

# The coupling method for inhomogeneous random intersection graphs

Katarzyna Rybarczyk\*

Adam Mickiewicz University  
ul. Umultowska 87,  
61-614 Poznań, Poland  
kryba@amu.edu.pl

Submitted: Apr 14, 2015; Accepted: Mar 29, 2017; Published: Apr 13, 2017  
Mathematics Subject Classifications: 05C80, 60C05

## Abstract

We present new results concerning threshold functions for a wide family of random intersection graphs. To this end we improve and generalize the coupling method introduced for random intersection graphs so that it may be used for a wider range of parameters. Using the new approach we are able to tighten the best known results concerning random intersection graphs and establish threshold functions for some monotone properties of inhomogeneous random intersection graphs. Considered properties are:  $k$ -connectivity, matching containment and hamiltonicity.

## 1 Introduction

Since their introduction by Karoński, Scheinerman, and Singer–Cohen [14] random intersection graphs have been attracting attention due to their interesting structure and wide applications. The random intersection graph model appears in problems concerning for example “gate matrix layout” for VLSI design (see e.g. [14]), cluster analysis and classification (see e.g. [12]), analysis of complex networks (see e.g. [2, 9]), secure wireless networks (see e.g. [4]) or epidemics ([8]). Several generalizations of the model have been proposed, mainly in order to adapt it to use for specific purposes. In this paper we consider the  $\mathcal{G}(n, m, \bar{p})$  model studied for example in [1, 3, 7, 15]. Alternative ways of generalizing the model defined in [14] are given for example in [9] and [12].

In a random intersection graph  $\mathcal{G}(n, m, \bar{p})$  there is a set of  $n$  vertices  $\mathcal{V} = \{v_1, \dots, v_n\}$ , an auxiliary set of  $m = m(n)$  features  $\mathcal{W} = \{w_1, \dots, w_{m(n)}\}$ , and a vector  $\bar{p} = \bar{p}(n) = (p_1, \dots, p_{m(n)})$  such that  $p_i \in (0, 1)$ , for each  $1 \leq i \leq m$ . To each vertex  $v \in \mathcal{V}$  we

---

\*Supported by National Science Center grant DEC-2011/01/B/ST1/03943.

attribute a set of its features  $W(v) \subseteq \mathcal{W}$  such that for each  $i$ ,  $1 \leq i \leq m$ ,  $w_i \in W(v)$  with probability  $p_i$  independently of all other features and vertices. If  $w \in W(v)$  then we say that  $v$  has chosen  $w$ . Any two vertices  $v, v' \in \mathcal{V}$  are connected by an edge in  $\mathcal{G}(n, m, \bar{p})$  if  $W(v)$  and  $W(v')$  intersect. If  $\bar{p}(n) = (p, \dots, p)$  for some  $p \in (0, 1)$  then  $\mathcal{G}(n, m, \bar{p})$  is a random intersection graph defined in [14]. We denote it by  $\mathcal{G}(n, m, p)$ .

The random intersection graph model is very flexible and its properties change a lot if we alter the parameters. For example  $\mathcal{G}(n, m, p)$  for some ranges of parameters behaves similarly to a random graph with independent edges (see [11, 17]) but in some cases it exhibit large dependencies between edge appearance (see for example [14, 19]). It was proved in [18] that in both cases  $\mathcal{G}(n, m, p)$  may be coupled with a random graph with independent edges so that with probability tending to 1 as  $n \rightarrow \infty$ , a graph with independent edges is a subgraph of  $\mathcal{G}(n, m, p)$ . It is also explained how this coupling may be used to obtain sharp results on threshold functions for  $\mathcal{G}(n, m, p)$ . Such properties as connectivity, a Hamilton cycle containment or a matching containment are given as examples. In general, the coupling technique provides a very elegant method to obtain bounds on threshold functions for random intersection graphs for a large class of properties.

The result presented in [18] is not tight for some values of  $n$ ,  $m$  and  $p$ . Therefore it cannot be generalized in a straightforward manner to  $\mathcal{G}(n, m, \bar{p})$  with arbitrary  $\bar{p}(n)$ . In particular the method does not give sharp results for  $np$  tending to a constant. In this article we modify and extend the techniques used in [18] in order to overcome these constraints. First of all, to get the general result of this paper, we couple  $\mathcal{G}(n, m, \bar{p})$  with an auxiliary random graph which does not have fully independent edges. Therefore we need to prove some additional facts about the auxiliary random graph model. Moreover we need tight bounds on the minimum degree threshold function for  $\mathcal{G}(n, m, \bar{p})$ . Due to edge dependencies, estimation of moments of the random variable counting vertices with a given degree in  $\mathcal{G}(n, m, \bar{p})$  is complicated. Therefore we suggest a different approach to resolve the problem. We divide  $\mathcal{G}(n, m, \bar{p})$  into subgraphs so that the solution of a coupon collector problem combined with the method of moments provide the answer. The new approach to the coupling method allows us to obtain better results on threshold functions for  $\mathcal{G}(n, m, p)$  and through those means resolve open problems unresolved in [18].

In conclusion, we provide a general method to establish bounds on threshold functions for many properties for  $\mathcal{G}(n, m, \bar{p})$ . By means of the method we are able to obtain sharp thresholds (for definition of a sharp threshold see Section 1.6 in [13]) for  $k$ -connectivity, perfect matching containment and hamiltonicity for the general model. Last but not least we considerably improve known results concerning  $\mathcal{G}(n, m, p)$ .

All limits in the paper are taken as  $n \rightarrow \infty$ . Throughout this paper we use the standard asymptotic notation  $o(\cdot)$ ,  $O(\cdot)$ ,  $\Omega(\cdot)$ ,  $\Theta(\cdot)$ ,  $\sim$ ,  $\ll$ , and  $\gg$  defined as in [13]. By  $\text{Bin}(n, p)$  and  $\text{Po}(\lambda)$  we denote the binomial distribution with parameters  $n$ ,  $p$  and the Poisson distribution with expected value  $\lambda$ , respectively. We also use the phrase “with high probability” to say with probability tending to one as  $n$  tends to infinity. All inequalities hold for  $n$  large enough. If it does not influence the reasoning, for clarity we omit  $\lfloor \cdot \rfloor$  and  $\lceil \cdot \rceil$ .

## 2 Main Results

In the article we compare random intersection graph  $\mathcal{G}(n, m, \bar{p})$  with a union of a random graph with independent edges  $G_2(n, \hat{p}_2)$  and a random graph  $G_3(n, \hat{p}_3)$  constructed on the basis of a random 3-uniform hypergraph with independent hyperedges. Generally, for any  $\hat{p} = \hat{p}(n) \in [0, 1]$  and  $i = 2, \dots, n$ , let  $H_i(n, \hat{p})$  be an  $i$ -uniform hypergraph with the vertex set  $\mathcal{V}$  in which each  $i$ -element subset of  $\mathcal{V}$  is added to the hyperedge set independently with probability  $\hat{p}$ .  $G_i(n, \hat{p})$  is a graph with the vertex set  $\mathcal{V}$  and  $\{v, v'\}$ ,  $v, v' \in \mathcal{V}$ , is an edge in  $G_i(n, \hat{p})$  if there exists a hyperedge in  $H_i(n, \hat{p})$  containing  $v$  and  $v'$ .

Before we present the main results of this paper, we introduce some additional notations which will be used repeatedly. Then we will give an intuitive description of the introduced quantities. This might help understand the main ideas behind the theorems.

Let  $\bar{p} = (p_1, \dots, p_m)$  be such that  $p_i \in (0, 1)$ , for all  $1 \leq i \leq m$ . We will use the following quantities

$$\begin{aligned}
 S_1 &= \sum_{i=1}^m np_i (1 - (1 - p_i)^{n-1}); \\
 S_2 &= \sum_{i=1}^m np_i \left(1 - \frac{1 - (1 - 2p_i)^n}{2np_i}\right); \\
 S_3 &= \sum_{i=1}^m np_i \left(\frac{1 - (1 - 2p_i)^n}{2np_i} - (1 - p_i)^{n-1}\right); \\
 S_{1,t} &= \sum_{i=1}^m t \binom{n}{t} p_i^t (1 - p_i)^{n-t}, \text{ for } t = 2, 3, \dots, n.
 \end{aligned} \tag{2.1}$$

By definition, the edge set of  $\mathcal{G}(n, m, \bar{p})$  is a union of edge sets of  $m$  cliques. The  $i$ -th clique,  $1 \leq i \leq m$ , consists of all the vertices which have chosen feature  $w_i$ . We will say that these cliques *are forming*  $\mathcal{G}(n, m, \bar{p})$ . The size of the  $i$ -th clique forming  $\mathcal{G}(n, m, \bar{p})$ ,  $1 \leq i \leq m$ , has binomial distribution  $\text{Bin}(n, p_i)$ . Moreover cliques (not only their sizes) are independent. Replacing edge sets of cliques forming  $\mathcal{G}(n, m, \bar{p})$  by sets of (to some extent independent) edges will be the key idea used in the coupling constructed in the proof of Theorem 2.1.

The quantity  $S_1$  is the expected value of the sum of sizes of these cliques forming  $\mathcal{G}(n, m, \bar{p})$  which contain at least two vertices. This value, among others, will be crucial in determining the number of isolated vertices in  $\mathcal{G}(n, m, \bar{p})$ . This is because a vertex in  $\mathcal{G}(n, m, \bar{p})$  is isolated if and only if it is not contained in any of the cliques of size at least 2.  $S_2$  is the expected value of the sum of sizes of all cliques forming  $\mathcal{G}(n, m, \bar{p})$  minus the number of cliques of odd size.  $S_3$  is the expected value of the number of cliques of odd size with at least 3 vertices (i.e. the number of cliques of odd size, excluding cliques of size 1).  $S_1$ ,  $S_2$  and  $S_3$  are, among others, present in the statement of Theorem 2.1. In this theorem properties of  $\mathcal{G}(n, m, \bar{p})$  are compared with properties of two other random graphs whose edge probabilities depend on  $S_1$ ,  $S_2$  and  $S_3$ . The first graph used in Theorem 2.1 is a random graph with approximately  $S_2/2$  independent edges. The second

one is  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  where  $G_2(n, \hat{p}_2)$  has approximately  $(S_1 - 3S_3)/2$  edges and  $G_3(n, \hat{p}_3)$  has approximately  $3S_3$  edges. The last quantity defined in (2.1) is  $S_{1,t}$ . It is the expected value of the sum of sizes of these cliques forming  $\mathcal{G}(n, m, \bar{p})$  which contain exactly  $t$  vertices. Moreover  $S_{1,t}/t$  is the number of features which were chosen by exactly  $t$  vertices. Values  $S_{1,t}$ ,  $t \geq 2$ , are mainly used in studying the minimum degree of  $\mathcal{G}(n, m, \bar{p})$ . For a more formal description of the relations between  $S_1$ ,  $S_2$ ,  $S_3$ ,  $S_{1,t}$  and the sizes of cliques forming  $\mathcal{G}(n, m, \bar{p})$ , we refer the reader to (4.5).

In what follows we consider monotone graph properties of random graphs. For the family  $\mathcal{G}$  of all graphs with the vertex set  $\mathcal{V}$ , we call  $\mathcal{A} \subseteq \mathcal{G}$  a property if it is closed under isomorphism. Moreover  $\mathcal{A}$  is increasing if  $G \in \mathcal{A}$  implies  $G' \in \mathcal{A}$  for all  $G' \in \mathcal{G}$  such that  $E(G) \subseteq E(G')$ . Examples of increasing properties are:  $k$ -connectivity, containing a perfect matching and containing a Hamilton cycle. The following theorem is an extension of the result obtained in [18].

**Theorem 2.1.** *Let  $\bar{p} = (p_1, \dots, p_m)$  be such that  $p_i \in (0, 1)$ , for all  $1 \leq i \leq m$  and  $S_1$ ,  $S_2$ , and  $S_3$  be given by (2.1). For a function  $\omega$  tending to infinity let*

$$\begin{aligned} \hat{p} &= \frac{S_2 - \omega\sqrt{S_2} - 2S_2^2n^{-2}}{2\binom{n}{2}}; \\ \hat{p}_2 &= \begin{cases} \frac{S_1 - 3S_3 - \omega\sqrt{S_1} - 2S_1^2n^{-2}}{2\binom{n}{2}}, & \text{for } S_3 \gg \sqrt{S_1} \text{ and } \omega \ll S_3/\sqrt{S_1}; \\ \frac{S_1 - \omega\sqrt{S_1} - 2S_1^2n^{-2}}{2\binom{n}{2}}, & \text{for } S_3 = O(\sqrt{S_1}); \end{cases} \\ \hat{p}_3 &= \begin{cases} \frac{S_3 - \omega\sqrt{S_1} - 6S_3^2n^{-3}}{\binom{n}{3}}, & \text{for } S_3 \gg \sqrt{S_1} \text{ and } \omega \ll S_3/\sqrt{S_1}; \\ 0, & \text{for } S_3 = O(\sqrt{S_1}). \end{cases} \end{aligned} \tag{2.2}$$

If  $S_1 \rightarrow \infty$  and  $S_1 = o(n^2)$  then for any increasing property  $\mathcal{A}$ .

$$\liminf_{n \rightarrow \infty} \Pr \{G_2(n, \hat{p}) \in \mathcal{A}\} \leq \limsup_{n \rightarrow \infty} \Pr \{\mathcal{G}(n, m, \bar{p}) \in \mathcal{A}\}, \tag{2.3}$$

$$\liminf_{n \rightarrow \infty} \Pr \{G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) \in \mathcal{A}\} \leq \limsup_{n \rightarrow \infty} \Pr \{\mathcal{G}(n, m, \bar{p}) \in \mathcal{A}\}. \tag{2.4}$$

*Remark 2.1.* Assumption  $S_1 \rightarrow \infty$  is natural since for  $S_1 = o(1)$  with high probability  $\mathcal{G}(n, m, \bar{p})$  is an empty graph.

*Remark 2.2.* The expected number of edges in  $G_3(n, \hat{p}_3)$  is asymptotically equal to  $3 \cdot S_3$ . If  $S_3 = O(\sqrt{S_1})$  then by Markov's inequality with high probability the number of edges in  $G_3(n, \hat{p}_3)$  is at most  $\omega\sqrt{S_1}$  (for any  $\omega \rightarrow \infty$ ). Thus  $S_2 = S_1 - S_3 = S_1 + O(\omega\sqrt{S_1})$  and the bound provided by (2.3) is as good as the one taking into consideration the edges from  $G_3(n, \hat{p}_3)$ .

*Remark 2.3.* The proof of Theorem 2.1 might be simply rewritten (see (4.12)) for  $S_1 =$

$\Omega(n^2)$  with

$$\begin{aligned} \hat{p} &= 1 - \exp\left(-\frac{S_2 - \omega\sqrt{S_2}}{2\binom{n}{2}}\right); \\ \hat{p}_2 &= \begin{cases} 1 - \exp\left(-\frac{S_1 - 3S_3 - \omega\sqrt{S_1}}{2\binom{n}{2}}\right), & \text{for } S_3 \gg \sqrt{S_1} \text{ and } \omega \ll S_3/\sqrt{S_1}; \\ 1 - \exp\left(-\frac{S_1 - \omega\sqrt{S_1}}{2\binom{n}{2}}\right), & \text{for } S_3 = O(\sqrt{S_1}); \end{cases} \\ \hat{p}_3 &= \begin{cases} 1 - \exp\left(-\frac{S_3 - \omega\sqrt{S_1}}{\binom{n}{3}}\right), & \text{for } S_3 \gg \sqrt{S_1} \text{ and } \omega \ll S_3/\sqrt{S_1}; \\ 0, & \text{for } S_3 = O(\sqrt{S_1}). \end{cases} \end{aligned}$$

Denote by  $\mathcal{C}_k$ ,  $\mathcal{PM}$  and  $\mathcal{HC}$  the following graph properties: a graph is  $k$ -connected, has a perfect matching and has a Hamilton cycle, respectively. We will use Theorem 2.1 to establish threshold functions for  $\mathcal{C}_k$ ,  $\mathcal{PM}$  and  $\mathcal{HC}$  in  $\mathcal{G}(n, m, \bar{p})$ . By  $\mathcal{C}_k$  we denote the vertex connectivity. From the proof it follows that the threshold function for the edge connectivity is the same as this for  $\mathcal{C}_k$ .

For any sequence  $c_n$  attaining a limit (possibly  $\pm\infty$ ) we write

$$f(c_n) = \begin{cases} 0 & \text{for } c_n \rightarrow -\infty; \\ e^{-e^{-c}} & \text{for } c_n \rightarrow c \in (-\infty, \infty); \\ 1 & \text{for } c_n \rightarrow \infty. \end{cases} \quad (2.5)$$

**Theorem 2.2.** *Let  $\max_{1 \leq i \leq m} p_i = o((\ln n)^{-1})$  and  $S_1$  and  $S_{1,2}$  be given by (2.1).*

(i) *If  $S_1 = n(\ln n + c_n)$ , where  $c_n$  attains a limit (possibly  $\pm\infty$ ), then*

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(n, m, \bar{p}) \in \mathcal{C}_1 \} = f(c_n),$$

where  $f(c_n)$  is given by (2.5).

(ii) *Let  $k$  be a positive integer and  $a_n = \frac{S_{1,2}}{S_1}$ . If*

$$S_1 = n(\ln n + (k-1) \ln \ln n + c_n),$$

then

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(n, m, \bar{p}) \in \mathcal{C}_k \} = \begin{cases} 0 & \text{for } c_n \rightarrow -\infty \text{ and } \liminf_{n \rightarrow \infty} a_n = a \in (0, 1]; \\ 1 & \text{for } c_n \rightarrow \infty. \end{cases}$$

Assumption  $\max_{1 \leq i \leq m} p_i = o((\ln n)^{-1})$  is necessary to avoid large cliques formed by vertices which chose the same feature. In the case  $\max_{1 \leq i \leq m} p_i = \Omega((\ln n)^{-1})$ , the expected value of the number of isolated vertices depends more on other properties of the sequence  $\bar{p} = \bar{p}(n)$ . The problem is explained in more detail in Section 5. Moreover in the case

$\liminf_{n \rightarrow \infty} a_n = 0$  the value  $\Pr \{ \mathcal{G}(n, m, \bar{p}) \in \mathcal{C}_k \}$  depends also on some properties of the sequence  $\bar{p} = \bar{p}(n)$ . The problem is studied in more detail in the case  $\bar{p} = (p, \dots, p)$  (see Theorems 2.5 and 2.6 and their proofs).

A straightforward corollary of Theorem 2.2(i) is that for  $S_1 = n(\ln n + c_n)$ ,  $c_n \rightarrow -\infty$  and any  $k = 1, 2, \dots, n$

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(n, m, \bar{p}) \in \mathcal{C}_k \} = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(n, m, \bar{p}) \in \mathcal{HC} \} = 0.$$

**Theorem 2.3.** *Let  $\max_{1 \leq i \leq m} p_i = o((\ln n)^{-1})$  and  $S_1$  be given by (2.1). If  $S_1 = n(\ln n + c_n)$ , where  $c_n$  attains a limit (possibly  $\pm\infty$ ), then*

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(2n, m, \bar{p}(2n)) \in \mathcal{PM} \} = f(c_{2n}),$$

where  $f(\cdot)$  is given by (2.5).

**Theorem 2.4.** *Let  $\max_{1 \leq i \leq m} p_i = o((\ln n)^{-1})$ ,  $S_1$  and  $S_{1,2}$  be given by (2.1) and  $a_n = \frac{S_{1,2}}{S_1}$ . If  $S_1 = n(\ln n + \ln \ln n + c_n)$ , then*

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(n, m, \bar{p}) \in \mathcal{HC} \} = \begin{cases} 0 & \text{for } c_n \rightarrow -\infty \text{ and } \liminf_{n \rightarrow \infty} a_n = a \in (0, 1]; \\ 1 & \text{for } c_n \rightarrow \infty. \end{cases}$$

Simple corollaries of Theorems 2.2–2.4 give threshold functions for properties in  $\mathcal{G}(n, m, p)$ , a special case of  $\mathcal{G}(n, m, \bar{p})$ .

**Corollary 2.1.** *Let  $m \gg \ln^2 n$  and  $p(1 - (1 - p)^{n-1}) = (\ln n + c_n)/m$ , where  $c_n$  attains a limit (possibly  $\pm\infty$ ). Then*

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(n, m, p) \in \mathcal{C}_1 \} = f(c_n)$$

and

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(2n, m, p) \in \mathcal{PM} \} = f(c_{2n}),$$

where  $f(\cdot)$  is given by (2.5).

In particular we may state the following extension of the result from [20].

