

POISSON APPROXIMATION OF THE NUMBER OF CLIQUES IN RANDOM INTERSECTION GRAPHS

KATARZYNA RYBARCZYK,* *Adam Mickiewicz University*
DUDLEY STARK,** *Queen Mary, University of London*

Abstract

A random intersection graph $G(n, m, p)$ is defined on a set \mathcal{V} of n vertices. There is an auxiliary set \mathcal{W} consisting of m objects, and each vertex $v \in \mathcal{V}$ is assigned a random subset of objects $\mathcal{W}_v \subseteq \mathcal{W}$ such that $w \in \mathcal{W}_v$ with probability p , independently for all $v \in \mathcal{V}$ and all $w \in \mathcal{W}$. Given two vertices $v_1, v_2 \in \mathcal{V}$, we set $v_1 \sim v_2$ if and only if $\mathcal{W}_{v_1} \cap \mathcal{W}_{v_2} \neq \emptyset$. We use Stein's method to obtain an upper bound on the total variation distance between the distribution of the number of h -cliques in $G(n, m, p)$ and a related Poisson distribution for any fixed integer h .

Keywords: Stein's method; Poisson approximation; random intersection graph

2010 Mathematics Subject Classification: Primary 60F05
Secondary 05C80

1. Introduction

In a random intersection graph there is a set of vertices \mathcal{V} and an auxiliary set of objects \mathcal{W} . Each vertex $v \in \mathcal{V}$ is assigned a subset of objects $\mathcal{W}_v \subseteq \mathcal{W}$. Two vertices v_1 and v_2 are adjacent if and only if $\mathcal{W}_{v_1} \cap \mathcal{W}_{v_2} \neq \emptyset$. A general model of the random intersection graph, in which each vertex is assigned a subset of objects $\mathcal{W}_v \subseteq \mathcal{W}$ chosen uniformly from all d -element subsets, where the cardinality d is determined according to the arbitrarily given probability distribution $\mathcal{P}_{(m)}$, was introduced in [9].

Random intersection graphs have been studied in relation to many applications, including the 'gate matrix layout' for VLSI design (see, e.g. [13]), cluster analysis and classification (see, e.g. [9], [10], and [11]), analysis of complex networks (see, e.g. [15] and [16]), network user profiling (see [14]), secure wireless networks (see, e.g. [4] and [7]), and epidemics (see [3] and [5]).

The random intersection graph $G(n, m, p)$ as first defined in [13] is the special case of the above model when $\mathcal{P}_{(m)}$ has binomial distribution $\text{Bin}(m, p)$, i.e. it is defined to be a graph with $|\mathcal{V}| = n$ vertices, $|\mathcal{W}| = m$ objects, and $P(w \in \mathcal{W}_v) = p$, independently for all $v \in \mathcal{V}$ and $w \in \mathcal{W}$. We set

$$m = \lfloor n^\alpha \rfloor \tag{1}$$

for some constant $\alpha > 0$. Given a fixed integer $h \geq 3$ and a graph H on h vertices, define the

Received 18 September 2008; revision received 2 June 2010.

* Postal address: Faculty of Mathematics and Computer Science, Adam Mickiewicz University, ul. Umultowska 87, 61-614 Poznań, Poland.

** Postal address: School of Mathematical Sciences, Queen Mary, University of London, London E1 4NS, UK.

Email address: d.s.stark@qmul.ac.uk

threshold of appearance of copies of H to be a function $\tau(H)$ such that

$$\lim_{n \rightarrow \infty} P(H \subseteq G(n, m, p)) = \begin{cases} 1 & \text{if } p/\tau(H) \rightarrow 0, \\ 0 & \text{if } p/\tau(H) \rightarrow \infty, \end{cases}$$

where $H \subseteq G(n, m, p)$ is the event that $G(n, m, p)$ contains a copy of H . It was shown in [13] that, for a clique K_h of size h , as long as $mp^2 = o(1)$,

$$\tau(K_h) = \begin{cases} n^{-1}m^{-1/h} & \text{for } \alpha \leq 2h/(h-1), \\ n^{-1/(h-1)}m^{-1/2} & \text{for } \alpha \geq 2h/(h-1). \end{cases}$$

It is worth mentioning that in the case in which $h = 3$ and $\alpha \geq 3$ the function $\tau(K_3)$ coincides with the threshold function for having a giant component τ_g (see [1]). However, when $\alpha < 3$, we have $\tau(K_3) \ll \tau_g$. Therefore, we may expect that near τ_g the clustering coefficient of $G(n, m, p)$ should be higher than in $G(n, p)$, where $G(n, p)$ is the random Erdős–Rényi graph with n vertices and edge probability p . It is known that it tends to a constant for $\alpha \leq 1$ (see [6]).

We are interested in showing that the number of copies of cliques K_h in $G(n, m, p)$ is approximately Poisson when p is near $\tau(K_h)$. Thus, we will assume that, for some constant $\delta > 0$, we have

$$\alpha = \frac{2\delta h}{h-1} \tag{2}$$

and that, for another constant $c > 0$,

$$p(n) \sim \begin{cases} cn^{-1}m^{-1/h} & \text{for } 0 < \delta < 1, \\ cn^{-(h+1)/(h-1)} & \text{for } \delta = 1, \\ cn^{-1/(h-1)}m^{-1/2} & \text{for } \delta > 1. \end{cases} \tag{3}$$

For m and p defined by (1), (2), and (3), it is easily checked that

$$mp^2 = o(1) \tag{4}$$

for all $\delta > 0$ and

$$mp \rightarrow \infty \tag{5}$$

for $\delta > \frac{1}{2}$.

All limits in this paper are taken as $n \rightarrow \infty$. Throughout this paper, we use the following standard notation: $a_n \sim b_n$ if $a_n/b_n \rightarrow 1$; $a_n \asymp b_n$ if there exist constants $c > 0$ and $C > 0$ such that $c \leq |a_n/b_n| \leq C$ for all $n \geq 1$; $a_n = O(b_n)$ if there exists a constant $C > 0$ such that $|a_n/b_n| \leq C$ for all $n \geq 1$; and $a_n = o(b_n)$ if $a_n/b_n \rightarrow 0$.

For any given graph H , we use the standard notation $V(H)$ and $E(H)$ to denote its set of vertices and edges, respectively. In what follows we will often use the following definition. For any subset of vertices $V' \subseteq V(H)$, we will denote by $H[V']$ an induced subgraph of H with a set of vertices V' and a set of edges $\{(v_1, v_2) \in E(H) : v_1, v_2 \in V'\}$.

