Sharp threshold functions for random intersection graphs via a coupling method.

Katarzyna Rybarczyk

Faculty of Mathematics and Computer Science,
Adam Mickiewicz University, 60–769 Poznań, Poland

kryba@amu.edu.pl

Submitted: Nov 25, 2009; Accepted: Feb 7, 2011; Published: Feb 14, 2011
Mathematics Subject Classification: 05C80
keywords: random intersection graphs, threshold functions,
connectivity, Hamilton cycle, perfect matching, coupling

Abstract

We present a new method which enables us to find threshold functions for many properties in random intersection graphs. This method is used to establish sharp threshold functions in random intersection graphs for k–connectivity, perfect matching containment and Hamilton cycle containment.

1 Introduction

In a general random intersection graph $\overline{\mathcal{G}}(n, m, \mathcal{P}_{(m)})$, as defined in [9], each vertex v from a vertex set $\mathcal{V}$ ($|\mathcal{V}| = n$) is assigned independently a subset of features $W_v \subseteq \mathcal{W}$ from an auxiliary set of features $\mathcal{W}$ ($|\mathcal{W}| = m$). Namely, for any vertex $v \in \mathcal{V}$, independently of all other vertices, first a cardinality of W_v is chosen according to the probability distribution $\mathcal{P}_{(m)} = (P_0, \ldots, P_m)$, and then the set W_v is picked uniformly at random from all subsets of $\mathcal{W}$ having the chosen cardinality. Two vertices v and v' are adjacent in a general intersection graph $\overline{\mathcal{G}}(n, m, \mathcal{P}_{(m)})$ if and only if W_v and $W_{v'}$ intersect. In this article we concentrate on the widely studied random intersection graph model $\mathcal{G}(n, m, p)$ first defined in [11, 17] which is a special case of the one above-mentioned. However the obtained results may be extended to a wider subclass of the $\overline{\mathcal{G}}(n, m, \mathcal{P}_{(m)})$ model, which will be also discussed. In $\mathcal{G}(n, m, p)$, as defined in [11, 17], the cardinality of W_v has the binomial distribution $\mathrm{Bin}(m, p)$, i.e. $\Pr\{w \in W_v\} = p$ independently for all $v \in \mathcal{V}$ and $w \in \mathcal{W}$. Usually, it is assumed that $m = n^{\alpha}$ for some constant $\alpha > 0$ (see for example [2, 6, 8, 11, 16, 17, 18]). However the main theorem of this article does not require this additional assumption.

Obviously, $\Pr\{\{v,v'\} \in E(\mathcal{G}(n,m,p))\} = 1-(1-p^2)^m$ for any distinct $v,v' \in \mathcal{V}$. Therefore one could expect that there is some relation between $\mathcal{G}(n,m,p)$ and a random graph $G(n,\hat{p})$ with edges appearing independently with probability $\hat{p}$ for $\hat{p}$ approximately $1-(1-p^2)^m$. It follows from the results on subgraph containment as presented in [11, 16], in general, these are not equivalence relations since the structures of $\mathcal{G}(n,m,p)$ and $G(n,\hat{p})$ differ significantly. However it was shown in [8] that for large m (i.e. $m = n^\alpha$ and $\alpha > 6$), dependencies between edge appearances in $\mathcal{G}(n,m,p)$ are small and the models have asymptotically the same properties. The equivalence theorem is extended to $m = n^\alpha$ and $\alpha \geq 3$ (see [15]), but for $m = n^\alpha$ and $\alpha < 3$ it is not true in general (see for example [11, 16]). In the context of the results stated above it seems intriguing that for $m = n^\alpha$ and $\alpha > 1$ the threshold functions of connectivity and phase transition in $\mathcal{G}(n,m,p)$ and $G(n,\hat{p})$ coincide (see [2, 7, 17]) even though the models differ a lot (for example the expected number of triangles in $\mathcal{G}(n,m,p)$ significantly exceeds the expected number of triangles in $G(n,\hat{p})$ for $\alpha < 3$). One of the aims of this paper is to get an improved understanding of the phenomena by a closer insight into the structure of $\mathcal{G}(n,m,p)$ and to use this knowledge to establish sharp threshold functions for other important properties of $\mathcal{G}(n,m,p)$.

Our work is partially inspired by the result of Efthymiou and Spirakis [6]. However the method significantly differs from the one used in [6] and therefore it enables us to obtain sharper threshold functions for the property of Hamilton cycle containment than those from [6].

The article is organised as follows. In Section 2 we present and prove the main theorem which relates $\mathcal{G}(n,m,p)$ to $G(n,\hat{p})$. In Section 3 the theorem is used to study properties of $\mathcal{G}(n,m,p)$. In particular, an alternative short proof of the connectivity theorem shown in [17] is given. Moreover, results concerning sharp threshold functions for Hamilton cycle containment, perfect matching containment and k-connectivity are proved. The method introduced here is strong enough to give some partial results on the threshold functions for other properties of $\mathcal{G}(n,m,p)$. However we present here graph properties for which the threshold functions obtained by our method are tight at least for $m = n^\alpha$ and $\alpha > 1$. In Section 4 extensions of the results to a wider subclass of the general random intersection graph model are presented. Moreover some interesting questions related to the main theorem are discussed.

