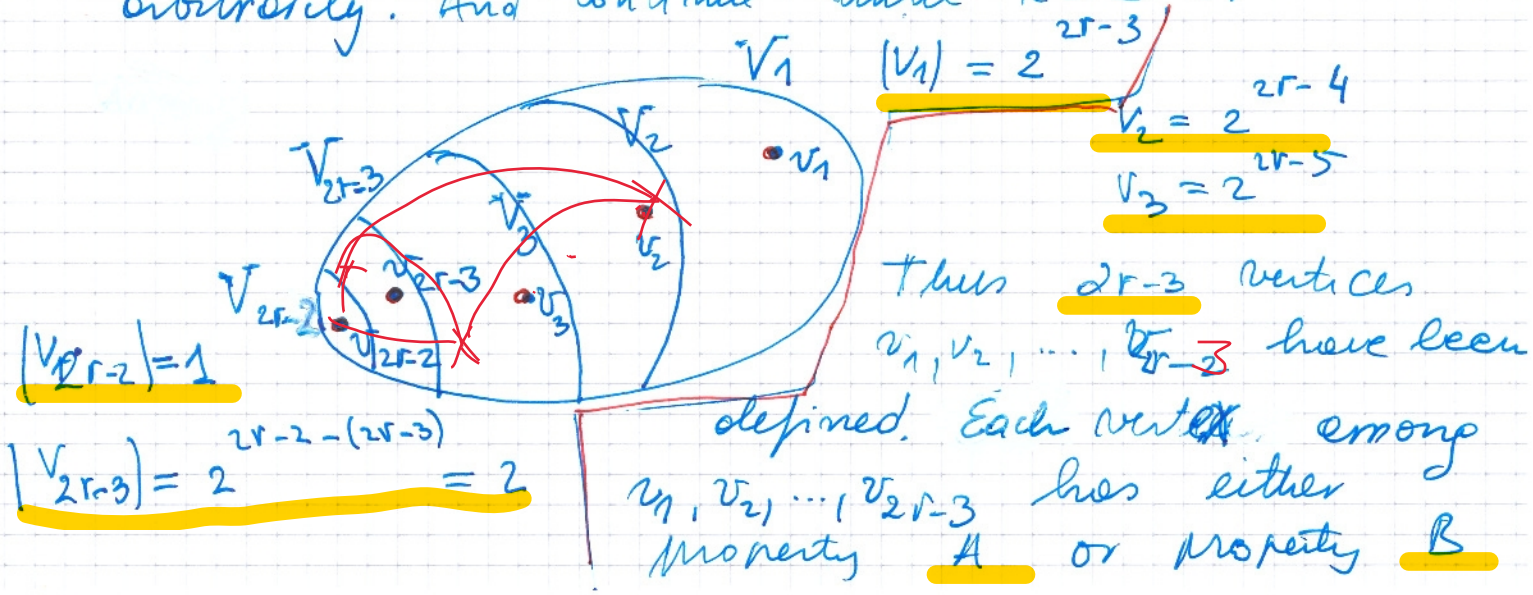
(ii) $V_i \subseteq V_{i-1} \setminus \{v_{i-1}\}$, $i = 2, \dots, 2r-2$

(iii) v_{i-1} is adjacent either to all vertices in V_i or to no vertex in V_i , $i = 2, \dots, 2r-2$.

let $V_1 \subseteq V(G)$ be any set of 2^{2r-3} vertices and pick $v_1 \in V_1$ arbitrarily.

Then (i) holds for $i=1$ while (ii) and (iii) hold trivially. Suppose now that V_{i-1} and $v_{i-1} \in V_{i-1}$ have been chosen so as to satisfy (i)-(iii) for $i-1$, where $1 < i \leq 2r-2$.

since $|V_{i-1} \setminus \{v_{i-1}\}| = 2^{2r-1-i} - 1$ is odd V_{i-1} has a subset V_i satisfying (i)-(iii); we pick $v_i \in V_i$ arbitrarily. And continue until $i = 2r-2$.



A: v_i is adjacent to all vertices in V_{i+1}

B: v_i is adjacent to none of the vertices in V_{i+1}

By the pigeon hole principle there exist $\lceil \frac{2r-3}{2} \rceil = r-1$ vertices among v_1, \dots, v_{r-3} holding the same property (A or B). These $r-1$ vertices together with v_{r-2} build either a K_r (if the $r-1$ vertices have property A) or a \overline{K}_r (if the $r-1$ vertices have property B).

Def The smallest integer $n \in \mathbb{N}$ associated with r in Theorem 7.3 is the Ramsey number $R(r)$ of r . We have shown that $R(r) \leq 2^{\frac{2r-3}{2}}$.