The total variation distance between a random variable taking nonnegative integer values and the random variable P_λ with Poisson distribution with parameter λ is defined to be

$$d_{TV}(X, P_\lambda) = \frac{1}{2} \sum_{k=0}^{\infty} \left| P(X = k) - \frac{e^{-\lambda} \lambda^k}{k!} \right|.$$

Theorem 1. Let $G(n, m, p)$ be a random intersection graph defined with m and p given in terms of n by (1), (2), and (3), and let $h \geq 3$ be a fixed integer. Let X_n be the random variable counting the number of instances of K_h in $G(n, m, p)$.

(i) If $\delta < 1$ then $\lambda_n = \mathbb{E} X_n \sim c^h/h!$ and

$$d_{\text{TV}}(X_n, P_{\lambda_n}) = O(n^{-2\delta/(h-1)}).$$

(ii) If $\delta = 1$ then $\lambda_n = \mathbb{E} X_n \sim (c^h + c^{h(h-1)})/h!$ and

$$d_{\text{TV}}(X_n, P_{\lambda_n}) = O(n^{-2/(h-1)}).$$

(iii) If $\delta > 1$ then $\lambda_n = \mathbb{E} X_n \sim c^{h(h-1)}/h!$ and

$$d_{\text{TV}}(X_n, P_{\lambda_n}) = O(n^{(1-\delta)h-2/(h-1)} + n^{-1}).$$

In (iii), the bound $n^{((1-\delta)h-2/(h-1))}$ is larger than the bound n^{-1} if and only if

$$\delta < \frac{h^2 - 3}{h(h-1)}.$$

We do not consider $\lambda_n \rightarrow \infty$ in this paper, only convergent λ_n .

In comparison, for Erdős–Rényi graphs $G(n, \hat{p})$, in which edges appear independently and with probability \hat{p} , the bound obtained by Stein's method (presented in Section 2) for the total variation distance between the number of cliques of size h and the Poisson distribution with parameter $\lambda = \binom{n}{h} \hat{p}^{\binom{h}{2}}$ is of order $O(n^{-1})$ for all fixed $h \geq 3$ when λ is bounded. In light of the equivalence theorem proved in [8], it is no surprise that the above result agrees with the well-known result for Erdős–Rényi graphs $G(n, \hat{p})$ when $\alpha > 6$ and $\hat{p} \sim mp^2$, since the abovementioned equivalence theorem says that, for $\alpha > 6$ and $\hat{p} \sim mp^2$, the graphs $G(n, m, p)$ and $G(n, \hat{p})$ have asymptotically the same properties. Moreover, this result is consistent with the hypothesis from [8] of the equivalence of the models for $\alpha > 3$ and their inequivalence for $\alpha < 3$. From the applied point of view, it would be interesting to obtain similar results for the general random intersection graph models defined as in [9]. It seems that similar methods to those used in this paper should give an analogous result for the random intersection graph model with probability distributions defined as in [6] (random intersection graphs with tunable degree distributions), i.e. for the binomial distribution with the parameter which is itself a random variable. Unfortunately, in the other interesting case of the uniform random intersection graph (in which $\mathcal{P}_{(m)}$ is the degenerate distribution), which is used for example to model wireless sensor networks [2], [4], [7], it seems it will not be straightforward to prove the analogous result.

The paper is organised as follows. We describe Stein's method in Section 2. In Sections 3, 4, and 5 we discuss the notion of a clique cover. Namely, in Section 3 definitions are given, in Section 4 lemmas describing relations between various clique covers are presented, and in Section 5 obtained results are used to describe the clique covers most likely to occur in random intersection graphs. In the last two sections these results are used to complete the proof. In Section 6 expectations and probabilities of the inclusion of one and two copies of K_h are calculated, whereas in Section 7 Stein's method is applied.

2. Stein’s method

Stein’s method is a way of showing that the distribution of a random variable is close to some target distribution. We will use a version of Stein’s method for Poisson approximation given in terms of dependency graphs.

Suppose that $X = \sum_{\beta \in \Gamma} I_\beta$ is a sum over an index set Γ of indicator random variables. A *dependency graph* L is a graph with vertex set Γ having the property that whenever $A \subseteq \Gamma$ and $B \subseteq \Gamma$ satisfy the property that there are no edges between A and B in L , it follows that $\{I_\beta : \beta \in A\}$ and $\{I_\gamma : \gamma \in B\}$ are mutually independent sets of random variables. We define $\lambda = EX$, $\pi_\beta = EI_\beta$ for $\beta \in \Gamma$, and P_λ to be a random variable with Poisson distribution with parameter λ . From Theorem 6.23 of [12] we have

$$d_{TV}(X, P_\lambda) \leq \min(\lambda^{-1}, 1) \left(\sum_{\beta \in V(L)} \pi_\beta^2 + \sum_{\beta\gamma \in E(L)} (E(I_\beta I_\gamma) + \pi_\beta \pi_\gamma) \right), \tag{6}$$

where the sum $\sum_{\beta\gamma \in E(L)}$ means summing over ordered pairs (β, γ) such that $\{\beta, \gamma\} \in E(L)$.

Denote by $K_{\mathcal{V}}$ the complete graph on the set of vertices \mathcal{V} . In our application of Stein’s method $\Gamma = \{K_{\mathcal{V}'} : \mathcal{V}' \subseteq \mathcal{V}, |\mathcal{V}'| = h\}$ and, for each $\beta \in \Gamma$, $I_\beta = \mathbf{1}(\beta \subseteq G(n, m, p))$, where $\mathbf{1}(\beta \subseteq G(n, m, p))$ is the indicator random variable which equals 1 when $\beta \subseteq G(n, m, p)$ and equals 0 otherwise. With these definitions, X counts the number of h -cliques in $G(n, m, p)$. Note that

$$\pi_\beta = P(\beta \subseteq G(n, m, p))$$

and

$$E(I_\beta I_\gamma) = P(\beta \subseteq G(n, m, p), \gamma \subseteq G(n, m, p)) = P(\beta \cup \gamma \subseteq G(n, m, p)), \tag{7}$$

where $\beta \cup \gamma$ is the graph with vertex set $V(\beta \cup \gamma) = V(\beta) \cup V(\gamma)$ and edge set $E(\beta \cup \gamma) = E(\beta) \cup E(\gamma)$.