All limits in the paper are taken as $n \to \infty$. Throughout the paper we use the notation $a_n = o(b_n)$ if $a_n/b_n \to 0$ and $a_n \sim b_n$ if $a_n/b_n \to 1$. Also by $\mathrm{Bin}(n,p)$ and $\mathrm{Po}(\lambda)$ we denote the binomial distribution with parameters n, p and the Poisson distribution with expected value λ, respectively. Moreover if a random variable X is stochastically dominated by Y we write $X \prec Y$. We also use the phrase "with high probability" to say with probability tending to one as n tends to infinity.

2 Main Result

Recall that for the family $\mathcal{G}$ of all graphs with a vertex set $\mathcal{V}$, we call $\mathcal{A} \subseteq \mathcal{G}$ an increasing property if $\mathcal{A}$ is closed under isomorphism and $G \in \mathcal{A}$ implies $G' \in \mathcal{A}$ for all $G' \in \mathcal{G}$

such that $E(G) \subseteq E(G')$. The theorem stated below relates $\mathcal{G}(n,m,p)$ to $G(n,\hat{p})$ for increasing properties. A motivation for the investigation in a comparison was the fact that, for $m = n^{\alpha}$ and $\alpha > 1$, if p and $\hat{p}$ are connectivity threshold functions of $\mathcal{G}(n,m,p)$ and $G(n,\hat{p})$, respectively, then $\Pr\{\{v,v'\} \in E(\mathcal{G}(n,m,p))\} \sim 1-(1-p^2)^m \sim mp^2 \sim \hat{p}$ (see [17]). In the proof of the theorem we explain that this is due to the fact that $np \to 0$. Surprisingly, in some cases the comparison also gives tight results for $np \not\to 0$, however with $\hat{p}$ differing from $1-(1-p^2)^m$. This is due to the fact that as $np \to \infty$ the number of large cliques in $\mathcal{G}(n,m,p)$ increases compared to $G(n,\hat{p})$ and thus both models have significantly different edge structures. Basically, as $np \to \infty$ and $\hat{p} = (1+o(1))mp/n$, $\mathcal{G}(n,m,p)$ has more edges than $G(n,\hat{p})$, however both models have the same number of isolated vertices. In the theorem we have the case $nmp \to \infty$ instead of $np \to \infty$, since the thesis also holds true in this case. However as $nmp \to \infty$ and $np \not\to \infty$ the results obtained using the theorem will not be tight.

Theorem 1. *Let $\mathcal{A}$ be an increasing property, $mp^2 < 1$, and*

$$\hat{p}_- = \begin{cases} mp^2\left(1-(n-2)p-\frac{mp^2}{2}\right) & \text{for } np = o(1); \\ \frac{mp}{n}\left(1-\frac{\omega}{\sqrt{mnp}}-\frac{2}{np}-\frac{mp}{2n}\right) & \text{for } nmp \to \infty \\ & \text{and some } \omega \to \infty, \omega = o(\sqrt{mnp}). \end{cases} \tag{1}$$

If

$$\Pr\{G(n,\hat{p}_-) \in \mathcal{A}\} \to 1,$$

then

$$\Pr\{\mathcal{G}(n,m,p) \in \mathcal{A}\} \to 1. \tag{2}$$

The main ingredient of the proof is a comparison of $\mathcal{G}(n,m,p)$ and $G(n,\hat{p})$ using intermediate auxiliary graphs. The comparison is made by a sequence of couplings and measuring the distance between distributions of auxiliary graph valued random variables. First we introduce necessary definitions and notation.

Let M be a random variable with range in the set of non-negative integers (in the simplest case M is a given positive integer with probability one). By $\mathcal{G}_*(n,M)$ we denote a random graph with vertex set $\mathcal{V}$ and edge set constructed by sampling M times with repetition elements from the set of all two element subsets of $\mathcal{V}$. A subset $\{v,v'\}$ is an edge of $\mathcal{G}_*(n,M)$ if and only if it is sampled at least once. If M equals a constant t with probability one, has the binomial distribution, or the Poisson distribution, we write $\mathcal{G}_*(n,t)$, $\mathcal{G}_*(n,\mathrm{Bin}(\cdot,\cdot))$, or $\mathcal{G}_*(n,\mathrm{Po}(\cdot))$, respectively.

