

Finding Hamilton cycles in random intersection graphs

Katarzyna Rybarczyk *

Adam Mickiewicz University, Poznań, Poland

received 15th Feb. 2017, revised 5th Oct. 2017, accepted 22nd Jan. 2018.

The construction of the random intersection graph model is based on a random family of sets. Such structures, which are derived from intersections of sets, appear in a natural manner in many applications. In this article we study the problem of finding a Hamilton cycle in a random intersection graph. To this end we analyse a classical algorithm for finding Hamilton cycles in random graphs (algorithm HAM) and study its efficiency on graphs from a family of random intersection graphs (denoted here by $\mathcal{G}(n, m, p)$). We prove that the threshold function for the property of HAM constructing a Hamilton cycle in $\mathcal{G}(n, m, p)$ is the same as the threshold function for the minimum degree at least two. Until now, known algorithms for finding Hamilton cycles in $\mathcal{G}(n, m, p)$ were designed to work in very small ranges of parameters and, unlike HAM, used the structure of the family of random sets.

Keywords: random intersection graphs, Hamilton cycle, algorithm

1 Introduction

Since its introduction by Karoński et al. (1999) the random intersection graph model and its generalisations have proven to have many applications. To mention just a few: “gate matrix layout” for VLSI design (see e.g. Karoński et al. (1999)), cluster analysis and classification (see e.g. Godehardt and Jaworski (2003)), analysis of complex networks (see e.g. Bloznelis (2013); Deijfen and Kets (2009)), secure wireless networks (see e.g. Bloznelis et al. (2009)) or epidemics (Brittom et al. (2008)). For more details we refer the reader to survey papers Bloznelis et al. (2015a,b).

In the random intersection graph model $\mathcal{G}(n, m, p)$ to each vertex from the vertex set \mathcal{V} ($|\mathcal{V}| = n$) we assign a random set of its features $\mathcal{W}(v)$ from an auxiliary set \mathcal{W} ($|\mathcal{W}| = m(n)$). For each $w \in \mathcal{W}$ and $v \in \mathcal{V}$ we have $w \in \mathcal{W}(v)$ with probability p , $p = p(n) \in (0, 1)$, independently of all other elements from \mathcal{V} and \mathcal{W} . We connect vertices v and v' by an edge if sets $\mathcal{W}(v)$ and $\mathcal{W}(v')$ intersect.

In the course of last years various properties of random intersection graph were studied. However, little is still known about algorithms which might efficiently construct or find structures in random intersection graphs. In this article we address the problem of efficiently finding a Hamilton cycle in $\mathcal{G}(n, m, p)$. The problem of determining a threshold function for the property of having a Hamilton cycle in the random intersection graph has already been studied by several authors, for example: Efthymiou and Spirakis

*supported by NCN (Narodowe Centrum Nauki) grant 2014/13/D/ST1/01175

(2005), Rybarczyk (2011) (both for the model considered in this paper), and Bloznelis and Radavičius (2011) (for the uniform random intersection graph model). Here we analyse algorithmic aspects of the problem of finding a Hamilton cycle in $\mathcal{G}(n, m, p)$. Raptopoulos and Spirakis (2005) proposed a randomised algorithm which with probability tending to 1 as $n \rightarrow \infty$ in polynomial time finds a Hamilton cycle in $\mathcal{G}(n, m, p)$ if $m/(n \ln n) \rightarrow 0$ and $p = \frac{\ln n + \omega(n)}{m}$ with $\omega(n)$ tending slowly to infinity. Our aim is to show that there exists an algorithm which efficiently finds a Hamilton cycle in $\mathcal{G}(n, m, p)$ for a wider range of parameters n, m, p . To this end we take algorithm HAM introduced by Bollobás et al. (1987) and analyse its performance on $\mathcal{G}(n, m, p)$.

All limits in the paper are taken as $n \rightarrow \infty$. Throughout the paper we use standard asymptotic notation $a_n = o(b_n)$ if $a_n/b_n \rightarrow 0$ as $n \rightarrow \infty$ and $a_n = O(b_n)$ if there exists a constant C such that $|a_n| \leq C|b_n|$, for large n . By $\text{Bin}(n, p)$ we denote the binomial distribution with parameters n, p . We also use the phrase “with high probability” to say with probability tending to one as n tends to infinity. We consequently omit $[\cdot]$ and $\lceil \cdot \rceil$ for the sake of clarity of presentation. All inequalities hold for n large enough.

2 Main results

Algorithm HAM, as presented by Bollobás et al. (1987), is designed to search for a Hamilton cycle in any graph with the minimum degree at least 2. It is assumed that the vertex set of the input graph is ordered, i.e. $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ (for simplicity we write $v_i < v_j$ if $i < j$). Algorithm HAM utilises the rotation technique introduced by Pósa (1976). The rotation technique uses the fact that, given a path P of length k in a graph \mathcal{G} (i.e. $P = (u_1, \dots, u_k)$), if there is an edge $e = \{u_k, u_i\}$, $2 \leq i \leq k-2$, we may construct another path of length k denoted by

$$\text{ROTATE}(P, e) = (u_1, u_2, \dots, u_i, u_k, u_{k-1}, \dots, u_{i+1}).$$

We state algorithm HAM following Bollobás et al. (1987) (see Figure 1). In most cases we keep notation consistent with the one introduced by Bollobás et al. (1987). Algorithm HAM, starting with a path consisting of one vertex v_1 , in each step extends the path (if its end has a neighbour outside the path or both endpoints are neighbours) or, if extension is not possible, it searches for new paths of the same length using Pósa’s rotation technique. In the latter case the algorithm explores all possible rotations in the BFS–type manner, i.e. given a considered path $P = (u_1, \dots, u_k)$ of length k , it explores all neighbours of u_k and u_1 and new paths resulting from rotations related to those neighbours. Then for each new path, one by one, HAM extends the path or, if an extension is not possible, it explores neighbours of the ends of this path and does rotations. The algorithm stops if either it finds a Hamilton cycle or it is not able to extend the path of length k . More precisely, it stops at stage k when it has explored without extension all paths which result from at most $2T + 1$ (T is a function of n and d) rotations of the initial path of length k .

In what follows whenever HAM is executed on $\mathcal{G}(n, m, p)$ with $mp^2 \leq 1$ we set

$$d = nmp^2.$$

The main result of this article is the following.

Theorem 1. *Let $\mathcal{G}(n, m, p)$ be a random intersection graph with $\ln n = o(m)$ and $mp^2 \leq 1$, then*

$$\lim_{n \rightarrow \infty} \Pr \{ \{ \delta(\mathcal{G}(n, m, p)) \geq 2 \} \cap \{ \text{HAM terminates unsuccessfully on } \mathcal{G}(n, m, p) \} \} = 0.$$

Algorithm HAM

Input: a connected graph $G = (\mathcal{V}, E)$ of minimum degree at least 2 and with ordered vertex set $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$, integer d .

begin

$k := 1$;

P_k – the path (v_1, u) , where $u = \min\{v : \{v_1, v\} \in E\}$

 (L1) **begin** {stage k begins here}

$Q_1 := P_k, s := 1, t := 1, \bar{\delta}(Q_1) := 0$

 ($\bar{\delta}(Q_s)$) – the number of rotations in the sequence constructing Q_s from Q_1)

repeat (u_0, u_1 – endpoints of Q_s , where $u_0 < u_1$);

for $i = 0, 1$ **do**

begin

$\{u_i, x_1\}, \dots, \{u_i, x_p\}$ – all edges incident to u_i and not contained in Q_s ,
 where $x_1 < \dots < x_p$;

for $j = 1$ **to** p **do**

if x_j is not on Q_s **then**

begin

$P_{k+1} := Q_s + \{u_i, x_j\}$; {extension}

$k := k + 1$; **goto** L1

end

else if $x_j = u_{1-i}$ **then**

begin

 let C be the cycle $Q_s + \{u_0, u_1\}$;

if C is a Hamiltonian cycle **then** terminate successfully

else

begin

 starting from u_0 , let u be the first vertex along Q_s , which is
 adjacent to some vertex not in C ; let v be the lowest numbered
 neighbour of u not in C and let u' and u'' be the neighbours
 of u on C , where $u' < u''$, then

$P_{k+1} := C + \{u, v\} - \{u, u'\}$; $k := k + 1$;

goto L1 {cycle extension}

end

end else

begin

$t := t + 1$; $Q_t := \text{ROTATE}(Q_s, \{u_i, x_j\})$; $\bar{\delta}(Q_t) := \bar{\delta}(Q_s) + 1$

end

end for{next j }

end for{next i }

$s := s + 1$;

until $\bar{\delta}(Q_s) \geq 2T + 1$ {where $T = \lceil \ln n / (\ln d - \ln \ln d) \rceil + 1$ }

 terminate unsuccessfully

end

end

Fig. 1: Algorithm HAM introduced by Bollobás et al. (1987)

We rigorously prove the above theorem in the case $m \geq n^{24/25}$. Then we present a sketch of the proof in the latter case.

Remark 1. We give the main theorem only for $\ln n = o(m)$ and $mp^2 \leq 1$. The latter cases would require considering more cases in the proofs however the reasoning would be easier. Namely, if $m = O(\ln n)$, in the relevant range of parameters, with high probability $\mathcal{G}(n, m, p)$ consists of $m = O(\ln n)$ independent very large cliques thus it is very close to a complete graph. Similarly if $mp^2 > 1$ then the probability of given two vertices being connected by an edge is a constant and with high probability almost all vertices have degree at least $n/1000$. This case would require, among others, setting $d = n$, in some places replacing d_0 by n , setting $T = 2$, and redefining property **P3** defined in the proof (as in this case $n/d_1 < 1$). Even though this case would require changes to some parts of the proof, in general the reasoning would follow the same lies and would be easier. Thus we do not include it for shortness.

Remark 2. Using the same arguments as Bollobás et al. (1987), we may show that with high probability the time complexity of HAM on $\mathcal{G}(n, m, p)$ is $O(n^{4+\varepsilon})$ (for any $\varepsilon > 0$). For completeness we give the proof in Appendix A.

Theorem 1 together with the probability of the event that the minimum degree of $\mathcal{G}(n, m, p)$ is at least 2, $\delta(\mathcal{G}(n, m, p)) \geq 2$, gives the probability of the property that HAM finds a Hamilton cycle in $\mathcal{G}(n, m, p)$. The proof of Theorem 2 is presented in Appendix B.

Theorem 2. If $\ln^2 n = o(m)$

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

where $c_n = o(\ln n)$, and for some $\varepsilon > 0$

$$a_n = \begin{cases} 1 & \text{for } m < (1 - \varepsilon) \frac{n \ln n}{\ln \ln n}; \\ \frac{np \ln n}{e^{np} - 1} & \text{for } m > (1 + \varepsilon) \frac{n \ln n}{\ln \ln n}, \end{cases}$$

then

$$\lim_{n \rightarrow \infty} \Pr \{ \delta(\mathcal{G}(n, m, p)) \geq 2 \} = \begin{cases} 0 & \text{for } c_n \rightarrow -\infty; \\ e^{-e^{-c}} & \text{for } c_n \rightarrow c \in (-\infty, \infty); \\ 1 & \text{for } c_n \rightarrow \infty. \end{cases}$$

In the above mentioned theorem we exclude the case $m = (1 + o(1))n \ln n / \ln \ln n$. This case may be analysed by the same methods as those presented in the proofs but would require several additional case studies and more cases in the definition of a_n . We omit it for the sake of clarity of presentation. The result may be also expanded to the case $c_n = \Omega(\ln n)$ by a simple coupling and monotonicity of the property $\{ \delta(\mathcal{G}(n, m, p)) \geq 2 \}$.

The remaining part of the article is organized as follows. In Section 3 we introduce some notation and gather facts which will be used repeatedly in the whole article. In the following three sections we present a proof of Theorem 1 in the case $mp^2 \leq 1$ and $m \geq n^{1-\varepsilon}$ for $\varepsilon = 1/25$. Namely, in Section 4 we study the properties of $\mathcal{G}(n, m, p)$, which are crucial for the analysis of algorithm HAM on $\mathcal{G}(n, m, p)$. Then in Section 5 we discuss the notion of deletable sets. In Section 6 we prove Theorem 1 in the case $mp^2 \leq 1$ and $m \geq n^{1-\varepsilon}$ using facts proved in the first part of the article. In Section 7 we give a sketch of the proof in the case where $m < n^{1-\varepsilon}$.