By using a probabilistic argument we will show that $R(r) \geq 2^{\frac{r}{2}}$. Thus $2^{\frac{r}{2}} \leq R(r) \leq 2^{\frac{2r-3}{2}}$.

Remark Ramsey's theorem may be rephrased as follows:
 If $H = K_r$ and G is a graph with sufficiently many vertices, then either G itself or its complement \overline{G} contains a subgraph which is isomorphic to H . Clearly this statement is true for any graph H because $H \subseteq K_{|H|}$ for any H .

Consider now the smallest $n \in \mathbb{N}$ such that every graph of order n either G contains H or \overline{G} contains H as a subgraph (indeed an isomorphic copy of H); this is the Ramsey number $R(H)$ of H .

Notice that if H has a few edges, we would expect that it is easier to find a "copy" of H in G or \overline{G} than if we were looking for a copy of K_r . So we would expect $R(H) \leq R(r)$ $R(r) = R(K_r)$

Def More generally, let $R(H_1, H_2)$ denote the smallest $n \in \mathbb{N}$ such that $H_1 \subseteq G$ or $H_2 \subseteq \bar{G}$ for every graph G on n vertices.

$R(K_r) = R(K_r, K_r)$ $H_1 = K_r$
 $H_2 = K_r$

at least

For most graphs H_1, H_2 only very rough estimates are known for $R(H_1, H_2)$, lower bounds provided by random graphs are in general sharper than any bounds provided by explicit construction.

The following is one of the few exact Ramsey numbers known.

Theorem 7.4 Let $s, t \in \mathbb{N}$ and let T be a tree of order t .

Then $R(T, K_s) = (s-1)(t-1) + 1$.

$t=2 \quad s=3$

No proof here but in the exercises

$R(T_2, K_3) = 2 \cdot 2 + 1 = 5$
 \square
 $\frac{3 \cdot 2}{n=4} = 3$
 \vdots

Back to random graphs:

$n=3$

Theorem 7.5 (Erdős 1947)

$\forall k \in \mathbb{N}, k \geq 3$, the Ramsey number $R(k)$ of k satisfies

$R(k) \geq 2^{k/2}$

Proof. For $k=3$ we trivially have $R(3) \geq 3 > 2^{3/2}$,

so let $k \geq 4$ hold. We show that $\forall n \leq 2^{k/2}$ and

$G \in \mathcal{G}(n, \frac{1}{2})$ the probability $P(\alpha(G) \geq k)$ and $P(\omega(G) \geq k)$

or both less than $\frac{1}{2}$. $P(\alpha(G) \geq k \vee \omega(G) \geq k) < \frac{1}{2} + \frac{1}{2} = 1$

For $p=q=\frac{1}{2}$ lemma 7.2 implies $P(\alpha(G) \geq k) \leq \binom{n}{k} \left(\frac{1}{2}\right)^{\binom{k}{2}}$

and $P(\omega(G) \geq k) \leq \binom{n}{k} \left(\frac{1}{2}\right)^{\binom{k}{2}}$

if $n \leq 2^{k/2}$ and using $k! > 2^k$ $\forall k \geq 4$ we get

$P[\alpha(G) \geq k] \leq \binom{n}{k} \left(\frac{1}{2}\right)^{\binom{k}{2}} \leq \frac{n(n-1) \dots (n-k+1)}{k!} \frac{1}{2^{\frac{k(k-1)}{2}}}$
 $< \frac{n(n-1) \dots (n-k+1)}{2^k} 2^{-\frac{k(k-1)}{2}} \leq \left(\frac{n}{2}\right)^k 2^{-\frac{k(k-1)}{2}} < \frac{1}{2}$

$$\leq \frac{2^{\frac{k^2}{2}}}{2^k} \cdot 2^{-\frac{1}{2}k(k-1)} = 2^{\frac{k^2}{2} - k - \frac{k^2}{2} + \frac{k}{2}} = 2^{-\frac{k}{2}} < \frac{1}{2}$$

Analogously $P[\omega(G) \geq k] < \frac{1}{2}$.

Putting things together we get:

$$P[\alpha(G) \geq k \text{ ~~or~~ } \omega(G) \geq k] \leq P[\omega(G) \geq k] + P[\alpha(G) \geq k] < \frac{1}{2} + \frac{1}{2} = 1 \text{ for } k \geq 3 \text{ and } n \leq 2^{k/2}$$

$\Rightarrow \exists G \in \mathcal{G}(n, p)$ s.t. $\alpha(G) < k$ and $\omega(G) < k$, i.e.