**Corollary 2.2.** *Let  $b_n$  be a sequence,  $\beta$  and  $\gamma$  be constants such that  $\beta\gamma(1 - e^{-\gamma}) = 1$ . If*

$$m = \beta n \ln n \quad \text{and} \quad p = \frac{\gamma}{n} \left( 1 + \frac{b_n}{\ln n} \right),$$

where  $b_n$  attains a limit (possibly  $\pm\infty$ ). Then

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(n, m, p) \in \mathcal{C}_1 \} = f \left( \left( 1 + \frac{e^{-\gamma}\gamma}{1 - e^{-\gamma}} \right) b_n \right)$$

and

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(2n, m, p(2n)) \in \mathcal{PM} \} = f \left( \left( 1 + \frac{e^{-\gamma}\gamma}{1 - e^{-\gamma}} \right) b_{2n} \right),$$

where  $f(\cdot)$  is given by (2.5).

With some additional effort one may show the following results for  $\mathcal{G}(n, m, p)$ .

**Theorem 2.5.** *Let  $m \gg \ln^2 n$  and*

$$p(1 - (1 - p)^{n-1}) = \frac{\ln n + \ln \left( \max \left\{ 1, \left( \frac{np \ln n}{e^{np} - 1} \right) \right\} \right) + c_n}{m}. \quad (2.6)$$

Then

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(n, m, p) \in \mathcal{HC} \} = \begin{cases} 0 & \text{for } c_n \rightarrow -\infty; \\ 1 & \text{for } c_n \rightarrow \infty. \end{cases}$$

**Theorem 2.6.** *Let  $m \gg \ln^2 n$ ,  $k$  be a positive integer, and*

$$a_n = (np)^{k-1} \left( \left( \frac{\ln n}{e^{np} - 1} \right)^{k-1} + \frac{\ln n}{e^{np} - 1} \right).$$

If

$$p(1 - (1 - p)^{n-1}) = \frac{\ln n + \ln \left( \max \{ 1, a_n \} \right) + c_n}{m},$$

then

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(n, m, p) \in \mathcal{C}_k \} = \begin{cases} 0 & \text{for } c_n \rightarrow -\infty; \\ 1 & \text{for } c_n \rightarrow \infty. \end{cases}$$

One of the questions posed in [18] is concerned with the range of  $m = m(n)$  for which the threshold function for  $\mathcal{C}_k$  in  $\mathcal{G}(n, m, p)$  is the same as the one for  $\delta(\mathcal{G}(n, m, p)) \geq 1$  (where by  $\delta(G)$  we denote the minimum degree of graph  $G$ ). Another question is concerned with the range of  $m = m(n)$  for which the threshold function for  $\mathcal{C}_k$  in  $\mathcal{G}(n, m, p)$  is the same as the one for  $\mathcal{C}_k$  in  $G_2(n, \hat{p})$  with  $\hat{p} = mp^2$ . Theorem 2.6 gives a final answer to these questions.

**Corollary 2.3.** *Let  $k$  be a positive integer. If*

$$p(1 - (1 - p)^{n-1}) = \begin{cases} \frac{\ln n + c_n}{m}, & \text{for } \ln^2 n \ll m \ll \frac{n \ln n}{\ln \ln n}; \\ \frac{\ln n + (k-1) \ln \ln n + c_n}{m}, & \text{for } m = \Omega(n \ln n); \end{cases}$$

then

$$\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(n, m, p) \in \mathcal{C}_k \} = \begin{cases} 0 & \text{for } c_n \rightarrow -\infty; \\ 1 & \text{for } c_n \rightarrow \infty. \end{cases}$$

### 3 Road map of the proofs

The remaining part of the article is organised as follows. Section 4 is devoted to a construction of a coupling of  $\mathcal{G}(n, m, \bar{p})$  and  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$ . Later in this section the coupling is used to prove Theorem 2.1. Moreover in the subsequent sections the coupling is applied in the proofs of other theorems. In Section 5 we find a threshold

function for the minimum degree at least  $k$  in  $\mathcal{G}(n, m, \bar{p})$ . To this end we construct a coupling of a construction of  $\mathcal{G}(n, m, \bar{p})$  and a coupon collector process. Section 6 is devoted to presenting properties of  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$ . Results gathered in Sections 5 and 6 are the building blocks of the proofs of Theorems 2.2–2.6. In Section 7 we complete the proofs of Theorems 2.2–2.6.

The article contains numerous lemmas and partial results which lead to the proofs of the main results. We will now shortly explain which lemmas are crucial in proving which theorems. The proofs of all Theorems 2.2–2.6 consist of two parts. The first part establishes an upper bound on the value  $\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(n, m, \bar{p}) \in \mathcal{A} \}$ . The second part gives a lower bound on  $\lim_{n \rightarrow \infty} \Pr \{ \mathcal{G}(n, m, \bar{p}) \in \mathcal{A} \}$  (for  $\mathcal{A}$  being  $\mathcal{C}_k$ ,  $\mathcal{PM}$ , or  $\mathcal{HC}$  – depending on a theorem).

All upper bounds follow by the fact that each of the considered properties impose a minimum degree condition. Namely, a graph in order to be  $k$ -connected has to have the minimum degree at least  $k$ . Moreover, a perfect matching is only possible in a graph with no isolated vertices (the minimum degree at least 1) and a Hamilton cycle requires the minimum degree at least 2. Therefore the upper bounds follow by Lemmas 5.1, 5.2 and 5.3 obtained in Section 5.

Establishing lower bounds requires more work, as the minimum degree condition is necessary but not sufficient. The first step towards proving this part of theorems is an analysis of the structure of  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$ .

As we want to use Theorem 2.1 to show Theorems 2.2, 2.3, and 2.4, we need to find a lower bound on the value  $\lim_{n \rightarrow \infty} \Pr \{ G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) \in \mathcal{A} \}$  for  $\mathcal{A}$  being  $\mathcal{C}_k$ ,  $\mathcal{PM}$ , and  $\mathcal{HC}$ , respectively. A large part of Section 6 is devoted to establishing these lower bounds. In particular, Lemma 6.1 gives some useful, basic properties of  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  which are applied in the proofs of almost all the remaining lemmas of this section, i.e. also in the proofs of all Theorems 2.2–2.6. These properties, among others, are used to show lower bounds on  $\lim_{n \rightarrow \infty} \Pr \{ G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) \in \mathcal{A} \}$ . These lower bounds are stated in Lemmas 6.4 and 6.6. They follow by Lemma 6.1, Lemma 6.2, and an analysis of some couplings discussed in Lemmas 6.3 and 6.5. Lemmas 6.3 and 6.5 provide properties of graph  $G(n)$  for which there exists a coupling such that with high probability  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) \subseteq G(n)$ .

Obtaining lower bounds needed in Theorems 2.5 and 2.6 is more complex. In this case the proofs are based on Lemmas 6.3 and 6.5, some additional properties of  $\mathcal{G}(n, m, p)$ , and properties of the coupling established in the proof of Theorem 2.1.

## 4 Coupling

In this section we present a proof of Theorem 2.1. In the proof we use auxiliary random graph models  $\mathbb{G}_{*i}(n, M)$ ,  $i = 2, 3, \dots, n$ , in which  $M$  is a random variable with non-negative integer values. For  $i = 2, \dots, n$ ,  $\mathbb{G}_{*i}(n, M)$  is constructed on the basis of a random hypergraph  $\mathbb{H}_{*i}(n, M)$ .  $\mathbb{H}_{*i}(n, M)$  is a random hypergraph with the vertex set  $\mathcal{V}$  in which a hyperedge set is constructed by sampling  $M$  times with repetition elements from the set of all  $i$ -element subsets of  $\mathcal{V}$  (all sets which are chosen several times are added



only once to the hyperedge set).  $\mathbb{G}_{*i}(n, M)$  is a graph with the vertex set  $\mathcal{V}$  in which  $v, v' \in \mathcal{V}$  are connected by an edge if  $\{v, v'\}$  is contained in at least one of the hyperedges of  $\mathbb{H}_{*i}(n, M)$ . For simplicity of notation, if  $M$  equals a constant  $t$  with probability one or has Poisson distribution, we write  $\mathbb{G}_{*i}(n, t)$  or  $\mathbb{G}_{*i}(n, \text{Po}(\cdot))$ , respectively. Recall that similarly  $G_i(n, \hat{p})$  is constructed on the basis of  $H_i(n, \hat{p})$  – a hypergraph with independent hyperedges (see definitions at the beginning of Section 2).

In this article we treat random graphs as random variables. By a coupling  $(\mathbb{G}_1, \mathbb{G}_2)$  of two random variables  $\mathbb{G}_1$  and  $\mathbb{G}_2$  we mean a choice of a probability space in which we define a random vector  $(\mathbb{G}'_1, \mathbb{G}'_2)$  such that  $\mathbb{G}'_1$  and  $\mathbb{G}'_2$  have the same distributions as  $\mathbb{G}_1$  and  $\mathbb{G}_2$ , respectively. For simplicity of notation we will not differentiate between  $(\mathbb{G}'_1, \mathbb{G}'_2)$  and  $(\mathbb{G}_1, \mathbb{G}_2)$ . For two random graphs  $\mathbb{G}_1$  and  $\mathbb{G}_2$  we write

$$\mathbb{G}_1 \preceq \mathbb{G}_2$$

if there exists a coupling  $(\mathbb{G}_1, \mathbb{G}_2)$ , such that in the probability space of the coupling  $\mathbb{G}_1$  is a subgraph of  $\mathbb{G}_2$  with probability 1. Moreover, we write

$$\mathbb{G}_1 = \mathbb{G}_2,$$

if  $\mathbb{G}_1$  and  $\mathbb{G}_2$  have the same probability distribution (equivalently there exists a coupling  $(\mathbb{G}_1, \mathbb{G}_2)$  such that  $\mathbb{G}_1 = \mathbb{G}_2$  with probability one). For two sequences of random graphs  $\mathbb{G}_1 = \mathbb{G}_1(n)$  and  $\mathbb{G}_2 = \mathbb{G}_2(n)$  we write

$$\mathbb{G}_1 \preceq_{1-o(1)} \mathbb{G}_2,$$

if there exists a sequence of couplings  $(\mathbb{G}_1(n), \mathbb{G}_2(n))$ , such that in the probability space of the coupling  $\mathbb{G}_1(n)$  is a subgraph of  $\mathbb{G}_2(n)$  with probability  $1 - o(1)$ , respectively.

Note that, for any  $\lambda$ , in  $\mathbb{H}_{*i}(n, \text{Po}(\lambda))$  each hyperedge appears independently with probability  $1 - \exp(-\lambda/\binom{n}{i})$ . Thus

$$\mathbb{G}_{*i}(n, \text{Po}(\lambda)) = G_i \left( n, 1 - \exp \left( -\frac{\lambda}{\binom{n}{i}} \right) \right). \quad (4.1)$$

We gather here a few useful facts concerning couplings of random graphs. For proofs see [17, 18].

**Proposition 4.1.** *Let  $M_n$  be a sequence of random variables and let  $t_n$  be a sequence of positive integers.*

(i) *If  $\Pr \{M_n \geq t_n\} = o(1)$  then  $\mathbb{G}_{*i}(n, M_n) \preceq_{1-o(1)} \mathbb{G}_{*i}(n, t_n)$ .*

(ii) *If  $\Pr \{M_n \leq t_n\} = o(1)$  then  $\mathbb{G}_{*i}(n, t_n) \preceq_{1-o(1)} \mathbb{G}_{*i}(n, M_n)$ .*

**Proposition 4.2.** *Let  $(\mathbb{G}_i)_{i=1, \dots, m}$  and  $(\mathbb{G}'_i)_{i=1, \dots, m}$  be sequences of independent random graphs. If*

$$\mathbb{G}_i \preceq \mathbb{G}'_i, \text{ for all } i = 1, \dots, m,$$

*then*

$$\bigcup_{i=1}^m \mathbb{G}_i \preceq \bigcup_{i=1}^m \mathbb{G}'_i.$$

**Proposition 4.3.** Let  $\mathbb{G}_1 = \mathbb{G}_1(n)$ ,  $\mathbb{G}_2 = \mathbb{G}_2(n)$ , and  $\mathbb{G}_3 = \mathbb{G}_3(n)$  be random graphs. If

$$\mathbb{G}_1 \preceq_{1-o(1)} \mathbb{G}_2 \quad \text{and} \quad \mathbb{G}_2 \preceq_{1-o(1)} \mathbb{G}_3$$

then

$$\mathbb{G}_1 \preceq_{1-o(1)} \mathbb{G}_3.$$

**Proposition 4.4.** Let  $\mathbb{G}_1 = \mathbb{G}_1(n)$  and  $\mathbb{G}_2 = \mathbb{G}_2(n)$  be two random graphs, such that

$$\mathbb{G}_1 \preceq_{1-o(1)} \mathbb{G}_2. \tag{4.2}$$

Then for any increasing property  $\mathcal{A}$

$$\liminf_{n \rightarrow \infty} \Pr \{ \mathbb{G}_1(n) \in \mathcal{A} \} \leq \limsup_{n \rightarrow \infty} \Pr \{ \mathbb{G}_2(n) \in \mathcal{A} \}.$$

*Proof.* Define event  $\mathcal{E} := \{ \mathbb{G}_1 \subseteq \mathbb{G}_2 \}$  on a probability space of a coupling  $(\mathbb{G}_1, \mathbb{G}_2)$  existing by (4.2). Then for any increasing property  $\mathcal{A}$

$$\begin{aligned} \Pr \{ \mathbb{G}_2 \in \mathcal{A} \} &\geq \Pr \{ \mathbb{G}_2 \in \mathcal{A} | \mathcal{E} \} \Pr \{ \mathcal{E} \} \\ &\geq \Pr \{ \mathbb{G}_1 \in \mathcal{A} | \mathcal{E} \} \Pr \{ \mathcal{E} \} \\ &= \Pr \{ \{ \mathbb{G}_1 \in \mathcal{A} \} \cap \mathcal{E} \} \\ &= \Pr \{ \mathbb{G}_1 \in \mathcal{A} \} + \Pr \{ \mathcal{E} \} - \Pr \{ \{ \mathbb{G}_1 \in \mathcal{A} \} \cup \mathcal{E} \} \\ &\geq \Pr \{ \mathbb{G}_1 \in \mathcal{A} \} + \Pr \{ \mathcal{E} \} - 1 \\ &= \Pr \{ \mathbb{G}_1 \in \mathcal{A} \} + o(1). \end{aligned}$$

The result follows by taking  $n \rightarrow \infty$ . □

In the proof of Theorem 2.1 we will use notations introduced in (2.1). First, we give a formal explanation of the relations between  $S_1$ ,  $S_2$ ,  $S_3$ , and  $S_{1,t}$  and the sizes of cliques forming  $\mathcal{G}(n, m, \bar{p})$ . In addition we will give some useful simple relations between  $S_1$ ,  $S_2$ ,  $S_3$ , and  $S_{1,t}$ .

Let  $w_i \in \mathcal{W}$ ,  $1 \leq i \leq m$ . Denote by  $V_i$  the set of vertices which have chosen feature  $w_i$  (i.e.  $V_i = \{v \in \mathcal{V} : w_i \in W(v)\}$ ). Let

$$V'_i = \begin{cases} V_i & \text{for } |V_i| \geq 2; \\ \emptyset & \text{otherwise.} \end{cases} \tag{4.3}$$

Moreover, for each  $i$ ,  $1 \leq i \leq m$ , let

$$\begin{aligned} X_i &= |V_i|; \\ Y_i &= |V'_i|; \\ Z_i &= \mathbb{I}_{\{Y_i \text{ is odd}\}}, \end{aligned} \tag{4.4}$$

where  $\mathbb{I}_A$  is an indicator random variable of the event  $A$ . Note that  $X_i$ ,  $1 \leq i \leq m$ , are independent random variables with the binomial distributions  $\text{Bin}(n, p_i)$ ,  $1 \leq i \leq m$ . Therefore by (2.1)

$$\begin{aligned}
 S_1 &= \mathbb{E} \sum_{i=1}^m (X_i - \mathbb{I}_{\{X_i=1\}}) = \mathbb{E} \sum_{i=1}^m Y_i; \\
 S_2 &= \mathbb{E} \sum_{i=1}^m (X_i - \mathbb{I}_{\{X_i \text{ is odd}\}}) = \mathbb{E} \sum_{i=1}^m (Y_i - \mathbb{I}_{\{Y_i \text{ is odd}\}}) = \mathbb{E} \sum_{i=1}^m (Y_i - Z_i); \\
 S_3 &= \mathbb{E} \sum_{i=1}^m (\mathbb{I}_{\{X_i \text{ is odd}\}} - \mathbb{I}_{\{X_i=1\}}) = \mathbb{E} \sum_{i=1}^m \mathbb{I}_{\{Y_i \text{ is odd}\}} = \mathbb{E} \sum_{i=1}^m Z_i; \\
 S_{1,t} &= \mathbb{E} \sum_{i=1}^m X_i \mathbb{I}_{\{X_i=t\}} = \mathbb{E} \sum_{i=1}^m Y_i \mathbb{I}_{\{Y_i=t\}} = \mathbb{E} \sum_{i=1}^m t \cdot \mathbb{I}_{\{X_i=t\}}, \quad \text{for } t = 2, 3, \dots, n.
 \end{aligned} \tag{4.5}$$

Therefore

$$S_1 = \sum_{t=2}^n S_{1,t} \quad \text{and} \quad S_2 = S_1 - S_3. \tag{4.6}$$

In addition a calculation shows that, for large  $n$ ,  $\mathbb{E}Y_i \geq 6\mathbb{E}Z_i$  (see Appendix A). Therefore

$$6S_3 \leq S_1 \quad \text{and} \quad \frac{5}{6}S_1 \leq S_2 \leq S_1. \tag{4.7}$$

*Proof of Theorem 2.1.* First we will show (2.4) in the case where  $S_3 \gg \sqrt{S_1}$ . Then we will explain shortly how to adjust the proof to prove Theorem 2.1 in the remaining cases.

Recall that the definition of  $V'_i$  is given by (4.3). For each  $i$ , let  $\mathcal{G}[V'_i]$  be a graph with the vertex set  $\mathcal{V}$  and edge set containing these edges from  $\mathcal{G}(n, m, \bar{p})$  which have both ends in  $V'_i$  (i.e. its edges form a clique with the vertex set  $V'_i$ ). In the proof we will use the fact that  $\bigcup_{i=1}^m \mathcal{G}[V'_i] = \mathcal{G}(n, m, \bar{p})$ . First, for each  $i$  we will construct a coupling of  $\mathcal{G}[V'_i]$  with an auxiliary random graph and then we will couple a union of the auxiliary random graphs with  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$ .