3 Notation and important facts

For any graph \mathcal{G} we denote by $E(\mathcal{G})$ its edge set. If we mention an intersection graph \mathcal{G} we mean not only the graph but also the underlying structure of feature sets attributed to vertices. For any intersection graph \mathcal{G} with a vertex set \mathcal{V} and feature set \mathcal{W} , if $w \in \mathcal{W}(v)$ then we say that vertex v chose feature w and w was chosen by v . For $v \in \mathcal{V}$, $w \in \mathcal{W}$, $\mathcal{S} \subseteq \mathcal{V}$, and $\mathcal{R} \subseteq \mathcal{W}$ in random intersection graph \mathcal{G} we use the following notation

$$\mathcal{W}_{\mathcal{G}}(v) = \{w \in \mathcal{W} : v \text{ chose } w\}, \quad \mathcal{W}_{\mathcal{G}}(\mathcal{S}) = \bigcup_{v \in \mathcal{S}} \mathcal{W}_{\mathcal{G}}(v);$$

i.e. the set of features chosen by v (and, respectively, chosen by vertices from \mathcal{S});

$$\mathcal{V}_{\mathcal{G}}(w) = \{v \in \mathcal{V} : w \text{ was chosen by } v\}, \quad \mathcal{V}_{\mathcal{G}}(\mathcal{R}) = \bigcup_{w \in \mathcal{R}} \mathcal{V}_{\mathcal{G}}(w);$$

i.e. the set of vertices which chose w (and, respectively, chose features from \mathcal{R});

$$\mathcal{W}'_{\mathcal{G}}(v) = \{w \in \mathcal{W}(v) : |\mathcal{V}_{\mathcal{G}}(w)| \geq 2\}, \quad \mathcal{W}'_{\mathcal{G}}(\mathcal{S}) = \bigcup_{v \in \mathcal{S}} \mathcal{W}'_{\mathcal{G}}(v);$$

i.e. the set of features chosen by v (and, respectively, chosen by vertices from \mathcal{S}) which contribute to at least one edge in \mathcal{G} ;

$$\mathcal{W}''_{\mathcal{G}}(\mathcal{S}) = \{w \in \mathcal{W}(\mathcal{S}) : |\mathcal{V}_{\mathcal{G}}(w) \cap \mathcal{S}| \geq 2\};$$

i.e. the set of features chosen by at least two vertices from \mathcal{S} .

Moreover for any graph \mathcal{G} with vertex set \mathcal{V} and edge set $E(\mathcal{G})$ and for any $\mathcal{S} \subseteq \mathcal{V}$ we set

$$\mathcal{N}_{\mathcal{G}}(v) = \{v' \in \mathcal{V} \setminus \{v\} : \{v, v'\} \in E(\mathcal{G})\}, \quad \mathcal{N}_{\mathcal{G}}(\mathcal{S}) = \bigcup_{v \in \mathcal{S}} \mathcal{N}_{\mathcal{G}}(v) \setminus \mathcal{S};$$

i.e. the set of neighbours of v in \mathcal{G} (and, respectively, neighbours of \mathcal{S} in \mathcal{G}).

In addition

$$\begin{aligned} W_{\mathcal{G}}(v) &= |\mathcal{W}_{\mathcal{G}}(v)|, & W_{\mathcal{G}}(\mathcal{S}) &= |\mathcal{W}_{\mathcal{G}}(\mathcal{S})|, \\ V_{\mathcal{G}}(w) &= |\mathcal{V}_{\mathcal{G}}(w)|, & V_{\mathcal{G}}(\mathcal{R}) &= |\mathcal{V}_{\mathcal{G}}(\mathcal{R})|, \\ W'_{\mathcal{G}}(v) &= |\mathcal{W}'_{\mathcal{G}}(v)|, & W'_{\mathcal{G}}(\mathcal{S}) &= |\mathcal{W}'_{\mathcal{G}}(\mathcal{S})|, \\ W''_{\mathcal{G}}(\mathcal{S}) &= |\mathcal{W}''_{\mathcal{G}}(\mathcal{S})|, & N_{\mathcal{G}}(v) &= |\mathcal{N}_{\mathcal{G}}(v)|, \\ N_{\mathcal{G}}(v) &= |\mathcal{N}_{\mathcal{G}}(v)|, & N_{\mathcal{G}}(\mathcal{S}) &= |\mathcal{N}_{\mathcal{G}}(\mathcal{S})|, \end{aligned}$$

and $\deg(v) = \deg_{\mathcal{G}}(v) = N_{\mathcal{G}}(v)$.

In the above mentioned notation we consequently omit subscript \mathcal{G} when it is clear from the context which \mathcal{G} we have in mind.

Alternatively, in order to obtain an instance \mathcal{G} of $\mathcal{G}(n, m, p)$, we may first construct an instance \mathcal{B} of a random bipartite graph $\mathcal{B}(n, m, p)$ with bipartition $(\mathcal{V}, \mathcal{W})$ in which each edge $\{v, w\}$, $v \in \mathcal{V}$ and $w \in \mathcal{W}$, appears independently with probability p . We connect vertices $v \in \mathcal{V}$ and $v' \in \mathcal{V}$ by an edge in $\mathcal{G}(n, m, p)$ if they have a common neighbour in $\mathcal{B}(n, m, p)$. An instance \mathcal{G} of $\mathcal{G}(n, m, p)$ constructed in this manner from \mathcal{B} we will call *associated with \mathcal{B}* . Note that $\mathcal{W}(v)$, $v \in \mathcal{V}$, is the set of neighbours of vertex v in $\mathcal{B}(n, m, p)$. Moreover $\mathcal{V}(w)$ is the set of neighbours of $w \in \mathcal{W}$ in $\mathcal{B}(n, m, p)$. In addition $W(v)$ and $V(w)$ are degrees of $v \in \mathcal{V}$ and $w \in \mathcal{W}$ in $\mathcal{B}(n, m, p)$.

Moreover, let

$$d_0 = mp(1 - (1 - p)^{n-1}), \quad d_1 = nmp^2.$$

Note that d_0 is the expected value of $W'_{\mathcal{G}(n, m, p)}(v)$ and d_1 is almost the expected degree of a vertex in $\mathcal{G}(n, m, p)$.

The following relations between d_0 , d_1 , and mp will be frequently used without mentioning them directly. In the range of parameters we are interested in (*i.e.* n , m , and p such that the probability of event $\{\delta(\mathcal{G}(n, m, p)) \geq 2\}$ does not tend to 0), we have

$$d_0 \geq (1 + o(1)) \ln n.$$

For the proof see for example Lemma 5.1 in Rybarczyk (2017).

Moreover by definition

$$\begin{aligned} d_0 &\leq mp, \\ d_0 &\leq nmp^2 = d_1, \\ d_0 &\leq mp \min\{np, 1\}, \\ d_1 &\leq 2d_0 \max\{np, 1\}. \end{aligned}$$

In addition,

if $np > 40$ then $0.9mp \leq d_0 \leq mp$ and

if $np \leq 40$ then $d_0 \leq d_1 \leq 40.1d_0$.

To get the first inequality for $np > 40$ we note that

$$d_0 \geq mp(1 - e^{-(n-1)p}) \geq mp(1 - e^{-39}) \geq 0.9mp.$$

In the proof of the second inequality for $np \leq 40$ we have, for large n ,

$$40.1d_0 \geq 40.1d_1 \cdot \frac{n-1}{n} \cdot \frac{1 - e^{-(n-1)p}}{(n-1)p} \geq (1 + o(1))40.1d_1 \frac{1 - e^{-40}}{40} \geq d_1$$

since $f(x) = (1 - e^{-x})/x$ is decreasing for $x > 0$.

We will frequently use Chernoff's inequality in the following form (see for example Theorem 2.1 in Janson et al. (2001)).

Lemma 1. *Let X be a random variable with the binomial distribution and expected value μ , then for any $0 < \varepsilon < 1$*

$$\Pr \{X \leq \varepsilon \mu\} \leq \exp(-\psi(\varepsilon)\mu),$$

where

$$\psi(\varepsilon) = \varepsilon \ln \varepsilon + 1 - \varepsilon.$$

4 Graph properties

In this and the following section we assume that $m \geq n^{1-\varepsilon}$ (for $\varepsilon = 1/25$) and $mp^2 \leq 1$ (i.e. $n \geq d_1$).

The efficiency of algorithm HAM on a random graph is due to the fact that random graphs, in general, have good expansion properties. Below we list the properties which will be used in the analysis of algorithm HAM on $\mathcal{G}(n, m, p)$.

Given an intersection graph \mathcal{G} associated with a bipartite graph \mathcal{B} , integers $d_0 = d_0(n)$, $d_1 = d_1(n)$, $m = m(n)$, a real $p = p(n) \in (0, 1)$, sets $\text{SMALL} \subseteq \mathcal{V}$ and $\text{LARGE} = \mathcal{V} \setminus \text{SMALL}$, and a constant $b_1 > 0$ we define the following properties:

P0 $\delta(\mathcal{G}) \geq 2$;

P1 $|\text{SMALL}| \leq n^{1/3}$ and
if $d_0 \geq 2 \ln n$ then $|\text{SMALL}| = \emptyset$;

P2 There are no $v, v' \in \text{SMALL}$ at distance 4 or less apart;

P3 For all $\mathcal{S} \subseteq \text{LARGE}$, if $|\mathcal{S}| \leq n/d_1$ then $N(\mathcal{S}) \geq b_1 d_1 |\mathcal{S}|$.

P4 For all $v \in \mathcal{V}$

$$W(v) \leq 4mp, \quad W'(v) \leq 4d_0, \quad N(v) \leq 12d_1.$$

P5 For all $w \in \mathcal{W}$

$$V(w) \leq \frac{\ln n}{\ln \ln n} \max\{np, 4\}.$$

Lemma 2. *Let $\mathcal{G}(n, m, p)$ be a random intersection graph with $p \leq \sqrt{(3 \ln n)/m}$ and*

$$d_0 \geq (1 + o(1)) \ln n.$$

Moreover let

$$\text{SMALL} = \{v \in \mathcal{V} : W'(v) \leq 0.1d_0\}, \quad \text{LARGE} = \{v \in \mathcal{V} : W'(v) > 0.1d_0\}.$$

If $m \geq n^{1-\varepsilon}$ ($\varepsilon = 1/25$) then with high probability $\mathcal{G}(n, m, p)$ has properties **P1–P5** with $b_1 = 0.001$.

It is easy to prove by the first moment method that if $p > \sqrt{3 \ln n/m}$ then with high probability $\mathcal{G}(n, m, p)$ is a complete graph. Therefore if $p > \sqrt{3 \ln n/m}$ then with high probability HAM finds a Hamilton cycle in $\mathcal{G}(n, m, p)$ without any problem. This is why we may restrict ourselves to the case $p \leq \sqrt{3 \ln n/m}$.

Proof of Lemma 2: Recall that all inequalities hold for large n .

In the proofs we assume that $\ln n = o(m)$. If we make some additional assumptions concerning relation between n and m , we mention it at the beginning of the proof of the considered property. Moreover, we set $b_1 = 0.001$.

P1 Note that $W'(v)$ has the binomial distribution $\text{Bin}(m, d_0/m)$. Therefore for $d_0 \geq (1 - o(1)) \ln n$ by Chernoff's inequality

$$\begin{aligned} \mathbb{E}|\text{SMALL}| &\leq n \Pr \{W'(v) \leq 0.1 \cdot d_0\} \\ &\leq n \exp(-\psi(0.1) \cdot d_0) \\ &\leq \exp(\ln n - 0.669 \ln n) = o(n^{1/3}). \end{aligned}$$

Thus by Markov's inequality we get that with high probability $|\text{SMALL}| \leq n^{1/3}$. Moreover the same calculation implies that if $d_0 \geq 2 \ln n$ then $\mathbb{E}|\text{SMALL}| = o(1)$ *i.e.* with high probability $\text{SMALL} = \emptyset$.

P2 In the proof of **P2** we assume additionally that $m \geq n^{1-\varepsilon}$ (for $\varepsilon = 1/25$). First recall that if $d_0 \geq 2 \ln n$ then with high probability $\text{SMALL} = \emptyset$, *i.e.* the statement holds trivially.

