Bull. Korean Math. Soc.

POISSON APPROXIMATION OF INDUCED SUBGRAPH COUNTS IN AN INHOMOGENEOUS RANDOM INTERSECTION GRAPH MODEL

Yilun Shang

Abstract. In this paper, we consider a class of inhomogeneous random intersection graphs by assigning random weight to each vertex and two vertices are adjacent if they choose some common elements. In the inhomogeneous random intersection graph model, vertices with larger weights are more likely to acquire many elements. We show the Poisson convergence of the number of induced copies of a fixed subgraph as the number of vertices n and the number of elements m, scaling as $m = \lfloor \beta n^{\alpha} \rfloor$ $(\alpha, \beta > 0)$, tend to infinity.

1. Introduction

Let n and m be two positive integers. Take $V = \{v_1, \cdots, v_n\}$ to be a set of n vertices and $W = \{w_1, \cdots, w_m\}$ a set of m elements. Let $\{\theta_i\}_{i=1}^{n}$ be a sequence of independent and identically distributed positive random variables with distribution F, where F is assumed to have mean 1 if its mean is finite. Let f_n be a function taking values in the interval $[0, 1]$. A bipartite graph $B(n, m, F, f_n)$ with vertex sets V and W is constructed by joining $v_i \in V$ and $w \in W$ (denoted $v_i \sim w$) independently in such way that $p_i := \mathbb{P}(v_i \sim w) = f_n(\theta_i)$. The inhomogeneous random intersection graph $G(n, m, F, f_n)$ on the vertex set V is obtained by adding an edge between two vertices v_i and v_j if and only if they have a common neighbor $w \in W$ in $B(n, m, F, f_n)$. By introducing the inhomogeneous weight sequence $\{\theta_i\}$, the model preferably generalizes the binomial random intersection graph [9], where F is a degenerate distribution centered at 1, and is able to interpret phenomena in real-life networks. An example is in social networks, where θ_i can be viewed as a measure of the social activity of an individual i; a vertex with larger weight is more likely to join many groups and thereby acquires many social contacts. Various monotone properties such as degree, independent set, connectivity, and hamiltonicity, of inhomogeneous random intersection graphs have been investigated; see e.g. [6, 10, 11, 14] and the recent brief review [4]. We mention that the inhomogeneity is considered solely for V here since the heterogeneous actors (i.e. vertices

1991 *Mathematics Subject Classification.* Primary 60F05 05C80 62E17.

Key words and phrases. Random graph, intersection graph, Poisson approximation, Stein's method, subgraph count.

in V) are arguably of more interest from the perspective of social network applications.

Let H be any fixed graph on $h \geq 2$ vertices and having at least one edge. Let $X(H)$ stand for the number of induced copies of H that can be found in $G(n, m, F, f_n)$. Here, we are interested in Poisson approximation of the distribution of $X(H)$, which has been extensively studied, for example, in the case of classical Erdős-Rényi random graphs [7], stochastic block model [5], and binomial random intersection graphs [12, 13]. Our main tool is the Stein-Chen method which shows the asymptotic distribution of the type of sums of random variables appearing in varied combinatorial problems [3, 7] in terms of the total variation distance. The total variation distance between a random variable X and a Poisson distributed random variable Poi(λ) with mean λ is defined as

$$d_{TV}(X, \mathrm{Poi}(\lambda)) = \frac{1}{2}\sum_{k=0}^{\infty} |\mathbb{P}(X = k) - e^{-\lambda}\lambda^k/k!|;$$

see e.g. [8] for more applications. Let $aut(H)$ denote the number of automorphisms of H. Since there are exactly $N := \binom{n}{h} h!/aut(H)$ copies of H in the complete graph K_n of order n, we collect all of them in a set $\mathcal{H} := \{H_1, \cdots, H_N\}$ and write $X(H) = \sum_{i=1}^{N} X_i$, where X_i stands for the indicator random variable of the event that H_i is an induced subgraph in $G(n, m, F, f_n)$.

Clearly, the property that $X(H) = k$ is not monotone. To show the approximate Poisson distribution of $X(H)$, we will rely on the powerful concept of clique cover introduced in [9]. By examining how each copy of H appears in $G(n, m, F, f_n)$ using clique covers, we present a Poisson approximation result for subgraph counts in Section 2, and the proofs are given in Section 3.