For each  $i$ ,  $1 \leq i \leq m$ , we construct independently a coupling of  $\mathbb{G}_{*2}(n, \frac{Y_i - 3Z_i}{2}) \cup \mathbb{G}_{*3}(n, Z_i)$  and  $\mathcal{G}[V'_i]$ . Given  $Y_i = y_i$  and  $Z_i = z_i$ , for each  $i$  independently, we generate instances of  $\mathbb{G}_{*2}(n, \frac{y_i - 3z_i}{2})$  and  $\mathbb{G}_{*3}(n, z_i)$ . Let  $Y'_i = y'_i$  be the number of non-isolated vertices in the constructed instance of  $\mathbb{G}_{*2}(n, \frac{y_i - 3z_i}{2}) \cup \mathbb{G}_{*3}(n, z_i)$ . By definition  $y'_i \leq y_i$ . Set now  $V'_i$  to be the union of the set of non-isolated vertices of the constructed instance of  $\mathbb{G}_{*2}(n, \frac{y_i - 3z_i}{2}) \cup \mathbb{G}_{*3}(n, z_i)$  and  $y_i - y'_i$  vertices chosen uniformly at random from the remaining vertices. This coupling implies

$$\mathbb{G}_{*2}\left(n, \frac{Y_i - 3Z_i}{2}\right) \cup \mathbb{G}_{*3}(n, Z_i) \preceq \mathcal{G}[V'_i]. \tag{4.8}$$

Now let

$$M_2 = \sum_{1 \leq i \leq m} \frac{Y_i - 3Z_i}{2} \quad \text{and} \quad M_3 = \sum_{1 \leq i \leq m} Z_i.$$

Graphs  $\mathbb{G}_{*2}(n, \frac{Y_i - 3Z_i}{2}) \cup \mathbb{G}_{*3}(n, Z_i)$ ,  $1 \leq i \leq m$ , are independent and  $\mathcal{G}[V'_i]$ ,  $1 \leq i \leq m$ , are independent. Therefore by Proposition 4.2, the definition of  $\mathcal{G}(n, m, \bar{p})$ , and the definitions of  $\mathbb{G}_{*2}(n, \cdot)$  and  $\mathbb{G}_{*3}(n, \cdot)$  we have

$$\begin{aligned} \mathbb{G}_{*2}(n, M_2) \cup \mathbb{G}_{*3}(n, M_3) &= \bigcup_{1 \leq i \leq m} (\mathbb{G}_{*2}(n, \frac{Y_i - 3Z_i}{2}) \cup \mathbb{G}_{*3}(n, Z_i)) \\ &\preceq \bigcup_{1 \leq i \leq m} \mathcal{G}[V'_i] \\ &= \mathcal{G}(n, m, \bar{p}). \end{aligned} \tag{4.9}$$

By (4.5)

$$\mathbb{E}M_2 = \frac{S_1 - 3S_3}{2} \quad \text{and} \quad \mathbb{E}M_3 = S_3.$$

We will prove that  $M_2$  and  $M_3$  with high probability are concentrated around their expected values. Since for each  $i$ ,  $1 \leq i \leq m$ ,  $X_i$  is a random variable with the binomial distribution  $\text{Bin}(n, p_i)$  and  $\mathbb{E}(X_i \mathbb{I}_{\{X_i=1\}}) = \mathbb{E}\mathbb{I}_{\{X_i=1\}}$  we have

$$\begin{aligned} \text{Var} \sum_{i=1}^m Y_i &= \sum_{i=1}^m \text{Var}(X_i - \mathbb{I}_{\{X_i=1\}}) \\ &= \sum_{i=1}^m (\text{Var}X_i + \text{Var}\mathbb{I}_{\{X_i=1\}} - 2(\mathbb{E}(X_i \mathbb{I}_{\{X_i=1\}}) - \mathbb{E}X_i \mathbb{E}\mathbb{I}_{\{X_i=1\}})) \\ &\leq \sum_{i=1}^m (\mathbb{E}X_i + \mathbb{E}\mathbb{I}_{\{X_i=1\}} - 2\mathbb{E}\mathbb{I}_{\{X_i=1\}} + 2\mathbb{E}X_i \mathbb{E}\mathbb{I}_{\{X_i=1\}}) \\ &= \sum_{i=1}^m (\mathbb{E}X_i - \mathbb{E}\mathbb{I}_{\{X_i=1\}} + 2\mathbb{E}X_i \mathbb{E}\mathbb{I}_{\{X_i=1\}}) \\ &= \sum_{i=1}^m (\mathbb{E}Y_i + 2(np_i)^2(1 - p_i)^{n-1}) \\ &\leq \sum_{i=1}^m (\mathbb{E}Y_i + 3(np_i - np_i(1 - p_i)^{n-1})) \\ &= 4S_1. \end{aligned}$$

In the penultimate line we use the fact that, for large  $n$ , function  $(2nx + 3)(1 - x)^{n-1}$  is decreasing for  $x \in [0, 1)$ . Similarly

$$\begin{aligned} \text{Var} \sum_{i=1}^m Z_i &= \sum_{i=1}^m \text{Var}(\mathbb{I}_{\{X_i \text{ odd}\}} - \mathbb{I}_{\{X_i=1\}}) \\ &= \sum_{i=1}^m (\text{Var}\mathbb{I}_{\{X_i \text{ odd}\}} + \text{Var}\mathbb{I}_{\{X_i=1\}} - 2(\mathbb{E}(\mathbb{I}_{\{X_i \text{ odd}\}} \mathbb{I}_{\{X_i=1\}}) - \mathbb{E}\mathbb{I}_{\{X_i \text{ odd}\}} \mathbb{E}\mathbb{I}_{\{X_i=1\}})) \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i=1}^m (\mathbb{E}\mathbb{I}_{\{X_i \text{ odd}\}} + \mathbb{E}\mathbb{I}_{\{X_i=1\}} - 2\mathbb{E}\mathbb{I}_{\{X_i=1\}} + 2\mathbb{E}\mathbb{I}_{\{X_i \text{ odd}\}}\mathbb{E}\mathbb{I}_{\{X_i=1\}}) \\
&= \sum_{i=1}^m (\mathbb{E}\mathbb{I}_{\{X_i \text{ odd}\}} - \mathbb{E}\mathbb{I}_{\{X_i=1\}} + 2\mathbb{E}\mathbb{I}_{\{X_i \text{ odd}\}}\mathbb{E}\mathbb{I}_{\{X_i=1\}}) \\
&\leq \sum_{i=1}^m (\mathbb{E}X_i - \mathbb{E}\mathbb{I}_{\{X_i=1\}} + 2\mathbb{E}X_i\mathbb{E}\mathbb{I}_{\{X_i=1\}}) \\
&\leq 4S_1.
\end{aligned}$$

Therefore by Chebyshev's inequality, for any function  $\omega' = \omega'(n)$  tending to infinity

$$\begin{aligned}
\Pr \left\{ \left| \sum_{i=1}^m Y_i - S_1 \right| \geq \omega' \sqrt{S_1} \right\} &\leq \frac{4}{(\omega')^2} = o(1), \\
\Pr \left\{ \left| \sum_{i=1}^m Z_i - S_3 \right| \geq \omega' \sqrt{S_1} \right\} &\leq \frac{4}{(\omega')^2} = o(1).
\end{aligned} \tag{4.10}$$

Thus with high probability for  $1 \ll \omega' \ll S_3/\sqrt{S_1}$

$$\begin{aligned}
M_2 &\geq \frac{S_1 - 3S_3 - 4\omega' \sqrt{S_1}}{2}, \\
M_3 &\geq S_3 - \omega' \sqrt{S_1}.
\end{aligned}$$

Note that (4.7) and  $\omega' \ll S_3/\sqrt{S_1}$  imply that r.h.s. is positive for large  $n$ .

Therefore by Proposition 4.1 and (4.9)

$$\begin{aligned}
\mathbb{G}_{*2} \left( n, \frac{S_1 - 3S_3 - 4\omega' \sqrt{S_1}}{2} \right) \cup \mathbb{G}_{*3} \left( n, S_3 - \omega' \sqrt{S_1} \right) \\
\preceq_{1-o(1)}^{(*)} \mathbb{G}_{*2} (n, M_2) \cup \mathbb{G}_{*3} (n, M_3) \preceq \mathcal{G} (n, m, \bar{p}). \tag{4.11}
\end{aligned}$$

We may assume that in the above coupling  $\mathbb{G}_{*2} \left( n, \frac{S_1 - 3S_3 - 4\omega' \sqrt{S_1}}{2} \right)$  and  $\mathbb{G}_{*3} \left( n, S_3 - \omega' \sqrt{S_1} \right)$  are independent. The main reason for this is the fact that even though  $M_2$  and  $M_3$  are dependent (i.e. also  $\mathbb{G}_{*2} (n, M_2)$  and  $\mathbb{G}_{*3} (n, M_3)$  are dependent), the choices of hyperedges of  $\mathbb{H}_{*i}(n, \cdot)$  (related to  $\mathbb{G}_{*i}(n, \cdot)$ ) in distinct draws are independent. We describe the coupling (\*) below.

Set  $m_2^- = \frac{S_1 - 3S_3 - 4\omega' \sqrt{S_1}}{2}$  and  $m_3^- = S_3 - \omega' \sqrt{S_1}$ . We construct the coupling (\*) from (4.11) in the following way. Assume we have independent instances of  $\mathbb{H}_{*2}(n, m_2^-)$  and  $\mathbb{H}_{*3}(n, m_3^-)$  (i.e.  $\mathbb{G}_{*2} (n, m_2^-)$  and  $\mathbb{G}_{*3} (n, m_3^-)$ ). Moreover let  $M_2 = m_2$  and  $M_3 = m_3$ .

– If  $m_2 > m_2^-$  (or  $m_3 > m_3^-$  resp.) then we perform  $m_2 - m_2^-$  (or  $m_3 - m_3^-$ ) additional draws and add hyperedges to  $\mathbb{H}_{*2}(n, m_2^-)$  ( $\mathbb{H}_{*3}(n, m_3^-)$  resp.) to get an instance of  $\mathbb{G}_{*2} (n, M_2)$  ( $\mathbb{G}_{*3} (n, M_3)$  resp.) with  $M_2 = m_2$  ( $M_3 = m_3$  resp.).

– If  $m_2 \leq m_2^-$  ( $m_3 \leq m_3^-$  resp.) then we delete from  $\mathbb{H}_{*2}(n, m_2^-)$  ( $\mathbb{H}_{*3}(n, m_3^-)$  resp.) hyperedges created by the last draws to get exactly  $m_i$ ,  $i = 2, 3$ , draws.

Let  $M'_2$  and  $M'_3$  be independent random variables with the Poisson distributions

$$\text{Po}\left(\frac{S_1 - 3S_3 - 5\omega'\sqrt{S_1}}{2}\right) \quad \text{and} \quad \text{Po}\left(S_3 - 2\omega'\sqrt{S_1}\right), \text{ respectively.}$$

Then by a sharp concentration of the Poisson distribution

$$\begin{aligned} \Pr\left\{M'_2 \leq \frac{S_1 - 3S_3 - 4\omega'\sqrt{S_1}}{2}\right\} &= 1 - o(1) \quad \text{and} \\ \Pr\left\{M'_3 \leq S_3 - \omega'\sqrt{S_1}\right\} &= 1 - o(1). \end{aligned}$$

Therefore by Proposition 4.1 and (4.1)

$$\begin{aligned} &G_2\left(n, 1 - e\left(-\frac{S_1 - 3S_3 - 5\omega'\sqrt{S_1}}{2\binom{n}{2}}\right)\right) \cup G_3\left(n, 1 - e\left(-\frac{S_3 - 2\omega'\sqrt{S_1}}{\binom{n}{3}}\right)\right) \\ &= \mathbb{G}_{*2}(n, M'_2) \cup \mathbb{G}_{*3}(n, M'_3) \\ &\preceq_{1-o(1)} \mathbb{G}_{*2}\left(n, \frac{S_1 - 3S_3 - 4\omega'\sqrt{S_1}}{2}\right) \cup \mathbb{G}_{*3}\left(n, S_3 - \omega'\sqrt{S_1}\right) \\ &\preceq_{1-o(1)} \mathcal{G}(n, m, \bar{p}). \end{aligned} \tag{4.12}$$

For  $x \geq 0$  we have  $1 - e^{-x} \geq x - \frac{x^2}{2}$ , therefore

$$\begin{aligned} \hat{p}_2 &= \frac{S_1 - 3S_3 - 5\omega'\sqrt{S_1} - \frac{2S_1^2}{n^2}}{2\binom{n}{2}} \leq 1 - e\left(-\frac{S_1 - 3S_3 - 5\omega'\sqrt{S_1}}{2\binom{n}{2}}\right); \\ \hat{p}_3 &\leq \frac{S_3 - 2\omega'\sqrt{S_1} - \frac{6S_3^2}{n^3}}{\binom{n}{3}} \leq 1 - e\left(-\frac{S_3 - 2\omega'\sqrt{S_1}}{\binom{n}{3}}\right). \end{aligned}$$

Therefore using standard couplings of  $G_2(n, \cdot)$  and  $H_3(n, \cdot)$  finally we get

$$\begin{aligned} &G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) \\ &\preceq G_2\left(n, 1 - e\left(-\frac{S_1 - 3S_3 - 5\omega'\sqrt{S_1}}{2\binom{n}{2}}\right)\right) \cup G_3\left(n, 1 - e\left(-\frac{S_3 - 2\omega'\sqrt{S_1}}{\binom{n}{3}}\right)\right) \\ &= \mathbb{G}_{*2}\left(n, \text{Po}\left(\frac{S_1 - 3S_3 - 5\omega'\sqrt{S_1}}{2}\right)\right) \cup \mathbb{G}_{*3}\left(n, \text{Po}\left(S_3 - 2\omega'\sqrt{S_1}\right)\right) \\ &\preceq_{1-o(1)} \mathcal{G}(n, m, \bar{p}). \end{aligned} \tag{4.13}$$

The result follows by Propositions 4.3 and 4.4.

An analogous reasoning gives (2.3) and (2.4) in the case  $S_3 = O(\sqrt{S_1})$ . One need to replace (4.8) by

$$\mathbb{G}_{*2} \left( n, \frac{Y_i - Z_i}{2} \right) \preceq \mathcal{G}[V'_i] \quad \text{and} \quad \mathbb{G}_{*2} \left( n, \frac{Y_i - 3Z_i}{2} \right) \preceq \mathcal{G}[V'_i],$$

respectively, and modify the proof accordingly.  $\square$

## 5 Vertex degrees in $\mathcal{G}(n, m, \bar{p})$

For any graph  $G$  denote by  $\delta(G)$  the minimum vertex degree in  $G$ .

**Lemma 5.1.** *Let  $c_n$  be a sequence of real numbers such that  $c_n$  attains a limit (possibly  $\pm\infty$ ),  $\bar{p} = (p_1, \dots, p_m)$  such that  $\max_{1 \leq i \leq m} p_i = o((\ln n)^{-1})$ , and  $S_1$  be given by (2.1).*

*If*

$$S_1 = n(\ln n + c_n)$$

*then*

$$\lim_{n \rightarrow \infty} \Pr \{ \delta(\mathcal{G}(n, m, \bar{p})) \geq 1 \} = f(c_n),$$

*where  $f(\cdot)$  is given by (2.5).*

Note that the condition  $\max_{1 \leq i \leq m} p_i = o((\ln n)^{-1})$  is necessary. Otherwise the number of vertices of a given degree depends more on the fluctuations of the values of vector  $\bar{p}$ . For example let  $m = n^2$  and  $\bar{p} = (p_1, p_2, \dots, p_m)$  equal

$$\left( \underbrace{\frac{b_n}{\ln n}, \dots, \frac{b_n}{\ln n}}_{\ln n(\ln n + c_n)/b_n}, \frac{1}{nm}, \dots, \frac{1}{nm} \right) \quad \text{or} \\ \left( \frac{1}{\sqrt{\ln n}}, \sqrt{\frac{\ln n + c_n}{nm}}, \dots, \sqrt{\frac{\ln n + c_n}{nm}} \right),$$

for some  $b_n, c_n = \Theta(1)$ . In both cases

$$S_1 = \sum_{i=1}^m np_i (1 - (1 - p_i)^{n-1}) = n(\ln n + c_n + o(1)).$$

On the other hand the expected number of vertices of degree 0 in  $\mathcal{G}(n, m, \bar{p})$  is

$$n \prod_{i=1}^m ((1 - p_i) + p_i(1 - p_i)^{n-1}) = n \prod_{i=1}^m ((1 - p_i)(1 - (1 - p_i)^{n-1})),$$

since a vertex  $v$  is isolated if for each  $i$ ,  $1 \leq i \leq m$ ,  $v$  does not choose  $w_i$  or  $v$  chooses  $w_i$  but no other vertex does. Therefore this expected value equals

$$(1 + o(1)) \exp \left( -c_n - \frac{b_n}{2} \right) \quad \text{or} \quad (1 + o(1)) \exp(-c_n),$$

respectively.

**Lemma 5.2.** Let  $c_n$  be a sequence of real numbers,  $k$  be a positive integer, and  $\bar{p} = (p_1, \dots, p_m)$  such that  $\max_{1 \leq i \leq m} p_i = o((\ln n)^{-1})$ . Moreover let  $S_1$  and  $S_{1,t}$ ,  $t = 2, \dots, k$ , be given by (2.1).

(i) If

$$S_1 = n \left( \ln n + (k-1) \ln \left( \max \left\{ 1, \left( \frac{S_{1,2}}{S_1} \ln n \right) \right\} \right) \right) + c_n$$

then

$$\lim_{n \rightarrow \infty} \Pr \{ \delta(\mathcal{G}(n, m, \bar{p})) \geq k \} = 0 \quad \text{for } c_n \rightarrow -\infty.$$

(ii) If

$$S_1 - \sum_{t=3}^k S_{1,t} = n \left( \ln n + (k-1) \ln \left( \max \left\{ 1, \left( \frac{S_{1,2}}{S_1} \ln n \right) \right\} \right) \right) + c_n$$

then

$$\lim_{n \rightarrow \infty} \Pr \{ \delta(\mathcal{G}(n, m, \bar{p})) \geq k \} = 1 \quad \text{for } c_n \rightarrow \infty.$$

Here and in the proof we assume that  $\sum_{t=3}^2 S_{1,t} = 0$ .

If we assume that  $p_i = p = p(n)$ , for all  $1 \leq i \leq m$ , then with a little more work the result of Lemma 5.2 may be improved.

**Lemma 5.3.** Let  $p = p(n) \in (0, 1)$ ,  $k$  be a positive integer,  $m = m(n)$  be such that  $m \gg \ln^2 n$ , and

$$a_n = (np)^{k-1} \left( \left( \frac{e^{-np} \ln n}{1 - e^{-np}} \right)^{k-1} + \frac{e^{-np} \ln n}{1 - e^{-np}} \right).$$

If

$$p(1 - (1 - p)^{n-1}) = \frac{\ln n + \ln(\max\{1, a_n\}) + c_n}{m},$$

then

$$\lim_{n \rightarrow \infty} \Pr \{ \delta(\mathcal{G}(n, m, p)) \geq k \} = \begin{cases} 0 & \text{for } c_n \rightarrow -\infty; \\ 1 & \text{for } c_n \rightarrow \infty. \end{cases}$$

*Proof of Lemma 5.1.* Assume first that  $c_n = \Omega(\ln n)$  and  $c_n \geq 0$ , for large  $n$ . Then for any  $c'_n = o(\ln n)$  there exists  $\bar{p}' = (p'_1, \dots, p'_m)$  such that  $p'_i \leq p_i$ , for all  $1 \leq i \leq m$ , and  $\sum_{i=1}^m np'_i(1 - (1 - p'_i)^{n-1}) = n(\ln n + c'_n)$  (as  $f(x) = x(1 - (1 - x)^{n-1})$  is a continuous, increasing function for  $x \in (0, 1)$ ). Therefore  $\mathcal{G}(n, m, \bar{p}') \preceq \mathcal{G}(n, m, \bar{p})$  and  $\Pr \{ \delta(\mathcal{G}(n, m, \bar{p}')) \geq 1 \} \leq \Pr \{ \delta(\mathcal{G}(n, m, \bar{p})) \geq 1 \}$ .

Now assume that  $c_n = \Omega(\ln n)$  and  $c_n \leq 0$ , for large  $n$ . Then for any  $c'_n = o(\ln n)$  we may find  $p = o((\ln n)^{-1})$  and  $m'$  such that  $m'np(1 - (1 - p)^{n-1}) = n(c'_n - c_n)$ . Let  $\bar{p}' = (p'_1, \dots, p'_{m+m'})$  be such that  $p'_i = p_i$ , for  $1 \leq i \leq m$  and  $p'_i = p$ , for  $m+1 \leq i \leq m+m'$ . Then  $\sum_{i=1}^{m+m'} np'_i(1 - (1 - p'_i)^{n-1}) = n(\ln n + c'_n)$ ,  $\mathcal{G}(n, m, \bar{p}) \preceq \mathcal{G}(n, m+m', \bar{p}')$ , and  $\Pr \{ \delta(\mathcal{G}(n, m, \bar{p})) \geq 1 \} \leq \Pr \{ \delta(\mathcal{G}(n, m+m', \bar{p}')) \geq 1 \}$ .

Therefore we may restrict ourselves to the case  $c_n = o(\ln n)$ .



Imagine an experiment consisting of an infinite sequence of draws. In each draw we pick independently and uniformly at random one element from the vertex set  $\mathcal{V}$ . If a vertex is picked in the  $i$ -th draw we call it *collected* in the  $i$ -th draw. The outcome of the experiment is an infinite sequence with elements from the vertex set  $\mathcal{V}$ . We will call this experiment the vertex collector process.

Moreover assume that independently we pick a vector

$$(Y_1, Y_2, \dots, Y_m)$$

chosen so that  $Y_i$  are independent random variables with the distribution of the random variables defined in (4.4).

Now we will construct an instance of  $\mathcal{G}(n, m, \bar{p})$  basing on those two random independent incomes. In fact we will construct a random family of sets  $\{V'_i : i = 1, \dots, m\}$  with  $|V'_i| = Y_i$ , for all  $i = 1, \dots, m$ , which is equivalent to constructing an instance of  $\mathcal{G}(n, m, \bar{p})$ .