$\forall n \leq 2^{k/2} \exists$ a graph G with n vertices which contains neither a stable set on k vertices nor a clique on k vertices

$$\Rightarrow R(k) \geq 2^{k/2}$$

□

In the context of random graphs each of the familiar graph invariants X like average degree, connectivity, growth, chromatic number, and so on, can be interpreted as a non-negative random variable X on $\mathcal{G}(n, p)$,

$$X: \mathcal{G}(n, p) \rightarrow [0, \infty)$$

$$G \rightarrow X(G)$$

\rightarrow average degree, connectivity, ...

The expectation is given as

$$E(X) = \sum_{G \in \mathcal{G}(n, p)} P(\{G\}) \cdot X(G)$$

A

If $X \in \mathbb{Z}_+$ we get $E(X) = \sum_{k \in \mathbb{Z}_+} k P[X(G) = k] =$

$$= \sum_{k \in \mathbb{N}} P[X \geq k]$$

$$A = P(X(G)=1) + 2P(X(G)=2) + 3P(X(G)=3) + \dots$$

$$= P(X(G)=1) + P(X(G)=2) + P(X(G)=3) + \dots$$

$\rightarrow P(X(G) \geq 1)$
 $\rightarrow P(X(G) \geq 2)$

$$P(X(G) \geq \alpha) + \dots \rightarrow P(X(G) \geq \alpha)$$

(10)

Computing the ^{expectation} of a random variable X can be a simple and effective way to establish the existence of a graph G such that (a) $X(G) < \alpha$ for some fixed $\alpha > 0$, and moreover, (b) G has some desired property P .

Indeed: if $E(X)$ is small, then $X(G)$ cannot be large for more than a few graphs in $\mathcal{G}(n, p)$, because $X(G) \geq 0, \forall G \in \mathcal{G}(n, p)$. Hence $X(G)$ is small for many graphs in $\mathcal{G}(n, p)$ and it is reasonable to expect that property P holds for some of them.

This idea is the background of lots of non-constructive existence proofs using random graphs. This ^{reasoning} is at the heart of the so-called probabilistic method. Quantifying we get the Markov's inequality:

Lemma 7.6 (Markov's inequality)

Let $X \geq 0$ be a random variable on $\mathcal{G}(n, p)$. and $\alpha > 0$.

$$\text{Then } P[X \geq \alpha] \leq \frac{E(X)}{\alpha}$$

Proof: $E(X) = \sum_{G \in \mathcal{G}(n, p)} P(\{G\}) \cdot X(G) \geq \sum_{\substack{G \in \mathcal{G}(n, p): \\ X(G) \geq \alpha}} P(\{G\}) \cdot \alpha = P(X \geq \alpha) \cdot \alpha$

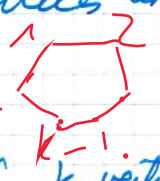
Remark: Since our probability spaces are finite the expectation can be often computed by using some double counting argument.

We illustrate this ^{idea} in the computation of the expected number of cycles of some given length $l \geq 3$ in a random graph $G \in \mathcal{G}(n, p)$ $E(X) = ?$

Let $X: \mathcal{G}(n, p) \rightarrow \mathbb{N}$ be a random variable which assigns the number $X(G)$ of k -cycles to a graph $G \in \mathcal{G}(n, p)$, i.e.

$X(G)$ is the number of subgraphs of G isomorphic to C_k

The number of potential such cycles, i.e. the cardinality of \mathcal{C}_k where \mathcal{C}_k is the set of all k -cycles with vertices in V is given as $|\mathcal{C}_k| = \frac{(n)_k}{2k}$ because there are n



$(n)_k := n(n-1) \dots (n-k+1)$ ways to choose a sequence of k vertices in V and each cycle is identified by $2k$ of those sequences.

Lemma 7.7 The expected number of k -cycles in $\mathcal{G}(n, p)$ is

$$E(X) = \frac{(n)_k}{2k} \cdot p^k$$

Proof Consider $C \in \mathcal{C}_k$ (fixed) its indicator random variable

$$X_C: \mathcal{G}(n, p) \rightarrow \{0, 1\} \text{ defined by } X_C: G \rightarrow \begin{cases} 1 & \text{if } C \subseteq G \\ 0 & \text{if } C \not\subseteq G \end{cases}$$

Since X_C takes only 1 as a positive value we have

$$E(X_C) = P[X_C = 1] = \sum_{G \in \mathcal{G}(n, p)} P[G] \cdot X_C(G) =$$

$$= \sum_{\substack{G \in \mathcal{G}(n, p) \\ X_C(G) = 1}} P[G] = \sum_{\substack{G \in \mathcal{G}(n, p) \\ C \subseteq G}} P[G] = P[G \supseteq C]$$

$$= p^k$$

Then $\forall G, X(G) = \sum_{C \in \mathcal{C}_k} X_C(G)$, or shortly $X = \sum_{C \in \mathcal{C}_k} X_C$.