We define a dependency graph L with a vertex set Γ such that if $\beta = K_{\mathcal{V}'}$ and $\gamma = K_{\mathcal{V}''}$, then $\{\beta, \gamma\} \in E(L)$ if and only if $\mathcal{V}' \cap \mathcal{V}'' \neq \emptyset$. Since subgraphs with disjoint vertex sets appear independently in $G(n, m, p)$, this is a well-defined dependency graph.

3. Clique covers

In order to estimate π_β as well as $E(I_\beta I_\gamma)$, we use the notion of clique covers of a graph introduced in [13].

Let H be any subgraph of $K_{\mathcal{V}}$. A *clique cover* \mathcal{C} is a set of subsets of $V(H)$ such that, firstly, $|C| \geq 2$ for all $C \in \mathcal{C}$ and, secondly, such that, for any $\{v_1, v_2\} \in E(H)$, there exists $C \in \mathcal{C}$ such that $v_1, v_2 \in C$. Note that, from the definition, the cliques induced in $K_{\mathcal{V}}$ by the sets in \mathcal{C} cover all the edges of H . To illustrate this, let H be a triangle. The nine clique covers of H are depicted in Figure 1. Note that edges may be contained in more than one set from a clique cover \mathcal{C} , but must be contained in at least one of them.

In a random intersection graph the set of vertices associated with a particular object $w \in \mathcal{W}$ always form a clique in $G(n, m, p)$, and the set of edges in $G(n, m, p)$ is the union of the m edge sets of the cliques generated by the elements of \mathcal{W} . We say that $H \subseteq G(n, m, p)$ is given by a clique cover $\mathcal{C} = \{C_1, \dots, C_t\}$ if there is a family of disjoint nonempty subsets $\{W_1, \dots, W_t\}$ of \mathcal{W} such that, firstly, for all $1 \leq i \leq t$, each element of W_i is an object assigned to all the vertices of C_i and no other vertices from $V(H)$, and, secondly, such that each $w \in \mathcal{W} \setminus \bigcup_{i=1}^t W_i$ is an

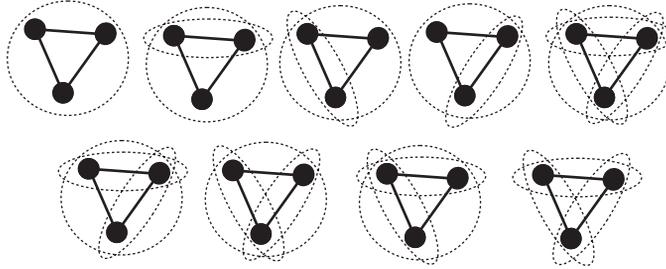


FIGURE 1: All the clique covers of a triangle.

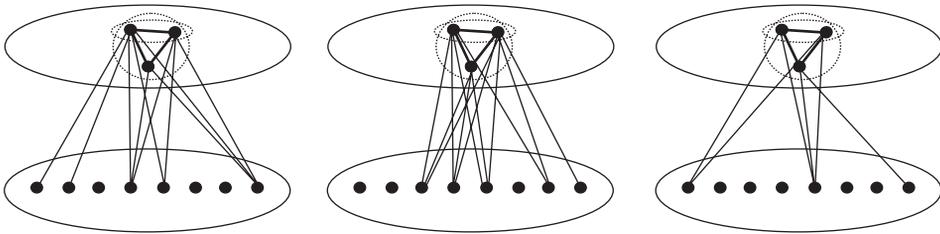


FIGURE 2: Examples of a triangle being given by a clique cover consisting of one three-element set and one two-element set.

object assigned to at most one vertex from $V(H)$. Note that there is a large number of ways in which a subgraph can be given by a particular clique cover. Figure 2 shows three ways in which a triangle can be given by a clique cover containing exactly one set of size 3 and one set of size 2.

Note that a graph H is given by exactly one clique cover and there are only a finite number of clique covers covering H . Therefore, if we let $\pi(H, \mathcal{C})$ denote the probability that H is given by a clique cover $\mathcal{C} = \{C_1, \dots, C_t\}$ then

$$\pi_H = \sum_{\text{Clique covers } \mathcal{C} \text{ of } H} \pi(H, \mathcal{C}).$$

In view of the preceding paragraph, the asymptotic behaviour of π_H is determined by the clique covers \mathcal{C} of H for which $\pi(H, \mathcal{C})$ dominates asymptotically. To estimate $\pi(H, \mathcal{C})$, we will use

$$\pi(H, \mathcal{C}) \sim \prod_{i=1}^t mp^{|C_i|} \tag{8}$$

from page 138 of [13] (we have used the fact that $|C_i| \geq 2$ in the abovementioned definition of a clique cover).

We will call a clique cover an *edge cover* if it contains only two-element sets. We will call a clique cover a *cover by one set* if it consists of only one set.

4. Operations on clique covers

For particular subgraphs H (which we will later take to be of the form β or $\beta \cup \gamma$ for $\beta, \gamma \in \Gamma$ to apply the bounds of Section 2), we will be interested in finding clique covers \mathcal{C} for which

$\pi(H, \mathcal{C})$ is of the largest asymptotic order of magnitude. To do this, we need information about the convergence of the fraction $\pi(H, \mathcal{C})/\pi(H, \mathcal{C}')$ for different clique covers \mathcal{C} and \mathcal{C}' of H .

Given a clique cover \mathcal{C} of H , we say that \mathcal{C}' is obtained from \mathcal{C} by merging two sets if, for some $C_1, C_2 \in \mathcal{C}$, \mathcal{C}' is obtained from \mathcal{C} by replacing C_1 and C_2 with $C_1 \cup C_2$. We say that \mathcal{C}' is obtained from \mathcal{C} by an edge cover of a set if, for some $C \in \mathcal{C}$, \mathcal{C}' is obtained from \mathcal{C} by replacing C with the set of edges in $H[C]$.