Set a vector

$$(Y_1, Y_2, \dots, Y_m) = (y_1, y_2, \dots, y_m).$$

Now, independently, we create an infinite sequence of draws described above. We divide this sequence of draws into subsequences. We will call them phases. In the first phase we perform as many draws as necessary to collect  $y_1$  distinct vertices from  $\mathcal{V}$ . We set  $V'_1$  to be the set consisting of the vertices collected in the first phase and  $T_1$  to be the number of draws made in the first phase. Then we continue and in the second phase we perform as many draws as needed to collect  $y_2$  distinct vertices from  $\mathcal{V}$ . Note that the number of draws and vertices collected in the second phase are independent of the first phase. As before  $V'_2$  is the set of collected vertices and  $T_2$  is the number of draws in the second phase. We continue in the same manner till the  $m$ -th phase. After constructing  $V'_m$  we proceed with draws without constructing any other sets.

Note that sets  $V'_1, \dots, V'_m$  constructed in this manner are independent, their sizes are independent random variables  $Y_1, \dots, Y_m$ , and each  $V'_i$ ,  $1 \leq i \leq m$ , is chosen uniformly at random from all subsets of  $\mathcal{V}$  of size  $Y_i$ . In the coupling of the vertex collector process and the construction of  $\mathcal{G}(n, m, \bar{p})$  described above, the construction of  $\mathcal{G}(n, m, \bar{p})$  is finished in  $\sum_{i=1}^m T_i$  draws. Now we will prove that for some  $T_- = T_-(n)$  and  $T_+ = T_+(n)$  with high probability

$$T_- \leq \sum_{i=1}^m T_i \leq T_+.$$

The values  $T_-$  and  $T_+$  will be specified later on. Recall that by (4.10), for any  $1 \ll \omega \ll \sqrt{S_1}/\ln n$ ,

$$\Pr \left\{ S_1 - \omega \sqrt{S_1} \leq \sum_{i=1}^m Y_i \leq S_1 + \omega \sqrt{S_1} \right\} = 1 - o(1)$$

Therefore we are left with studying the random variable

$$\sum_{i=1}^m T_i - \sum_{i=1}^m Y_i = \sum_{i=1}^m (T_i - Y_i).$$

If  $Y_i = y_i = 0$  then  $T_i = 0$  and  $T_i - Y_i = 0$ . Assume now that  $Y_i = y_i \geq 2$ . If, in the  $i$ -th phase, we have already collected  $j < y_i$  vertices then the number of draws to collect the  $j + 1$ -st vertex has geometric distribution with parameter  $\frac{n-j}{n}$ . Therefore

$$\mathbb{E}T_i = \left( \sum_{j=0}^{y_i-1} \frac{n}{n-j} \right).$$

and

$$\begin{aligned} \mathbb{E}(T_i - Y_i | Y_i = y_i) &= \left( \sum_{j=0}^{y_i-1} \frac{n}{n-j} \right) - y_i \\ &= \sum_{j=0}^{y_i-1} \frac{j}{n-j} \\ &\leq \begin{cases} \sum_{j=0}^{y_i-1} \frac{2j}{n} = \frac{y_i(y_i-1)}{n} & \text{for } 2 \leq y_i < n/2; \\ n \ln n & \text{for } y_i \geq n/2. \end{cases} \end{aligned}$$

Moreover, for large  $n$ ,

$$\Pr \left\{ Y_i \geq \frac{n}{2} \right\} \leq \binom{n}{\frac{n}{2}} p_i^{\frac{n}{2}} \leq \left( \frac{enp_i}{\frac{n}{2}} \right)^{\frac{n}{2}} \leq \frac{p_i^2}{\ln n}.$$

Thus

$$\mathbb{E}(T_i - Y_i) \leq \sum_{y_i=2}^{n/2} \frac{y_i(y_i-1)}{n} \binom{n}{y_i} p_i^{y_i} (1-p_i)^{n-y_i} + n \ln n \Pr \left\{ Y_i \geq \frac{n}{2} \right\} \leq 2np_i^2.$$

Therefore

$$\begin{aligned} \mathbb{E} \left( \sum_{i=1}^m (T_i - Y_i) \right) &\leq \sum_{i=1}^m 2np_i^2 \\ &= \sum_{i=1}^m 2 \max \left\{ p_i, \frac{1}{n} \right\} \min \{ np_i, n^2 p_i^2 \} \\ &\leq 4 \max \left\{ p_1, \dots, p_m, \frac{1}{n} \right\} \sum_{i=1}^m \frac{1}{2} \min \{ np_i, n^2 p_i^2 \} \\ &= o \left( \frac{1}{\ln n} \right) \sum_{i=1}^m np_i (1 - (1-p_i)^{n-1}) \\ &= o \left( \frac{S_1}{\ln n} \right), \end{aligned}$$

where  $S_1$  is defined by (2.1). In order to get the penultimate line consider two cases  $p(n-1) \leq 1/2$  and  $p(n-1) > 1/2$ .

Moreover by Markov's inequality for any  $\omega$

$$\Pr \left\{ \sum_{i=1}^m (T_i - Y_i) \geq \frac{S_1}{\omega \ln n} \right\} \leq \mathbb{E} \left( \sum_{i=1}^m (T_i - Y_i) \right) \frac{\omega \ln n}{S_1}.$$

Therefore with high probability

$$\sum_{i=1}^m (T_i - Y_i) \leq \frac{S_1}{\omega \ln n}, \text{ for any } 1 \ll \omega \ll \frac{S_1}{(\ln n \mathbb{E}(\sum_{i=1}^m (T_i - Y_i)))}. \quad (5.1)$$

This combined with (4.10) gives that with high probability

$$T_- \leq \sum_{i=1}^m T_i \leq T_+, \quad (5.2)$$

where

$$T_- = S_1 - \omega \sqrt{S_1} \quad \text{and} \quad T_+ = S_1 + \omega \sqrt{S_1} + \frac{S_1}{\omega \ln n}$$

for some  $1 \ll \omega \ll \min\{S_1/(\ln n \mathbb{E}(\sum_{i=1}^m (T_i - Y_i))), \sqrt{S_1}/\ln n\}$ . Therefore with high probability the number of draws needed to construct  $\mathcal{G}(n, m, \bar{p})$  is between  $T_-$  and  $T_+$ .

Now we will describe a relation between the minimum degree of  $\mathcal{G}(n, m, \bar{p})$  and the number of collected vertices in the first  $T_-$  and  $T_+$  draws. If the construction of  $\mathcal{G}(n, m, \bar{p})$  is finished between  $T_-$  and  $T_+$  draws and each vertex has been collected at least once in the first  $T_-$  draws then there is no isolated vertex in  $\mathcal{G}(n, m, \bar{p})$ . Moreover the number of isolated vertices in  $\mathcal{G}(n, m, \bar{p})$  is at least the number of non collected vertices at time  $T_+$ .

In the probability space of the coupling of the vertex collector process with the construction of  $\mathcal{G}(n, m, \bar{p})$  define events:

- $\mathcal{A}_-$  – all vertices are collected in at most  $T_-$  draws;
- $\mathcal{A}_+$  – all vertices are collected in at most  $T_+$  draws;
- $\mathcal{A}$  –  $\delta(\mathcal{G}(n, m, \bar{p})) \geq 1$ ;
- $\mathcal{B}$  – the construction of  $\mathcal{G}(n, m, \bar{p})$  is finished between  $T_-$ -th and  $T_+$ -th draw.

Recall that  $\delta(\mathcal{G}(n, m, \bar{p})) \geq 1$  if and only if all vertices are collected before the construction of  $\mathcal{G}(n, m, \bar{p})$  is finished. Therefore

$$\mathcal{A}_+ \cap \mathcal{B} \subseteq \mathcal{A} \cap \mathcal{B} \subseteq \mathcal{A}_- \cap \mathcal{B}.$$

This implies

$$\Pr \{\mathcal{A}_+\} - \Pr \{\mathcal{B}^c\} \leq \Pr \{\mathcal{A}\} \leq \Pr \{\mathcal{A}_-\} + \Pr \{\mathcal{B}^c\}, \quad (5.3)$$

where for any event  $\mathcal{C}$  we denote by  $\mathcal{C}^c$  its complement.

Let  $S_1 = n(\ln n + c_n)$ , then

$$T_{\pm} = n(\ln n + c_n + o(1)).$$

By the classical results on the coupon collector problem [10], if  $T = T_n = n(\ln n + d_n)$ ,  $d_n \rightarrow d \in [-\infty, +\infty]$ , then the probability of the event that all vertices are collected in at most  $T$  draws tends to  $f(d_n)$  as  $n \rightarrow \infty$ . Therefore

$$\Pr \{\mathcal{A}_-\} \sim \Pr \{\mathcal{A}_+\} \sim f(c_n).$$

Moreover by (5.2)

$$\Pr \{\mathcal{B}^c\} = o(1).$$

Thus the lemma follows by substituting the above values to (5.3).  $\square$

The proofs of Lemmas 5.2 and 5.3 rely on couplings with the vertex collector process similar to the one used in the proof of Lemma 5.1 and a technique of dividing  $\mathcal{G}(n, m, \bar{p})$  into auxiliary subgraphs. We will give here the idea of the proofs of Lemmas 5.2 and 5.3 and present technical details in Appendix B.

First we gather some simple facts concerning the division of  $\mathcal{G}(n, m, \bar{p})$ . Let  $k \geq 2$  be an integer. For each  $2 \leq t \leq k$ , let  $\mathcal{G}_t(n, m, \bar{p})$  be a random graph with the vertex set  $\mathcal{V}$  and edge set consisting of these edges from  $\mathcal{G}(n, m, \bar{p})$ , which are contained in at least one of the sets  $\{V_i : |V_i| = t\}$ , where  $V_i$  is defined as in the proof of Theorem 2.1. Moreover let  $\mathcal{G}_{\geq k+1}(n, m, \bar{p})$  be the subgraph of  $\mathcal{G}(n, m, \bar{p})$  containing only these edges which are subsets of at least one of the sets  $\{V_i : |V_i| \geq k+1\}$ . We also define  $\mathcal{G}_{\neq t}(n, m, \bar{p})$  to be the subgraph of  $\mathcal{G}(n, m, \bar{p})$  containing only these edges which are subsets of at least one of the sets  $\{V_i : |V_i| \neq t\}$ . Then, for any  $k$  and  $2 \leq t \leq k$

$$\mathcal{G}(n, m, \bar{p}) = \bigcup_{t=2}^k \mathcal{G}_t(n, m, \bar{p}) \cup \mathcal{G}_{\geq k+1}(n, m, \bar{p}) = \mathcal{G}_{\neq t}(n, m, \bar{p}) \cup \mathcal{G}_t(n, m, \bar{p}).$$

Let

$$\begin{aligned} Q_t &= \sum_{i=1}^m \mathbb{I}_{Y_i=t}, \text{ for all } t = 2, \dots, k; \\ Q_{\geq k+1} &= \sum_{i=1}^m Y_i \mathbb{I}_{Y_i \geq k+1} = \sum_{i=1}^m Y_i - \sum_{t=2}^k t Q_t; \\ Q_{\neq t} &= \sum_{i=1}^m Y_i \mathbb{I}_{Y_i \neq t} = \sum_{i=1}^m Y_i - t Q_t. \end{aligned} \tag{5.4}$$

By definition  $Q_t$ ,  $2 \leq t \leq k$ , is the number of features which have been chosen by exactly  $t$  vertices, i.e. the number of cliques forming  $\mathcal{G}(n, m, \bar{p})$ , which have exactly  $t$  vertices. Thus  $Q_t$ ,  $2 \leq t \leq k$ , is also the number of cliques forming  $\mathcal{G}_t(n, m, \bar{p})$  i.e.  $\mathcal{G}_t(n, m, \bar{p}) = \mathbb{G}_{*t}(n, Q_t)$  (where  $\mathbb{G}_{*t}(n, \cdot)$  is defined at the beginning of Section 4). Moreover by (4.5)

$$\mathbb{E}Q_t = \frac{S_{1,t}}{t} \text{ for all } t = 2, \dots, k.$$

Moreover  $Q_{\geq k+1}$  is the sum of sizes of the cliques forming  $\mathcal{G}(n, m, \bar{p})$  which have at least  $k+1$  vertices, i.e. the sum of sizes of the cliques forming  $\mathcal{G}_{\geq k+1}(n, m, \bar{p})$ . Therefore by (4.5) and (4.6)

$$\mathbb{E}Q_{\geq k+1} = S_1 - \sum_{t=2}^k S_{1,t}.$$

Last but not least, the expected value of the sum of sizes of the cliques forming  $\mathcal{G}_{\neq t}(n, m, \bar{p})$  is

$$\mathbb{E}Q_{\neq t} = S_1 - S_{1,t}.$$

For all  $1 \leq t \leq k$ ,  $Q_t$  is a sum of independent Bernoulli random variables and  $S_{1,t} \leq S_1$  (see (4.6)). Therefore by Chebyshev's inequality and by (4.10) for any  $\omega \rightarrow \infty$  with high probability

$$\max \left\{ 0, \frac{S_{1,t} - \omega\sqrt{S_1}}{t} \right\} \leq Q_t \leq \frac{S_{1,t} + \omega\sqrt{S_1}}{t}, \text{ for all } t = 2, 3, \dots, k \quad (5.5)$$

$$Q_{\geq k+1} \geq S_1 - \sum_{t=2}^k S_{1,t} - k\omega\sqrt{S_1}, \quad (5.6)$$

$$S_1 - S_{1,t} - 2\omega\sqrt{S_1} \leq Q_{\neq t} \leq S_1 - S_{1,t} + 2\omega\sqrt{S_1}, \text{ for all } t = 2, 3, \dots, k. \quad (5.7)$$

Moreover, for any  $t = 2, 3, \dots, k$ , if  $S_{1,t} \rightarrow \infty$  then with high probability

$$\frac{S_{1,t} - \omega\sqrt{S_{1,t}}}{t} \leq Q_t \leq \frac{S_{1,t} + \omega\sqrt{S_{1,t}}}{t}. \quad (5.8)$$

Now we are ready to give sketches of the proofs of Lemmas 5.2 and 5.3. Technical details of the proofs are presented in Appendix B.

*Sketch of the proof of Lemma 5.2 (i).* We prove the lemma in the case  $c_n = o(\ln n)$  and  $a_n = S_{1,2}/S_1 \gg 1/\ln n$ , i.e.  $S_{1,2} \rightarrow \infty$ . In the latter cases the result follows by Lemma 5.1. We use the division  $\mathcal{G}(n, m, \bar{p}) = \mathcal{G}_2(n, m, \bar{p}) \cup \mathcal{G}_{\neq 2}(n, m, \bar{p})$ . If there exists a vertex which is isolated in  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  and has degree at most  $k-1$  in  $\mathcal{G}_2(n, m, \bar{p})$  then  $\delta(\mathcal{G}(n, m, \bar{p})) \leq k-1$ . In order to bound the number of vertices which are isolated in  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  and have degree at most  $k-1$  in  $\mathcal{G}_2(n, m, \bar{p})$  we will use a coupling.

First of all, note that using (5.8) and the same technique as in the proof of Theorem 2.1 we may show that

$$\mathcal{G}_2(n, m, \bar{p}) \preceq_{1-o(1)} G_2(n, \hat{p}_{2+}) \quad \text{with} \quad \hat{p}_{2+} = \frac{S_{1,2} + 2\omega\sqrt{S_{1,2}}}{2\binom{n}{2}}.$$

Moreover we may construct a coupling of the vertex collector process described in the proof of Lemma 5.1 and a construction  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$ . This coupling is analogous to the one described in the proof of Lemma 5.1. The only change is that during the construction of  $\mathcal{G}(n, m, \bar{p})$  we omit phases related to the cliques of size 2. By (5.1) and (5.7) with high

probability the construction of  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  takes at most  $T_+ = S_1 - S_{1,2} + 2\omega\sqrt{S_1} + \frac{S_1}{\omega \ln n}$  draws in the vertex collector process.

The only problem might be caused by the dependency between graphs  $\mathcal{G}_2(n, m, \bar{p})$  and  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$ . This dependency is due to the fact that  $Q_2$  and  $Q_{\neq 2}$  are dependent. On the other hand the choices of cliques forming  $\mathcal{G}_2(n, m, \bar{p})$  and  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  are independent. Therefore ideas used in proving (4.11) might also be used here to show that we may couple  $\mathcal{G}_2(n, m, \bar{p}) \cup \mathcal{G}_{\neq 2}(n, m, \bar{p})$  with  $G_2(n, \hat{p}_{2+})$  and an independent vertex collector process in such a manner that with high probability  $\mathcal{G}_2(n, m, \bar{p}) \subseteq G_2(n, \hat{p}_{2+})$  and the construction of  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  is finished in at most  $T_+$  draws.

With the second moment method we may show that if

$$S_1 = n \left( \ln n + (k-1) \ln \left( \max \left\{ 1, \left( \frac{S_{1,2}}{S_1} \ln n \right) \right\} \right) + c_n \right), \quad c_n \rightarrow -\infty,$$

then with high probability there exists a vertex of degree at most  $k-1$  in  $G_2(n, \hat{p}_{2+})$  which has not been collected in  $T_+$  draws of the vertex collector process. Therefore, by the constructed coupling, with high probability there exists a vertex which has degree at most  $k-1$  in  $\mathcal{G}_2(n, m, \bar{p})$  and is isolated in  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$ . Hence

$$\lim_{n \rightarrow \infty} \Pr \{ \delta(\mathcal{G}(n, m, \bar{p})) \geq k \} = 0. \quad \square$$

*Sketch of the proof of Lemma 5.2 (ii).* We use a similar reasoning as in the proof of (i). We may concentrate on the case  $c_n = O(\ln n)$  and  $S_{1,2}/S_1 \gg 1/\ln n$ . Let  $k \geq 2$ , then

$$\mathcal{G}_2(n, m, \bar{p}) \cup \mathcal{G}_{\geq k+1}(n, m, \bar{p}) \subseteq \mathcal{G}(n, m, \bar{p})$$

As before, we may construct a coupling of  $\mathcal{G}_2(n, m, \bar{p}) \cup \mathcal{G}_{\geq k+1}(n, m, \bar{p})$  with  $G_2(n, \hat{p}_{2-})$  and an independent vertex collector process with at least  $T_-$  draws where

$$\hat{p}_{2-} = \frac{S_{1,2} - 2\omega\sqrt{S_{1,2}} - 2S_{1,2}^2 n^{-2}}{2\binom{n}{2}}, \quad T_- = \max \left\{ 0, S_1 - \sum_{t=2}^k S_{1,t} - k\omega\sqrt{S_1} \right\}.$$

In this coupling with high probability  $G_2(n, \hat{p}_{2-}) \subseteq \mathcal{G}_2(n, m, \bar{p})$  and the construction of  $\mathcal{G}_{\geq k+1}(n, m, \bar{p})$  takes at least  $T_-$  draws.

By the first moment method we may show that if

$$S_1 - \sum_{t=3}^k S_{1,t} = n \left( \ln n + (k-1) \ln \left( \max \left\{ 1, \left( \frac{S_{1,2}}{S_1} \ln n \right) \right\} \right) + c_n \right), \quad c_n \rightarrow \infty,$$

then with high probability each vertex has degree at least  $k$  in  $G_2(n, \hat{p}_{2-})$  or is collected in  $T_-$  draws in the vertex collector process. Therefore, by the coupling, with high probability each vertex has degree at least  $k$  in  $\mathcal{G}_2(n, m, \bar{p})$  or is not isolated in  $\mathcal{G}_{\geq k+1}(n, m, \bar{p})$ . This implies that with high probability each vertex has degree at least  $k$  in  $\mathcal{G}_2(n, m, \bar{p}) \cup \mathcal{G}_{\geq k+1}(n, m, \bar{p})$  and

$$\lim_{n \rightarrow \infty} \Pr \{ \delta(\mathcal{G}(n, m, \bar{p})) \geq k \} = 1 \quad \text{for } c_n \rightarrow \infty. \quad \square$$

*Sketch of the proof of Lemma 5.3.* The main ideas behind the proof of Lemma 5.3 are the same as in the proof of Lemma 5.2. As before, we may restrict ourselves to the case  $S_1 \sim n \ln n$  (recall that  $p(1 - (1 - p)^{n-1}) = S_1/nm$ ). We divide the proof into cases, depending on the value  $np$  (recall that the expected size of  $V_i$  is  $\mathbb{E}X_i = np$ ,  $1 \leq i \leq m$ ). For convenience in the case  $\bar{p} = (p, \dots, p)$  we replace  $\bar{p}$  by  $p$  in notations  $\mathcal{G}_t(n, m, \bar{p})$ ,  $\mathcal{G}_{\geq k+1}(n, m, \bar{p})$ ,  $\mathcal{G}_{\neq t}(n, m, \bar{p})$ .

If  $S_{1,2} = O(S_1/\ln n)$  and  $S_{1,t} = O(S_1/\ln n)$  then the result follows by Lemma 5.2.