Now assume that $d_0 \leq 2 \ln n$, therefore either (for $np \leq 40$)

$$d_1 \leq 40.1d_0 = O(\ln n)$$

or (for $np > 40$)

$$d_1 = \frac{n}{m}(mp)^2 \leq n^\varepsilon(1.2d_0)^2 = O(n^\varepsilon \ln^2 n).$$

If there exist two vertices $v, v' \in \text{SMALL}$, which are at distance t , $t \leq 4$, in \mathcal{G} , then there exists a path $vw_1v_1w_2 \dots v_{t-1}w_{t-1}w_tv'$ of length $2t$ in \mathcal{B} with which \mathcal{G} is associated and in intersection graph \mathcal{G} the set $\mathcal{W}'(\{v, v'\}) \setminus \{w_1, \dots, w_t\}$ is of cardinality at most $0.2d_0$. The number of possible paths of the form $vw_1v_1w_2 \dots v_{t-1}w_{t-1}w_tv'$ is upper bounded by $n^{t+1}m^t$. Probability that the path $vw_1v_1w_2 \dots v_{t-1}w_{t-1}w_tv'$ is present in $\mathcal{B}(n, m, p)$ is p^{2t} . Moreover, under assumption that the path $vw_1v_1w_2 \dots v_{t-1}w_{t-1}w_tv'$ is present in $\mathcal{B}(n, m, p)$, the cardinality of the set $\mathcal{W}'(\{v, v'\}) \setminus \{w_1, \dots, w_t\}$ has the binomial distribution $\text{Bin}(m-t, 2(d_0/m) - p^2)$ with the expected value $\mu = (m-t)(2(d_0/m) - p^2) = (1 - o(1))2d_0$ as $d_0/m \geq p(1 - e^{-(n-1)p}) \geq p(1 - e^{-1}) \min\{1, (n-1)p\} \gg p^2$ for $p \leq \sqrt{3 \ln n/m} = o(1)$. Since $0.2d_0 = 0.1 \cdot (1 + o(1))\mu$, by Chernoff's inequality the probability that there exist two SMALL vertices at distance at most 4 is upper bounded by

$$\begin{aligned} \sum_{t=1}^4 n^{t+1} m^t p^{2t} \exp(-\psi(0.1(1+o(1)))(1+o(1))2d_0) \\ = O(1)nd_1^4 n^{-4/3} = O(n^{-1/3} n^{4\varepsilon} \ln^8 n) = o(1). \end{aligned}$$

Now we will prove property **P5** as it will be needed in the proof of **P3**.

P5 Let $k = \frac{\ln n}{\ln \ln n} \max\{4, np\}$. Recall that $V(w)$ has the binomial distribution $\text{Bin}(n, p)$ and that we

have $p \leq \sqrt{3 \ln n/m}$. Therefore

$$\begin{aligned}
\Pr \{ \exists_{w \in \mathcal{W}} V(w) \geq k \} &\leq m \binom{n}{k} p^k \\
&\leq m \left(\frac{enp}{k} \right)^{k-2} (np)^2 \\
&\leq m \exp \left((k-2) \left(1 + \ln(np) - \ln \frac{\ln n}{\ln \ln n} - \ln \max\{4, np\} \right) \right) \\
&\quad \cdot n^2 \frac{3 \ln n}{m} \\
&\leq n^{-3} \cdot n^2 \cdot 3 \ln n = o(1).
\end{aligned}$$

P3 The structure of $\mathcal{G}(n, m, p)$ differs a lot depending on the value np . Therefore we will need to divide the proof into cases:

$$np \leq 40 \quad \text{and} \quad |\mathcal{S}| \leq \frac{n}{d_1^4}; \quad (1)$$

$$np \leq 40 \quad \text{and} \quad \frac{n}{d_1^4} \leq |\mathcal{S}| \leq \frac{n}{d_1}; \quad (2)$$

$$np > 40. \quad (3)$$

First assume that (1) is fulfilled. We will first prove that in this case with high probability for any $\mathcal{K} \subseteq \mathcal{V}$, if $|\mathcal{K}| \leq n/d_1^3$ we have $W''(\mathcal{K}) \leq 2|\mathcal{K}|$.

$$\begin{aligned}
\Pr \left\{ \exists_{\mathcal{K} \subseteq \mathcal{V}, |\mathcal{K}| \leq n/d_1^3} W''(\mathcal{K}) \geq 2|\mathcal{K}| \right\} &\leq \sum_{k=2}^{n/d_1^3} \binom{n}{k} \binom{m}{2k} \left[\binom{k}{2} p^2 \right]^{2k} \\
&\leq \sum_{k=2}^{n/d_1^3} \left[\frac{en}{k} \cdot \frac{e^2 m^2}{2^2 k^2} \cdot \frac{k^4}{4} p^4 \right]^k \\
&\leq \sum_{k=2}^{n/d_1^3} \left[\frac{e^3}{16} \cdot \frac{k}{n} (nmp^2)^2 \right]^k \\
&\leq \sum_{k=2}^{n/d_1^3} \left(\frac{e^3}{16} \cdot \frac{1}{d_1} \right)^k = o(1).
\end{aligned}$$

Moreover if $np \leq 40$, and **P5** is fulfilled, then

$$\forall_{w \in \mathcal{W}} V(w) \leq \frac{\ln n}{\ln \ln n} \max\{4, np\} \leq \frac{40 \ln n}{\ln \ln n}. \quad (4)$$

Assume that \mathcal{G} is an intersection graph such that (4) is true and

$$\forall_{\mathcal{K} \subseteq \mathcal{V}, |\mathcal{K}| \leq n/d_1^3} W''(\mathcal{K}) \leq 2|\mathcal{K}|, \quad (5)$$

are fulfilled. In \mathcal{G} , let $\mathcal{S} \subseteq \text{LARGE}$ be such that $|\mathcal{S}| \leq n/d_1^4$. We will find a lower bound for $W'(\mathcal{S})$. Let \mathcal{B} be the bipartite graph associated with \mathcal{G} . Then $W'(\mathcal{S})$ is at least the number of edges in \mathcal{B} between \mathcal{S} and $\mathcal{W}'(\mathcal{S}) \setminus \mathcal{W}''(\mathcal{S})$. There are $\sum_{v \in \mathcal{S}} W'(v)$ edges between \mathcal{S} and $\mathcal{W}'(\mathcal{S})$ in \mathcal{B} and at most $\sum_{w \in \mathcal{W}''(\mathcal{S})} V(w)$ of them are incident to some features from $\mathcal{W}''(\mathcal{S})$, therefore by (4) and (5)

$$\begin{aligned} W'(\mathcal{S}) &\geq \sum_{v \in \mathcal{S}} W'(v) - \sum_{w \in \mathcal{W}''(\mathcal{S})} V(w) \\ &\geq 0.1d_0|\mathcal{S}| - 2|\mathcal{S}| \frac{40 \ln n}{\ln \ln n} \geq 0.1(1 - o(1))d_0|\mathcal{S}|. \end{aligned}$$

Set $\mathcal{K} = \mathcal{S} \cup \mathcal{N}(\mathcal{S})$. Assume that $N(\mathcal{S}) \leq b_1 d_1 |\mathcal{S}|$. Then, using (5), as $\mathcal{W}'(\mathcal{S}) \subseteq \mathcal{W}''(\mathcal{K})$,

$$\begin{aligned} W'(\mathcal{S}) &\leq W''(\mathcal{K}) \leq 2|\mathcal{K}| \leq 2(|\mathcal{S}| + b_1 d_1 |\mathcal{S}|) \\ &\leq 2(1 + b_1 \cdot 41 \cdot d_0)|\mathcal{S}| = 82(1 + o(1))b_1 d_0 |\mathcal{S}|. \end{aligned}$$

Combining the above inequalities we get

$$0.1(1 - o(1))d_0|\mathcal{S}| \leq 82(1 + o(1))b_1 d_0 |\mathcal{S}|.$$

Thus

$$\frac{1}{1000} = b_1 \geq (1 + o(1)) \frac{1}{820},$$

a contradiction. Therefore in case (1) we have with high probability

$$N(\mathcal{S}) \geq b_1 d_1 |\mathcal{S}|.$$

Now assume that (2) is fulfilled and \mathcal{S} is any subset of \mathcal{V} , not necessarily a subset of LARGE . Note that $W(\mathcal{S})$, $|\mathcal{S}| = s$, has the binomial distribution $\text{Bin}(m, 1 - (1 - p)^s)$. Therefore, for large n , $W(\mathcal{S})$ stochastically dominates a random variable X_s with the binomial distribution $\text{Bin}(m, sp(1 - sp))$ (recall that $sp \leq np/d_1 = 1/mp = o(1)$). Therefore

$$\begin{aligned} &\sum_{\mathcal{S}, n/d_1^4 \leq |\mathcal{S}| \leq n/d_1} \Pr \{W(\mathcal{S}) < 0.05|\mathcal{S}|mp\} \\ &\leq \sum_{s=n/d_1^4}^{n/d_1} \binom{n}{s} \Pr \{X_s < 0.05smp\} \\ &\leq \sum_{s=n/d_1^4}^{n/d_1} (ed_1^4)^s \exp \left(-\psi \left(\frac{0.05}{1 - sp} \right) smp(1 - sp) \right) \\ &\leq \sum_{s=n/d_1^4}^{n/d_1} \exp (s(1 + 4 \ln(mp) + 4 \ln(np) - 0.8mp)) = o(1). \end{aligned} \tag{6}$$

Now we will show a similar bound in the case where (3) is fulfilled. In this case, if $\mathcal{S} \subseteq \text{LARGE}$, $|\mathcal{S}| = s$, then for all $v \in \mathcal{S}$

$$W(v) \geq W'(v) \geq 0.1d_0 \geq 0.09mp.$$

Moreover, for any $v \in \mathcal{V}$, given $W(v)$, the set $\mathcal{W}(v)$ is uniformly distributed among all subsets of \mathcal{W} of cardinality $W(v)$. Therefore, if for all $v \in \mathcal{S}$ we have $W(v) \geq 0.09pm$ then the probability that $W(\mathcal{S})$ is at most $0.05smp$ is at most

$$\binom{m}{0.05smp} \left(\frac{0.05smp}{m} \right)^{0.09smp} = (e^{1.25} \cdot 0.05 \cdot sp)^{0.04smp}.$$

Recall that $sp \leq np/d_1 = 1/mp = o(1)$. Therefore

$$\begin{aligned} & \sum_{\mathcal{S} \subseteq \mathcal{V}, |\mathcal{S}| \leq n/d_1} \Pr \{ \{W(\mathcal{S}) \leq 0.05|\mathcal{S}|mp\} \cap \{\mathcal{S} \subseteq \text{LARGE}\} \} \\ & \leq \sum_{\mathcal{S} \subseteq \mathcal{V}, |\mathcal{S}| \leq n/d_1} \Pr \{ \{W(\mathcal{S}) \leq 0.05|\mathcal{S}|mp\} \cap \{ \forall v \in \mathcal{S} W(v) \geq 0.09mp \} \} \\ & \leq \sum_{s=1}^{n/d_1} \binom{n}{s} (e^{1.25} \cdot 0.05 \cdot sp)^{0.04smp} \\ & \leq \sum_{s=1}^{n/d_1} \exp(s(\ln n + 0.04mp(1.25 + \ln 0.05 + \ln sp))) = o(1). \end{aligned} \tag{7}$$

Now we will find a bound on the probability that for some $\mathcal{S} \subseteq \text{LARGE}$ the size of its neighbourhood $N(\mathcal{S})$ is smaller than $b_1|\mathcal{S}|d_1$ with $b_1 = 0.001$. In both cases, (2) and (3), the proof is similar.