We now define a quantity k_0 which appears in the next two lemmas. We will show in Lemma 1 and Lemma 2 below that k_0 is roughly the set size below which replacing sets by edge covers increases $\pi(H, \mathcal{C})$ and above which merging sets increases $\pi(H, \mathcal{C})$. In view of (1), (2), and (3), there is a real number a such that $p(n) \asymp n^{-a}$ defined by

$$a = \begin{cases} \frac{h-1+2\delta}{h-1} & \text{for } \delta \leq 1, \\ \frac{1+\delta h}{h-1} & \text{for } \delta \geq 1. \end{cases}$$

We define the quantity k_0 by

$$k_0 = \frac{\alpha}{2a - \alpha}.$$

Using (2), we note that

$$k_0 = \begin{cases} \frac{\delta h}{h - \delta h - 1 + 2\delta} & \text{for } \delta \leq 1, \\ \delta h & \text{for } \delta \geq 1. \end{cases}$$

Moreover, we have

$$0 < k_0 < h \quad \text{for } \delta < 1, \quad k_0 = h \quad \text{for } \delta = 1, \quad k_0 > h \quad \text{for } \delta > 1. \tag{9}$$

Lemma 1. *Let H be any fixed graph, let m and p be given by (1), (2), and (3), and let \mathcal{C} be any clique cover of H . Moreover, let \mathcal{C}' be a clique cover of H obtained from \mathcal{C} by replacing any $C \in \mathcal{C}$ for which $|C| = k > 2$ with an edge cover of $H[C]$. Then, if $k_0 \geq 3$,*

$$\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} = \begin{cases} o(1) & \text{for } 2 < k < k_0, \\ O(1) & \text{for } k = k_0. \end{cases}$$

Moreover, if $H[C]$ is a clique then

$$\begin{aligned} \frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} &= o(1) \quad \text{for } 2 < k < k_0, \\ \frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} &\asymp 1 \quad \text{for } k = k_0, \\ \frac{\pi(H, \mathcal{C}')}{\pi(H, \mathcal{C})} &= o(1) \quad \text{for } k > \max\{k_0, 2\}. \end{aligned}$$

Proof. We will compare the probabilities of the clique covers $\pi(H, \mathcal{C})$ and $\pi(H, \mathcal{C}')$, where \mathcal{C}' is obtained from \mathcal{C} by replacing $C \in \mathcal{C}$, $|C| = k$, with an edge cover of $H[C]$. Let e be the number of edges in $H[C]$. Then,

$$mp^k \asymp n^{\alpha-ka}$$

and

$$(mp^2)^{\binom{k}{2}} \asymp n^{k(k-1)(\alpha-2a)/2}.$$

Thus, from (8),

$$\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} \asymp \frac{mp^k}{(mp^2)^e} \leq \frac{mp^k}{(mp^2)^{\binom{k}{2}}} = n^{f(k)/2},$$

where

$$\begin{aligned} f(k) &= 2\alpha - 2ka - k(k-1)(\alpha-2a) \\ &= (2a-\alpha)k^2 - (4a-\alpha)k + 2\alpha \\ &= (2a-\alpha)(k-2)(k-k_0); \end{aligned}$$

thus,

$$f(k) = 0 \quad \text{if and only if} \quad k = 2 \text{ or } k = k_0$$

and, since $2a - \alpha > 0$,

$$f(k) < 0 \quad \text{for } k \in (2, k_0), k_0 > 2,$$

and

$$f(k) > 0 \quad \text{for } k > \max\{k_0, 2\}.$$

Moreover, if $e = \binom{k}{2}$ then

$$\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} \asymp n^{f(k)/2}.$$

We call a clique cover \mathcal{C} a *simple clique cover* if, for any two different $C_1, C_2 \in \mathcal{C}$, we have $|C_1 \cap C_2| \leq 1$.

Lemma 2. *Let H be any fixed graph, let m and p be given by (1), (2), and (3), and let \mathcal{C} be a simple clique cover of H . Moreover, suppose that $|C_1 \cap C_2| = 1$ and assume that in H , for $C_1, C_2 \in \mathcal{C}$, there are exactly k edges $\{e_1, \dots, e_k\}$ with one vertex in C_1 and one vertex in C_2 . If \mathcal{C}' is a clique cover of H obtained from \mathcal{C} by replacing C_1 and C_2 with $C_1 \cup C_2$ and then merging all sets containing at least one edge from $\{e_1, \dots, e_k\}$ with $C_1 \cup C_2$, then*

$$\frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} = \begin{cases} o(1) & \text{for } k > \frac{1}{2}k_0 - \frac{1}{2}, \\ O(1) & \text{for } k = \frac{1}{2}k_0 - \frac{1}{2}. \end{cases}$$

Proof. Let H be any graph on h_1 vertices, let $p \asymp n^{-a}$, $m = n^\alpha$, and let $k_1 + 1 \geq 2$ and $k_2 + 1 \geq 2$ be the sizes of sets C_1 and C_2 in a clique cover \mathcal{C} , respectively. Moreover, let $\{e_1, \dots, e_k\}$ be the set of all edges with one vertex in C_1 and one vertex in C_2 . Since \mathcal{C} is simple for each e_i , there is exactly one set $C_{e_i} \in \mathcal{C}$ containing e_i . In addition, for $i \neq j$, $C_{e_i} \neq C_{e_j}$ (otherwise, $|C_{e_i} \cap C_{e_j}| \geq 2$ or $|C_{e_i} \cap C_2| \geq 2$). From the statement of the lemma, \mathcal{C}' is obtained from \mathcal{C} by replacing $C_1, C_2, C_{e_1}, \dots, C_{e_k}$ with one set $C = C_1 \cup C_2 \cup C_{e_1} \cup \dots \cup C_{e_k}$. From

the assumption that \mathcal{C} is simple we have $|C_{e_i} \cap (C_1 \cup C_2)| = 2$ for all $1 \leq i \leq k$; thus,

$$\begin{aligned} \left| C_1 \cup C_2 \cup \bigcup_{i=1}^k C_{e_i} \right| &= \left| C_1 \cup C_2 \cup \bigcup_{i=1}^k (C_{e_i} \setminus (C_1 \cup C_2)) \right| \\ &\leq |C_1 \cup C_2| + \sum_{i=1}^k (|C_{e_i}| - 2) \\ &= k_1 + k_2 + 1 + \sum_{i=1}^k (|C_{e_i}| - 2) \end{aligned}$$

and

$$\frac{1}{p^{|C_1 \cup C_2 \cup \bigcup_{i=1}^k C_{e_i}|}} \leq \frac{1}{p^{k_1+k_2+1} \prod_{i=1}^k p^{|C_{e_i}|-2}}.$$