Otherwise  $S_{1,2} \gg S_1/\ln n$  or  $S_{1,k} \gg S_1/\ln n$  (i.e.  $np \leq \ln \ln n + (k-1) \ln \ln \ln n - \omega$  for some  $\omega = \omega(n) \rightarrow \infty$ ). For  $c_n \rightarrow -\infty$  we use either the division  $\mathcal{G}_{\neq k}(n, m, p) \cup \mathcal{G}_k(n, m, p)$  or the division  $\mathcal{G}_{\neq 2}(n, m, p) \cup \mathcal{G}_2(n, m, p)$ . For  $c_n \rightarrow \infty$  we use the division  $\bigcup_{t=2}^k \mathcal{G}_2(n, m, p) \cup \mathcal{G}_{\geq k+1}(n, m, p)$ .  $\square$

## 6 Structural properties of $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$

In this section we give several structural results concerning  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$ . Some of the presented facts are analogous to those known for  $G_2(n, \hat{p}_2)$ . Then we will combine obtained results with the coupling constructed in the proof of Theorem 2.1 in order to obtain the lower bounds on  $\lim_{n \rightarrow \infty} \Pr\{\mathcal{G}(n, m, \bar{p}) \in \mathcal{A}\}$ , for  $\mathcal{A}$  being  $\mathcal{C}_k$ ,  $\mathcal{PM}$  or  $\mathcal{HC}$ .

For any graph  $G$  and any set  $S \subseteq V(G)$  denote by  $N_G(S)$  the set of neighbours of vertices from  $S$  contained in  $V(G) \setminus S$ . For simplicity we write  $N_2(\cdot) = N_{G_2(n, \hat{p}_2)}(\cdot)$  and  $N_3(\cdot) = N_{G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)}(\cdot)$ . Note that given an instance of  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  we have  $|N_2(S)| \leq |N_3(S)|$ .

**Lemma 6.1.** *Let  $k$  and  $C$  be positive integers and*

$$\begin{aligned} \hat{p}_2 + \frac{n}{2}\hat{p}_3 &= \frac{\theta_n \ln n}{n}, \quad \text{where } \liminf_{n \rightarrow \infty} \theta_n = \theta > \frac{1}{2} \text{ and } \theta_n = O(1), \\ \hat{p}_2 &= \frac{\theta'_n \ln n}{n}, \quad \text{where } \liminf_{n \rightarrow \infty} \theta'_n = \theta' \geq \frac{1}{2} \text{ and } \theta'_n = O(1). \end{aligned} \tag{6.1}$$

Let moreover  $\gamma$  be a positive real number such that  $1 - \theta' < \gamma < 1$ . Then with high probability  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  has the following properties:

(i)  $\mathcal{B}_1$  – for all  $S \subseteq \mathcal{V}$  such that  $n^\gamma \leq |S| \leq \frac{1}{4}n$

$$|N_3(S)| > 2|S|.$$

(ii)  $\mathcal{B}_2$  – for all  $S \subseteq \mathcal{V}$  such that  $n^\gamma \leq |S| \leq \frac{2}{3}n$

$$\begin{aligned} |N_3(S)| &> \min\{|S|, 4n \ln \ln n / \ln n\} \\ &> \min\{|S|, \alpha(G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3))\}, \end{aligned}$$

where  $\alpha(G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3))$  is the stability number of  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$ .

(iii)  $\mathcal{B}_{3,k}$  – for all  $S \subseteq \mathcal{V}$ , if  $1 \leq |S| \leq n^\gamma$  and all vertices in  $S$  have degree at least  $4k + 14$  in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  we have

$$|N_3(S)| \geq 2k|S|.$$

(iv)  $\mathcal{B}_{4,C}$  – any two vertices of degree at most  $C$  are at distance at least 5.

(v)  $\mathcal{B}_5$  – contains a path of length at least

$$\left(1 - \frac{9 \ln 2}{\ln n}\right) n.$$

*Proof.* (i) and (ii) By Chernoff's inequality (for the proof see for example Theorem 2.1 in [13]) for any  $\delta = \delta(n) = o(1)$  and any binomial random variable  $X = X_n$  we have

$$\Pr \{X \leq \delta \mathbb{E}X\} \leq \exp(-\mathbb{E}X(\delta \ln \delta + 1 - \delta)) = \exp(-(1 + o(1))\mathbb{E}X).$$

First consider the case  $n^\gamma \leq |S| \leq 4n \ln \ln n / \ln n$ . Let  $s = |S|$ . Then  $|N_2(S)|$  has binomial distribution  $\text{Bin}(n - s, 1 - (1 - \hat{p}_2)^s)$  with the expected value  $\mathbb{E}|N_2(S)| \gg 2|S|$ . Denote by  $X$  the number of sets  $S$  of cardinality  $n^\gamma \leq |S| \leq 4n \ln \ln n / \ln n$  such that  $|N_2(S)| \leq 2|S|$ . Using Chernoff's inequality we get

$$\begin{aligned} \mathbb{E}X &\leq \sum_{s=n^\gamma}^{4n \ln \ln n / \ln n} \binom{n}{s} \exp(-(1 + o(1))(n - s)(1 - (1 - \hat{p}_2)^s)) \\ &\leq \sum_{s=n^\gamma}^{n/(\ln n \ln \ln n)} \exp\left(s \left(1 + \ln \frac{n}{s}\right) - (1 + o(1))s n \hat{p}_2\right) \\ &\quad + \sum_{s=n/(\ln n \ln \ln n)}^{4n \ln \ln n / \ln n} \exp\left(s \left(1 + \ln \frac{n}{s}\right) - n \frac{\theta'}{2 \ln \ln n}\right) \\ &\leq \sum_{s=n^\gamma}^{n/(\ln n \ln \ln n)} \exp(s(1 + (1 - \gamma) \ln n - (1 + o(1))\theta' \ln n)) \\ &\quad + \sum_{s=n/(\ln n \ln \ln n)}^{4n \ln \ln n / \ln n} \exp\left(s \left((1 + o(1)) \ln \ln n - \frac{\theta' \ln n}{8(\ln \ln n)^2}\right)\right) \\ &= o(1). \end{aligned}$$

Therefore with high probability for all sets  $S \subseteq \mathcal{V}$ ,  $n^\gamma \leq |S| \leq 4n \ln \ln n / \ln n$ , we have

$$|N_2(S)| \geq 2|S|.$$

Now consider the case  $|S| \geq 4n \ln \ln n / \ln n$ . Let  $r = 4n \ln \ln n / \ln n$ . Let moreover  $\overline{K}_r$  and  $\overline{K}_{r,r}$  be the complement of the complete graph on  $r$  vertices and the complement of



the complete bipartite graph with each set of bipartition of cardinality  $r$ . Denote by  $X_r$  and  $X_{r,r}$  the number of  $\overline{K}_r$  and  $\overline{K}_{r,r}$  in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$ , respectively. Then

$$\begin{aligned} \mathbb{E}X_r &= \binom{n}{r} (1 - \hat{p}_2)^{\binom{r}{2}} (1 - \hat{p}_3)^{\binom{r}{3} + \binom{r}{2}(n-r)} \\ &\leq \left( \frac{en}{r} \exp\left(-\frac{r}{2}(\hat{p}_2 + (n-r)\hat{p}_3) + o(1)\right) \right)^r \\ &\leq \left( \frac{e \ln n}{4 \ln \ln n} \exp(-2\theta_n \ln \ln n) \right)^r = o(1) \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}X_{r,r} &\leq \binom{n}{r}^2 (1 - \hat{p}_2)^{r^2} (1 - \hat{p}_3)^{2\binom{r}{2}r + r^2(n-2r)} \\ &\leq \left( \frac{en}{r} \exp\left(-\frac{r}{2}(\hat{p}_2 + (n-2r)\hat{p}_3) + o(1)\right) \right)^{2r} = o(1). \end{aligned}$$

Therefore with high probability  $X_r = 0$  which implies that with high probability

$$\alpha(G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)) \leq r.$$

Moreover with high probability  $X_{r,r} = 0$  thus with high probability for any  $S \subseteq \mathcal{V}$  such that  $r \leq |S|$  we have

$$N_3(S) \geq n - |S| - r.$$

(Otherwise  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  would contain  $\overline{K}_{r,r}$ .)

Therefore with high probability for any  $S \subseteq \mathcal{V}$  such that  $r \leq |S| \leq 2n/3$

$$N_3(S) \geq \min\{|S|, r\} \geq \min\{|S|, \alpha(G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3))\}.$$

Moreover with high probability for any  $S \subseteq \mathcal{V}$  such that  $r \leq |S| \leq n/4$  we have

$$N_3(S) \geq n - |S| - r \geq 2|S|.$$

This finishes the proof of (i) and (ii).

(iii) For any two disjoint sets  $S \subseteq \mathcal{V}$  and  $S' \subseteq \mathcal{V}$  we denote by  $e(S, S')$  the number of edges in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  with one end in  $S$  and one end in  $S'$  and by  $e(S)$  the number of edges in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  with both ends in  $S$ . We will bound the numbers  $e(S)$  and  $e(S, S')$  for  $S \subseteq \mathcal{V}$  and  $S' \subseteq \mathcal{V} \setminus S$  such that  $1 \leq |S| \leq n^\gamma$  and  $|S'| = O(|S|)$ . Let  $k$  be a positive integer,  $S \subseteq \mathcal{V}$ ,  $S' \subseteq \mathcal{V} \setminus S$ ,  $|S| = s$ ,  $1 \leq s \leq n^\gamma$  and  $|S'| = 2ks$ . A pair  $\{v, v'\}$  ( $v \in S$  and  $v' \in S'$ ) is an edge in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  if at least one of the three following events occurs:

- $\{v, v'\}$  is an edge in  $G_2(n, \hat{p}_2)$ ,
- there is a hyperedge  $\{v, v', v''\}$  with  $v'' \in \mathcal{V} \setminus (S \cup S')$  in  $H_3(n, \hat{p}_3)$  associated with  $G_3(n, \hat{p}_3)$ ,

- there is a hyperedge  $\{v, v', v''\}$  with  $v'' \in S \cup S'$  in  $H_3(n, \hat{p}_3)$  associated with  $G_3(n, \hat{p}_3)$ .

The union of the two first events occurs with probability at most  $\hat{p}_2 + (n - (2k + 1)s)\hat{p}_3 = O(\ln n/n)$  independently for all  $v \in S$  and  $v' \in S'$ . Moreover each hyperedge of  $H_3(n, \hat{p}_3)$  with all vertices in  $S \cup S'$  appears independently with probability  $\hat{p}_3 = O(\ln n/n^2)$  and generates two edges between  $S$  and  $S'$  in  $G_3(n, \hat{p}_3)$ . Therefore

$$\begin{aligned} & \Pr \left\{ \begin{array}{l} \exists_{S, S' \subseteq \mathcal{V}} e(S, S') \geq 4(k+1)|S| \\ |S'| = 2k|S| \\ 1 \leq |S| \leq n^\gamma \end{array} \right\} \\ & \leq \sum_{s=1}^{n^\gamma} \binom{n}{s} \binom{n-s}{2ks} \left[ \binom{2ks^2}{(2k+2)s} (\hat{p}_2 + (n - (2k+1)s)\hat{p}_3)^{(2k+2)s} \right. \\ & \quad \left. + \binom{\binom{s}{2}2ks + \binom{2ks}{2}s}{(k+1)s} \hat{p}_3^{(k+1)s} \right] \\ & \leq \sum_{s=1}^{n^\gamma} \left[ \left( O(1) \left(\frac{n}{s}\right)^{2k+1} \left(\frac{s \ln n}{n}\right)^{2k+2} \right)^s \right. \\ & \quad \left. + \left( O(1) \left(\frac{n}{s}\right)^{2k+1} s^{k+1} \left(\frac{\ln n}{n^2}\right)^{k+1} \right)^s \right] \\ & \leq \sum_{s=1}^{n^\gamma} \left[ \left( O(1) \frac{\ln^{2k+2} n}{n^{1-\gamma}} \right)^s + \left( O(1) \frac{\ln^{k+1} n}{n} \right)^s \right] \\ & = o(1). \end{aligned}$$

Similarly

$$\begin{aligned} & \Pr \left\{ \exists_{S \subseteq \mathcal{V}} \begin{array}{l} e(S) \geq 5|S| \\ 1 \leq |S| \leq n^\gamma \end{array} \right\} \\ & \leq \sum_{s=1}^{n^\gamma} \binom{n}{s} \left[ \binom{\binom{s}{2}}{2s} (\hat{p}_2 + (n-s)\hat{p}_3)^{2s} + \binom{\binom{s}{3}}{s} \hat{p}_3^s \right] \\ & \leq \sum_{s=1}^{n^\gamma} \left( O(1) \frac{\ln^2 n}{n^{1-\gamma}} \right)^s + \left( O(1) \frac{\ln n}{n^{1-\gamma}} \right)^s \\ & = o(1). \end{aligned}$$

Therefore with high probability for any set  $S$  ( $1 \leq |S| \leq n^\gamma$ ) of vertices of degree at least  $4k + 14$  in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  we have

$$|N_3(S)| \geq 2k|S|.$$

Otherwise for  $S' = N_3(S)$  there would be  $|S'| = |N_3(S)| \leq 2k|S|$  and  $e(S, N_3(S)) + 2e(S) \geq (4k + 14)|S|$ , i.e.  $e(S, N_3(S)) \geq 4(k+1)|S|$  or  $e(S) \geq 5|S|$ .

(iv) The probability that in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  there are two vertices of degree at most  $C$  at distance at most 4 is at most

$$\begin{aligned} & n^2 \sum_{l=0}^3 n^l (\hat{p}_2 + n\hat{p}_3)^{l+1} \\ & \cdot \left( \sum_{i=0}^C \sum_{j=0}^{\lfloor i/2 \rfloor} \binom{\binom{n-1-l}{2}}{j} \hat{p}_3^j \binom{n-1-l}{i-2j} \hat{p}_2^{i-2j} \right)^2 \\ & \cdot (1 - \hat{p}_2)^{n-1-i-l} (1 - \hat{p}_3)^{\binom{n-1-l}{2} - \binom{i+l}{2}} \\ & = O(1)n(\ln n)^{2C+4} e^{-2n(\hat{p}_2 + \frac{n}{2}\hat{p}_3)} = o(1). \end{aligned}$$

(v) Follows by Theorem 8.1 in [6]. □

**Lemma 6.2.** *Let  $k$  be a positive integer and  $\hat{p}_2 = \hat{p}_2(n) \in (0, 1)$ ,  $\hat{p}_3 = \hat{p}_3(n) \in (0, 1)$  be such that*

$$\hat{p}_2 + \frac{n}{2}\hat{p}_3 = \frac{\ln n + (k-1)\ln \ln n + c_n}{n},$$

(i) *If  $k = 1$  and  $c_n$  attains a limit (possibly  $\pm\infty$ ) then*

$$\Pr \{ \delta(G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)) \geq 1 \} \rightarrow f(c_n),$$

where  $f(\cdot)$  is defined by (2.5).

(ii) *If  $c_n \rightarrow \infty$  then*

$$\Pr \{ \delta(G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)) \geq k \} \rightarrow 1.$$

*Proof.* (i) Follows by Theorem 3.10 in [16].

(ii) Let  $X_t$  be the number of vertices of degree  $0 \leq t \leq k-1$  in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$ . Then

$$\begin{aligned} \mathbb{E}X_t & \leq n \binom{n-1}{t} \sum_{i=0}^t \binom{\binom{t}{2}}{i} \hat{p}_3^i \hat{p}_2^{\max\{t-2i, 0\}} (1 - \hat{p}_2)^{n-t-1} (1 - \hat{p}_3)^{\binom{n-1}{2} - \binom{t}{2}} \\ & = O(1)n(\ln n)^t \exp(-\ln n + (k-1)\ln \ln n + c_n) = o(1). \end{aligned}$$

Thus with high probability  $X_t = 0$  for all  $t \leq k-1$ . □

**Lemma 6.3.** *Let  $k$  be a positive integer and  $\hat{p}_2$  and  $\hat{p}_3$  be as in (6.1). Let  $\mathbb{G}(n)$  be a random graph such that*

$$G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) \preceq_{1-o(1)} \mathbb{G}(n) \tag{6.2}$$

and in the probability space of the coupling with high probability all vertices of degree at most  $4k+13$  in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  are at distance at least 5 in  $\mathbb{G}(n)$ . If  $\Pr \{ \delta(\mathbb{G}(n)) \geq k \}$  is bounded away from zero by a constant then

$$\Pr \{ \mathbb{G}(n) \in \mathcal{C}_k | \delta(\mathbb{G}(n)) \geq k \} \rightarrow 1, \tag{6.3}$$

$$\Pr \{ \mathbb{G}(2n) \in \mathcal{PM} | \delta(\mathbb{G}(2n)) \geq 1 \} \rightarrow 1. \tag{6.4}$$

In particular, if we substitute  $\mathbb{G}(n) = G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  then by Lemmas 6.1(iv), 6.2, and 6.3 we obtain the following result.

**Lemma 6.4.** *Let  $\hat{p}_2$  and  $\hat{p}_3$  fulfil (6.1).*

(i) *If  $\hat{p}_2 + \frac{n}{2}\hat{p}_3 = (\ln n + c_n)/n$  for  $c_n$  attaining a limit (possibly  $\pm\infty$ ) then*

$$\begin{aligned} \Pr \{G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) \in \mathcal{C}_1\} &\rightarrow f(c_n); \\ \Pr \{G_2(2n, \hat{p}_2(2n)) \cup G_3(2n, \hat{p}_3(2n)) \in \mathcal{PM}\} &\rightarrow f(c_{2n}), \end{aligned}$$

where  $f(\cdot)$  is defined by (2.5).

(ii) *If  $\hat{p}_2 + \frac{n}{2}\hat{p}_3 = (\ln n + (k-1)\ln \ln n + c_n)/n$  and  $c_n \rightarrow \infty$  then*

$$\Pr \{G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) \in \mathcal{C}_k\} \rightarrow 1.$$

*Proof of Lemma 6.3.* Denote by  $\mathbb{G}(n)_{\delta \geq k}$  a graph  $\mathbb{G}(n)$  under condition  $\delta(\mathbb{G}(n)) \geq k$ . Let  $\mathcal{A}$  be a graph property. In what follows we will use the fact that if  $\Pr \{\delta(\mathbb{G}(n)) \geq k\}$  is bounded away from zero and  $\mathbb{G}(n)$  has  $\mathcal{A}$  with high probability, then also  $\mathbb{G}(n)_{\delta \geq k}$  has property  $\mathcal{A}$  with high probability.