For any $\mathcal{S} \subseteq \mathcal{V}$, given the value $W(\mathcal{S})$, $N(\mathcal{S})$ has the binomial distribution with parameters $n - |\mathcal{S}|$ and $1 - (1-p)^{W(\mathcal{S})}$, i.e. $\text{Bin}(n - |\mathcal{S}|, 1 - (1-p)^{W(\mathcal{S})})$. If $W(\mathcal{S}) \geq 0.05|\mathcal{S}|mp$ then, for large n , $N(\mathcal{S})$ stochastically dominates a random variable with the binomial distribution $\text{Bin}(n - \frac{n}{d_1}, 0.04|\mathcal{S}|mp^2)$ (note that $|\mathcal{S}|mp^2 \leq 1$ as $|\mathcal{S}| \leq n/d_1$). Therefore

$$\begin{aligned} & \Pr \{ \{N(\mathcal{S}) \leq 0.001|\mathcal{S}|d_1\} \cap \{\mathcal{S} \subseteq \text{LARGE}\} \} \\ & \leq \Pr \{ N(\mathcal{S}) \leq 0.001|\mathcal{S}|d_1 \mid W(\mathcal{S}) \geq 0.05|\mathcal{S}|mp \} + \Pr \{ \{W(\mathcal{S}) < 0.05|\mathcal{S}|mp\} \cap \{\mathcal{S} \subseteq \text{LARGE}\} \} \\ & \leq \exp \left(-\psi \left(\frac{0.025}{1 - d_1^{-1}} \right) (1 - d_1^{-1})0.04|\mathcal{S}|nmp^2 \right) + \Pr \{ \{W(\mathcal{S}) < 0.05|\mathcal{S}|mp\} \cap \{\mathcal{S} \subseteq \text{LARGE}\} \}. \end{aligned}$$

For $|\mathcal{S}| \geq n/d_1^4$

$$\psi \left(\frac{0.025}{1 - d_1^{-1}} \right) (1 - d_1^{-1})0.04d_1 \geq 0.03d_1.$$

For $np > 40$

$$\psi \left(\frac{0.025}{1 - d_1^{-1}} \right) (1 - d_1^{-1})0.04nmp^2 > 1.4mp > 1.4 \cdot 0.9 \ln n > 1.2 \ln n.$$

Thus finally in the case (2), using (6), we get

$$\begin{aligned}
& \Pr \left\{ \exists_{\mathcal{S} \subseteq \mathcal{V}, n/d_1^4 \leq |\mathcal{S}| \leq n/d_1} \{N(\mathcal{S}) \leq 0.001|\mathcal{S}|d_1\} \cap \{\mathcal{S} \subseteq \text{LARGE}\} \right\} \\
& \leq \sum_{\mathcal{S} \subseteq \mathcal{V}, n/d_1^4 \leq |\mathcal{S}| \leq n/d_1} \Pr \left\{ \{N(\mathcal{S}) \leq 0.001|\mathcal{S}|d_1\} \cap \{\mathcal{S} \subseteq \text{LARGE}\} \right\} \\
& \leq \sum_{\mathcal{S} \subseteq \mathcal{V}, n/d_1^4 \leq |\mathcal{S}| \leq n/d_1} \Pr \left\{ N(\mathcal{S}) \leq 0.001|\mathcal{S}|d_1 \mid W(\mathcal{S}) \geq 0.05|\mathcal{S}|mp \right\} \\
& \quad + \sum_{\mathcal{S} \subseteq \mathcal{V}, n/d_1^4 \leq |\mathcal{S}| \leq n/d_1} \Pr \{W(\mathcal{S}) < 0.05|\mathcal{S}|mp\} \\
& \leq \sum_{s=n/d_1^4}^{n/d_1} \binom{n}{s} \exp(-0.3sd_1) + o(1) \\
& \leq \sum_{s=n/d_1^4}^{n/d_1} \exp \left(s \left(1 + \ln \frac{n}{s} - 0.3d_1 \right) \right) + o(1) \\
& \leq \sum_{s=n/d_1^4}^{n/d_1} \exp (s(1 + 4 \ln d_1 - 0.3d_1)) + o(1) = o(1)
\end{aligned}$$

Similarly in case (3), using (7),

$$\begin{aligned}
& \Pr \left\{ \exists_{\mathcal{S} \subseteq \mathcal{V}, |\mathcal{S}| \leq n/d_1} \{N(\mathcal{S}) \leq 0.001|\mathcal{S}|d_1\} \cap \{\mathcal{S} \subseteq \text{LARGE}\} \right\} \\
& \leq \sum_{s=1}^{n/d_1} \binom{n}{s} \exp(-1.2s \ln n) + o(1) \\
& \leq \sum_{s=1}^{n/d_1} \exp (s(\ln n - 1.2 \ln n)) + o(1) = o(1)
\end{aligned}$$

Finally we get that in all cases with high probability

$$\forall_{\mathcal{S} \subseteq \text{LARGE}, |\mathcal{S}| \leq n/d_1} N(\mathcal{S}) \geq 0.001|\mathcal{S}|d_1.$$

P4 We will show the statement only for $N(v)$. $W(v)$ and $W'(v)$ have the binomial distribution $\text{Bin}(m, p)$ and $\text{Bin}(m, d_0/m)$, resp. therefore the proofs of the remaining statements are similar but easier. Note that if $d_1 > n/12$ then the statement holds trivially. In the latter case, given $W(v) \leq 4mp$, $N(v)$ is stochastically dominated by a binomial random variable with the binomial distribution $\text{Bin}(n, 4d_1/n)$ (as $1 - (1-p)^{W(v)} \leq 1 - (1-p)^{4mp} \leq 4mp^2$). Having in mind that $d_1 \geq d_0 \geq (1 + o(1)) \ln n$ and

$$mp \geq d_0 \geq (1 + o(1)) \ln n$$

$$\begin{aligned} \Pr \{ \exists v \in \mathcal{V} N(v) \geq 12d_1 \} &\leq n (\Pr \{ N(v) \geq 12d_1 \mid W(v) \leq 4mp \} + \Pr \{ W(v) \geq 4mp \}) \\ &\leq n \binom{n}{12d_1} \left(\frac{4d_1}{n} \right)^{12d_1} + n \binom{m}{4mp} p^{4mp} \\ &\leq n \left(\frac{e}{3} \right)^{12d_1} + n \left(\frac{e}{4} \right)^{4mp} \\ &\leq \exp(\ln n + 12(1 - \ln 3)d_1) + \exp(\ln n + 4(1 - \ln 4)mp) = o(1) \end{aligned}$$

□

5 Deletable sets

In this section we will mainly analyse an instance \mathcal{G} of the random intersection graph. \mathcal{G} will have properties listed in the previous section and will be such that HAM does not find a Hamilton cycle in \mathcal{G} . We will first create a random subgraph \mathcal{G}_q of \mathcal{G} by randomly deleting connections between vertices and attributes in the bipartite graph \mathcal{B} associated with \mathcal{G} . We will bound the probability that execution of HAM on \mathcal{G}_q proceeds the same steps as on \mathcal{G} and \mathcal{G}_q has good expansion properties. As a tool we will use a so called *deletable set* of edges. Namely we will bound the probability that the edges from $E(\mathcal{G}) \setminus E(\mathcal{G}')$ form a deletable set.

In the analysis of HAM we will use some additional notation. In most cases it is consistent with Bollobás et al. (1987). If HAM terminates unsuccessfully on stage k then

$\mathcal{H}(G)$ is the set of paths $P^{(1)}, \dots, P^{(M)}$ and edges of the form $\{u_0, u_1\}$, where:

- $P_1 = P^{(1)}, \dots, P^{(M)} = P_k$ is the sequence of paths constructed by HAM, where $P^{(i+1)}$ is obtained from $P^{(i)}$ by a simple extension, cycle extension, or rotation and
- $\{u_0, u_1\}$ are endpoints of the paths on which HAM executes a cycle extension;

$\text{END}(G) = \{v \in \mathcal{V} : \text{there exists a path } Q_s \text{ on stage } k \text{ with } v \text{ as an endpoint and } \bar{\delta}(Q_s) = t \text{ for some } 1 \leq t \leq T\};$

and for $x \in \text{END}(G)$

$\text{END}(G, x) = \{v \in \mathcal{V} : \text{there exists a path } Q_s \text{ on stage } k \text{ with } v \text{ and } x \text{ as endpoints and } \bar{\delta}(Q_s) = t \text{ for some } 1 \leq t \leq 2T\};$

Let \mathcal{G} be such that its vertex set is divided into two disjoint sets LARGE and SMALL and HAM with input value d terminates unsuccessfully on \mathcal{G} . Given a constant $b_2 > 0$, we call $X \subseteq E(\mathcal{G})$ *deletable with b_2* if

- D1** no edge of X is incident to a vertex from SMALL;
- D2** for any $v \in \text{LARGE}$ there are at most $b_2 d$ edges from X incident to v ;
- D3** $X \cap \mathcal{H}(\mathcal{G}) = \emptyset$.

From the description of the algorithm it follows that

$$|\mathcal{H}(\mathcal{G})| \leq n(2T + 2), \quad (8)$$

where T is defined as in the algorithm, *i.e.* for $d = \Omega(\ln n)$ we have $T = o(\ln n)$.

Following the lines of the proof of Lemma 3.2 from Bollobás et al. (1987) one may show the following lemma (see the proof in Appendix D).

Lemma 3. *Let $\{\mathcal{G}_n\}_{n=1,2,\dots}$ be a sequence of graphs on n vertices such that for large n HAM with input value d , $d = d(n) = \Omega(\ln n)$, terminates unsuccessfully in stage k on \mathcal{G}_n . Moreover let $\mathcal{G}'_n \subseteq \mathcal{G}_n$, b_3 ($b_3 > 0$) and b_4 ($0 \leq b_4 < \min\{b_3, 1/3\}$) be constants and $\delta_0 = \max\{1, b_4 \min\{\delta(\mathcal{G}'_n), d\}\}$. If for large n*

(i) $\delta(\mathcal{G}'_n) \geq 2$;

(ii) *There exists a set $\mathcal{V}' \subseteq \mathcal{V}$ such that:*

- for every v there are at most δ_0 vertices from $\mathcal{V} \setminus \mathcal{V}'$ at distance at most 2 from v ;
- for all $\mathcal{S} \subseteq \mathcal{V}'$ such that $|\mathcal{S}| \leq n/d$, $N_{\mathcal{G}'_n}(\mathcal{S}) \geq b_3 d |\mathcal{S}|$;

(iii) $\mathcal{H}(\mathcal{G}_n) \subseteq E(\mathcal{G}'_n)$;

then HAM terminates unsuccessfully in stage k on \mathcal{G}'_n and for any constant $b_5 < (b_3 - b_4)/2$ and large n

$$\begin{aligned} |\text{END}(\mathcal{G}')| &\geq b_5 n; \\ |\text{END}(\mathcal{G}', x)| &\geq b_5 n, \quad \text{for any } x \in \text{END}(\mathcal{G}'). \end{aligned}$$

Remark 3. Let LARGE and SMALL be defined as in Lemma 2. Let \mathcal{G} be an instance of $\mathcal{G}(n, m, p)$ with properties **P0–P5** and let $d = d_1$. Moreover let X be deletable with $b_2 < b_1$ in \mathcal{G} , $\mathcal{G}' = (\mathcal{V}, E(\mathcal{G}) \setminus X)$, and let $\mathcal{V}' = \text{LARGE}$. Then **P0**, **P3**, **D1**, and **D2** imply (i) for n large. Moreover, **D3** implies (iii), and **D1**, **D2**, **P2**, and **P3** imply (ii) with $b_4 = 0$ and $b_3 < b_1 - b_2$.

Assume that we have an instance \mathcal{B} of $\mathcal{B}(n, m, p)$ and intersection graph \mathcal{G} associated with \mathcal{B} . Denote by \mathcal{B}_q a random subgraph of \mathcal{B} obtained by deleting each edge of \mathcal{B} independently with probability $q = \lambda/n$, where λ is a positive constant. Moreover let \mathcal{G}_q be a random subgraph of \mathcal{G} associated with \mathcal{B}_q . For any \mathcal{G} and its subgraph \mathcal{G}_q let X_q be the set of edges included in \mathcal{G} but absent in \mathcal{G}_q . Consider a probability space in which we first pick \mathcal{G} according to the probability distribution of $\mathcal{G}(n, m, p)$ and then create \mathcal{G}_q . The outcome of the experiment is a pair of graphs $(\mathcal{G}, \mathcal{G}_q)$. Denote by $\mathcal{G}_q(n, m, p)$ the random graph \mathcal{G}_q constructed in this manner and by $X_q(n, m, p)$ the set X_q . If an edge was present in \mathcal{G} (\mathcal{B} , resp.) and is absent in \mathcal{G}_q (\mathcal{B}_q , resp) we will say that the edge was *deleted* in \mathcal{G}_q (\mathcal{B}_q , resp).