For any random variables G_1 and G_2 with values in a countable set A, by the total variation distance we mean

$$\begin{aligned} d_{TV}(G_1,G_2) &= \max_{A' \subseteq A} |\Pr\{G_1 \in A'\} - \Pr\{G_2 \in A'\}| \\ &= \frac{1}{2}\sum_{a \in A} |\Pr\{G_1 = a\} - \Pr\{G_2 = a\}|. \end{aligned}$$

By a coupling (G_1, G_2) of two random variables G_1 and G_2 we mean a choice of a probability space on which a random vector (G_1', G_2') is defined and G_1' and G_2' have the same distributions as G_1 and G_2, respectively. For simplicity of notation we will not differentiate between (G_1', G_2') and (G_1, G_2). For two graph valued random variables G_1 and G_2 we write

$$G_1 \preceq G_2 \quad \text{or} \quad G_1 \preceq_{1-o(1)} G_2,$$

if there exists a coupling (G_1, G_2), such that under the coupling G_1 is a subgraph of G_2 with probability 1 or $1 - o(1)$, respectively. Moreover, we write

$$G_1 = G_2,$$

if G_1 and G_2 have the same probability distribution (equivalently there exists a coupling (G_1, G_2) such that $G_1 = G_2$ with probability one).

It is simple to construct suitable couplings which implies the following fact.

Fact 1. *(i) Let M_n be a sequence of random variables and let a_n be a sequence of numbers. If*

$$\Pr\{M_n \geq a_n\} = o(1) \quad (\Pr\{M_n \leq a_n\} = o(1)),$$

then

$$\mathcal{G}_*(n, M_n) \preceq_{1-o(1)} \mathcal{G}_*(n, a_n) \quad (\mathcal{G}_*(n, a_n) \preceq_{1-o(1)} \mathcal{G}_*(n, M_n)).$$

(ii) If a random variable M is stochastically dominated by M' (i.e. $M \prec M'$), then

$$\mathcal{G}_*(n, M) \preceq \mathcal{G}_*(n, M').$$

The proof of the next fact is analogous to the proof of Fact 2 in [15].

Fact 2. *Let $(G_i)_{i=1,\ldots,m}$ and $(G_i')_{i=1,\ldots,m}$ be sequences of independent random graphs. If*

$$G_i \preceq G_i', \text{ for all } i = 1, \ldots, m$$

then

$$\bigcup_{i=1}^{m} G_i \preceq \bigcup_{i=1}^{m} G_i'.$$

Proof of Theorem 1. Let $w \in \mathcal{W}$. Denote by V_w the set of vertices which have chosen feature w and put $X_w = |V_w|$. Let $\mathcal{G}[V_w]$ be a graph with vertex set $\mathcal{V}$ and edge set containing those edges which have both ends in V_w (i.e. its edges form a clique with the vertex set V_w). We can construct a coupling $(\mathcal{G}_*(n, \lfloor X_w/2 \rfloor), \mathcal{G}[V_w])$ which implies

$$\mathcal{G}_*(n, \lfloor X_w/2 \rfloor) \preceq \mathcal{G}[V_w],$$

in the following way. Given the value of X_w, first we generate an instance G_w of $\mathcal{G}_*(n, \lfloor X_w/2 \rfloor)$. Let Y_w be the number of non-isolated vertices in G_w. By definition Y_w is at most X_w, therefore V_w may be chosen to be a union of the set of non–isolated vertices in G_w and $X_w - Y_w$ vertices chosen uniformly at random from the remaining ones.

Graphs $\mathcal{G}_*(n, \lfloor X_w/2 \rfloor)$, $w \in \mathcal{W}$, are independent, and $\mathcal{G}[V_w]$, $w \in \mathcal{W}$, are independent. Thus by Fact 2 and the definition of $\mathcal{G}(n,m,p)$, we have

$$\bigcup_{w\in\mathcal{W}} \mathcal{G}_*(n, \lfloor X_w/2 \rfloor) \preceq \bigcup_{w\in\mathcal{W}} \mathcal{G}[V_w] = \mathcal{G}(n,m,p).$$

Since X_w, $w \in \mathcal{W}$, are independent random variables and $\mathcal{G}[V_w]$, $w \in \mathcal{W}$, are independent as well, by the above equation and the definition of $\mathcal{G}_*(n,\cdot)$,

$$\mathcal{G}_*\left(n, \textstyle\sum_{w\in\mathcal{W}} \lfloor X_w/2 \rfloor\right) = \bigcup_{w\in\mathcal{W}} \mathcal{G}_*(n, \lfloor X_w/2 \rfloor) \preceq \mathcal{G}(n,m,p). \tag{3}$$

Now consider the following two cases
CASE 1: $np = o(1)$.
Notice that

$$\sum_{w\in\mathcal{W}} \mathbb{I}_w \prec \sum_{w\in\mathcal{W}} \lfloor X_w/2 \rfloor,$$

where

$$\mathbb{I}_w = \begin{cases} 1, & \text{if } X_w \geq 2; \\ 0, & \text{otherwise.} \end{cases}$$