2. Results

To show the approximate Poisson distribution of $X(H)$, we need the following powerful concept of clique cover, which helps to quantify the different ways copies of H can appear in $G(n, m, F, f_n)$. A clique cover $\mathcal{C}$ of H is a family of non-empty subsets of vertex set $V(H)$ such that, each induces a clique of H, and for any edge $\{v_1, v_2\} \in E(H)$, there exists $C \in \mathcal{C}$ such that $v_1, v_2 \in C$. By definition, the cliques induced by sets from $\mathcal{C}$ exactly cover the edges of H, and each edge is allowed to be covered by more than one sets. We say that $\mathcal{C}$ is proper if $|C| \geq 2$ for all $C \in \mathcal{C}$. Let $\mathcal{C} = \{C_1, C_2, \cdots, C_t\}$ be a clique cover of H. We say that a copy H_0 of H in $G(n, m, F, f_n)$ is induced by the clique cover $\mathcal{C}$ of H_0 if there is a set of disjoint non-empty subsets $\{W_1, W_2, \cdots, W_t\}$ of W such that for all $1 \leq i \leq t$ each element of W_i is linked to all vertices of C_i and no other vertices from $V(H_0)$ in $B(n, m, F, f_n)$, and each element $w \in W \backslash \cup_{i=1}^{t} W_i$ is linked to at most one vertex from $V(H_0)$. Denote by $\mathcal{C}(H)$ the finite set of proper clique covers of H. It is easy to see that if H_0 is an induced copy of H in $G(n, m, F, f_n)$, then it is induced by exactly one proper clique cover from $\mathcal{C}(H_0)$.

Given a clique cover $\mathcal{C} = \{C_1, C_2, \cdots, C_t\}$ of H, for $S \subseteq V(H)$ we define the following two restricted clique covers $\mathcal{C}[S] = \{C_i \cap S : |C_i \cap S| \geq 1, i = 1, \cdots, t\}$ and $\mathcal{C}'[S] = \{C_i \cap S : |C_i \cap S| \geq 2, i = 1, \cdots, t\}$. By definition, these restricted clique covers are multisets since it is possible that $C_i \cap S = C_j \cap S$ for $i \neq j$. In particular, when $S = V(H)$, we write $\mathcal{C}$ and $\mathcal{C}'$ for short, respectively. Moreover, $|\mathcal{C}[S]|$ stands for the number of cliques in $\mathcal{C}[S]$ with multiplicity, while $\sum \mathcal{C}[S] := \sum_{\substack{|C_i \cap S| \geq 1 \\ 1 \leq i \leq t}} |C_i \cap S|$ denotes the sum of clique sizes in $\mathcal{C}[S]$. Similar notation applies for $\mathcal{C}'[S]$.

To obtain interesting structural properties, the number of elements m is typically taken to be $m = \lfloor \beta n^\alpha \rfloor$ for some constants $\alpha, \beta > 0$; see e.g. [9, 15]. We assume this form for m in what follows. Denote by $X(H, \mathcal{C}, S)$ the number of copies of $H[S]$ induced by $\mathcal{C}[S]$. When F is degenerated and concentrated at 1, it is shown in [9] that, assuming $mp_1^2 = o(1)$, $\mathbb{E}(X(H, \mathcal{C}, S)) = \Theta(\psi(H, \mathcal{C}, S))$, where

$$
\begin{aligned}
(1) \qquad & \psi(H, \mathcal{C}, S) \\
&= \min \left\{ n^{|S|} m^{|\mathcal{C}[S]|} \prod_{\substack{v_i \in C \\ C \in \mathcal{C}[S]}} p_i, n^{|S|} m^{|\mathcal{C}'[S]|} \prod_{\substack{v_i \in C \\ C \in \mathcal{C}'[S]}} p_i \right\} \\
&= \min \left\{ \beta^{|\mathcal{C}[S]|} n^{|S| + \alpha |\mathcal{C}[S]|} \prod_{\substack{v_i \in C \\ C \in \mathcal{C}[S]}} p_i, \beta^{|\mathcal{C}'[S]|} n^{|S| + \alpha |\mathcal{C}'[S]|} \prod_{\substack{v_i \in C \\ C \in \mathcal{C}'[S]}} p_i \right\}.
\end{aligned}
$$

Moreover, we define

$$
(2) \qquad \eta_2(H, \mathcal{C}, S) = \begin{cases} \frac{|S| + \alpha |\mathcal{C}[S]|}{\sum \mathcal{C}[S]}, & \text{if } \alpha < \frac{|S|}{\sum \mathcal{C}[S] - |\mathcal{C}[S]|} \text{ or } \sum \mathcal{C}[S] = |\mathcal{C}[S]|; \\ \frac{|S| + \alpha |\mathcal{C}'[S]|}{\sum \mathcal{C}'[S]}, & \text{otherwise.} \end{cases}
$$

It is straightforward to see that $\mathbb{E}(X(H, \mathcal{C}, S)) = \Theta(1)$ when $p_1 = \gamma n^{-\eta_2(H, \mathcal{C}, S)}$ for some constant $\gamma > 0$ if F is degenerated and concentrated at 1 [13]; while in general $\mathbb{E}(X(H, \mathcal{C}, S)) = \Theta_p(1)$ when $p_i = f_n(\theta_i) = \gamma \theta_i n^{-\eta_2(H, \mathcal{C}, S)} \wedge 1$ for $\gamma > 0$ under the analogous assumption presented in Theorem 1 below. Here and in what follows, we adopt the standard asymptotic notations $O(\cdot)$, $o(\cdot)$, $\Theta(\cdot)$, etc. and their probabilistic versions $O_p(\cdot)$, $o_p(\cdot)$, $\Theta_p(\cdot)$, etc. defined in [7]. In particular, $a_n = O_p(b_n)$ as $n \to \infty$ if for every $\varepsilon > 0$ there are constants c and n_0 satisfying $\mathbb{P}(|a_n| \leq cb_n) > 1 - \varepsilon$ for all $n \geq n_0$; $a_n = o_p(b_n)$ as $n \to \infty$ if for every $\varepsilon > 0$, $\mathbb{P}(|a_n| < \varepsilon b_n) \to 1$; $a_n = \Theta_p(b_n)$ as $n \to \infty$ if for every $\varepsilon > 0$ there are constants $c_1, c_2 > 0$ and n_0 satisfying $\mathbb{P}(c_1 b_n \leq a_n \leq c_2 b_n) > 1 - \varepsilon$ for all $n \geq n_0$.