By linearity of expectation we get

$$E(X) = \sum_{C \in \mathcal{C}_k} E(X_C) = \sum_{C \in \mathcal{C}_k} P(C \subseteq G) = \frac{(n)_k}{2k} \cdot p^k$$



7.2 The probabilistic method

Idea: In order to prove the existence of an object with some desired property, one defines a probability space on some ^{considered} definitely non-empty class of objects and then shows that an elem. of this space has the desired property with positive probability.

We illustrate the idea in the proof of one of the earliest results using this method, Erdős's theorem on graph with large girth and large chromatic number.

Theorem 7.8 (Erdős 1959)

$\forall k \in \mathbb{N}, \exists$ a graph G with $g(G) > k$ and $\chi(G) > k$.

We call a cycle of length $\leq k$ a short cycle. and a set of $\frac{|G|}{k}$ or more vertices big. so in order to prove Th. 7.8 it is enough to find a graph without short cycles and without big independent sets of vertices; then the color classes, all of them are small, and we need more than k colors to color G .

How to find such a graph? If we choose p small enough a random graph in $\mathcal{G}(n, p)$ is unlikely to contain any (short) cycles. If we choose p large enough then G is unlikely to have big independent sets.

Question: Do there 2 ranges of p overlap?

i.e. can we choose p such that for some $n \in \mathbb{N}$

$$P[g \leq k] < \frac{1}{2} \text{ and } P[\alpha \geq \frac{n}{k}] < \frac{1}{2} ?$$

If yes then $\mathcal{G}(n, p)$ contains at least one graph without short cycles and without big independent sets.

Such a choice is impossible!

(13)

$p < \frac{1}{n}$ should hold to make short cycles unlikely
but this choice of p makes cycles of all unlikely
 $\Rightarrow G$ will be bipartite (w.h.p. = with high probability)
and will have at least $\frac{n}{2}$ independent vertices
to a big indep. set

However: in order to make big indep. sets unlikely
we need $p > \frac{1}{n}$. We will set $p = n^{-\epsilon}$ for $\epsilon > 0$

If ϵ is small enough then there will be only a few
short cycles in G and if in every of them a vertex
is deleted the remaining graph H will have no
short cycles and its independence number is at most
that of G . Since H is not much smaller than G ,
its chromatic number will still be large!

Lemma 7.9 let $k > 0, k \in \mathbb{Z}$, and let $p = p(n)$ be a
function of n s. that $p \geq (6k \ln n) n^{-1}$ for n large.
Then $\lim_{n \rightarrow \infty} P[\alpha \geq \frac{1}{2} n/k] = 0$.
prob of edge existence depends on no. of vertices!

Proof: $\forall n, r \in \mathbb{N}$ with $n \geq r \geq 2, \forall G \in \mathcal{G}(n, p)$:

Lemma 7.2 implies

$$\begin{aligned} P[\alpha \geq r] &\leq \binom{n}{r} p^{\binom{r}{2}} \leq n^r p^{\binom{r}{2}} = \left(n p^{\frac{r-1}{2}} \right)^r \\ &\leq \left(n (1-p)^{\frac{r-1}{2}} \right)^r \leq \left(n e^{-\frac{p(r-1)}{2}} \right)^r \quad (*) \\ &\quad \text{because } 1-p \leq e^{-p} + p \end{aligned}$$

If $p \geq (6k \ln n) n^{-1}$ and $r \geq \frac{1}{2} n/k \Rightarrow \frac{r}{2} \geq \frac{3}{2} \frac{k \ln n \cdot n}{n} \cdot \frac{n}{4k}$

We get $n e^{-p(r-1)/2} = n e^{-\frac{pr}{2} + \frac{p}{2}} \leq n^{-\frac{3}{2} \ln n + \frac{p}{2}}$

$$\leq n n^{-3/2} e^{1/2} = \sqrt{\frac{e}{n}} \xrightarrow{n \rightarrow \infty} 0$$

Thus with $p \geq \frac{6k \ln k}{n}$ for n large we get for 14

$$r := \left\lfloor \frac{1}{2} \frac{n}{k} \right\rfloor$$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\alpha \geq \frac{1}{2} \frac{n}{k} \right) = \lim_{n \rightarrow \infty} \mathbb{P} \left[\alpha \geq r \right] \leq$$

$$\leq \lim_{n \rightarrow \infty} \left(\sqrt{\frac{e}{n}} \right)^{\left\lfloor \frac{1}{2} \frac{n}{k} \right\rfloor} \leq \lim_{n \rightarrow \infty} \left(\frac{3}{n} \right)^{\frac{1}{2} \frac{n}{k}} = 0 \quad \square$$