From now on we assume that  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  and  $\mathbb{G}(n)$  are defined on the same probability space existing by (6.2). We call a vertex  $v \in \mathcal{V}$  *small* if its degree in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  is at most  $4k+13$ . Otherwise we call  $v$  a *large* vertex. Let  $S \subseteq \mathcal{V}$  and  $|S| \leq n^\gamma$ . Denote by  $S^+$  and  $S^-$  the subset of large and small vertices in  $S$ , respectively. Then by Lemma 6.1(iii) with high probability

$$|N_{\mathbb{G}(n)_{\delta \geq k}}(S^+)| \geq |N_3(S^+)| \geq 2k|S^+|.$$

Moreover in  $\mathbb{G}(n)_{\delta \geq k}$  all vertices in  $S^-$  have degree at least  $k$  and with high probability are at distance at least 5. Therefore with high probability in  $\mathbb{G}(n)_{\delta \geq k}$  no two vertices in  $S^-$  are connected by an edge. Moreover they have no common neighbour. Therefore with high probability  $|N_{\mathbb{G}(n)_{\delta \geq k}}(S^-)| \geq k|S^-|$ . We are left with the question how much  $N_{\mathbb{G}(n)_{\delta \geq k}}(S^-)$  overlaps with  $N_{\mathbb{G}(n)_{\delta \geq k}}(S^+) \cup S^+$ . This will give us the bound on the size of  $N_{\mathbb{G}(n)_{\delta \geq k}}(S^+ \cup S^-)$ . If there was a vertex  $v$  in  $S^+$  and two vertices  $v'$  and  $v''$  in  $S^-$  such that  $v'$  is a neighbour of  $v$  or its neighbour and  $v''$  is a neighbour of  $v$  or its neighbour then  $v'$  and  $v''$  would be connected by a path of length at most 4. Therefore with high probability there are no such three vertices, i.e. for each  $v \in S^+$  there is at most one vertex  $v' \in S^-$  such that  $v'$  is a neighbour of  $v$  or of its neighbour. This implies that with high probability at most  $|S^+|$  vertices from  $S^-$  have neighbours in  $N_{\mathbb{G}(n)_{\delta \geq k}}(S^+) \cup S^+$ . Recall that each of the remaining vertices from  $S^-$  has  $k$  neighbours outside  $N_{\mathbb{G}(n)_{\delta \geq k}}(S^+) \cup S^+$  and their neighbour sets are disjoint. Thus with high probability

$$\forall S \subseteq \mathcal{V}, 1 \leq |S| \leq n^\gamma |N_{\mathbb{G}(n)_{\delta \geq k}}(S)| \geq |N_{\mathbb{G}(n)_{\delta \geq k}}(S^+)| + k \max\{|S^-| - |S^+|, 0\} \geq k|S|.$$

If we combine this with Lemma 6.1(i) and (ii) we get that with high probability

$$|N_{\mathbb{G}(n)_{\delta \geq k}}(S)| \geq k, \quad \text{for all } S \subseteq \mathcal{V}, 1 \leq |S| \leq \frac{n}{2}; \quad (6.5)$$

$$|N_{\mathbb{G}(n)_{\delta \geq 1}}(S)| \geq \min\{|S|, 4n \ln \ln n / \ln n\} \geq \min\{|S|, \alpha(\mathbb{G}(n)_{\delta \geq k})\}, \quad (6.6)$$

$$\text{for all } S \subseteq \mathcal{V}, 1 \leq |S| \leq \frac{2n}{3};$$

$$|N_{\mathbb{G}(n)_{\delta \geq 2}}(S)| \geq 2|S|, \quad \text{for all } S \subseteq \mathcal{V}, 1 \leq |S| \leq \frac{n}{4}. \quad (6.7)$$

Finally (6.3) follows immediately by (6.5). Moreover if (6.6) is fulfilled then  $\mathbb{G}(n)_{\delta \geq 1}$  has a perfect matching (for the proof see for example Lemma 1 in [5]). Therefore (6.4) follows. (6.7) will be used later to establish a threshold function for a Hamilton cycle.  $\square$

**Lemma 6.5.** *Let  $k$  be a positive integer,  $\theta_n$  and  $\theta'_n$  be sequences such that  $\liminf_{n \rightarrow \infty} \theta_n = \theta > \frac{1}{2}$ ,  $\liminf_{n \rightarrow \infty} \theta'_n = \theta' \geq \frac{1}{2}$ ,  $\theta, \theta' = O(1)$ , and*

$$\hat{p}_2 + \frac{n}{2}\hat{p}_3 = \frac{\theta_n \ln n}{n}, \quad \hat{p}_2 = \frac{\theta'_n \ln n}{n} \quad \text{and} \quad \hat{p}_4 = \frac{576}{n}.$$

Let moreover  $\mathbb{G}(n)$  be a random graph such that:

- (i)  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) \preceq_{1-o(1)} \mathbb{G}(n)$ ;
- (ii) with high probability  $\delta(\mathbb{G}(n)) \geq 2$ ;
- (iii) in a probability space existing by (i) all vertices of degree at most 21 in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  are at distance at least 5 in  $\mathbb{G}(n)$ .

Then

$$\Pr \{ \mathbb{G}(n) \cup G_2(n, \hat{p}_4) \in \mathcal{HC} \} \rightarrow 1.$$

In particular, by Lemmas 6.1(iv) and 6.2 we get the following simple corollary of Lemma 6.5.

**Lemma 6.6.** *Let  $\hat{p}_2$  and  $\hat{p}_3$  be as in Lemma 6.5. If moreover*

$$\hat{p}_2 + \frac{n}{2}\hat{p}_3 = (\ln n + \ln \ln n + c_n)/n \quad \text{where } c_n \rightarrow \infty$$

then

$$\Pr \{ G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) \in \mathcal{HC} \} \rightarrow 1.$$

*Proof of Lemma 6.5.* We will use the ideas from the proof of Theorem 8.9 [6]. Let  $t = 9n/\ln n$  and  $\hat{p}_{4,0} = 64 \ln n/n^2$ . Then  $t\hat{p}_{4,0} = \hat{p}_4$ . For any graph  $G$  let  $l(G)$  be the length of the longest path in  $G$  and  $l(G) = n$  if  $G$  has a Hamilton cycle. We say that  $G$  has property  $\mathcal{Q}$  if

$$G \text{ is connected} \quad \text{and} \quad |N_G(S)| \geq 2|S|, \quad \text{for all } S \subseteq \mathcal{V}, |S| \leq n/4.$$

In the proof we assume that  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  and  $\mathbb{G}(n)$  are defined on the probability space of the coupling existing by (i). Let  $\mathbb{G}_0 = \mathbb{G}(n)$  and

$$\mathbb{G}_i = \mathbb{G}_{i-1} \cup G_2(n, \hat{p}_{4,0}), \quad \text{for } 1 \leq i \leq t.$$

By Lemma 6.3, assumptions (i)–(iii), and (6.7) with high probability  $\mathbb{G}(n)$  has property  $\mathcal{Q}$ . Moreover by Lemma 6.1(v) and the coupling from assumption (i) with high probability

$$l(\mathbb{G}_0) \geq \left(1 - \frac{9 \ln 2}{\ln n}\right) n.$$

It may be shown (see (8.7) in [6]) that

$$\Pr \{l(\mathbb{G}_i) = n - t + i - 1 | l(\mathbb{G}_{i-1}) = n - t + i - 1 \text{ and } \mathbb{G}_{i-1} \text{ has } \mathcal{Q}\} \leq (1 - \hat{p}_{4,0})^{n^{2/32}} \leq n^{-2}.$$

Thus

$$\Pr \{l(\mathbb{G}_t) = n\} \geq 1 - \frac{t}{n^2} - o(1) = 1 - o(1).$$

Since

$$\mathbb{G}_t \preceq \mathbb{G}(n) \cup G_2(n, t\hat{p}_{4,0}),$$

this finishes the proof. □

## 7 Threshold functions

*Proof of Theorems 2.2–2.4.* First let  $S_1 = n(\ln n + c_n)$  and  $\hat{p}_2$  and  $\hat{p}_3$  be given by (2.2). Then by Lemma 5.1

$$\lim_{n \rightarrow \infty} \Pr \{\mathcal{G}(n, m, \bar{p}) \in \mathcal{C}_1\} \leq \lim_{n \rightarrow \infty} \Pr \{\delta(\mathcal{G}(n, m, \bar{p})) \geq 1\} = f(c_n).$$

Moreover for  $c_n = o(\ln n)$  by (2.2) and (4.7) we have

$$\hat{p}_2 + \frac{n}{2}\hat{p}_3 = \frac{\theta_n \ln n}{n} \text{ and } \hat{p}_2 = \frac{\theta'_n \ln n}{n} \text{ with } \theta_n \rightarrow 1 \text{ and } \theta'_n \rightarrow 1/2.$$

Therefore by Theorem 2.1 and Lemma 6.4(i)

$$\lim_{n \rightarrow \infty} \Pr \{\mathcal{G}(n, m, \bar{p}) \in \mathcal{C}_1\} \geq \lim_{n \rightarrow \infty} \Pr \{G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) \in \mathcal{C}_1\} = f(c_n)$$

In the case  $c_n = \Omega(\ln n)$  we may use the couplings described at the beginning of the proof of Lemma 5.1. This implies Theorem 2.2(i). Similarly Theorem 2.2(ii) follows by Theorem 2.1, Lemma 6.4(ii), and Lemma 5.2; Theorem 2.3 by Theorem 2.1, Lemma 6.4(i), and Lemma 5.1; Theorem 2.4 by Theorem 2.1, Lemma 6.6 and Lemma 5.2. □

In the following proofs we will assume that  $c_n = O(\ln n)$ . In the other cases the theorems follow by Lemma 5.1 or Theorem 2.1 (2.3) combined with known results concerning  $G_2(n, \hat{p})$ .

*Proof of Theorem 2.5.* Let  $c_n \rightarrow -\infty$ , then by Lemma 5.3

$$\Pr \{ \mathcal{G}(n, m, p) \in \mathcal{HC} \} \leq \Pr \{ \delta(\mathcal{G}(n, m, p)) \geq 2 \} \rightarrow 0.$$

Let now  $c_n \rightarrow \infty$ . We will consider two cases:

**CASE 1:**  $m = \Omega(n/\ln n)$ .

Let

$$m' = m \left( 1 + \frac{700}{\ln n} \right)^{-1} \quad \text{and} \quad m'' = \frac{700}{\ln n} m'.$$

Then for  $p$  given by (2.6)

$$m'p(1 - (1 - p)^{n-1}) = \ln n + \ln \max \left\{ 1, \frac{np \ln n}{e^{np} - 1} \right\} - 700 + o(1) + c_n$$

and

$$nm'p^2 = \begin{cases} O(\ln n) & \text{for } np = O(1), \\ \frac{n}{m'}(m'p)^2 = O(\ln^3 n) & \text{for } np \rightarrow \infty. \end{cases} \quad (7.1)$$

Let  $\omega \rightarrow \infty$  slowly enough. Define

$$\hat{p}_2 = \begin{cases} \frac{m'p(1 - (1 - p)^{n-1}) - 3m'p \left( \frac{1 - (1 - 2p)^n}{2np} - (1 - p)^{n-1} \right) - \omega \sqrt{\frac{\ln n}{n}}}{n}, \\ \quad \text{for } nm'p \left( \frac{1 - (1 - 2p)^n}{2np} - (1 - p)^{n-1} \right) \gg \sqrt{n \ln n}; \\ \frac{m'p(1 - (1 - p)^{n-1}) - \omega \sqrt{\frac{\ln n}{n}}}{n}, \quad \text{otherwise;} \end{cases}$$

$$\hat{p}_3 = \begin{cases} \frac{m'p \left( \frac{1 - (1 - 2p)^n}{2np} - (1 - p)^{n-1} \right) - \omega \sqrt{\frac{\ln n}{n}}}{\frac{n^2}{6}}, \\ \quad \text{for } nm'p \left( \frac{1 - (1 - 2p)^n}{2np} - (1 - p)^{n-1} \right) \gg \sqrt{n \ln n}; \\ 0, \quad \text{otherwise;} \end{cases}$$

$$\hat{p}_4 = \frac{m''p \left( 1 - \frac{1 - (1 - 2p)^n}{2np} \right) - \omega \sqrt{\frac{1}{n}}}{n}.$$

Moreover by (2.1) and (4.7), for large  $n$ ,

$$\hat{p}_4 = \frac{m''p \left( 1 - \frac{1 - (1 - 2p)^n}{2np} \right) - \omega \sqrt{\frac{1}{n}}}{n} = \frac{\frac{700}{\ln n} \left( 1 + \frac{700}{\ln n} \right)^{-1} S_2 - \omega \sqrt{n}}{n^2} = \frac{(1 + o(1)) 700 \cdot \frac{5}{6} \cdot S_1}{n^2 \ln n} \geq \frac{576}{n}.$$

Let now  $\mathbb{G}(n) = \mathcal{G}(n, m', p)$ . Then by Theorem 2.1

$$\begin{aligned} G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) &\leq_{1-o(1)} \mathcal{G}(n, m', p) = \mathbb{G}(n); \\ G_2(n, \hat{p}_4) &\leq_{1-o(1)} \mathcal{G}(n, m'', p). \end{aligned} \quad (7.2)$$

Moreover by Lemma 5.3 with high probability  $\delta(\mathbb{G}(n)) \geq 2$ . Therefore assumptions (i) and (ii) of Lemma 6.5 are satisfied.

We are left with proving that assumption (iii) is satisfied. Let  $C$  be a positive integer. We will show that in the probability space of the coupling (7.2) with high probability

any two vertices of degree at most  $C$  in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  are at distance at least 5 in  $\mathcal{G}(n, m', p) = \mathbb{G}(n)$ . Therefore we need to study the couplings described in the proof of Theorem 2.1. If in the definitions of  $S_1$ ,  $S_3$ ,  $M_2$  and  $M_3$  we replace  $m$  by  $m'$  then proceeding as in the proof of Theorem 2.1, we may show that

$$\begin{aligned}
& G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3) \\
&= \mathbb{G}_{*2} \left( n, \text{Po} \left( \binom{n}{2} \ln(1 - \hat{p}_2)^{-1} \right) \right) \cup \mathbb{G}_{*3} \left( n, \text{Po} \left( \binom{n}{3} \ln(1 - \hat{p}_3)^{-1} \right) \right) \\
&\stackrel{(i)}{\preceq} \mathbb{G}_{*2} \left( n, \text{Po} \left( \frac{S_1 - 3S_3 - 5\omega' \sqrt{S_1}}{2} \right) \right) \cup \mathbb{G}_{*3} \left( n, \text{Po} \left( S_3 - 2\omega' \sqrt{S_1} \right) \right) \\
&\stackrel{(ii)}{\preceq}_{1-o(1)} \mathbb{G}_{*2} \left( n, \frac{S_1 - 3S_3 - 4\omega' \sqrt{S_1}}{2} \right) \cup \mathbb{G}_{*3} \left( n, S_3 - \omega' \sqrt{S_1} \right) \\
&\stackrel{(iii)}{\preceq}_{1-o(1)} \mathbb{G}_{*2}(n, M_2) \cup \mathbb{G}_{*3}(n, M_3) = \bigcup_{1 \leq i \leq m'} (\mathbb{G}_{*2}(n, \frac{Y_i - 3Z_i}{2}) \cup \mathbb{G}_{*3}(n, Z_i)) \\
&\stackrel{(iv)}{\preceq} \mathcal{G}(n, m', p)
\end{aligned}$$

for some  $\omega' = \Theta(\omega)$ .

In the probability space of the above couplings, we will find a relation between the degree of a vertex  $v \in \mathcal{V}$  in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  and the number of these features  $w_i \in W(v)$  that are chosen by at least two vertices (i.e. the number of these features that contribute to at least one edge in  $\mathcal{G}(n, m', p)$ ). For any  $v \in \mathcal{V}$  in  $\mathcal{G}(n, m', p)$  let

$$W'(v) = \{w_i \in W(v) : |V'_i| \geq 2\},$$

where  $V'_i$  is defined in (4.3). We will find a relation between the degree of a vertex  $v$  in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  and  $|W'(v)|$  in  $\mathcal{G}(n, m', p)$ .

To this end we will need an insight into the construction of the intermediary couplings (i)–(iv). Recall that in the coupling (iv), the construction of  $\mathbb{G}_{*2}(n, M_2) \cup \mathbb{G}_{*3}(n, M_3)$  proceeds in  $m'$  rounds. In the  $i$ -th round  $\mathbb{G}_{*2}(n, (Y_i - 3Z_i)/2) \cup \mathbb{G}_{*3}(n, Z_i)$  is constructed in  $\frac{Y_i - 3Z_i}{2} + Z_i$  draws. In each draw a hyperedge of size 2 or 3 in  $\mathbb{H}_{*2}((Y_i - 3Z_i)/2) \cup \mathbb{H}_{*3}(Z_i)$  is chosen. Then  $V'_i$  (in  $\mathcal{G}(n, m', p)$ ) is formed from non isolated vertices of  $\mathbb{H}_{*2}((Y_i - 3Z_i)/2) \cup \mathbb{H}_{*3}(Z_i)$  and possibly some extra vertices. The number of extra vertices equals  $Y_i$  minus the number of non isolated vertices in  $\mathbb{H}_{*2}((Y_i - 3Z_i)/2) \cup \mathbb{H}_{*3}(Z_i)$  (i.e. also in  $\mathbb{G}_{*2}(n, (Y_i - 3Z_i)/2) \cup \mathbb{G}_{*3}(n, Z_i)$ ). We will show that with high probability for each  $v \in \mathcal{V}$  there is at most one  $w_i \in W'(v)$  such that  $v$  is isolated in  $\mathbb{H}_{*2}((Y_i - 3Z_i)/2) \cup \mathbb{H}_{*3}(Z_i)$ . This will prove that with high probability for each  $v \in \mathcal{V}$  the number of draws in which a hyperedge containing  $v$  is chosen is at least  $|W'(v)| - 1$ .

Given  $v \in \mathcal{V}$  and  $1 \leq i \leq m'$ , denote by  $\mathcal{A}_{v,i,1}$  the event that  $v$  is an isolated vertex in  $\mathbb{G}_{*2}(n, (Y_i - 3Z_i)/2) \cup \mathbb{G}_{*3}(n, Z_i)$  and by  $\mathcal{A}_{v,i,2}$  the event that  $v \in V'_i$ . We will use these events to compare the number of hyperedges in  $\bigcup_i \mathbb{H}_{*2}((Y_i - 3Z_i)/2) \cup \mathbb{H}_{*3}(Z_i)$  containing  $v$  with the number of features  $w_i$  such that  $v \in V'_i$  (i.e. with the number  $|W'(v)|$ ). Denote by  $\mathcal{A}_v$  the event that while constructing  $\mathbb{G}_{*2}(n, M_2) \cup \mathbb{G}_{*3}(n, M_3)$  there are less than



$|W'(v)| - 1$  draws in which a hyperedge containing  $v$  is chosen. Then

$$\mathcal{A}_v \subseteq \bigcup_{i \neq j} (\mathcal{A}_{v,i,1} \cap \mathcal{A}_{v,i,2}) \cap (\mathcal{A}_{v,j,1} \cap \mathcal{A}_{v,j,2}).$$

For  $i \neq j$ ,  $(\mathcal{A}_{v,i,1} \cap \mathcal{A}_{v,i,2})$  and  $(\mathcal{A}_{v,j,1} \cap \mathcal{A}_{v,j,2})$  are independent and

$$\Pr \{ \mathcal{A}_{v,i,1} \cup \mathcal{A}_{v,i,2} \} = 1,$$

thus

$$\begin{aligned} \Pr \{ \mathcal{A}_v \} &\leq \binom{m'}{2} (\Pr \{ \mathcal{A}_{v,1,1} \cap \mathcal{A}_{v,1,2} \})^2 \\ &= \binom{m'}{2} (\Pr \{ \mathcal{A}_{v,1,1} \} + \Pr \{ \mathcal{A}_{v,1,2} \} - 1)^2 \\ &\leq \binom{m'}{2} \left( (1-p)^n + np(1-p)^{n-1} \right. \\ &\quad \left. + \sum_{y=2}^{n-1} \binom{n}{y} p^y (1-p)^{n-y} \left( 1 - \frac{1}{n} \right)^y + p^n \cdot 0 \right. \\ &\quad \left. + \sum_{y=2}^n \binom{n}{y} p^y (1-p)^{n-y} \frac{y}{n} - 1 \right)^2 \\ &= \binom{m'}{2} \left( \sum_{y=0}^n \binom{n}{y} p^y (1-p)^{n-y} \left( 1 - \frac{1}{n} \right)^y - p^n \right. \\ &\quad \left. + \sum_{y=1}^n \binom{n}{y} p^y (1-p)^{n-y} \frac{y}{n} - 1 \right)^2 \\ &\leq \binom{m'}{2} \left( \left( 1 - \frac{p}{n} \right)^n + p - 1 \right)^2 \leq (m')^2 p^4. \end{aligned}$$

Therefore by (7.1)

$$\Pr \left\{ \bigcup_{v \in \mathcal{V}} \mathcal{A}_v \right\} \leq n(m')^2 p^4 = O\left(\frac{\ln^6 n}{n}\right) = o(1). \quad (7.3)$$

Thus in the probability space of the couplings, with high probability for all  $v \in \mathcal{V}$  in the construction of  $\mathbb{G}_{*2}(n, M_2) \cup \mathbb{G}_{*3}(n, M_3)$  the number of draws in which a hyperedge containing  $v$  is chosen is at least  $|W'(v)| - 1$ .

Now we will find a relation between the number of draws in the construction of  $\mathbb{G}_{*2}(n, M_2) \cup \mathbb{G}_{*3}(n, M_3)$  and the number of draws in the construction of

$$G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$$

$$= \mathbb{G}_{*2} \left( n, \text{Po} \left( \binom{n}{2} \ln(1 - \hat{p}_2)^{-1} \right) \right) \cup \mathbb{G}_{*3} \left( n, \text{Po} \left( \binom{n}{3} \ln(1 - \hat{p}_3)^{-1} \right) \right).$$

Note that in the couplings (i)–(iii) in order to construct a graph  $\mathbb{G}_{*2}(n, M_2) \cup \mathbb{G}_{*3}(n, M_3)$  from  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  one performs some additional draws of hyperedges in auxiliary hypergraphs  $\mathbb{H}_{*2}(n, \cdot) \cup \mathbb{H}_{*3}(n, \cdot)$ . By the sharp concentration of the Poisson distribution and (4.10) with high probability the number of additional draws is at most  $K\omega\sqrt{n \ln n}$  for some constant  $K$ . Under the condition that the number of additional draws is at most  $K\omega\sqrt{n \ln n}$ , probability that a hyperedge containing  $v$  is chosen in at least 3 of the additional draws is at most

$$\binom{K\omega\sqrt{n \ln n}}{3} \left( \frac{2}{n} \right)^3 = o(n^{-1}).$$

Therefore, by the union bound, with high probability in additional draws for each vertex  $v \in \mathcal{V}$  a hyperedge containing  $v$  is chosen at most 2 times. Combining this with (7.3) we obtain that with high probability for all  $v \in \mathcal{V}$  in the construction of  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  the number of draws in which a hyperedge containing  $v$  is chosen is at least  $|W'(v)| - 3$ .