The random variable $Z_1 = \sum_{w\in\mathcal{W}} \mathbb{I}_w$ has the binomial distribution $\mathrm{Bin}(m,q)$, where $q = \Pr\{X_w \geq 2\}$, therefore by Fact 1(ii),

$$\mathcal{G}_*(n, \mathrm{Bin}(m,q)) \preceq \mathcal{G}_*\left(n, \textstyle\sum_{w\in\mathcal{W}} \lfloor X_w/2 \rfloor\right). \tag{4}$$

Let M_1 and M_2 be random variables with the binomial distribution $\mathrm{Bin}(m,q)$ and the Poisson distribution $\mathrm{Po}(mq)$, respectively. A simple calculation shows that in $\mathcal{G}_*(n, M_1)$ each edge appears independently with probability $1 - \exp(-mq/\binom{n}{2})$ (see [8]). Therefore by properties of the total variation distance and the Poisson approximation of binomial random variables (see [8] and [1] or [15]), we have

$$\begin{aligned} d_{TV}&\left(\mathcal{G}_*(n, \mathrm{Bin}(m,q)), G\left(n, 1 - \exp(-mq/\tbinom{n}{2})\right)\right) \\ &= d_{TV}\left(\mathcal{G}_*(n, M_1), \mathcal{G}_*(n, M_2)\right) \leq 2 d_{TV}(M_1, M_2) \leq 2q \leq 2\binom{n}{2} p^2 = o(1). \end{aligned} \tag{5}$$

Moreover $q \geq \Pr\{X_w = 2\} = \binom{n}{2} p^2 (1-p)^{n-2}$ and $1 - \exp(-x) \geq x - x^2/2$ for $x < 1$ (recall that $mp^2 < 1$ by the assumptions of the theorem), thus

$$p_- = mp^2 \left(1 - (n-2)p - \frac{mp^2}{2}\right) \leq 1 - \exp(-mq/\tbinom{n}{2}).$$

Therefore by a standard coupling of $G(n,\cdot)$ we obtain

$$G(n, p_-) \preceq G\left(n, 1 - \exp(-mq/\tbinom{n}{2})\right). \tag{6}$$

CASE 2: $nmp \to \infty$**.**
Notice that

$$\frac{Z_2}{2} - m \prec \sum_{w \in \mathcal{W}} \lfloor X_w/2 \rfloor,$$

where $Z_2 = \sum_{w \in \mathcal{W}} X_w$ has the binomial distribution $\mathrm{Bin}\,(nm, p)$. By Fact 1(ii),

$$\mathcal{G}_*\left(n, \frac{Z_2}{2} - m\right) \preceq \mathcal{G}_*\left(n, \sum_{w \in \mathcal{W}} \lfloor X_w/2 \rfloor\right). \tag{7}$$

By Chernoff's bound for the Poisson distribution (see [14] Lemma 1.2) for any function $\omega \to \infty$, $\omega = o(\sqrt{nmp})$,

$$\Pr\left\{\frac{Z_2}{2} - m \leq \frac{nmp}{2}\left(1 - \frac{\omega}{2\sqrt{nmp}} - \frac{2}{np}\right)\right\} = \Pr\left\{Z_2 \leq nmp - \frac{\omega\sqrt{mnp}}{2}\right\} = o(1).$$

Moreover, the same bound applied to a random variable Z_3 with the Poisson distribution $\mathrm{Po}\left(\frac{nmp}{2}\left(1 - \frac{\omega}{\sqrt{nmp}} - \frac{2}{np}\right)\right)$ gives

$$\Pr\left\{Z_3 \geq \frac{nmp}{2}\left(1 - \frac{\omega}{2\sqrt{nmp}} - \frac{2}{np}\right)\right\} = \Pr\left\{Z_3 \geq \mathbb{E}Z_3 + \frac{\omega\sqrt{nmp}}{4}\right\} = o(1).$$

Therefore, using twice Fact 1(i), we get

$$\mathcal{G}_*\left(n, \mathrm{Po}\left(\tfrac{nmp}{2}\left(1 - \tfrac{\omega}{\sqrt{nmp}} - \tfrac{2}{np}\right)\right)\right) \preceq_{1-o(1)} \mathcal{G}_*\left(n, \frac{Z_2}{2} - m\right). \tag{8}$$

Recall that, for any λ, in $\mathcal{G}_*\,(n, \mathrm{Po}\,(\lambda))$ each edge appears independently with probability $1 - \exp(-\lambda/\binom{n}{2})$ (see [8]). Therefore

$$G\left(n, 1 - \exp\left(-\tfrac{mp}{n-1}\left(1 - \tfrac{\omega}{\sqrt{nmp}} - \tfrac{2}{np}\right)\right)\right) = \mathcal{G}_*\left(n, \mathrm{Po}\left(\tfrac{nmp}{2}\left(1 - \tfrac{\omega}{\sqrt{nmp}} - \tfrac{2}{np}\right)\right)\right). \tag{9}$$