Lemma 4. *Let \mathcal{G} be an intersection graph on n vertices and let LARGE and SMALL be defined as in Lemma 2. If \mathcal{G} has properties **P0–P5** and HAM terminates unsuccessfully on \mathcal{G} then*

$$\Pr \left\{ X_q(n, m, p) \text{ is deletable with } b_2 = 0.5b_1 \mid \mathcal{G}(n, m, p) = \mathcal{G} \right\} \geq (1 + o(1)) \exp(-2\lambda(2T + 2))$$

where $o(1)$ is uniformly bounded over all possible choices of \mathcal{G} .

Proof: Recall that X_q is deletable if it has properties **D1**, **D2**, and **D3**.

Set features in \mathcal{W} in any order $\{w_1, \dots, w_m\}$. For any edge $\{v, v'\}$ let $w_{vv'}$ be the smallest (in order of \mathcal{W}) feature in $\mathcal{W}''(\{v, v'\})$ (neighbour of both v and v' in \mathcal{B}).

For any $v \in \mathcal{V}$, $v' \in \mathcal{N}(v)$ and $w \in \mathcal{W}'(v)$ define events:

$A_{vv'}$ – event that neither $\{v, w_{vv'}\}$ nor $\{w_{vv'}, v'\}$ was deleted in \mathcal{B}_q ;

$B_{vv'}$ – event that $\{v', w_{vv'}\}$ was deleted in \mathcal{B}_q ;

C_{vw} – event that $\{v, w\}$ was deleted in \mathcal{B}_q .

Consider events

$$\begin{aligned} A &= \bigcap_{v \in \text{SMALL}, v' \in \mathcal{N}(v)} A_{vv'} \cap \bigcap_{\{v, v'\} \in \mathcal{H}(\mathcal{G})} A_{vv'}, \\ B &= \bigcup_{v \in \text{LARGE}} \bigcup_{\mathcal{S} \subseteq \mathcal{N}(v), |\mathcal{S}| = 0.1b_1d_0} \bigcap_{v' \in \mathcal{S}} B_{vv'}, \\ C &= \bigcup_{v \in \text{LARGE}} \bigcup_{\mathcal{R} \subseteq \mathcal{W}'(v), |\mathcal{R}| = 0.1 \frac{\ln \ln n}{\ln n} b_1d_0} \bigcap_{w \in \mathcal{R}} C_{vw}, \end{aligned}$$

and event

$$A \cap (B \cup C)^c = A \cap B^c \cap C^c,$$

where, for any event D , by D^c we denote the complement of event D . First we will prove that if event $A \cap B^c \cap C^c$ occurs then X_q is deletable with $b_2 = 0.5b_1 = 0.0005$. $\bigcap_{v \in \text{SMALL}, v' \in \mathcal{N}(v)} A_{vv'}$ implies **D1** and $\bigcap_{\{v, v'\} \in \mathcal{H}(\mathcal{G})} A_{vv'}$ implies **D3**. We are left with showing that $B^c \cap C^c$ implies **D2**.

Let $v \in \text{LARGE}$. Note that if an edge between v and $\mathcal{N}(v)$ is deleted in \mathcal{G}_q then there is a deleted edge in \mathcal{B}_q between $\mathcal{W}'(v)$ and $\mathcal{N}(v)$ or between v and $\mathcal{W}'(v)$.

Let $\mathcal{S} \subseteq \mathcal{N}(v)$ be the largest set such that, for all $v' \in \mathcal{S}$, all edges between v' and $\mathcal{W}'(v)$ were deleted in \mathcal{B}_q . Event B^c implies that $|\mathcal{S}| \leq 0.1b_1d_0 \leq 0.1b_1d_1$, therefore there are at most $0.1b_1d_1$ edges incident to v deleted in \mathcal{G}_q due to deletions of edges between $\mathcal{W}'(v)$ and $\mathcal{N}(v)$ in \mathcal{B}_q .

Let $\mathcal{R} \subseteq \mathcal{W}'(v)$ be the largest set such that, for all $w \in \mathcal{R}$, $\{v, w\}$ was deleted in \mathcal{B}_q . Event C^c implies that $|\mathcal{R}| \leq 0.1 \frac{\ln \ln n}{\ln n} b_1d_0$. Property **P5** implies that

$$\begin{aligned} V_{\mathcal{G}_q}(\mathcal{R}) &\leq |\mathcal{R}| \frac{\ln n}{\ln \ln n} \max\{np, 4\} \\ &\leq 0.1 \frac{\ln \ln n}{\ln n} b_1d_0 \frac{\ln n}{\ln \ln n} \max\{np, 4\} \\ &\leq 0.4b_1d_0 \max\{np, 1\} \leq 0.4b_1mp \min\{np, 1\} \max\{np, 1\} = 0.4b_1d_1. \end{aligned}$$

Therefore there are at most $0.4b_1d_1$ edges between v and $\mathcal{N}(v)$ deleted in \mathcal{G}_q due to deletions of edges between v and $\mathcal{W}'(v)$ in \mathcal{B}_q .

In conclusion, $B^c \cap C^c$ and **P5** imply that for any vertex $v \in \text{LARGE}$ there are at most $0.5b_1d_1$ edges incident to v in X_q .

Now we will bound the probability that, given \mathcal{G} with properties **P0–P5**, $A \cap (B \cup C)^c$, occurs. Let $\Pr_{\mathcal{G}} \{\cdot\} = \Pr \{\cdot \mid \mathcal{G}(n, m, p) = \mathcal{G}\}$.

In what follows we will use the fact that, given $\mathcal{G} = \mathcal{G}(n, m, p)$, the edges in \mathcal{B}_q are deleted independently. Moreover, if we know that some set of edges of \mathcal{B} , say E , was not deleted in \mathcal{B}_q , then for any other set of edges E' of \mathcal{B} (possibly intersecting with E) the probability that all edges from E' were not deleted (were deleted, resp.) is at least $(1 - q)^{|E'|}$ (at most $q^{|E'|}$, resp.).

Now we will bound the number of edges incident to vertices from SMALL. Recall that by **P1**, if $d_0 \geq 2 \ln n$ then $|\text{SMALL}| = 0$. Therefore we need only to consider the case $d_0 \leq 2 \ln n$. If $m \geq n^{1-\varepsilon}$ and $d_0 \leq 2 \ln n$ then $d_1 = O(n^\varepsilon \ln^2 n)$ (see discussion in the proof of **P2**). Therefore properties **P1** and **P4** imply that there are at most

$$|\text{SMALL}| \cdot 12d_1 = O(n^{1/3} n^{1/25} \ln^2 n) = o(n)$$

edges in \mathcal{G} incident to vertices from SMALL. In addition, by (8), $|\mathcal{H}(\mathcal{G})| \leq n(2T + 2)$. Therefore

$$\Pr_{\mathcal{G}} \{A\} \geq ((1 - q)^2)^{o(n)} \cdot (1 - q)^{2|\mathcal{H}(\mathcal{G})|} = (1 - o(1)) \exp(-2\lambda(2T + 2)).$$

Moreover by **P4** and the fact that $p \leq \sqrt{3 \ln n/m} = o(1)$

$$\begin{aligned} \Pr_{\mathcal{G}} \{B|A\} &\leq \sum_{v \in \text{LARGE}} \sum_{S \subseteq \mathcal{N}(v), |S|=0.1b_1d_0} \Pr \left\{ \bigcap_{v' \in S} B_{vv'} \mid A \right\} \\ &\leq n \binom{12d_1}{0.1b_1d_0} q^{0.1b_1d_0} = n \left(\frac{12e}{0.1b_1} \cdot 2 \max\{np, 1\} \cdot \frac{\lambda}{n} \right)^{0.1b_1d_0} \\ &\leq n (O(1) \max\{p, n^{-1}\})^{0.1b_1(1+o(1)) \ln n} = o(1). \end{aligned}$$

and

$$\begin{aligned} \Pr_{\mathcal{G}} \{C|A\} &\leq \sum_{v \in \text{LARGE}} \sum_{\mathcal{R} \subseteq \mathcal{W}'(v), |\mathcal{R}|=0.1b_1 \frac{\ln \ln n}{\ln n} d_0} \Pr_{\mathcal{G}} \left\{ \bigcap_{w \in \mathcal{R}} C_{vw} \mid A \right\} \\ &\leq n \binom{4d_0}{0.1b_1 \frac{\ln \ln n}{\ln n} d_0} q^{0.1b_1 \frac{\ln \ln n}{\ln n} d_0} \\ &\leq n \left(\frac{e4}{0.1b_1} \cdot \frac{\ln n}{\ln \ln n} \cdot \frac{\lambda}{n} \right)^{0.1b_1 \frac{\ln \ln n}{\ln n} d_0} = o(1). \end{aligned}$$

Therefore, given an instance \mathcal{G} of $\mathcal{G}(n, m, p)$ with properties **P0–P5** we have

$$\begin{aligned} \Pr_{\mathcal{G}} \{A \cap (B \cup C)^c\} &= \Pr_{\mathcal{G}} \{A\} (1 - \Pr_{\mathcal{G}} \{B \cup C|A\}) \\ &\geq \Pr_{\mathcal{G}} \{A\} (1 - \Pr_{\mathcal{G}} \{B|A\} - \Pr_{\mathcal{G}} \{C|A\}) \geq (1 - o(1)) \exp(-2\lambda(2T + 2)), \end{aligned}$$

where $o(1)$ is uniformly bounded over all possible choices of \mathcal{G} with properties **P0–P5**. \square

6 Probability of HAM terminating unsuccessfully

Recall that we consider pairs $(\mathcal{G}, \mathcal{G}_q)$, where \mathcal{G} is chosen according to the probability distribution of $\mathcal{G}(n, m, p)$ and \mathcal{G}_q is a random subgraph of \mathcal{G} . In previous sections we already bounded the probability

that in the case when HAM terminates unsuccessfully on \mathcal{G} during the execution of HAM on \mathcal{G}_q a linear number of longest paths is created. In this section we will additionally prove that it is highly unlikely that among the large number of paths in \mathcal{G}_q none of them close a cycle in \mathcal{G} . This will lead us to the bound on the probability that HAM terminates unsuccessfully on $\mathcal{G}(n, m, p)$.

Define LARGE and SMALL as in Lemma 2. For a pair $(\mathcal{G}(n, m, p), \mathcal{G}_q(n, m, p))$ analysed in the previous section define events

D_1 – $\mathcal{G}(n, m, p)$ has properties **P1–P5** with $b_1 = 1/1000$;

$D_2 - \delta(\mathcal{G}(n, m, p)) \geq 2$ ($\mathcal{G}(n, m, p)$ has property **P0**);

D_3 – HAM terminates unsuccessfully on $\mathcal{G}(n, m, p)$;

E_1 – HAM terminates unsuccessfully on $\mathcal{G}_q(n, m, p)$;

We consider also two other events contained in E_1

E_2 – there is no edge $\{v, v'\} \in X_q(n, m, p)$ such that $v \in \text{END}(\mathcal{G}_q(n, m, p))$ and $v' \in \text{END}(\mathcal{G}_q(n, m, p), v)$;

E_3 – $|\text{END}(\mathcal{G}_q(n, m, p))| \geq n/5000$ and $|\text{END}(\mathcal{G}_q(n, m, p), v)| \geq n/5000$, for all $v \in \text{END}(\mathcal{G}_q(n, m, p))$.

Let moreover

$$D = D_1 \cap D_2 \cap D_3 \quad \text{and} \quad E = E_1 \cap E_2 \cap E_3.$$

Note that by Lemma 3 and Remark 3, D and the event that $X_q(n, m, p)$ is deletable with constant $b_2 = 0.5b_1 = 1/2000$ imply event E . Therefore by Lemma 4

$$\begin{aligned} \Pr\{D \cap E\} &\geq \Pr\{E|D\} \Pr\{D\} \\ &\geq \Pr\{X_q(n, m, p) \text{ is deletable with } b_2 = 0.5b_1 | D\} \Pr\{D\} \\ &\geq (1 - o(1)) \exp(-2\lambda(2T + 2)) \Pr\{D\}. \end{aligned} \tag{9}$$

Now we will find an upper bound on the probability $\Pr\{D \cap E\}$. Note that $\mathcal{G}_q(n, m, p)$ is distributed as $\mathcal{G}(n, m, p')$ associated with $\mathcal{B}(n, m, p')$, where $p' = p(1 - q)$.