Hence,

$$\begin{aligned} \frac{\pi(H, \mathcal{C})}{\pi(H, \mathcal{C}')} &\sim \frac{mp^{k_1+1} mp^{k_2+1} \prod_{i=1}^k mp^{|C_{e_i}|}}{mp^{|C_1 \cup C_2 \cup \bigcup_{i=1}^k C_{e_i}|}} \\ &= O(1) \frac{m^2 p^{k_1+k_2+2}}{mp^{k_1+k_2+1}} \prod_{i=1}^k \frac{mp^{|C_{e_i}|}}{p^{|C_{e_i}|-2}} \\ &= O(1) mp (mp^2)^k \\ &= O(1) n^{\alpha-a+k(\alpha-2a)} \\ &= O(1) n^{g(k)}, \end{aligned}$$

where $g(k) = \alpha - a + k(\alpha - 2a)$. Since $2a - \alpha > 0$,

$$\begin{aligned} g(k) \leq 0 &\iff \alpha - a + k(\alpha - 2a) \leq 0 \\ &\iff \alpha - a \leq k(2a - \alpha) \\ &\iff k \geq \frac{\alpha - a}{2a - \alpha} = \frac{\alpha/2 - a + \alpha/2}{2a - \alpha} = \frac{1}{2}k_0 - \frac{1}{2}. \end{aligned}$$

5. Optimal clique covers

We will call a clique cover \mathcal{C} of a graph H an *optimal clique cover* if $\pi(H, \mathcal{C}')/\pi(H, \mathcal{C}) = O(1)$ for all other clique covers \mathcal{C}' of H . Note that, since there is a finite number of clique covers, an optimal clique cover exists, and that

$$\pi_H \sim \sum_{\mathcal{C} \text{ optimal clique cover}} \pi(H, \mathcal{C}). \tag{10}$$

It follows from Lemma 3 that we may assume that optimal clique covers are simple. Recall assumption (4).

Lemma 3. *Let \mathcal{C} be an optimal clique cover for graph H . There are no sets $C_1, C_2 \in \mathcal{C}$ such that $|C_1 \cap C_2| \geq 2$.*

Proof. Let $|C_1| \geq 2$ and $|C_2| \geq 2$. We create a new clique cover \mathcal{C}' from \mathcal{C} by replacing C_1 and C_2 with $C = C_1 \cup C_2$, for which

$$\pi(H, \mathcal{C}) = \pi(H, \mathcal{C}') \frac{mp^{|C_1|} mp^{|C_2|}}{mp^{|C|}} = \pi(H, \mathcal{C}') mp^{|C_1 \cap C_2|} = o(\pi(H, \mathcal{C}')).$$

We will also need the following lemma.

Lemma 4. *Let \mathcal{C} be an optimal clique cover of H . There is no set $C \in \mathcal{C}$ such that $C = C_1 \cup C_2$, $C_1 \cap C_2 = \emptyset$, and $C_1, C_2 \neq \emptyset$, and such that there are no edges between C_1 and C_2 in H .*

Proof. Assume that in \mathcal{C} there is a set C such that $C = C_1 \cup C_2$, $C_1 \cap C_2 = \emptyset$, and $C_1, C_2 \neq \emptyset$, and that there are no edges between C_1 and C_2 in H .

If $|C_1| = 1$ and $|C_2| = 1$, then $\mathcal{C}' = \mathcal{C} \setminus \{C\}$ is a clique cover of H and $\pi(H, \mathcal{C}) = \pi(H, \mathcal{C}') mp^2 = o(\pi(H, \mathcal{C}'))$.

Let $|C_1| = 1$ and $|C_2| \geq 2$. If \mathcal{C}' is obtained from \mathcal{C} by replacing C with C_2 , then obviously \mathcal{C}' is a clique cover and $\pi(H, \mathcal{C}) = \pi(H, \mathcal{C}') mp^{|C_1|} / mp^{|C_2|} = o(\pi(H, \mathcal{C}'))$.

Let $|C_1| \geq 2$ and $|C_2| \geq 2$. If \mathcal{C}' is obtained from \mathcal{C} by replacing C with C_1 and C_2 , then $\pi(H, \mathcal{C}) = \pi(H, \mathcal{C}') mp^{|C_1|} / m^2 p^{|C|} = o(\pi(H, \mathcal{C}'))$.

We now use our results from the previous section to find optimal clique covers for complete graphs. In Lemma 6 below we find optimal clique covers for unions of two complete graphs. Lemma 5 below can be deduced from the proof of Corollary 6 of [13].

Lemma 5. *Let β be a clique on h vertices. For m given by (1) and (2), and p given by (3), the only optimal clique covers of β are: (i) a cover by one set when $\delta < 1$; (ii) an edge cover and a cover by one set when $\delta = 1$; (iii) an edge cover when $\delta > 1$.*

Proof. Let \mathcal{C} be any simple clique cover.

(i) Suppose that $\delta < 1$. We know from (9) that $0 < k_0 < h$. If \mathcal{C} consists of only sets of size at most k_0 then from Lemma 1 we have $\pi(\beta, \mathcal{C}) = O(\pi(\beta, \mathcal{C}')) = o(\pi(\beta, \mathcal{C}''))$, where \mathcal{C}' is an edge cover and \mathcal{C}'' is a cover by one set. If \mathcal{C} contains a set C for which

$$|C| - 1 \geq \max\{k_0, 2\} - 1 > \frac{1}{2}k_0 - \frac{1}{2},$$

then we can merge C with other sets from \mathcal{C} repeatedly until we get the cover by one set \mathcal{C}'' , and from Lemma 2 we have $\pi(\beta, \mathcal{C}) = o(\pi(\beta, \mathcal{C}''))$. Thus, \mathcal{C}'' is the only optimal clique cover.

(ii) Suppose that $\delta = 1$. In this case $k_0 = h$. We know that if in any clique cover \mathcal{C} of β we replace all sets with at most $h - 1$ vertices by edge covers, we will create a clique cover \mathcal{C}' , for which $\pi(\beta, \mathcal{C}) = o(\pi(\beta, \mathcal{C}'))$. Thus, from Lemma 1, an edge cover and a cover with only one set on h vertices are the only optimal clique covers.

(iii) Suppose that $\delta > 1$. Then we know that $k_0 > h$ and so, by Lemma 1, for any \mathcal{C} different from an edge cover, $\pi(\beta, \mathcal{C}) = o(\pi(\beta, \mathcal{C}'))$, where \mathcal{C}' is an edge cover.