Since

$$\tfrac{mp}{n}\left(1 - \tfrac{\omega}{\sqrt{nmp}} - \tfrac{2}{np} - \tfrac{mp}{2n}\right) \leq 1 - \exp\left(-\tfrac{mp}{n-1}\left(1 - \tfrac{\omega}{\sqrt{nmp}} - \tfrac{2}{np}\right)\right),$$

a standard coupling of $G\,(n, \cdot)$ implies

$$G\,(n, p_-) \preceq G\left(n, 1 - \exp\left(-\tfrac{mp}{n-1}\left(1 - \tfrac{\omega}{\sqrt{nmp}} - \tfrac{2}{np}\right)\right)\right). \tag{10}$$

In equations (3)–(10) we have established relations between $G\,(n, p_-)$ and $\mathcal{G}\,(n, m, p)$ using intermediate auxiliary random graphs. From them we can deduce the assertion of the theorem.

First recall (see for example [8]) that if for some graph valued random variables G_1 and G_2

$$d_{TV}\,(G_1, G_2) = o(1),$$

then for any $a \in [0;1]$ and any graph property $\mathcal{A}$

$$\Pr\{G_1 \in \mathcal{A}\} \to a \quad \text{iff} \quad \Pr\{G_2 \in \mathcal{A}\} \to a.$$

Now let G_1 and G_2 be two random graphs such that

$$G_1 \preceq G_2 \quad \text{or} \quad G_1 \preceq_{1-o(1)} G_2. \tag{11}$$

Assume that for an increasing property $\mathcal{A}$,

$$\Pr\{G_1 \in \mathcal{A}\} \to 1.$$

Under the coupling (G_1, G_2) given by (11) define event $\mathcal{H} := \{G_1 \subseteq G_2\}$. Then

$$\begin{aligned}
1 \geq \Pr\{G_2 \in \mathcal{A}\} &\geq \Pr\{G_2 \in \mathcal{A}|\mathcal{H}\}\Pr\{\mathcal{H}\} \\
&\geq \Pr\{G_1 \in \mathcal{A}|\mathcal{H}\}\Pr\{\mathcal{H}\} \\
&= \Pr\{\{G_1 \in \mathcal{A}\} \cap \mathcal{H}\} \\
&= \Pr\{G_1 \in \mathcal{A}\} + \Pr\{\mathcal{H}\} - \Pr\{\{G_1 \in \mathcal{A}\} \cup \mathcal{H}\} \\
&\geq \Pr\{G_1 \in \mathcal{A}\} + \Pr\{\mathcal{H}\} - 1 \\
&= 1 + o(1),
\end{aligned}$$

which means that

$$\Pr\{G_2 \in \mathcal{A}\} \to 1.$$

Therefore the above facts concerning the total variation distance and the properties of couplings combined with equations (3), (4), (5) and (6) imply Theorem 1 in the case $np = o(1)$ and combined with equations (3), (7), (8), (9) and (10) imply the theorem in the case $nmp \to \infty$ □

3 Sharp threshold functions

Many graph properties in $G(n, \hat{p})$ follow the so called "minimum degree phenomenon". This means that with high probability the properties hold in $G(n, \hat{p})$ as soon as their necessary minimum degree condition is satisfied. In this section, using Theorem 1, we show that the "minimum degree phenomenon" also holds in the case of $\mathcal{G}(n, m, p)$ for $m = n^\alpha$ and $\alpha > 1$ and, to some extent, for $m = n^\alpha$ and $\alpha \leq 1$. Recall that while studying properties of $\mathcal{G}(n, m, p)$, it is standard to assume $m = n^\alpha$, and in this section we follow this convention. The properties considered are: k-connectivity, perfect matching containment and Hamilton cycle containment. All these properties are increasing and thus Theorem 1 may be used. Note that for p_k considered in the theorems if $\alpha > 1$ then $np \to 0$ and if $\alpha \leq 1$, then $np \to \infty$. The following theorems are proved.

Theorem 2. *Let $m = n^\alpha$ and*

$$p_1 = \begin{cases} \frac{\ln n + \omega}{m}, & \text{for } \alpha \leq 1; \\ \sqrt{\frac{\ln n + \omega}{nm}}, & \text{for } \alpha > 1. \end{cases}$$

(i) If $\omega \to -\infty$, then with high probability $\mathcal{G}(n, m, p_1)$ is disconnected and does not contain a perfect matching.

(ii) If $\omega \to \infty$, then with high probability $\mathcal{G}(n, m, p_1)$ is connected and contains a perfect matching.

Theorems 3 and 4 consider the same properties. However they are stated separately since in the case $\alpha > 1$ (Theorem 3) the obtained threshold functions are tight and for $\alpha \leq 1$ (Theorem 4) they may possibly be tightened by other methods.

Theorem 3. *Let $k \geq 1$ be a constant integer, $\alpha > 1$, $m = n^\alpha$ and*

$$p_k = \sqrt{\frac{\ln n + (k-1)\ln\ln n + \omega}{mn}}.$$

1. *(i) If $\omega \to -\infty$, then with high probability $\mathcal{G}(n, m, p_k)$ is not k-connected.*

   (ii) If $\omega \to \infty$, then with high probability $\mathcal{G}(n, m, p_k)$ is k-connected.