Now we will show the relation between the number of draws of hyperedges and the number of edges incident to  $v$  in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$ . The difference between the numbers is due to the fact that hyperedges containing given pair of vertices (edge) might be drawn several times in the hypergraph  $\mathbb{H}_{*2}(n, \cdot) \cup \mathbb{H}_{*3}(n, \cdot)$  but there is only one edge in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$ . Therefore the number of draws might exceed the number of edges. Note that the number of draws made in the construction of  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  is a random variable with the expected value at most  $S_1 = O(n \ln n)$ . Therefore by Markov's inequality with high probability the number of draws made in the construction of  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  is at most  $n \ln^2 n$ . In addition the probability that a pair  $\{v, v'\}$  is contained in a hyperedge chosen in a given draw is at most  $6/n(n-1)$ . Therefore the probability that there exists an edge in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  drawn at least 3 times during the construction is at most

$$\binom{n}{2} \binom{n \ln^2 n}{3} \left( \frac{6}{n(n-1)} \right)^3 + o(1) = o(1).$$

In conclusion, with high probability for all  $v \in \mathcal{V}$  the number of edges incident to  $v$  in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  is at least half of the number of drawn hyperedges containing  $v$ , which is with high probability at least  $(|W'(v)| - 3)/2$ . Thus in the probability space of the couplings with high probability for all  $v \in \mathcal{V}$  of degree at most  $C$  in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  we have  $|W'(v)| \leq 2C + 3$ .

Finally, the probability that there are two vertices  $v, v'$  such that  $|W'(v)| \leq 2C + 3$  and  $|W'(v')| \leq 2C + 3$  connected by a path of length at most 4 in  $\mathcal{G}(n, m', p)$  is at most

$$\begin{aligned} & n^2 \sum_{t=1}^4 (m')^t n^{t-1} p^{2t} \cdot \left( \sum_{l=0}^{2C+3} \binom{m' - t}{l} (p - p(1-p)^{n-1})^l [(1-p) + p(1-p)^{n-1}]^{m' - O(1)} \right)^2 \\ & = O(1)n(\ln n)^{O(1)} \exp(-2m'p(1 - (1-p)^{n-1})) = o(1). \end{aligned}$$

This shows that assumption (iii) of Lemma 6.5 is fulfilled.

Therefore by Lemma 6.5 with high probability

$$\mathbb{G}(n) \cup G_2 \left( n, \frac{576}{n} \right) \in \mathcal{HC}.$$

Thus in the case  $m = \Omega(n/\ln n)$  the theorem follows by (7.2) and a straightforward coupling of random intersection graphs.

$$\mathbb{G}(n) \cup G_2 \left( n, \frac{576}{n} \right) \preceq_{1-o(1)} \mathcal{G}(n, m', p) \cup \mathcal{G}(n, m'', p) \preceq \mathcal{G}(n, m, p).$$

**CASE 2:**  $m = o(n/\ln n)$ .

Note that in this case (2.6) and  $c_n = O(\ln n)$  imply that

$$p = \frac{\ln n + c'_n}{m}$$

for some  $c'_n \rightarrow \infty$ . Let  $m'$  be such that  $m$  divides  $m'$  and  $m' \sim n/\ln n$ . Take an instance of  $\mathcal{G}(n, m', p')$  with the set of features  $\mathcal{W}'$  of size  $m'$  and

$$p' = \frac{\ln n + c'_n}{m'}.$$

Divide  $\mathcal{W}'$  into  $m$  groups of  $m'/m$  features. Denote by  $A_i$  the set of vertices which have chosen features from the  $i$ -th,  $1 \leq i \leq m$ , group in  $\mathcal{G}(n, m', p')$ .  $|A_i|$  has binomial distribution  $\text{Bin}(n, p'')$  for some  $p'' \leq \frac{m'}{m} p' = p$ . Now take an instance of  $\mathcal{G}(n, m', p')$  and construct the sets  $V_i$ ,  $1 \leq i \leq m$ , in  $\mathcal{G}(n, m, p)$  by taking  $A_i$  and additionally adding to  $V_i$  each vertex from  $\mathcal{V} \setminus A_i$  independently with probability  $(p - p'')/(1 - p'')$ . This coupling implies

$$\mathcal{G}(n, m', p') \preceq \mathcal{G}(n, m, p).$$

From the considerations concerning the first case we have that with high probability  $\mathcal{G}(n, m', p') \in \mathcal{HC}$ . Therefore also  $\mathcal{G}(n, m, p) \in \mathcal{HC}$  with high probability.  $\square$

*Proof of Theorem 2.6.* The technique of the proof is analogous to the technique of the proof of Theorem 2.5. We consider two cases  $m = \Omega(n/\ln n)$  and  $m = o(n/\ln n)$ . In the first case the proof relies on Lemma 5.3 and Lemma 6.3. Moreover we have to use the fact shown in the proof of Theorem 2.5, that in the coupling constructed in the proof of Theorem 2.1 with high probability the vertices of degree bounded by a constant in  $G_2(n, \hat{p}_2) \cup G_3(n, \hat{p}_3)$  are at distance at least 5 in  $\mathcal{G}(n, m, p)$ . The proof of the case  $m = o(n/\ln n)$  is the same as in the proof of Theorem 2.5.  $\square$

## Appendix

### A Proof of (4.7)

For  $0 < p_i < 1/2$ , using the fact that  $\mathbb{I}_{\{X_i \text{ odd}\}} = (1 - (-1)^{X_i})/2$  and the properties of the binomial distribution  $\text{Bin}(n, p_i)$  we get

$$\mathbb{E}Y_i = \mathbb{E}Y_i \mathbb{I}_{\{Y_i \text{ odd}\}} + \mathbb{E}Y_i \mathbb{I}_{\{Y_i \text{ even}\}}$$

$$\begin{aligned}
&= (\mathbb{E}X_i\mathbb{I}_{\{X_i, \text{odd}\}} - \mathbb{E}\mathbb{I}_{\{X_i=1\}}) + \mathbb{E}X_i\mathbb{I}_{\{X_i, \text{even}\}} \\
&= \frac{np}{2} (1 + (1 - 2p)^{n-1} - 2(1 - p)^{n-1}) + \frac{np}{2} (1 - (1 - 2p)^{n-1}) \\
&\geq 2 \cdot \frac{np}{2} (1 + (1 - 2p)^{n-1} - 2(1 - p)^{n-1}) \\
&\geq 2\mathbb{E}Y_i\mathbb{I}_{\{Y_i, \text{odd}\}} \geq 2 \cdot 3\mathbb{E}\mathbb{I}_{\{Y_i, \text{odd}\}} = 6\mathbb{E}Z_i.
\end{aligned}$$

Moreover for  $1/2 \leq p_i < 1$  and  $n \geq 14$

$$\mathbb{E}Y_i = np(1 - (1 - p)^{n-1}) \geq 6 \geq 6\mathbb{E}Z_i.$$

## B Detailed proofs of Lemmas 5.2 and 5.3

*Proof of Lemma 5.2 (i).* Let

$$S_1 = n(\ln n + (k - 1) \ln(\max\{1, a_n \ln n\}) + c_n), \text{ where } a_n = \frac{S_{1,2}}{S_1} \text{ and } c_n \rightarrow -\infty.$$

If  $a_n = O(1/\ln n)$  then  $S_1 = n(\ln n + O(1) + c_n)$ . On the other hand, if  $c_n = \Omega(\ln n)$  then  $S_1 = n(\ln n + (1 + o(1))c_n)$  (because  $a_n \leq 1$ ). In both cases by Lemma 5.1

$$\Pr\{\delta(\mathcal{G}(n, m, \bar{p})) \geq k\} \leq \Pr\{\delta(\mathcal{G}(n, m, \bar{p})) \geq 1\} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Therefore we are left with the case  $a_n = S_{1,2}/S_1 \gg 1/\ln n$  and  $c_n = o(\ln n)$ . Then  $S_{1,2} \rightarrow \infty$ . We use the division

$$\mathcal{G}(n, m, \bar{p}) = \mathcal{G}_2(n, m, \bar{p}) \cup \mathcal{G}_{\neq 2}(n, m, \bar{p}).$$

Using ideas similar to those used in the proof of Theorem 2.1 (see (4.11) and the discussion below it) we may show that

$$\begin{aligned}
\mathcal{G}_2(n, m, \bar{p}) \cup \mathcal{G}_{\neq 2}(n, m, \bar{p}) &\preceq_{1-o(1)}^{(*)} \mathbb{G}_{*2} \left( n, \frac{S_{1,2} + \omega\sqrt{S_{1,2}}}{2} \right) \cup \mathcal{G}_{\neq 2}(n, m, \bar{p}) \\
&\preceq_{1-o(1)}^{(**)} G_2(n, \hat{p}_{2+}) \cup \mathcal{G}_{\neq 2}(n, m, \bar{p}), \quad (\text{B.1})
\end{aligned}$$

where  $\hat{p}_{2+} = (S_{1,2} + 2\omega\sqrt{S_{1,2}})/2 \binom{n}{2}$ ,  $\mathbb{G}_{*2}(n, (S_{1,2} + \omega\sqrt{S_{1,2}})/t)$  and  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  are independent, and also  $G_2(n, \hat{p}_{2+})$  and  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  are independent.

We will first present a construction of coupling (\*). Let  $Q_2$  be defined by (5.4). Then  $\mathcal{G}_2(n, m, \bar{p}) = \mathbb{G}_{*2}(n, Q_2)$ . Set  $(S_{1,2} + \omega\sqrt{S_{1,2}})/2 = q_+$ . Let  $Q_2 = q$ . If  $q \leq q_+$  then, in the coupling, first we choose independently with repetition  $q$  edges uniformly at random from all 2 element subsets of  $\mathcal{V}$  and add them to the edge sets of  $\mathbb{G}_{*2}(n, Q_2)$  and  $\mathbb{G}_{*2}(n, q_+)$  (if edges repeat we add only one of them). Then we choose independently with repetition  $q_+ - q$  additional edges uniformly at random from all 2 element subsets of  $\mathcal{V}$ . These additional edges are added only to the edge set of  $\mathbb{G}_{*2}(n, q_+)$ . If  $q > q_+$ , similarly, we first

choose with repetition  $q_+$  edges and add them to the both graphs. Then we choose  $q - q_+$  additional edges and add them only to  $\mathbb{G}_{*2}(n, Q_2)$ . By (5.5) with high probability  $Q_2 \leq q_+$  thus the coupling described above implies  $\mathcal{G}_2(n, m, \bar{p}) = \mathbb{G}_{*2}(n, Q_2) \preceq_{1-o(1)} \mathbb{G}_{*2}(n, q_+)$ . Moreover  $q_+$  is a number (not a random variable) and choices of edges in  $\mathbb{G}_{*2}(n, q_+)$  are independent of choices of edges in  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$ . Therefore  $\mathbb{G}_{*2}(n, q_+)$  and  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  are independent. The coupling (\*\*\*) is analogous to the one constructed in (4.13). Namely, we use the coupling  $\mathbb{G}_{*2}(n, q_+) \preceq_{1-o(1)} \mathbb{G}_{*2}(n, \text{Po}((S_{1,2} + 2\omega\sqrt{S_{1,2}})/2)) \preceq G_2(n, \hat{p}_{2+})$  (the last coupling follows by (4.1) and inequality  $1 - e^{-x} \leq x$ ).

Now we will study properties of  $G_2(n, \hat{p}_{2+}) \cup \mathcal{G}_{\neq 2}(n, m, \bar{p})$ . We will show that with high probability there exists a vertex which is isolated in  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  and has degree at most  $k - 1$  in  $G_2(n, \hat{p}_{2+})$ . The construction of  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  may be coupled with the vertex collector process described in the proof of Lemma 5.1. This coupling proceeds in the same way as the coupling of the construction of  $\mathcal{G}(n, m, \bar{p})$  and the vertex collector process, with one change: one omits the phases with  $Y_i = 2$ . By (5.1) and (5.7), in the coupling with high probability the construction of  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  is finished before

$$T_{2+} = S_1 - S_{1,2} + 2\omega\sqrt{S_1} + \frac{S_1}{\omega \ln n}$$

draws, for  $\omega \rightarrow \infty$ . Moreover, in the coupling, every vertex which has not been collected by the end of the process, is isolated in  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$ . Therefore we need to show that with high probability there exists a vertex of degree  $k - 1$  in  $G_2(n, \hat{p}_{2+})$  and is not collected in  $T_{2+}$  draws of the vertex collector process independent of  $G_2(n, \hat{p}_{2+})$ .

Take any probability space on which we define the vertex collector process on  $\mathcal{V}$  and  $G_2(n, \hat{p}_{2+})$ , in such a manner that they are independent. Let  $X_+$  be a random variable counting vertices which have not been collected during the vertex collector process in  $T_{2+}$  draws and have degree  $k - 1$  in  $G_2(n, \hat{p}_{2+})$ . If

$$S_1 = n(\ln n + (k - 1) \ln(a_n \ln n) + c_n)$$

then for  $\omega$  tending to infinity slowly enough

$$\begin{aligned} \frac{1}{n}T_{2+} + (n - k)\hat{p}_{2+} &= \frac{1}{n} \left( S_1 + O \left( \omega\sqrt{S_1} + \frac{S_1}{\omega \ln n} \right) \right) \\ &= \ln n + (k - 1) \ln(a_n \ln n) + c_n + o(1) \end{aligned}$$

and

$$\hat{p}_{2+} \sim \frac{S_{1,2}}{n^2} = a_n \frac{S_1}{n^2} \sim \frac{a_n \ln n}{n}.$$

Therefore

$$\begin{aligned} \mathbb{E}X_+ &= n \left( 1 - \frac{1}{n} \right)^{T_{2+}} \binom{n-1}{k-1} \hat{p}_{2+}^{k-1} (1 - \hat{p}_{2+})^{n-k} \\ &\sim n \frac{1}{(k-1)!} (a_n \ln n)^{k-1} \exp(-\ln n - (k-1) \ln(a_n \ln n) - c_n) \end{aligned}$$

$$\sim \frac{e^{-c_n}}{(k-1)!}$$

and (as  $n\hat{p}_{2+} \sim a_n \ln n \rightarrow \infty$ )

$$\begin{aligned} \mathbb{E}X_+(X_+ - 1) &\sim n^2 \left(1 - \frac{2}{n}\right)^{T_{2+}} (1 - \hat{p}_{2+}) \left(\binom{n-2}{k-1} \hat{p}_{2+}^{k-1} (1 - \hat{p}_{2+})^{n-k-1}\right)^2 \\ &\quad + n^2 \left(1 - \frac{2}{n}\right)^{T_{2+}} \hat{p}_{2+} \left(\binom{n-2}{k-2} \hat{p}_{2+}^{k-2} (1 - \hat{p}_{2+})^{n-k}\right)^2 \\ &\sim \left(\frac{e^{-c_n}}{(k-1)!}\right)^2. \end{aligned}$$

Thus by the second moment method with high probability  $X_+ > 0$  as  $c_n \rightarrow -\infty$ . By (5.1), (5.7), and the coupling of  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  with the vertex collector process we have that with high probability there is a vertex which is isolated in  $\mathcal{G}_{\neq 2}(n, m, \bar{p})$  and has degree at most  $k-1$  in  $G_2(n, \hat{p}_{2+})$ . Thus by coupling (B.1) with high probability there is a vertex of degree at most  $k-1$  in  $\mathcal{G}(n, m, \bar{p})$ , i.e.

$$\lim_{n \rightarrow \infty} \Pr \{\delta(\mathcal{G}(n, m, \bar{p})) \geq k\} = 0. \quad \square$$

*Proof of Lemma 5.2 (ii).* Let

$$S_1 - \sum_{t=3}^k S_{1,t} = n \left( \ln n + (k-1) \ln \left( \max \left\{ 1, \left( \frac{S_{1,2}}{S_1} \ln n \right) \right\} \right) + c_n \right) \text{ with } c_n \rightarrow \infty.$$

If  $S_1 \gg n \ln n$  then by Theorem 2.1 and the classical results concerning  $G_2(n, \hat{p})$  with  $\hat{p} \gg \ln n/n$  we have

$$1 = \lim_{n \rightarrow \infty} \Pr \{\delta(G_2(n, \hat{p})) \geq k\} \leq \lim_{n \rightarrow \infty} \Pr \{\delta(\mathcal{G}(n, m, \bar{p})) \geq k\}.$$

Now let  $S_1 = O(n \ln n)$ . If  $S_{1,2}/S_1 = O(1/\ln n)$  and  $S_1 = O(n \ln n)$  then  $S_1 - \sum_{t=3}^k S_{1,t} = n(\ln n + O(1) + c_n)$ . In this case we use the same reasoning as in the proof of Lemma 5.1. We construct a coupling of the vertex collector process and the construction of  $\mathcal{G}_{\geq k+1}(n, m, \bar{p})$ . By (5.1) and (5.6) the construction of  $\mathcal{G}_{\geq k+1}(n, m, \bar{p})$  does not finish before

$$T_- = S_1 - \sum_{t=3}^k S_{1,t} - k\sqrt{S_1} = n(\ln n + O(1) + c_n)$$

draws (recall that  $\sqrt{S_1} = O(\sqrt{n \ln n})$ ). Therefore by the coupling and the classical results on the vertex collector process [10] with high probability there is no isolated vertex in  $\mathcal{G}_{\geq k+1}(n, m, \bar{p})$ . Thus

$$\lim_{n \rightarrow \infty} \Pr \{\delta(\mathcal{G}(n, m, \bar{p})) \geq k\} = 1.$$

We are left with the case  $S_1 = O(n \ln n)$  (which implies  $c_n = O(\ln n)$ ) and  $a_n = S_{1,2}/S_1 \gg 1/\ln n$  (which implies  $S_{1,2} \rightarrow \infty$  and  $a_n \ln n \rightarrow \infty$ ). The proof uses the same techniques as the proof of Lemma 5.2(i). As in (B.1) we may construct couplings

$$G_2(n, \hat{p}_{2-}) \cup \mathcal{G}_{\geq k+1}(n, m, \bar{p}) \preceq_{1-o(1)} \mathcal{G}_2(n, m, \bar{p}) \cup \mathcal{G}_{\geq k+1}(n, m, \bar{p}) \preceq \mathcal{G}(n, m, \bar{p}), \quad (\text{B.2})$$

where  $\hat{p}_{2-} = (S_{1,2} - 2\omega\sqrt{S_{1,2}})/2 \binom{n}{2}$  and graphs  $G_2(n, \hat{p}_{2-})$  and  $\mathcal{G}_{\geq k+1}(n, m, \bar{p})$  are independent. Moreover we may couple a construction of  $\mathcal{G}_{\geq k+1}(n, m, \bar{p})$  with the vertex collector process in such a manner that with high probability  $\mathcal{G}_{\geq k+1}(n, m, \bar{p})$  is constructed in at least

$$T_- = S_1 - \sum_{t=3}^k S_{1,t} - k\sqrt{S_1} = n(\ln n + \ln(\max\{1, a_n \ln n\}) + c_n + o(1))$$

draws.