Now, let β and γ be two different cliques intersecting at at least one vertex. We need to find at least one optimal clique cover for $\beta \cup \gamma$ in order to estimate the order of magnitude of $E(I_\beta I_\gamma)$ (see (7)) appearing in (6).

We will say that a clique cover \mathcal{C}' of H is *at least as good* as another clique cover \mathcal{C} of H if $\pi(H, \mathcal{C}) = O(\pi(H, \mathcal{C}'))$.

Lemma 6. *Let β and γ be complete graphs on h vertices with $V(\beta) \cap V(\gamma) \neq \emptyset$. For m given by (1) and (2), and p given by (3), the optimal clique covers of $\beta \cup \gamma$ include: (i) a clique cover*

containing at least one set C for which $C \supseteq V(\beta)$ or $C \supseteq V(\gamma)$, and otherwise containing two element subsets when $\delta \leq 1$; (ii) either an edge cover or a cover by one set when $\delta > 1$.

Proof. Let \mathcal{C} be any simple clique cover of $\beta \cup \gamma$.

If all sets in \mathcal{C} are of size at most k_0 then from Lemma 1 an edge cover of $\beta \cup \gamma$ is at least as good as \mathcal{C} .

If in \mathcal{C} there is at least one set of size $|C| \geq \lfloor k_0 \rfloor + 1$, then either it is entirely contained in β or γ , or, from Lemma 4, it has at least one vertex in $V(\beta) \cap V(\gamma)$. In the latter case at least $(|C| - 1)/2 + 1$ vertices of C are in one of the sets $V(\beta)$ and $V(\gamma)$. Thus, without loss of generality, we may assume that

$$|C \cap V(\gamma)| \geq \frac{\lfloor k_0 \rfloor}{2} + 1 \geq \frac{k_0 - 1}{2} + 1,$$

i.e.

$$|C \cap V(\gamma)| - 1 \geq \frac{1}{2}k_0 - \frac{1}{2}.$$

Given any vertex $v \in V(\gamma)$ not contained in C , there is a set C_1 containing v and one vertex from C . Moreover, since $|C \cap V(\gamma)| - 1 \geq k_0/2 - \frac{1}{2}$, there are at least $k_0/2 - \frac{1}{2}$ edges in $\beta \cup \gamma$ joining $v \in C_1$ with vertices from C . Using Lemma 2, without losing optimality, we can merge C_1, C , and all the sets containing at least one edge joining vertices from C_1 and C . Without losing optimality of the cover, we may repeat the procedure until in the clique cover we have a large set containing all the vertices from $V(\gamma)$. This clique cover is at least as good as \mathcal{C} . For all the sets from \mathcal{C} of size larger than k_0 , if they were not merged before, we repeat the procedure above to obtain a clique cover which is at least as good as the previous one. This clique cover contains at least one set containing $V(\beta)$ or $V(\gamma)$. Finally, any remaining sets which are of size at most k_0 , if they exist and cannot be merged, can be replaced by edge covers without losing optimality using Lemma 1.

(i) Assume now that $\delta \leq 1$ and that \mathcal{C} is an edge cover. In this case, using the second part of Lemma 1, we can prove that a cover with one set $C = V(\gamma)$ and all other sets being two-element sets is at least as good as the edge cover.

(ii) Assume now that $\delta > 1$. A simple clique cover with two sets, C_1 and C_2 , containing $V(\beta)$ and $V(\gamma)$, respectively, may be replaced by an edge cover if $|C_1| < k_0$ and $|C_2| < k_0$ without losing optimality by Lemma 1. Moreover, since $|V(\beta \cup \gamma)| \leq 2h - 1, k_0 > h$, and we consider only simple clique covers, we may now assume without loss of generality that \mathcal{C} contains one large set C such that $C \supseteq V(\gamma)$ and otherwise contains only two-element subsets. Define x by $x = |C \cap V(\beta)|$ and s by

$$s = h - |V(\beta) \cap V(\gamma)|. \tag{11}$$

If $x = 1$ then we must have $s = h - 1$ and $C = V(\gamma)$. In this case from Lemma 1 we know that an edge cover is at least as good as \mathcal{C} . Suppose now that $x \geq 2$ (i.e. $\max\{h - s, 2\} \leq x \leq h$). Since an edge cover is optimal for a clique on h vertices, $\pi_\gamma \sim (mp^2)^{\binom{h}{2}}$ and

$$\pi(\beta \cup \gamma, \mathcal{C}) \sim (mp^2)^{\binom{h}{2} - \binom{x}{2}} mp^{s+x} \sim \pi_\gamma \frac{mp^x}{(mp^2)^{\binom{x}{2}}} p^s.$$

For $s = h - 1$, since $k_0 > h$, from the proof of Lemma 1 we know that $mp^x / (mp^2)^{\binom{x}{2}}$ is of the

largest order of magnitude when $x = 2$. Thus, because of (5), from Lemma 1 we have

$$\begin{aligned} \pi(\beta \cup \gamma, \mathcal{C}) &\sim (mp^2)^{\binom{h}{2}} p^{h-1} \\ &= O(1)(mp^2)^{\binom{h}{2}} mp^{h-1} \\ &= O(1)(mp^2)^2 \binom{h}{2} \frac{mp^h}{(mp^2)^{\binom{h}{2}}} \\ &= O(1)(mp^2)^2 \binom{h}{2} \\ &= O(\pi(\beta \cup \gamma, \mathcal{C}')), \end{aligned}$$

where \mathcal{C}' is an edge cover. Thus, an edge cover is at least as good as \mathcal{C} . For $1 \leq s \leq h - 2$, we have $h - s \leq x \leq h$ and from the proof of Lemma 1 we know that $mp^x / (mp^2)^{\binom{x}{2}}$ is of the largest order of magnitude when $x = h - s$ or $x = h$. If $x = h - s$ then $C = V(\gamma)$ and, from Lemma 1, an edge cover is at least as good as \mathcal{C} . If $x = h$ then \mathcal{C} is a cover by one set.

6. Probabilities of cliques and their unions

In this section we estimate the terms π_β and $E(I_\beta I_\gamma)$ appearing in (6). Recall that X_n is the random variable counting the number of instances of K_h in $G(n, m, p)$, i.e. $X_n = \sum_{\beta \in \Gamma} I_\beta$ is a sum over an index set $\{K_{\mathcal{V}'} : \mathcal{V}' \subseteq \mathcal{V}, |\mathcal{V}'| = h\}$, where I_β is an indicator random variable of the event $\beta \subseteq G(n, m, p)$.