2. *(i) If $\omega \to -\infty$, then with high probability $\mathcal{G}(n, m, p_2)$ does not contain a Hamilton cycle.*

   (ii) If $\omega \to \infty$, then with high probability $\mathcal{G}(n, m, p_2)$ contains a Hamilton cycle.

Theorem 4. *Let $k \geq 1$ be a constant integer, $\alpha \leq 1$, $m = n^\alpha$,*

$$p_k = \frac{\ln n + (k-1)\ln\ln n + \omega}{m}.$$

1. *(i) If $\omega \to -\infty$, then with high probability $\mathcal{G}(n, m, p_1)$ is not k-connected.*

   (ii) If $\omega \to \infty$, then with high probability $\mathcal{G}(n, m, p_k)$ is k-connected.

2. *(i) If $\omega \to -\infty$, then with high probability $\mathcal{G}(n, m, p_1)$ does not contain a Hamilton cycle.*

   (ii) If $\omega \to \infty$, then with high probability $\mathcal{G}(n, m, p_2)$ contains a Hamilton cycle.

Theorem 2 in its part concerning connectivity was obtained in [17]. However we state it here since it gives a global overview of the new method's implications and we are able to provide a new elegant proof of it. To the best of our knowledge the remaining results have not been proved before.

Proof of Theorems 2, 3 and 4. Denote

$$\hat{p}_k = \frac{\ln n + (k-1)\ln\ln n + \omega}{n}.$$

By some classical results (Erdős and Rényi [7], Bollobás and Thomason [5], Komlós and Szeméredi [12] and Bollobás [4])

1. (i) If $\omega \to -\infty$, then with high probability $G(n, \hat{p}_1)$ does not contain a perfect matching.
   (ii) If $\omega \to \infty$, then with high probability $G(n, \hat{p}_1)$ contains a perfect matching.
2. (i) If $\omega \to -\infty$, then with high probability $G(n, \hat{p}_k)$ is not k-connected.
   (ii) If $\omega \to \infty$, then with high probability $G(n, \hat{p}_k)$ is k-connected.
3. (i) If $\omega \to -\infty$, then with high probability $G(n, \hat{p}_2)$ does not contain a Hamilton cycle.
   (ii) If $\omega \to \infty$, then with high probability $G(n, \hat{p}_2)$ contains a Hamilton cycle.

Since k–connectivity, Hamilton cycle containment and perfect matching containment are all increasing properties, parts (ii) of Theorems 2, 3 and 4 follow by Theorem 1.

We are left with proving parts (i). The necessary condition for k–connectivity, perfect matching and Hamilton cycle containment are minimum degree at least k, 1 and 2, respectively. Therefore the following two lemmas imply parts (i) of the theorems.

Denote by $\delta(\mathcal{G}(n, m, p))$ the minimum degree of $\mathcal{G}(n, m, p)$.

Lemma 1. *Let $k \geq 1$ be a constant integer, $\alpha > 1$ and*

$$p_k = \sqrt{\frac{\ln n + (k-1)\ln\ln n + \omega}{nm}},$$

(i) If $\omega \to -\infty$ then with high probability $\delta(\mathcal{G}(n, m, p_k)) < k$

(ii) If $\omega \to \infty$ then with high probability $\delta(\mathcal{G}(n, m, p_k)) \geq k$

Lemma 2. *Let $\alpha \leq 1$ and*

$$p_1 = \frac{\ln n + \omega}{m}.$$

(i) If $\omega \to -\infty$ then with high probability $\delta(\mathcal{G}(n, m, p_1)) = 0$.

(ii) If $\omega \to \infty$ then with high probability $\delta(\mathcal{G}(n, m, p_1)) \geq 1$.

Lemma 2 was shown in [17]. Part (ii) of Lemma 1 is easily obtained by the first moment method (see for example [10]). Moreover, to prove the theorems, only part (i) is needed. Its proof is a standard application of the second moment method (see [10]) and we sketch it for completeness.

We assume that $\omega = o(\ln n)$. Since the property "minimum degree at least k" is increasing, the result for larger ω follows by a simple coupling argument applied to $\mathcal{G}(n, m, \cdot)$. The vertex degree analysis becomes complex for α near 1 due to edge dependencies. Therefore, to simplify arguments, instead of a random variable representing the degree of a vertex $v \in \mathcal{V}$, we study the auxiliary random variable

$$Z_v = |\{(v', w) : v \neq v' \in \mathcal{V}, w \in W_v \text{ and } w \in W_{v'}\}|.$$

Let

$$\xi_v = \begin{cases} 1, & \text{if } Z_v = k-1; \\ 0, & \text{otherwise;} \end{cases} \quad \text{and} \quad \xi = \sum_{v \in \mathcal{V}} \xi_v.$$

Clearly, if $\xi_v = 1$, then the degree of the vertex v is at most $k-1$. Therefore $\Pr\{\xi > 0\} \to 1$ implies part (i) of Lemma 1.