First we bound the probability of event F^c , where

$$F - \forall_{\mathcal{S}, |\mathcal{S}|=n/5000} W(\mathcal{S}) = W_{\mathcal{G}(n, m, p')}(\mathcal{S}) \geq 0.3m \min\left\{\frac{np'}{5000}, 1\right\} \text{ in } \mathcal{G}(n, m, p').$$

Assume that $n \leq m$. Recall that if $|\mathcal{S}| = n/5000$ then $W(\mathcal{S})$ has the binomial distribution $\text{Bin}(m, 1 - (1 - p')^{n/5000})$ and $1 - (1 - p')^{n/5000} \geq 0.5 \min\{np'/5000, 1\}$. As $mp' \geq (1 + o(1)) \ln n$ and $n \leq m$

$$\mathbb{E}W(\mathcal{S}) \geq 0.5m \min\left\{\frac{np'}{5000}, 1\right\} \geq \frac{n}{5000} 0.5 \min\left\{mp', \frac{5000m}{n}\right\} \geq \frac{n}{5000} \cdot 2500.$$

Therefore, if $n \leq m$ then

$$\begin{aligned}
\Pr \{F^c\} &= \Pr \left\{ \exists S \subseteq \mathcal{W}, |S| = \frac{n}{5000} W(S) \leq 0.3m \min \left\{ \frac{np'}{5000}, 1 \right\} \right\} \\
&\leq \binom{n}{n/5000} \exp \left(-\psi(0.6) 0.5m \min \left\{ \frac{np'}{5000}, 1 \right\} \right) \\
&\leq \exp \left(\frac{n}{5000} (1 + \ln 5000 - \psi(0.6) \cdot 2500) \right) \\
&\leq \exp \left(\frac{n}{5000} (1 + \ln 5000 - 233) \right) \leq \exp(-0.04n).
\end{aligned} \tag{10}$$

Now assume that $n \geq m$. Then $np' = (n/m)mp' \geq (1 + o(1)) \ln n$, therefore, for n large, we have $0.3m \min \left\{ \frac{np'}{5000}, 1 \right\} = 0.3m$. Let

$$V'' = |\{v \in \mathcal{V} : W(v) \leq 0.1mp'\}|.$$

Note that $\Pr \{W(v) \leq 0.1mp'\} \leq \exp(-\psi(0.1)mp') \leq n^{-2/3}$ and $W(v)$, $v \in \mathcal{V}$, are independent. V'' is therefore stochastically dominated by a random variable with the binomial distribution $\text{Bin}(n, n^{-2/3})$. In addition, given $W(v) = t \geq 0.1mp'$, the set $\mathcal{W}(v)$ is distributed uniformly over all subsets of \mathcal{W} of cardinality t (independently of all other vertices from \mathcal{V}). Thus, for $n \geq m$,

$$\begin{aligned}
\Pr \{F^c\} &\leq \Pr \{F^c | V'' \leq n^{1/2}\} + \Pr \{V'' \geq n^{1/2}\} \\
&\leq \binom{n}{\frac{n}{5000}} \binom{m}{0.3m} \left(\frac{0.3m}{m} \right)^{\left(\frac{n}{5000} - n^{1/2}\right) \cdot 0.1mp'} + \binom{n}{n^{1/2}} (n^{-2/3})^{n^{1/2}} \\
&\leq (e5000)^{\frac{n}{5000}} \left(\frac{e}{0.3} \right)^{0.3m} 0.3^{(1+o(1))\frac{n}{5000}mp'} + (en^{1/2}n^{-2/3})^{n^{1/2}} \\
&\leq \exp(-n^{1/2}).
\end{aligned}$$

Recall that $\mathcal{B}_q(n, m, p)$ is distributed as $\mathcal{B}(n, m, p')$. Therefore $\mathcal{B}(n, m, p)$ is distributed as a random bipartite graph $\mathcal{B}_q(n, m, p) \cup \mathcal{B}(n, m, q')$, where $\mathcal{B}_q(n, m, p)$ and $\mathcal{B}(n, m, q')$ are independent and we have $q' = pq/(1-p+pq) = (1+o(1))pq$ (recall that we assume that $\ln n = o(m)$ i.e. $p \leq \sqrt{3 \ln n/m} = o(1)$). If E_2 occurs then in $\mathcal{B}(n, m, q')$ for all $v \in \text{END}(\mathcal{G}')$ there is no edge between $\mathcal{W}(\text{END}(\mathcal{G}', v))$ and v , therefore

$$\begin{aligned}
\Pr \{E_2 | E_1 \cap E_3 \cap \bar{F}\} &\leq \prod_{v \in \text{END}(\mathcal{G}')} (1 - q')^{W(\text{END}(\mathcal{G}', v))} \\
&\leq \exp \left(-\frac{n}{5000} \cdot (1 + o(1))p \cdot \frac{\lambda}{n} \cdot 0.3m \min \left\{ \frac{np'}{5000}, 1 \right\} \right) \\
&\leq \exp \left(-(1 + o(1))\lambda \cdot \frac{3}{50000} \cdot \min \left\{ \frac{d_1}{5000}, mp \right\} \right) \\
&\leq \exp(-\lambda' \ln n)
\end{aligned}$$

for $\lambda' = \lambda \cdot 10^{-8}$.

Therefore by (10) and the above calculation

$$\Pr\{D \cap E\} \leq \Pr\{E_2 | E_1 \cap E_3 \cap F\} + \Pr\{F^c\} \leq (1 + o(1)) \exp\{-\lambda' \ln n\}.$$

Combining this equation with (9) we get

$$(1 - o(1)) \exp(-2\lambda(2T + 2)) \Pr\{D\} \leq (1 + o(1)) \exp\{-\lambda' \ln n\},$$

i.e.

$$\Pr\{D\} \leq (1 + o(1)) \exp\{-\lambda' \ln n + 2\lambda(2T + 2)\} = o(1).$$

Therefore by the above equation and Lemma 2

$$\begin{aligned} \Pr\{\{\delta(\mathcal{G}(n, m, p)) \geq 2\} \cap \{\text{HAM finishes unsuccessfully on } \mathcal{G}(n, m, p)\}\} \\ = \Pr\{D_2 \cap D_3\} \leq \Pr\{D_1 \cap D_2 \cap D_3\} + \Pr\{D_1^c\} = o(1). \end{aligned}$$

7 Proof for $m \leq n^{1-\varepsilon}$

We will not give the whole proof in the case $m \leq n^{1-\varepsilon}$ ($\varepsilon = 1/25$) and $mp^2 \leq 1$ as, in its main part, it follows the lines of the proof for $m \geq n^{1-\varepsilon}$. We just note several adjustments that we make.

We leave the definition of d but we redefine SMALL and LARGE.

$$\text{SMALL}_* = \{v \in \mathcal{V} : W'(v) \leq 6 \cdot 10^{-3} mp\}, \quad \text{LARGE}_* = \{v \in \mathcal{V} : W'(v) > 6 \cdot 10^{-3} mp\}.$$

We need to add that in this case

$$np = (n/m)mp \geq (1 + o(1))n^\varepsilon \ln n \quad \text{and} \quad mp = (1 + o(1))d_0. \quad (11)$$

We redefine the properties.

P0* $\delta(\mathcal{G}) \geq np/2 \geq n^\varepsilon \ln n/2$;

P1* $|\text{SMALL}_*| \leq n^\varepsilon$ and
if $mp \geq 2 \ln n$ then $|\text{SMALL}_*| = \emptyset$;

P3* For all $\mathcal{S} \subseteq \text{LARGE}_*$, if $|\mathcal{S}| \leq n/d_1$ then $N(\mathcal{S}) \geq b_1 d_1 |\mathcal{S}|$.

P4 For all $v \in \mathcal{V}$

$$W(v) \leq 4mp, \quad W'(v) \leq 4d_0, \quad N(v) \leq 12d_1.$$

P5 For all $w \in \mathcal{W}$

$$V(w) \leq \frac{\ln n}{\ln \ln n} \max\{np, 4\}.$$

Now we may prove the following analogue of Lemma 2

Lemma 5. Let $\mathcal{G}(n, m, p)$ be a random intersection graph and

$$d_0 \geq (1 + o(1)) \ln n.$$

Moreover let

$$SMALL_* = \{v \in \mathcal{V} : W'(v) \leq 6 \cdot 10^{-3} mp\}, \quad LARGE_* = \{v \in \mathcal{V} : W'(v) > 6 \cdot 10^{-3} mp\}.$$

If $m < n^{1-\varepsilon}$ ($\varepsilon = 1/25$) then with high probability $\mathcal{G}(n, m, p)$ has properties **P1***, **P3*** with $b_1 = 0.001$, **P4**, and **P5**.

Remark 4. If $m < n^{1-\varepsilon}$ then

$$\Pr \{ \{ \delta(\mathcal{G}(n, m, p)) \geq 2 \} \cap \{ \delta(\mathcal{G}(n, m, p)) \geq np/2 \}^c \} = o(1). \quad (12)$$

The proofs of Lemma 5 and Remark 4 are presented in Appendix C.

We also redefine the notion of a deletable set. Let \mathcal{G} be such that HAM terminates unsuccessfully on \mathcal{G} . Given constant $b_2 > 0$, we call $X \subseteq E(\mathcal{G})$ *deletable* with b_2 if

D1* for any $v \in SMALL_*$ there are at most $b_2 d_0$ edges from X incident to v ;

D2* for any $v \in LARGE_*$ there are at most $b_2 d_1$ edges from X incident to v ;

D3 $X \cap \mathcal{H}(G) = \emptyset$.

Now we will show that with the above definition we may give an analogue of Remark 3.

Lemma 6. Let \mathcal{G} be an instance of $\mathcal{G}(n, m, p)$ and $LARGE_*$ and $SMALL_*$ be defined as in Lemma 5. If \mathcal{G} has properties **P0***, **P1***, **P3***, **P4**, **P5** and X is deletable in \mathcal{G} with a constant $b_2 < b_1$. Then for $\mathcal{G}' = (\mathcal{V}, E(\mathcal{G}) \setminus X)$ and large n the assumptions of Lemma 3 with $\mathcal{V}' = LARGE_*$ are fulfilled with any constants $0 < b_3 < b_1 - b_2$ and $0 < b_4 < \min\{b_3, 1/3\}$.

Proof:

Recall that $mp = o(np)$, $d_0 = (1 + o(1))mp$, and $np = o(d_1)$. Therefore By **P0***, **P3***, **D1***, and **D2***

$$\delta(\mathcal{G}') \geq \min\{\delta(\mathcal{G}) - b_2 d_0, (b_1 - b_2)d_1\} \geq (1 + o(1))np/2 \geq (1 + o(1))n^\varepsilon \ln n/2.$$

This implies (i).

Now we show the first part of (ii). By the above calculation we have

$$\min\{\delta(\mathcal{G}'), d_1\} \geq (1 + o(1))n^\varepsilon \ln n/2,$$

therefore by **P0*** and **P1***

$$|SMALL_*| \leq n^\varepsilon = o(\min\{\delta(\mathcal{G}'), d_1\}) < b_4 \min\{\delta(\mathcal{G}'), d_1\}$$

for any constant b_4 and large n . This implies the first part of (ii). The second part of (ii) follows by **P3*** and **D2***. Finally (iii) follows by **D3**. \square

Now we will show an analogue of Lemma 4.

Lemma 7. Let $LARGE_*$ and $SMALL_*$ be defined as in Lemma 5 and \mathcal{G} be an intersection graph on n vertices with properties $\mathbf{P0}_*$, $\mathbf{P1}_*$, $\mathbf{P3}_*$, $\mathbf{P4}$, $\mathbf{P5}$ such that HAM finishes unsuccessfully on \mathcal{G} . Then

$$\Pr \left\{ X_q(n, m, p) \text{ is deletable with } b_2 = 0.5b_1 \mid \mathcal{G}(n, m, p) = \mathcal{G} \right\} \geq (1 + o(1)) \exp(-2\lambda(2T + 2))$$

where $o(1)$ is uniformly bounded over all possible choices of \mathcal{G} .