Lemma 7. *Suppose that m is given by (1) and (2), and that p is given by (3). Then, for each $\beta \in \Gamma$,*

$$\pi_\beta \sim \begin{cases} \frac{c^h}{n^h} & \text{for } \delta < 1, \\ \frac{c^h + c^{h(h-1)}}{n^h} & \text{for } \delta = 1, \\ \frac{c^{h(h-1)}}{n^h} & \text{for } \delta > 1. \end{cases} \tag{12}$$

Moreover, let λ_n be defined by

$$\lambda_n := EX_n = \binom{n}{h} \pi_\beta. \tag{13}$$

Then,

$$\lim_{n \rightarrow \infty} \lambda_n = \begin{cases} \frac{c^h}{h!} & \text{for } \delta < 1, \\ \frac{c^h + c^{h(h-1)}}{h!} & \text{for } \delta = 1, \\ \frac{c^{h(h-1)}}{h!} & \text{for } \delta > 1. \end{cases} \tag{14}$$

Proof. Let β be a clique on h vertices. Using (8), (10), and Lemma 5, we have

- if $\delta < 1$ and \mathcal{C} is a cover by one set, then

$$\pi_\beta \sim \pi(\beta, \mathcal{C}) \sim mp^h \sim \frac{c^h}{n^h};$$

- if $\delta = 1$, \mathcal{C} is a cover by one set, and \mathcal{C}' is an edge cover of β , then

$$\pi_\beta \sim \pi(\beta, \mathcal{C}) + \pi(\beta, \mathcal{C}') \sim mp^h + (mp^2)^{\binom{h}{2}} \sim \frac{c^h + c^{h(h-1)}}{n^h};$$

- if $\delta > 1$ and \mathcal{C} is an edge cover, then

$$\pi_\beta \sim \pi(\beta, \mathcal{C}) \sim (mp^2)^{\binom{h}{2}} \sim \frac{c^{h(h-1)}}{n^h}.$$

The limits in (14) follow immediately from (13) and (12).

In the following lemma we let π stand for π_γ for any $\gamma \in \Gamma$.

Lemma 8. *Suppose that m is given by (1) and (2), and that p is given by (3). Define s by (11). Then, for each $\beta, \gamma \in \Gamma$,*

$$P(\beta \subseteq G(n, m, p), \gamma \subseteq G(n, m, p)) = \begin{cases} O(\pi^2 + \pi p^{h-1}) & \text{if } \delta \leq 1 \text{ and } s = h - 1, \\ O(\pi p^s) & \text{if } \delta \leq 1 \text{ and } s < h - 1, \\ O(\pi^2 + mp^{2h-1}) & \text{if } \delta > 1 \text{ and } s = h - 1, \\ O(mp^{h+s} + \pi^2 n^{(h-s)(h-s-1)/(h-1)}) & \text{if } \delta > 1 \text{ and } s < h - 1. \end{cases}$$

Proof. Let \mathcal{C} be a cover of $\beta \cup \gamma$ described in Lemma 6, and let C be the largest set in \mathcal{C} . First assume that $\delta \leq 1$.

Case 1: $s = h - 1$ and $|C \cap V(\beta)| \leq 1$, which implies that $|V(\beta) \cap V(\gamma)| = 1$. Since C covers no edges of β , it must be the case that $\mathcal{C} \setminus \{C\}$ is a cover of β and, therefore,

$$\pi(\beta \cup \gamma, \mathcal{C}) = O(1)\pi^2.$$

Case 2: $s < h - 1$ or $|C \cap V(\beta)| \geq 2$. Without loss of generality, we may assume that $V(\gamma) \subseteq C$. Clearly, $\mathcal{C}[\beta] = \{C' \cap V(\beta) : C' \in \mathcal{C}\}$ is a clique cover of β and

$$\pi(\beta \cup \gamma, \mathcal{C}) = \pi(\beta, \mathcal{C}[\beta])p^s = O(1)\pi p^s.$$

Now assume that $\delta > 1$.

Case 1: $|C| = 2$, so \mathcal{C} is an edge cover. For $s = h - 1$,

$$\pi(\beta \cup \gamma, \mathcal{C}) = O(1)\pi^2.$$

For $s < h - 1$,

$$\pi(\beta \cup \gamma, \mathcal{C}) \sim \pi^2 (mp^2)^{-\binom{h-s}{2}} \asymp \pi^2 n^{(h-s)(h-s-1)/(h-1)},$$

where we have used (3).

Case 2: $|C| = h + s$. This is a cover by one set and

$$\pi(\beta \cup \gamma, \mathcal{C}) \sim mp^{h+s}.$$

7. Proof of Theorem 1

We will use (6) with the dependency graph L defined in Section 2. Observe that, since $\pi_\beta = \pi_\gamma$ for all β, γ , and $|V(L)| < |E(L)|$,

$$\sum_{\beta \in V(L)} \pi_\beta^2 \leq \sum_{\beta\gamma \in E(L)} \pi_\beta \pi_\gamma,$$

and, therefore, we only need to estimate the following two sums:

$$\sum_{\beta\gamma \in E(L)} \pi_\beta \pi_\gamma$$

and

$$\sum_{\beta \in V(L)} \sum_{s=1}^{h-1} \sum_{\substack{\gamma \in V(L) \\ |V(\beta) \cap V(\gamma)|=h-s}} P(\beta \subseteq G(n, m, p), \gamma \subseteq G(n, m, p)). \tag{15}$$

Note that, for given $\beta \in V(L)$,

$$\begin{aligned} |\{\gamma \in V(L) : |V(\beta) \cap V(\gamma)| = h - s\}| &= O(n^s), \\ |\{\gamma \in V(L) : |V(\beta) \cap V(\gamma)| \neq \emptyset\}| &= O(n^{h-1}), \end{aligned}$$

and

$$|E(L)| = |V(L)| |\{\gamma \in V(L) : |V(\beta) \cap V(\gamma)| \neq \emptyset\}| = O(n^{2h-1}).$$

Hence, using $n^h \pi = O(1)$ gives

$$\sum_{\beta\gamma \in E(L)} \pi_\beta \pi_\gamma = O(n^{2h-1} \pi^2) = O(n^{-1}).$$