Let $X_v = |W_v|$. By Chernoff's bound (see Theorem 2.1 in [10] or Lemma 1.1 in [14]),

$$\Pr\{x_- \le X_v \le x_+\} = 1 - o\left(n^{-2}\right) \quad \text{for } x_\pm = mp_k\left(1 \pm \sqrt{5\ln n/(mp_k)}\right).$$

Moreover, given $X_v = x$, Z_v has the binomial distribution $\operatorname{Bin}((n-1)x, p_k)$. Thus after careful calculation we get

$$\begin{aligned} \mathbb{E}\xi &= n\Pr\{Z_v = k-1\} \\ &= n\textstyle\sum_{x=x_-}^{x_+} \Pr\{Z_v = k-1 | X_v = x\}\Pr\{X_v = x\} + o\left(n^{-2}\right) \\ &\ge \tfrac{1}{(k-1)!}\exp\left(-\omega + o(1)\right)\left(1 + o(1)\right) \to \infty. \end{aligned} \tag{12}$$

Let $v, v' \in \mathcal{V}$ and $S = |W_v \cap W_{v'}|$. Given $i \in \{0,1,2\}$ and $x, x' \in [x_-; x_+ + 2]$ denote by $\mathcal{H}(x,x',i)$ the event $\{X_v = x+i, X_{v'} = x'+i, S = i\}$. A calculation shows that if $i \in \{0,1,2\}$ and $x, x' \in [x_-; x_+ + 2]$, then uniformly over all x, x'

$$\begin{aligned} \Pr\{\mathcal{H}(x,x',i)\} &= \Pr\{X_v = x+i\}\Pr\{X_{v'} = x'+i\}\Pr\{S = i | X_{v'} = x'+i, X_v = x+i\} \\ &= (1+o(1))\Pr\{X_v = x\}\Pr\{X_{v'} = x'\}\Pr\{S = i\}. \end{aligned}$$

Moreover, uniformly over all $x, x' \in [x_-; x_+ + 2]$, we have

$$\begin{aligned} &\Pr\{Z_v = k-1, Z_{v'} = k-1 | \mathcal{H}(x,x',i)\} \\ &\qquad = (1+o(1))\Pr\{Z_v = k-1 | X_v = x\}\Pr\{Z_{v'} = k-1 | X_{v'} = x\}. \end{aligned}$$

Denote $J = [x_- + 2, x_+]$. Since S has the binomial distribution $\operatorname{Bin}(m, p_k^2)$, and by Chernoff's bound applied to X_v and $X_{v'}$, we get

$$\begin{aligned} &\Pr\{X_v \notin J \text{ or } X_{v'} \notin J \text{ or } S \notin \{0,1,2\}\} \\ &\qquad \le \Pr\{X_v \notin J\} + \Pr\{X_{v'} \notin J\} + \Pr\{S \ge 3\} = o\left(n^{-2}\right). \end{aligned}$$

Finally by the above calculation and (12) for $v \ne v' \in \mathcal{V}$

$$\begin{aligned} \mathbb{E}\xi(\xi-1) =& n(n-1)\Pr\{Z_v = k-1, Z_{v'} = k-1\} \\ \le& n(n-1) \\ &\cdot \sum_{x=x_-}^{x_+}\sum_{x'=x_-}^{x_+}\sum_{i=0}^{2} \Pr\{Z_v = k-1, Z_{v'} = k-1 | \mathcal{H}(x,x',i)\}\Pr\{\mathcal{H}(x,x',i)\} \\ &+ n(n-1)\Pr\{X_v \notin J \text{ or } X_{v'} \notin J \text{ or } S \notin \{0,1,2\}\} \\ =& (1+o(1))\Pr\{Z_v = k-1\}\Pr\{Z_{v'} = k-1\} + o(1), \end{aligned}$$

which by the second moment method implies $\Pr\{\xi > 0\} \to 1$. □

4 Final remarks

The obtained results may be extended to a wider class of the general random intersection graph model $\overline{\mathcal{G}}(n,m,\mathcal{P}_{(m)})$. As an example we state here a uniform random intersection graph which is $\overline{\mathcal{G}}(n,m,\mathcal{P}_{(m)}) = \overline{\mathcal{G}}(n,m,\mathcal{P}_d)$ with probability distribution $\mathcal{P}_{(m)} = \mathcal{P}_d$ concentrated in $d = d(n)$, for some $d(n)$. More precisely in $\overline{\mathcal{G}}(n,m,\mathcal{P}_d)$, for all $v \in \mathcal{V}$, the set W_v is chosen uniformly at random from all d–element subsets of $\mathcal{W}$. By Lemma 4 from [3] Theorems 2 and 3 hold true, if we assume that $\alpha > 1$ and replace p_k by $d_k = mp_k$ and $\mathcal{G}(n,m,p_k)$ by $\overline{\mathcal{G}}(n,m,\mathcal{P}_{d_k})$.