Proof: We redefine events A and B . For any $v \in \mathcal{V}$ and $w \in \mathcal{W}$ let A_{vw}^* be event that $\{v, w\}$ was not deleted in \mathcal{B}_q . Moreover

$$\begin{aligned} A_* &= \bigcap_{v \in SMALL_*, w \in \mathcal{W}(v)} A_{vw}^* \cap \bigcap_{\{v, v'\} \in \mathcal{H}(\mathcal{G})} A_{vv'}, \\ B_* &= \bigcup_{v \in \mathcal{V}} \bigcup_{S \subseteq \mathcal{N}(v), |S|=0.1b_1d_0} \bigcap_{v' \in S} B_{vv'}, \\ C &= \bigcup_{v \in LARGE_*} \bigcup_{\mathcal{R} \subseteq \mathcal{W}'(v), |\mathcal{R}|=0.1 \frac{\ln \ln n}{\ln n} b_1 d_0} \bigcap_{w \in \mathcal{R}} C_{vw}. \end{aligned}$$

Recall that for $v \in SMALL_*$, if $mp \leq 2 \ln n$ then we have $W(v) \leq 6 \cdot 10^{-3} mp \leq 12 \cdot 10^{-3} \ln n$. Otherwise by $\mathbf{P1}_*$ $SMALL_* = \emptyset$. Therefore by $\mathbf{P1}_*$ and (8)

$$\Pr_{\mathcal{G}} \{A_*\} \geq (1 - q)^{n^{\varepsilon} \cdot 12 \cdot 10^{-3} \ln n} (1 - q)^{2|\mathcal{H}(\mathcal{G})|} \geq (1 + o(1)) \exp(-2\lambda(2T + 2)).$$

Moreover, similar bounds as for $\Pr_{\mathcal{G}} \{B|A\}$ and $\Pr_{\mathcal{G}} \{C|A\}$ in the case $m \geq n^{1-\varepsilon}$ give

$$\Pr_{\mathcal{G}} \{B_*|A_*\} = o(1) \quad \text{and} \quad \Pr_{\mathcal{G}} \{C|A_*\} = o(1).$$

Therefore

$$\Pr_{\mathcal{G}} \{A_* \cap (B_* \cup C)^c\} \geq (1 + o(1)) \exp(-2\lambda(2T + 2)).$$

We are left with showing that $A_* \cap (B_* \cup C)^c$ implies that X_q is deletable. The same arguments as before give that A_* implies $\mathbf{D3}$ and $(B_* \cup C)^c$ implies $\mathbf{D2}_*$. Moreover A_* with B_* imply $\mathbf{D1}_*$.

In the remaining part of the proof we only replace D_1 and D_2 by

$$D_{1*} - \mathcal{G}(n, m, p) \text{ has properties } \mathbf{P1}_*, \mathbf{P3}_*, \mathbf{P4}, \mathbf{P5} \text{ with } b_1 = 1/1000;$$

$$D_{2*} - \delta(\mathcal{G}(n, m, p)) \geq np/2 \text{ } \mathcal{G}(n, m, p) \text{ has property } \mathbf{P0}_*;$$

and at the end, using (12), we get

$$\begin{aligned} &\Pr \{ \{ \delta(\mathcal{G}(n, m, p)) \geq 2 \} \cap \{ \text{HAM finishes unsuccessfully on } \mathcal{G}(n, m, p) \} \} \\ &\leq \Pr \{ \{ \delta(\mathcal{G}(n, m, p)) \geq np/2 \} \cap \{ \text{HAM finishes unsuccessfully on } \mathcal{G}(n, m, p) \} \} \\ &\quad + \Pr \{ \{ \delta(\mathcal{G}(n, m, p)) \geq 2 \} \cap \{ \delta(\mathcal{G}(n, m, p)) \geq np/2 \}^c \} \\ &= \Pr \{ D_{2*} \cap D_3 \} + o(1) \\ &\leq \Pr \{ D_{1*} \cap D_{2*} \cap D_3 \} + \Pr \{ D_{1*}^c \} + o(1) = o(1). \end{aligned}$$

□

Appendix A

First consider the case $mp^2 \leq 1$. In this case by Lemma 2 (property **P4**) with high probability each vertex in $\mathcal{G}(n, m, p)$ has degree at most $12d$. Therefore in stage k , $1 \leq k \leq n$, each of the considered paths has at most $24d$ rotations (at most $12d$ rotations for each of its ends). As we consider only those paths which result from at most $2T + 1$ rotations, the total number of rotations made during the execution of stage k is at most $(12d)^{2T+2} = O(n^{2+\varepsilon})$, for any constant $\varepsilon > 0$. Summing over all stages we get that with high probability the algorithm makes $O(n^{3+\varepsilon})$ rotations. Considering the time needed to execute a rotation we get that with high probability HAM terminates in $O(n^{4+\varepsilon})$ time on $\mathcal{G}(n, m, p)$. In the case $mp^2 \geq 1$ we have $T \leq 2$ therefore in each stage the number of rotations is at most $O(n^2)$. Thus the algorithm terminates in $O(n^4)$ rounds.

Appendix B

Proof of Theorem 2:

The following result is a special case of Lemma 5.1 from Rybarczyk (2017).

Lemma 8. *If $\ln^2 n = o(m)$ and*

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

then

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

Now we will find the probability of existence of a vertex of degree 1. Recall that

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

and

$$a_n = \begin{cases} 1 & \text{for } m < (1 - \varepsilon)n \ln n / \ln \ln n; \\ \frac{np \ln n}{e^{np} - 1} & \text{for } m > (1 + \varepsilon)n \ln n / \ln \ln n. \end{cases}$$

First consider the case $m < (1 - \varepsilon)n \ln n / \ln \ln n$. In this case $p(1 - (1 - p)^{n-1}) = \frac{\ln n + c_n}{m}$ as $a_n = 1$. We will show that if $c_n \rightarrow c \in (-\infty, \infty]$ then with high probability there is no vertex of degree 1 in $\mathcal{G}(n, m, p)$. Note that in this case

$$np \geq np(1 - (1 - p)^{n-1}) = \frac{n(\ln n + c_n)}{m} > (1 + o(1)) \frac{\ln \ln n}{1 - \varepsilon}.$$

Therefore, for large n , $np > (1 + o(1)) \ln \ln n / (1 - \varepsilon) > (1 + \varepsilon) \ln \ln n$, $\ln np = o(np)$, and $mp = (1 + o(1)) \ln n$.

Note that $v \in \mathcal{V}$ has degree 1 in an intersection graph \mathcal{G} associated with a bipartite graph \mathcal{B} if there exist a feature w and vertex v' such that $\mathcal{V}(w) = \{v, v'\}$ and for any other feature $w' \in \mathcal{W} \setminus \{w\}$ pair $\{v, w'\}$

is not an edge in \mathcal{B} , or w' is adjacent only to v in \mathcal{B} , or $\mathcal{V}(w') = \{v, v'\}$. Therefore, for any vertex $v \in \mathcal{V}$ and $c_n \rightarrow c \in (-\infty, \infty]$,

$$\begin{aligned}
\Pr \{\deg(v) = 1\} &\leq \sum_{w \in \mathcal{W}} \sum_{v' \in \mathcal{V}} p^2 (1-p)^{n-2} ((1-p) + p(1-p)^{n-1} + p^2)^{(m-1)} \\
&\leq mp \cdot np \cdot \exp(-np + p) \\
&\quad \cdot \exp(-mp(1 - (1-p)^{n-1}) + O(mp^2 + p)) \\
&\leq \exp(\ln mp + \ln np - np - \ln n - c_n + o(1)) \\
&\leq \exp(\ln((1+o(1)) \ln n) - (1+o(1))np - \ln n - c_n + o(1)) \\
&\leq \exp(\ln \ln n - (1+\varepsilon) \ln \ln n - \ln n - c_n + o(1)) \\
&= o(n^{-1})
\end{aligned}$$

Thus with high probability there is no vertex of degree 1 in $\mathcal{G}(n, m, p)$. This combined with Lemma 8 completes the proof in this case.

Now assume that $m > (1 + \varepsilon)n \ln n / \ln \ln n$. We will use the method of moments to get a Poisson approximation of the number of vertices of degree 1 (see for example Corollary 6.8 in Janson et al. (2001)). In this case $np < (mp \ln \ln n) / ((1 + \varepsilon) \ln n)$ and, as a consequence, $a_n = (np \ln n) / (e^{np} - 1) \rightarrow \infty$ (note that $x / (e^x - 1)$ is decreasing for $x \in (0, \infty)$). Recall that $mp^2 = o(1)$, i.e. also $np^2 = o(1)$. Let k be a constant integer and $\{v_1, \dots, v_k\} \subseteq \mathcal{V}$. Note that $(m)_k$ is the number of choices of w_1, \dots, w_k , which might be chosen by vertices v_1, \dots, v_k , resp., $(n-k)p^2(1-p)^{n-2}$ is the probability that w_i , $1 \leq i \leq k$, was chosen only by v_i and one other vertex from $\mathcal{V} \setminus \{v_1, \dots, v_k\}$, and $((1-p)^k + kp(1-p)^{n-1})^{(m-k)}$ is the probability that each feature from $\mathcal{W} \setminus \{w_1, \dots, w_k\}$ either was not chosen by any vertex from $\{v_1, \dots, v_k\}$ or was chosen by exactly one vertex from $\{v_1, \dots, v_k\}$ and no other vertex. Therefore

$$\begin{aligned}
&\Pr \{\deg(v_1) = \dots = \deg(v_k) = 1\} \\
&\geq (m)_k ((n-k)p^2(1-p)^{n-2})^k ((1-p)^k + kp(1-p)^{n-1})^{(m-k)} \\
&\geq (1+o(1))(mp(1-(1-p)^{n-1}))^k \left(\frac{np(1-p)^{n-1}}{(1-(1-p)^{n-1})} \right)^k \\
&\quad \cdot (1-kp + kp(1-p)^{n-1})^{(m-k)} \\
&= (1+o(1)) \left(\frac{\ln n \cdot np}{(1-p)^{-(n-1)} - 1} \right)^k \\
&\quad \cdot \exp(-kmp(1-(1-p)^{n-1}) + O(mp^2 + p)) \\
&= (1+o(1)) \left(\frac{\ln n \cdot np}{e^{np} - 1} \right)^k \exp(-k(\ln n + \ln a_n + c_n)) \\
&= (1+o(1)) \left(\frac{e^{-c_n}}{n} \right)^k.
\end{aligned}$$

Now we will find an upper bound on $\Pr \{\deg(v_1) = \dots = \deg(v_k) = 1\}$. Below we set i , $0 \leq i \leq k/2$, to be the number of edges induced by $\{v_1, \dots, v_k\}$ in $\mathcal{G}(n, m, p)$. We note that the probability that for

a given feature w set $\mathcal{V}(w)$ is the same as some edge incident to a vertex from $\{v_1, \dots, v_k\}$ is upper bounded by kp^2 . Therefore

$$\begin{aligned}
& \Pr \{ \deg(v_1) = \dots = \deg(v_k) = 1 \} \\
& \leq \sum_{i=0}^{k/2} \binom{m}{k-i} \binom{k}{2i} (np^2(1-p)^{n-2})^{(k-i)} \\
& \quad \cdot ((1-p)^k + kp(1-p)^{n-1} + kp^2)^{(m-k)} \\
& \leq (1 - kp + kp(1-p)^{n-1} + (kp)^2 + kp^2)^{(m-k)} \\
& \quad \cdot \sum_{i=0}^{k/2} k^{2i} (mp(1 - (1-p)^{n-1}))^{k-i} \left(\frac{np(1-p)^{n-1}}{(1 - (1-p)^{n-1})} \right)^{k-i} \\
& = (1 + o(1)) \exp(-k(\ln n + \ln a_n + c_n)) \sum_{i=0}^{k/2} k^{2i} a_n^{k-2i} \\
& = (1 + o(1)) \left(\frac{e^{-c_n}}{n} \right)^k.
\end{aligned}$$

Therefore by the method of moments the number of vertices of degree 1 in $\mathcal{G}(n, m, p)$ tends to a random variable with the Poisson distribution $\text{Po}(e^{-c})$, where $c = \lim_{n \rightarrow \infty} c_n$. This combined with Lemma 8 finishes the proof in the second case. \square

Appendix C

Proof of Remark 4: In fact we will show that with high probability

$$\forall \mathcal{R} \subseteq \mathcal{W}, |\mathcal{R}| \leq p^{-1} V(\mathcal{R}) \geq \frac{np|\mathcal{R}|}{2} + 1. \quad (13)$$

Note that (13) implies that with high probability each vertex is either isolated or of degree at least $np/2$. Therefore (12) follows by (13).