We now use Lemma 8 to estimate (15). For $\delta \leq 1$, it holds that $np \asymp m^{-1/h} = n^{-2\delta/(h-1)}$. If $\delta \leq 1$ and $s = h - 1$, then

$$\begin{aligned} &\sum_{\beta \in V(L)} \sum_{\substack{\gamma \in V(L) \\ |V(\beta) \cap V(\gamma)|=h-s}} P(\beta \subseteq G(n, m, p), \gamma \subseteq G(n, m, p)) \\ &= O(n^{2h-1} \pi^2 + \pi n^{2h-1} p^{h-1}) \\ &= O(n^{-1} + (np)^{h-1}) \\ &= O(n^{-1} + n^{-2\delta}). \end{aligned}$$

If $\delta \leq 1$ and $1 \leq s < h - 1$, then

$$\begin{aligned} \sum_{\beta \in V(L)} \sum_{\substack{\gamma \in V(L) \\ |V(\beta) \cap V(\gamma)|=h-s}} P(\beta \subseteq G(n, m, p), \gamma \subseteq G(n, m, p)) &= O(n^h \pi n^s p^s) \\ &= O(n^{-2\delta s/(h-1)}) \\ &= O(n^{-2\delta/(h-1)}). \end{aligned}$$

For $\delta > 1$, we can check that $np \asymp n^{(h-2)/(h-1)}m^{-1/2} = n^{((1-\delta)h-2)/(h-1)} = o(1)$. If $\delta > 1$ and $s = h - 1$, then

$$\begin{aligned} & \sum_{\beta \in V(L)} \sum_{\substack{\gamma \in V(L) \\ |V(\beta) \cap V(\gamma)|=h-s}} P(\beta \subseteq G(n, m, p), \gamma \subseteq G(n, m, p)) \\ &= O(n^{2h-1}\pi^2 + n^{2h-1}mp^{2h-1}) \\ &= O(n^{-1} + n^{h+1}mp^{h+1}) \\ &= O(n^{-1} + (np)^{h+1}m) \\ &= O(n^{-1} + n^{((h+1)((1-\delta)h-2)+2\delta h)/(h-1)}) \\ &= O(n^{-1} + n^{(1-\delta)h-2/(h-1)}). \end{aligned}$$

If $\delta > 1$ and $1 \leq s < h - 1$, then

$$\begin{aligned} & \sum_{\beta \in V(L)} \sum_{\substack{\gamma \in V(L) \\ |V(\beta) \cap V(\gamma)|=h-s}} P(\beta \subseteq G(n, m, p), \gamma \subseteq G(n, m, p)) \\ &= O(n^{h+s}mp^{h+s} + n^{2h}\pi^2n^{s-h}n^{(h-s)(h-s-1)/(h-1)}) \\ &= O(n^{h+1}mp^{h+1} + n^{((h-s)(h-s-1)-(h-s)(h-1))/(h-1)}) \\ &= O(n^{(1-\delta)h-2/(h-1)} + n^{-(h-s)s/(h-1)}) \\ &= O(n^{(1-\delta)h-2/(h-1)} + n^{-1}). \end{aligned}$$

Using the estimates above in (6) along with (14) proves Theorem 1.

Acknowledgements

We would like to thank Jerzy Jaworski for his invaluable advice and support. K. Rybarczyk acknowledges partial support from the Ministry of Science and Higher Education, grant number N N206 2701 33, 2007–2010.

References

- [1] BEHRISCH, M. (2007). Component evolution in random intersection graphs. *Electron. J. Combinatorics* **14**, R17, 12pp.
- [2] BLACKBURN, S. R. AND GERKE, S. (2009). Connectivity of the uniform random intersection graph. *Discrete Math.* **309**, 5130–5140.
- [3] BLANCHARD, PH., KRUEGER, A., KRUEGER, T. AND MARTIN, P. (2005). The epidemics of corruption. Preprint. Available at <http://arxiv.org/abs/physics/0505031v1>.
- [4] BŁOZNELIS, M., JAWORSKI, J. AND RYBARCZYK, K. (2009). Component evolution in a secure wireless sensor network. *Networks* **53**, 19–26.
- [5] BRITTON, T., DEIJFEN, M., LAGERÅS, A. N. AND LINDHOLM, M. (2008). Epidemics on random graphs with tunable clustering. *J. Appl. Prob.* **45**, 743–756.
- [6] DEIJFEN, M. AND KETS, W. (2009). Random intersection graphs with tunable degree distribution and clustering. *Prob. Eng. Inf. Sci.* **23**, 661–674.
- [7] DI PIETRO, R. *et al.* (2006). Sensor networks that are provably resilient. In *Proc. IEEE Internat. Conf. on Security and Privacy in Communication Networks (SECURECOMM 2006)*, Baltimore.
- [8] FILL, J. A., SCHEINERMAN, E. R. AND SINGER-COHEN, K. B. (2000). 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.
- [9] GODEHARDT, E. AND JAWORSKI, J. (2002). Two models of random intersection graphs for classification. In *Exploratory Data Analysis in Empirical Research*, eds O. Opitz and M. Schwaiger, Springer, Berlin, pp. 67–81.

- [10] GODEHARDT, E., JAWORSKI, J. AND RYBARCZYK, K. (2007). Random intersection graphs and classification. In *Advances in Data Analysis*, eds R. Decker and H.-J. Lenz, Springer, Berlin, pp. 67–74.
- [11] GODEHARDT, E., JAWORSKI, J. AND RYBARCZYK, K. (2010). Isolated vertices in random intersection graphs. In *Advances in Data Analysis, Data Handling and Business Intelligence*, eds A. Fink *et al.*, Springer, Berlin, pp. 135–145.
- [12] JANSON, S., ŁUCZAK, T. AND RUCIŃSKI, A. (2000). *Random Graphs*. Wiley-Interscience, New York.
- [13] KAROŃSKI, M., SCHEINERMAN, E. R. AND SINGER-COHEN, K. B. (1999). On random intersection graphs: the subgraph problem. *Combinatorics Prob. Comput.* **8**, 131–159.
- [14] MARCHETTE, D. J. (2004). *Random Graphs for Statistical Pattern Recognition*. John Wiley, Hoboken, NJ.
- [15] NEWMAN, M. E. J., STROGATZ, S. H. AND WATTS, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118.
- [16] NEWMAN, M. E. J., WATTS, D. J. AND STROGATZ, S. H. (2002). Random graph models of social networks. *Proc. Nat. Acad. Sci. USA* **99**, 2566–2572.