As it clearly follows from Theorem 2, the couplings used in the proof of Theorem 1 are tight. However, in the case $np \to \infty$ they do not always give the best results (see Theorem 4). Notice that in the case $\alpha < 1$ it is easy to strengthen Lemma 2 by a simple application of Chernoff's bound.

Lemma 3. *Let $\alpha < 1$ and*

$$p_1 = \frac{\ln n + \omega}{m}.$$

If $\omega \to \infty$ then with high probability $\delta(\mathcal{G}(n,m,p_1)) \geq (1+o(1)) n \ln n / m$.

Therefore having in mind the "minimum degree phenomenon", we may conjecture that the threshold function given in Theorem 4 may be tightened. However we believe that to prove the following conjecture a new method has to be used.

Conjecture 1. Let $\alpha < 1$,

$$p = \frac{\ln n + \omega}{m},$$

and $\omega \to \infty$. Then with high probability $\mathcal{G}(n,m,p)$ is k-connected for any constant k and contains a Hamilton cycle.

This conjecture contains the assumption $\alpha < 1$. Probably the case $\alpha = 1$ is more complex. The thesis may be supported by the results concerning the degree distribution [18] and the phase transition [13] for $\alpha = 1$. Although they consider p near phase transition threshold, they show that, for some properties, there is a value of α for which an analysis of $\mathcal{G}(n,m,p)$ is complicated.

Acknowledgements

I would like to thank all colleagues attending our seminar and Mindaugas Bloznelis for their helpful remarks, which allowed me to improve the layout of the paper and remove some ambiguities. I am also grateful to the anonymous referee and my colleagues Erhard Godehardt and Michał Ren for making many helpful suggestions that improved the exposition.

References

[1] Barbour, A. D., Holst, L. and Janson, S. *Poisson Approximation.* Oxford University Press, 1992

[2] Behrisch, M. Component evolution in random intersection graphs. *Electron. J. Combin.* **14**, R17, 2007.

[3] Bloznelis, M., Jaworski, J. and Rybarczyk, K. Component evolution in a secure wireless sensor networks. *Networks* **53**, 19–26, 2009.

[4] Bollobás, B. The evolution of sparse graphs, *Graph Theory and Combinatorics (Cambridge 1983)*, 35-57, Academic Press, 1984.

[5] Bollobás, B. and Thomason, A. Random graphs of small order. *Random Graphs '83 , Proceedings, Poznań, 1983*, 47 – 97, 1985.

[6] Efthymiou, C. and Spirakis, P. G. On the existence of hamiltonian cycles in random intersection graphs. In: *Automata, Languages and Programming 32nd International Colloquium, ICALP 2005, Lisbon, Portugal,* 690–701, 2005.

[7] Erdős, P. and Rényi, A. On random graphs I. *Publ. Math. Debrecen* **6,** 290–297, 1959.

[8] Fill, J. A., Scheinerman, E. R. and Singer-Cohen, K. B. 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 Algorithms* **16,** 156–176, 2000.

[9] Godehardt, E. and Jaworski, J. Two models of random intersection graphs for classification. In: M. Schwaiger, O. Opitz, (eds.): *Exploratory Data Analysis in Empirical Research, (Proceedings of the 25th Annual Conference of the Gesellschaft für Klassifikation e.V., University of Munich, 2001).* Springer, 67–81, 2002.

[10] Janson, S., Łuczak, T. and Ruciński, A. *Random Graphs.* Wiley-Interscience, New York, 2000.

[11] Karoński, M., Scheinerman, E. R. and Singer-Cohen, K. B. On random intersection graphs: the subgraph problem. *Combin. Probab. Comput.* **8,** 131–159, 1999.

[12] Komlós, J. and Szeméredi, E. Limit distributions for the existence of hamilton cycles in a random graph. *Discrete Math.* **43,** 55 – 63, 1983.

[13] Lagerås, A. N. and Lindholm, M. *Electron. J. Combin.* **15** (1), N10, 2008.

[14] Penrose, M. *Random Geometric Graphs.* Oxford University Press, 2003.

[15] Rybarczyk, K. Equivalence of a random intersection graph and $G(n, p)$. *Random Structures and Algorithms* **38** (1-2), 205–234, 2011.

[16] Rybarczyk, K. and Stark, D. Poisson approximation of the number of cliques in random intersection graphs. *Journal of Applied Probability* **47** (3), 826–840, 2010.

[17] Singer-Cohen, K. B. *Random intersection graphs.* PhD thesis, Department of Mathematical Sciences, The Johns Hopkins University, 1995.

[18] Stark, D. The vertex degree distribution of random intersection graphs. *Random Structures and Algorithms* **24**, 249–258, 2004.