We will now show (13). Recall that $np \geq (1 + o(1))n^\varepsilon \ln n$. For $\mathcal{R} \subseteq \mathcal{W}$ with $|\mathcal{R}| = r \leq p^{-1}$, $V(\mathcal{R})$ has the binomial distribution $\text{Bin}(n, 1 - (1-p)^r)$ with $\mathbb{E}V(\mathcal{R}) \geq 3npr/5$ (as for large n we have $1 - (1-p)^r \geq 3pr/5$). Therefore

$$\begin{aligned}
\Pr \{ \exists \mathcal{R} \subseteq \mathcal{W}, |\mathcal{R}| \leq p^{-1} V(\mathcal{R}) \leq (npr/2) + 1 \} & \leq \sum_{r=1}^{p^{-1}} \binom{m}{r} \exp(-\psi((1 + o(1))5/6)3npr/5) \\
& \leq \sum_{r=1}^{p^{-1}} \exp(-r(0.008np - \ln m)) = o(1).
\end{aligned}$$

□

Proof of Lemma 5:**P1*** This proof is practically the same as the proof of **P1**.**P3*** As in the proof of **P3** in the case $np > 40$

$$\begin{aligned}
& \Pr \left\{ \exists \mathcal{S} \subseteq \text{LARGE}_*, |\mathcal{S}| \leq n/d_1, W(\mathcal{S}) \leq 3 \cdot 10^{-3} \cdot mp|\mathcal{S}| \right\} \\
& \leq \sum_{s=1}^{n/d_1} \binom{n}{s} \binom{m}{3 \cdot 10^{-3}mps} \left(\frac{3 \cdot 10^{-3}mps}{m} \right)^{6 \cdot 10^{-3}mps} \\
& \leq \sum_{s=1}^{n/d_1} n^s (e \cdot 3 \cdot 10^{-3}ps)^{3 \cdot 10^{-3}mps} \\
& \leq \sum_{s=1}^{n/d_1} \exp \left(s (\ln n + 3 \cdot 10^{-3}(1 + o(1)) \ln n (\ln ps + O(1))) \right) = o(1).
\end{aligned}$$

The above stated inequalities together with (13) imply that if $m < n^{1-\varepsilon}$ then with high probability for all $\mathcal{S} \subseteq \text{LARGE}_*$, $|\mathcal{S}| \leq n/d_1$, and n large

$$N(\mathcal{S}) = V(W(\mathcal{S})) - |\mathcal{S}| \geq \frac{np|W(\mathcal{S})|}{2} - |\mathcal{S}| \geq \frac{np}{2} \cdot 3 \cdot 10^{-3} \cdot mp|\mathcal{S}| - |\mathcal{S}| > 10^{-3}d_1|\mathcal{S}|.$$

P4 and **P5** The proofs of **P4** and **P5** presented in the proof of Lemma 2 are also valid in the case $m < n^{1-\varepsilon}$.

□

Appendix D

Proof of Lemma 3:

Set $\mathcal{G} = \mathcal{G}_n$ and $\mathcal{G}' = \mathcal{G}'_n$. As $\mathcal{H}(\mathcal{G}) \subseteq E(\mathcal{G}')$, the execution of HAM on \mathcal{G}' will be the same as on \mathcal{G} , i.e. HAM will start stage k in \mathcal{G}' with P_k (with endpoints u_0 and u_1) and terminate unsuccessfully in this stage.

Let

$$\mathcal{S}_t = \{v : v \in \mathcal{V}' \text{ and at stage } k \text{ there exists a path } Q_s \text{ with endpoints } u_0, v \text{ such that } \bar{\delta}(Q_s) = t\}$$

First we prove that $\mathcal{S}_1 \neq \emptyset$. $P_k = Q_1$. If $u_1 \notin \mathcal{V}'$ then u_1 has at least $\delta(\mathcal{G}')$ neighbours on Q_1 and at least $\delta(\mathcal{G}') - 1$ of them are not connected with u_1 by an edge of Q_1 . Therefore on Q_1 there are at least $\delta(\mathcal{G}') - 1 \geq 1$ new ends obtained by one rotation from Q_1 . If $\delta_0 = 1$ then all of these ends are in \mathcal{V}' (otherwise there would be a vertex with more than $\delta_0 = 1$ vertices at distance at most 2 in $\mathcal{V} \setminus \mathcal{V}'$). This implies $\mathcal{S}_1 \neq \emptyset$ in the case $\delta_0 = 1$. If $\delta_0 > 1$ (i.e. $\delta(\mathcal{G}') \geq 1/b_4 \geq 3$) then, by (ii), at least

$$(1 - b_4)\delta(\mathcal{G}') - 1 \geq \frac{2}{3} \cdot 3 - 1 = 1$$

of the ends of the paths obtained by one rotation from Q_1 are in \mathcal{V}' , i.e. $\mathcal{S}_1 \neq \emptyset$. If $u_1 \in \mathcal{V}'$ then u_1 has at least b_3d neighbours and at least $b_3d - 1$ new ends obtained by a rotation from Q_1 . Then at least $b_3d - b_4d - 1 \geq (b_3 - b_4)d - 1 \geq 1$ of these ends are in \mathcal{V}' , i.e. $\mathcal{S}_1 \neq \emptyset$.

Now assume that $1 \leq |\mathcal{S}_t| \leq n/d$ for some t , $1 \leq t \leq T$.

For each $v \in \mathcal{S}_t$ set a path $Q_{s(v)}$ with endpoints u_0 and v such that $\bar{\delta}(Q_{s(v)}) = t$.

Now look at the pairs (v, u) , where u is a neighbour of v in \mathcal{G}' , i.e. $u \in \mathcal{N}_{\mathcal{G}'}(v)$. For $\{v, u\}$ which is not an edge of $Q_{s(v)}$, denote by $x(v, u)$ the endpoint of $ROTATE(Q_{s(v)}, \{v, u\})$ other than u_0 . If $x(v, u) \in \mathcal{V}'$ then $x(v, u) \in \mathcal{S}_{t+1}$.

For any $u \in \mathcal{N}_{\mathcal{G}'}(\mathcal{S}_t)$ there are at most 2 vertices $x \in \mathcal{S}_{t+1}$ such that for some $v \in \mathcal{S}_t$ we have $x = x(v, u)$ and $\{x, u\} \in E(P_k)$.

Moreover, for any v there is only one $u \in \mathcal{N}_{\mathcal{G}'}(v)$ such that $\{v, u\} \in E(Q_{s(v)})$, at most b_4d vertices $u \in \mathcal{N}_{\mathcal{G}'}(v)$ such that $x(v, u) \notin \mathcal{V}'$ and at most t (where $t = \bar{\delta}(Q_{s(v)})$) vertices $u \in \mathcal{N}_{\mathcal{G}'}(v)$ such that $\{x, u\} \notin E(P_k)$ (since there are at most t edges on $Q_{s(v)}$ which are not in P_k).

Therefore

$$\begin{aligned} |\mathcal{S}_{t+1}| &= |\{x(v, u) : v \in \mathcal{S}_t, u \in \mathcal{N}_{\mathcal{G}'}(v) \text{ and } x(v, u) \in \mathcal{V}'\}| \\ &\geq |\{x(v, u) : v \in \mathcal{S}_t, u \in \mathcal{N}_{\mathcal{G}'}(v), x(v, u) \in \mathcal{V}', \{u, v\} \notin E(Q_{s(v)}), \{x(v, u), u\} \in E(P_k)\}| \\ &\geq |\{u \in \mathcal{N}_{\mathcal{G}'}(\mathcal{S}_t) : \exists v \in \mathcal{S}_t x(v, u) \in \mathcal{V}', \{u, v\} \notin E(Q_{s(v)}), \{x(v, u), u\} \in E(P_k)\}|/2 \\ &\geq (N_{\mathcal{G}'}(\mathcal{S}_t) - (t + 1 + b_4d)|\mathcal{S}_t|)/2 \\ &\geq ((b_3 - b_4)d - T - 1)|\mathcal{S}_t|/2 \\ &= (1 + o(1))(b_3 - b_4)d|\mathcal{S}_t|/2 \geq b_5d|\mathcal{S}_t| \end{aligned}$$

for any constant $b_5 < (b_3 - b_4)/2$.

$\mathcal{S}_1 \neq \emptyset$ and $|\mathcal{S}_{t+1}| \geq b_5d|\mathcal{S}_t|$, therefore for some $\tau \leq T - 1$ we have $|\mathcal{S}_\tau| \geq n/d$. Applying the same argument as in above calculations for any $\mathcal{S}' \subseteq \mathcal{S}_\tau$ such that $|\mathcal{S}'| = n/d$ we have that $|\mathcal{S}_{\tau+1}| \geq b_5d|\mathcal{S}'| \geq b_5n$. To prove the second part of the lemma, for any $x \in \text{END}(\mathcal{G}')$, it remains to apply the same arguments replacing u_0 by x and P_k by the path Q_σ with one endpoint x and $\bar{\delta}(Q_\sigma) \leq 2T$. \square

References

- M. Bloznelis. Degree and clustering coefficient in sparse random intersection graphs. *The Annals of Applied Probability*, 23(3):1254–1289, 2013.
- M. Bloznelis and I. Radavičius. A note on hamiltonicity of uniform random intersection graphs. *Lithuanian Mathematical Journal*, 51(2):155, 2011.
- M. Bloznelis, J. Jaworski, and K. Rybarczyk. Component evolution in a secure wireless sensor network. *Networks*, 53(1):19–26, 2009.
- M. Bloznelis, E. Godehardt, J. Jaworski, V. Kurauskas, and K. Rybarczyk. Recent progress in complex network analysis: Models of random intersection graphs. In B. Lausen, S. Krolak-Schwerdt, and M. Böhmer, editors, *Data Science, Learning by Latent Structures, and Knowledge Discovery*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 69–78. Springer Berlin Heidelberg, 2015a.

- M. Bloznelis, E. Godehardt, J. Jaworski, V. Kurauskas, and K. Rybarczyk. Recent progress in complex network analysis: Properties of random intersection graphs. In B. Lausen, S. Krolak-Schwerdt, and M. Böhmer, editors, *Data Science, Learning by Latent Structures, and Knowledge Discovery*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 79–88. Springer Berlin Heidelberg, 2015b.
- B. Bollobás, T. I. Fenner, and A. M. Frieze. An algorithm for finding hamilton paths and cycles in random graphs. *Combinatorica*, 7(4):327–341, 1987.
- 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.
- 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.
- C. Efthymiou and P. G. Spirakis. On the existence of Hamiltonian cycles in random intersection graphs. In *Automata, Languages and Programming 32nd International Colloquium, ICALP 2005, Lisbon, Portugal, July 11-15, 2005. Proceedings*, pages 690–701, 2005.
- 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.
- S. Janson, T. Łuczak, and A. Ruciński. *Random Graphs*. Wiley, 2001.
- M. Karoński, E. R. Scheinerman, and K. Singer-Cohen. On random intersection graphs: The subgraph problem. *Combinatorics, Probability and Computing*, 8:131–159, 1999.
- L. Pósa. Hamiltonian circuits in random graphs. *Discrete Mathematics*, 14(4):359 – 364, 1976.
- C. Raptopoulos and P. Spirakis. Simple and efficient greedy algorithms for hamilton cycles in random intersection graphs. In *Algorithms and Computation 16th International Symposium, ISAAC 2005, Sanya, Hainan, China, December 19-21, 2005. Proceedings*, pages 493–504, 2005.
- K. Rybarczyk. Sharp threshold functions for random intersection graphs via a coupling method. *The Electronic Journal of Combinatorics*, 18(1):P36, 2011.
- K. Rybarczyk. The coupling method for inhomogeneous random intersection graphs. *The Electronic Journal of Combinatorics*, 24(2):P2.10, 2017.