Consider any probability space on which we define the vertex collector process on  $\mathcal{V}$  and  $G_2(n, \hat{p}_{2-})$ , in such a manner that they are independent. Let  $X_-$  be a random variable defined on this probability space equal to the number of vertices which have not been collected during the vertex collector process with  $T_-$  draws and have degree at most  $k-1$  in  $G_2(n, \hat{p}_{2-})$ . If  $S_1 - \sum_{t=3}^k S_{1,t} = n(\ln n + (k-1)\ln(a_n \ln n) + c_n)$  then for  $\omega$  tending to infinity slowly enough

$$\begin{aligned} \frac{1}{n}T_- + (n-k)\hat{p}_{2-} &\geq \frac{1}{n} \left( S_1 - \sum_{t=3}^{k-1} S_{1,t} + O\left(\omega\sqrt{S_1} + S_1^2 n^{-2}\right) \right) \\ &= \ln n + (k-1)\ln(a_n \ln n) + c_n + o(1). \end{aligned}$$

Moreover

$$\hat{p}_{2-} \sim \frac{S_{1,2}}{n^2} = a_n \frac{S_1}{n^2} = \Theta\left(\frac{a_n \ln n}{n}\right)$$

and  $n\hat{p}_{2-} = \Theta(a_n \ln n)$  tends to infinity. Thus

$$\begin{aligned} \mathbb{E}X_- &= n \left(1 - \frac{1}{n}\right)^{T_-} \left( \sum_{i=0}^{k-1} \binom{n-1}{i} \hat{p}_{2-}^i (1 - \hat{p}_{2-})^{k-1-i} \right) (1 - \hat{p}_{2-})^{n-k} \\ &= O(1)n \frac{1}{(k-1)!} (a_n \ln n)^{k-1} \exp(-\ln n - (k-1)\ln(a_n \ln n) - c_n) \\ &= O(1) \frac{e^{-c_n}}{(k-1)!} \end{aligned}$$

Therefore with high probability  $X_- = 0$ , i.e. with high probability each vertex is collected in  $T_-$  draws or has degree at least  $k$  in  $G_2(n, \hat{p}_{2-})$ . Note that if each vertex is non-isolated in  $\mathcal{G}_{k+1}(n, m, \bar{p})$  or has degree at least  $k$  in  $\mathcal{G}_2(n, m, \bar{p})$ , then  $\delta(\mathcal{G}(n, m, \bar{p})) \geq k$ . Therefore coupling (B.2) implies that with high probability  $\delta(\mathcal{G}(n, m, \bar{p})) \geq k$ .  $\square$

Before we proceed with the proof of Lemma 5.3 we start with discussion on the structure of  $\mathcal{G}(n, m, p)$ . From now on for  $\bar{p} = (p, \dots, p)$  we replace  $\bar{p}$  by  $p$  in notations  $\mathcal{G}_t(n, m, \bar{p})$ ,  $\mathcal{G}_{\geq k+1}(n, m, \bar{p})$ ,  $\mathcal{G}_{\neq t}(n, m, \bar{p})$ .

Note that if  $np \gg \ln m$  then with high probability  $\bigcup_{t=2}^k \mathcal{G}_t(n, m, p)$  is empty. Therefore the number of vertices of degree at most  $k-1$  is the number of isolated vertices in  $\mathcal{G}_{\geq k+1}(n, m, p)$  (i.e. it might be approximated by the number of non-collected vertices in the vertex collector process). On the other hand if  $np = o(m^{-1/3})$  then with high probability  $\mathcal{G}(n, m, \bar{p}) = \mathcal{G}_2(n, m, p)$  and the number of vertices of degree at most  $k-1$  may be counted in the same manner as in a graph with independent edges. The proof of Lemma 5.3 provides the discussion about what happens for  $np$  in between.

*Proof of Lemma 5.3.* Let

$$a_n = a_n(p) = (np)^{k-1} \left( \left( \frac{\ln n}{e^{np} - 1} \right)^{k-1} + \frac{\ln n}{e^{np} - 1} \right).$$

First we will prove that we may restrict our attention to the case  $S_1 \sim n \ln n$  (i.e.  $c_n = o(\ln n)$ ). Recall that  $p(1 - (1-p)^{n-1}) = S_1/nm$ . Moreover, note that  $0 \leq \ln \max\{1, a_n\} \leq (k-1) \ln \ln n + O(1) = o(\ln n)$ . Therefore, by monotonicity and continuity of  $f(x) = x(1 - (1-x)^{n-1})$ , for any  $p$  with  $c_n = \Omega(\ln n)$ ,  $c_n \rightarrow -\infty$ , there exists  $p'$ ,  $p \leq p'$ , with  $c'_n = o(\ln n)$ ,  $c'_n \rightarrow -\infty$ , such that

$$p(1 - (1-p)^{n-1}) = \frac{\ln n + \ln \max\{1, a_n(p)\} + c_n}{m} \leq \frac{\ln n + \ln \max\{1, a_n(p')\} + c'_n}{m} = p'(1 - (1-p')^{n-1}).$$

Thus

$$\Pr \{ \mathcal{G}(n, m, p) \in \mathcal{C}_k \} \leq \Pr \{ \mathcal{G}(n, m, p') \in \mathcal{C}_k \}.$$

An analogous statement is true for  $p$  with  $c_n \rightarrow \infty$  and  $p' \leq p$  with  $c'_n = o(\ln n)$  and  $c'_n \rightarrow \infty$ . From now on we assume that  $c_n = o(\ln n)$ . Note that if  $m \gg \ln^2 n$  then  $p = O(p(1 - (1-p)^{n-1})) = O(\ln n/m) = o(1/\ln n)$ . This will allow us to use Lemmas 5.1 and 5.2.

We will also use some relations between  $a_n$  and values  $S_{1,t}$ ,  $t = 2, k$ . By (2.1)

$$S_{1,t} \sim m \frac{(np)^t}{(t-1)!} (1-p)^{n-t} = ((t-1)!)^{-1} \frac{(np)^{t-1} (1-p)^{n-t}}{1 - (1-p)^{n-1}} S_1, \quad \text{for } t = 2, \dots, k,$$

where  $S_1$  and  $S_{1,t}$  are defined by (2.1). Therefore for  $p = o(1)$

$$S_{1,t} \sim ((t-1)!)^{-1} \frac{(np)^{t-1}}{(1-p)^{-n} - 1} S_1, \quad \text{for } t = 2, \dots, k. \tag{B.3}$$

Moreover if  $np^2 = o(1)$  then

$$\left( \frac{S_{1,2}}{S_1} \ln n \right)^{k-1} + (k-1)! \frac{S_{1,k}}{S_1} \ln n \sim a_n. \tag{B.4}$$



We will divide the proof into cases depending on the value  $a_n$  (or equivalently  $np$ ).

**CASE 1.**  $a_n = O(1)$  (i.e.  $np \geq \ln \ln n + (k-1) \ln \ln \ln n + d$ , for some  $d \in \mathbb{R}$ ).

In this case  $S_{1,t}/S_1 = O(1/\ln n)$  for  $2 \leq t \leq k$ . Moreover  $\ln(\max\{1, a_n\}) = O(1)$ . Therefore  $S_1 = n(\ln n + O(1) + c_n)$  and  $S_1 - \sum_{t=3}^k S_{1,t} = n(\ln n + O(1) + c_n)$ . Thus the result follows by Lemma 5.2.

**CASE 2.**  $a_n \rightarrow \infty$  (i.e.  $np \leq \ln \ln n + (k-1) \ln \ln \ln n - \omega$ , for some  $\omega \rightarrow \infty$ ).

First let  $c_n \rightarrow -\infty$ .

Here we distinguish between  $a_n \sim ((k-1)!S_{1,k} \ln n)/S_1$  (i.e.  $\ln \ln n + \omega \leq np \leq \ln \ln n + (k-1) \ln \ln \ln n - \omega$ , for some  $\omega \rightarrow \infty$ ) and  $a_n = \Theta((S_{1,2} \ln n/S_1)^{k-1})$  (i.e.  $np \leq \ln \ln n + d$  for some  $d \in \mathbb{R}$ ).

For  $a_n \sim (k-1)!S_{1,k} \ln n/S_1$  and  $a_n \rightarrow \infty$  we use the following division and couplings.

$$\mathcal{G}(n, m, p) = \mathcal{G}_k(n, m, p) \cup \mathcal{G}_{\neq k}(n, m, p) \preceq_{1-o(1)} \mathbb{G}_{*k} \left( n, \frac{S_{1,k} + \omega \sqrt{S_{1,k}}}{k} \right) \cup \mathcal{G}_{\neq k}(n, m, p) \preceq_{1-o(1)} G_k(n, \hat{p}_{k+}) \cup \mathcal{G}_{\neq k}(n, m, p), \quad (\text{B.5})$$

where  $\hat{p}_{k+} = (S_{1,k} + 2\sqrt{S_{1,k}})/k \binom{n}{k}$  (for more details on the construction of the couplings see (B.1)). As in the proof of Lemma 5.2 the construction of  $\mathcal{G}_{\neq k}(n, m, p)$  in  $G_k(n, \hat{p}_{k+}) \cup \mathcal{G}_{\neq k}(n, m, p)$  may be coupled with the vertex collector process on  $\mathcal{V}$ . Moreover by (5.7) and (5.1) with high probability the construction of  $\mathcal{G}_{\neq k}(n, m, p)$  is finished in at most

$$T_{k+} = S_1 - S_{1,k} + 2\omega \sqrt{S_1} + \frac{S_1}{\omega \ln n}$$

draws, for  $\omega \rightarrow \infty$ . Take any probability space on which we define independent: the vertex collector process on  $\mathcal{V}$  and  $G_k(n, \hat{p}_{k+})$ . Let  $X_+$  be a random variable equal to the number of vertices which have not been collected during the vertex collector process in  $T_{k+}$  draws and have degree  $k-1$  in  $G_k(n, \hat{p}_{k+})$ . For  $\omega$  tending to infinity slowly enough

$$\begin{aligned} \frac{1}{n} T_{k+} + \binom{n-1}{k-1} \hat{p}_{k+} &= \frac{1}{n} \left( S_1 + O \left( \omega \sqrt{S_1} + \frac{S_1}{\omega \ln n} \right) \right) \\ &= \ln n + \ln a_n + c_n + o(1). \end{aligned}$$

and by (B.4)

$$p_{k+} \sim \frac{S_{1,k}}{k \binom{n}{k}} \sim \frac{\frac{a_n S_1}{(k-1)! \ln n}}{n \binom{n-1}{k-1}} \sim \frac{a_n}{(k-1)! \binom{n-1}{k-1}}$$

Thus

$$\mathbb{E}X_+ = n \left( 1 - \frac{1}{n} \right)^{T_{k+}} \binom{n-1}{k-1} \hat{p}_{k+} (1 - \hat{p}_{k+})^{\binom{n-1}{k-1} - 1}$$

$$\begin{aligned}
&\sim n \cdot \binom{n-1}{k-1} \cdot \frac{a_n}{(k-1)! \binom{n-1}{k-1}} \cdot \exp\left(-\frac{T_{k+}}{n} - \binom{n-1}{k-1} p_{k+}\right) \\
&\sim n \frac{1}{(k-1)!} a_n \exp(-\ln n - \ln a_n - c_n) \\
&\sim \frac{e^{-c_n}}{(k-1)!}
\end{aligned}$$

and similarly

$$\begin{aligned}
\mathbb{E}X_+(X_+ - 1) &\sim n^2 \left(1 - \frac{2}{n}\right)^{T_{k+}} \\
&\cdot \left[ \left( \binom{n-2}{k-1} \hat{p}_{k+} \right)^2 (1 - \hat{p}_{k+})^{2\binom{n-1}{k-1} - \binom{n-2}{k-2} - 2} \right. \\
&\quad \left. + \binom{n-2}{k-2} \hat{p}_{k+} (1 - \hat{p}_{k+})^{2\binom{n-1}{k-1} - \binom{n-2}{k-2} - 1} \right] \\
&\sim \left( \frac{e^{-c_n}}{(k-1)!} \right)^2.
\end{aligned}$$

Thus by the second moment method with high probability  $X_+ > 0$  as  $c_n \rightarrow -\infty$  (i.e. with high probability there exists a vertex, which has degree at most  $k-1$  in  $G_k(n, \hat{p}_{k+})$  or has not been collected in  $T_{k+}$  draws). Thus by (5.1), (5.7), (B.5), and the coupling of  $\mathcal{G}_{\neq k}(n, m, p)$  with the vertex collector process with high probability

$$\lim_{n \rightarrow \infty} \Pr \{ \delta(\mathcal{G}(n, m, p)) \geq k \} = 0.$$

We are left with  $a_n = \Theta((S_{1,2} \ln n / S_1)^{k-1})$ . Then

$$S_1 = n(\ln n + (k-1) \ln \left( \max \left\{ 1, \frac{S_{1,2} \ln n}{S_1} \right\} \right)) + O(1) + c_n.$$

Therefore the result follows by Lemma 5.2(i)

Now assume that  $c_n \rightarrow \infty$ .

Here we use the couplings

$$\begin{aligned}
&\bigcup_{t=2}^k G_t(n, \hat{q}_t) \cup \mathcal{G}_{\geq k+1}(n, m, p) \\
&\preceq_{1-o(1)} \bigcup_{t=2}^k \mathbb{G}_{*t} \left( n, \max \left\{ 0, \frac{S_{1,t} - \omega \sqrt{S_{1,t}}}{t} \right\} \right) \cup \mathcal{G}_{\geq k+1}(n, m, p) \tag{B.6} \\
&\preceq_{1-o(1)} \bigcup_{t=2}^k \mathcal{G}_t(n, m, p) \cup \mathcal{G}_{\geq k+1}(n, m, p) \\
&= \mathcal{G}(n, m, p)
\end{aligned}$$

where

$$\hat{q}_t = \begin{cases} 0, & \text{for } S_{1,t} = O(\omega\sqrt{S_1}); \\ \frac{S_{1,t} - 2\omega\sqrt{S_1} - t!S_{1,t}^2n^{-2}}{t\binom{n}{t}}, & \text{otherwise.} \end{cases}$$

Therefore

$$\binom{n-1}{t-1}\hat{q}_t \geq S_{1,t} - O(\omega\sqrt{S_1}) \tag{B.7}$$

and by (2.1)

$$\hat{q}_t \leq \frac{S_{1,t}}{t\binom{n}{t}} = O(mp^t e^{-np}).$$

Take any probability space on which we may define independent: the vertex collector process and  $\bigcup_{t=2}^k G_t(n, \hat{q}_t)$ . Let  $X_-$  be a random variable equal to the number of vertices which have not been collected during the vertex collector process in

$$T_- = S_1 - \sum_{t=3}^k S_{1,t} - k\sqrt{S_1} \tag{B.8}$$

draws and have degree at most  $k-1$  in  $\bigcup_{t=2}^k G_t(n, \hat{q}_t)$ . Note that if  $v$  has degree at most  $k-1$  in  $\bigcup_{t=2}^k G_t(n, \hat{q}_t)$ , then for some  $0 \leq k_0 \leq k-1$  and some sequence of integers  $r_2, \dots, r_t$  such that  $\sum_{t=2}^k (t-1)r_t = k_0$

- (i) there is a set  $\mathcal{V}' \subseteq \mathcal{V} \setminus \{v\}$  of  $k_0$  vertices such that for each  $t$  there are  $r_t$  hyperedges in  $H_t(n, \hat{q}_t)$  (generating  $G_t(n, \hat{q}_t)$ ) containing  $v$  and contained in  $\mathcal{V}' \cup \{v\}$ .
- (ii) and all hyperedges in  $\bigcup_{t=2}^k H_t(n, \hat{q}_t)$  containing  $v$  are subsets of  $\mathcal{V}' \cup \{v\}$ .

Let  $r = \sum_{t=2}^k r_t$ . Given a sequence  $r_2, \dots, r_t$  such that  $\sum_{t=2}^k (t-1)r_t = k_0$ , if  $c_n = O(\ln n)$  then event (i) occurs with probability at most

$$\begin{aligned} & \binom{n-1}{k_0} \prod_{t=2}^k \binom{k_0-1}{t-1}^{r_t} (O(1)(mp^t)e^{-np})^{r_t} \\ &= O(1)n^{k_0}m^r p^{k_0+r} (e^{-np})^r \\ &= O(1)(np)^{k_0} (mpe^{-np})^r \\ &= O(1)(np)^{k_0} \left( \frac{\ln n}{(1-e^{-np})} e^{-np} \right)^r \\ &= O(a_n). \end{aligned}$$

In the penultimate equality we have used the fact that  $mp(1-e^{-np}) \sim \ln n$ . There are  $O(1)$  sequences of integers  $r_2, \dots, r_t$  such that  $\sum_{t=2}^k (t-1)r_t \leq k-1$ , thus the probability that there exists such sequence  $r_2, \dots, r_t$  that (i) and (ii) occur is at most

$$O(a_n) \prod_{t=2}^k (1 - \hat{q}_t)^{\binom{n-1}{t-1} - \binom{k-1}{t-1}}.$$

Moreover by (B.7) and (B.8)

$$\begin{aligned} \frac{1}{n}T_- + \sum_{t=2}^k \binom{n-1}{t-1} \hat{q}_t &\geq \frac{1}{n} \left( S_1 - O\left(\omega\sqrt{S_1} + S_1^2 n^{-2}\right) \right) \\ &= \ln n + \ln a_n + c_n + o(1). \end{aligned}$$

Therefore finally

$$\mathbb{E}X_- = n \left(1 - \frac{1}{n}\right)^{T_-} O(a_n) \prod_{t=2}^k (1 - \hat{q}_t)^{\binom{n-1}{t-1} - \binom{k-1}{t-1}} = o(1).$$

Hence with high probability  $X_- = 0$ . Thus by (5.5), (B.6) and the coupling of the construction of  $\mathcal{G}_{\geq k+1}(n, m, p)$  with the vertex collector process with high probability  $\delta(\mathcal{G}(n, m, p)) \geq k$ .  $\square$

## Acknowledgements

I would like to thank Jerzy Jaworski for fruitful discussions and his help in improving the presentation of the article. I am also grateful to the reviewers for their thoroughness and their invaluable comments.

## References

- [1] A.D. Barbour and G. Reinert. The shortest distance in random multi-type intersection graphs. *Random Structures & Algorithms*, 39(2):179–209, 2011.
- [2] M. Bloznelis. Degree and clustering coefficient in sparse random intersection graphs. *Annals of Applied Probability*, 23:1254–1289, 2013.
- [3] M. Bloznelis and J. Damarackas. Degree distribution of an inhomogeneous random intersection graph. *The Electronic Journal of Combinatorics*, 20(3):#P3, 2013.
- [4] M. Bloznelis, J. Jaworski, and K. Rybarczyk. Component evolution in a secure wireless sensor network. *Networks*, 53(1):19–26, 2009.
- [5] M. Bloznelis and T. Łuczak. Perfect matchings in random intersection graphs. *Acta Mathematica Hungarica*, 138(1–2):15–33, 2013.
- [6] B. Bollobás. *Random Graphs*. Academic Press, 1985.
- [7] M. Bradonjić, A. Hagberg, N. W. Hengartner, N. Lemons, and A. G. Percus. Component evolution in general random intersection graphs. In *Algorithms and Models for the Web-Graph*, LNCS 6516, pages 36–43. Springer, 2010.
- [8] T. Britton, M. Deijfen, A. N. Lagerås, and M. Lindholm. Epidemics on random graphs with tunable clustering. *Journal of Applied Probability*, 45(3):743–756, 2008.

- [9] M. Deijfen and W. Kets. Random intersection graphs with tunable degree distribution and clustering. *Probability in the Engineering and Information Sciences*, 23:661–674, 2009.
- [10] P. Erdős and A Rényi. On a classical problem of probability theory,. *Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, 6:215–220, 1961.
- [11] J. A. Fill, E. R. Scheinerman, and K. B. Singer-Cohen. Random intersection graphs when  $m = \omega(n)$ : An equivalence theorem relating the evolution of the  $G(n, m, p)$  and  $G(n, p)$  models. *Random Structures and Algorithms*, 16:156–176, 2000.
- [12] E. Godehardt and J. Jaworski. Two models of random intersection graphs for classification. In O. Opitz and M. Schwaiger, editors, *Studies in Classification, Data Analysis and Knowledge Organization*, volume 22, pages 67–81. Springer, 2003.
- [13] S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. Wiley, 2001.
- [14] M. Karoński, E. R. Scheinerman, and K.B. Singer-Cohen. On random intersection graphs: The subgraph problem. *Combinatorics, Probability and Computing*, 8:131–159, 1999.
- [15] S. Nikolettseas, C. Raptopoulos, and P. Spirakis. Large independent sets in general random intersection graphs. *Theoretical Computer Science*, 406:215–224, 2008.
- [16] K. Rybarczyk. O pewnych zastosowaniach hipergrafów losowych (On some applications of random intersection graphs). Master’s thesis, Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poznań, 2005.
- [17] K. Rybarczyk. Equivalence of the random intersection graph and  $G(n,p)$ . *Random Structures and Algorithms*, 38:205–234, 2011.
- [18] K. Rybarczyk. Sharp threshold functions for random intersection graphs via a coupling method. *The Electronic Journal of Combinatorics*, 18(1):#P36, 2011.
- [19] K. Rybarczyk and D. Stark. Poisson approximation of the number of cliques in the random intersection graph. *Journal of Applied Probability*, 47(3):826 – 840, 2010.
- [20] K. B. Singer. *Random intersection graphs*. PhD thesis, Department of Mathematical Sciences, The Johns Hopkins University, 1995.