Next, define $\eta_1(H, \mathcal{C}) = \min_{\emptyset \neq S \subseteq V(H)} \eta_2(H, \mathcal{C}, S)$ and $\eta_0 = \eta_0(H) = \max_{\mathcal{C} \in \mathcal{C}(H)} \eta_1(H, \mathcal{C})$. A clique cover $\mathcal{C} \in \mathcal{C}(H)$ is said to be strictly α-balanced if $\eta_2(H, \mathcal{C}, S) > \eta_2(H, \mathcal{C}, V(H))$ for all $\emptyset \neq S \subset V(H)$. Let $\mathcal{C}_0(H) = \{\mathcal{C} \in \mathcal{C}(H) : \eta_1(H, \mathcal{C}) = \eta_0\}$. Following [13], we say H is strictly α-balanced if any $\mathcal{C} \in \mathcal{C}_0(H)$

is strictly α-balanced. Note that η_0, η_1, and η_2 all depend on α. The strictly α-balance condition is a non-trivial extension to the strictly balance condition which concerns the subgraph density in the Erdős-Rényi graph scenario [7].

Theorem 1. *Let H be a fixed graph and $m = \lfloor \beta n^\alpha \rfloor$ for $\alpha, \beta > 0$. Suppose that F has finite mean, $p_i = \gamma\theta_i n^{-\eta_0} \wedge 1$ for $\gamma > 0$, and that for any $\mathcal{C} \in \mathcal{C}(H)$, $m \cdot \min_{C\in\mathcal{C}}\{(\max_{v_i\in V(H)\setminus C} p_i) \cdot (\max_{v_i\in C} p_i)\} = o_p(1)$. If H is strictly α-balanced, then*

$$d_{TV}(X(H), \mathrm{Poi}(\lambda)) = o(1),$$

where $\lambda = \frac{1}{aut(H)} \sum_{\mathcal{C}\in\mathcal{C}_0(H)} \beta^{|\mathcal{C}|}\gamma^{\sum \mathcal{C}} \prod_{v_i\in C\in\mathcal{C}} \theta_i$.

Note that when F is a degenerate distribution concentrated on 1 and $\beta = 1$, the inhomogeneous random intersection graph is reduced to the binomial case considered in [13] and the main result therein can be recovered. The parameter λ quantifies the limit of the number of copies of H induced by clique covers in $\mathcal{C}_0(H)$. In the next section, we will prove the result using Stein's method and the notion of dependency graph [7]. Similar techniques have been used in e.g. [9, 13] for binomial random intersection graph models.

3. Proofs

In the following, we assume that F has finite mean. Hence, θ_i is finite with probability 1 and the Markov inequality yields $p_i = o_p(1)$ for $1 \le i \le n$. The following result shows the asymptotic independence of the numbers of elements.

Lemma 1. *Suppose that $\mathcal{C} = \{C_1, \cdots, C_r\}$ is a clique cover of H. For $1 \le j \le r$, let M_j be the number of elements in W that are connected to all vertices from C_j and no vertex from $V(H)\setminus C_j$ in $G(n, m, F, f_n)$ and let $\tilde{M}_j$ be a Poisson random variable $\mathrm{Poi}(m \prod_{v_i\in C_j} p_i)$. If $m \cdot \min_{1\le j\le r}\{(\max_{v_i\in V(H)\setminus C_j} p_i) \cdot (\prod_{v_i\in C_j} p_i)\} = o_p(1)$, then*

$$\mathbb{P}(\cap_{j=1}^r\{M_j = a_j\}) \sim_p \prod_{j=1}^r \mathbb{P}(\tilde{M}_j = a_j)$$

holds uniformly for all $a_j \le A = A(n)$ satisfying $A = o(\sqrt{m})$ and $Ap_1 = o_p(1)$.

Proof. Given $\{\theta_i\}_{i=1}^n$, we define $q_j = \prod_{v_i\in C_j} p_i \prod_{v_i\in V(H)\setminus C_j}(1-p_i)$ for $1 \le j \le r$. Clearly, q_j represents the probability that an element w is connected to all vertices from C_j but no other vertices from $V(H)$ in $B(n, m, F, f_n)$. Define $q_0 = 1 - \sum_{j=1}^r q_j$ and $a_0 = m - \sum_{j=1}^r a_j$. Note that $q_j \le \min_{1\le j\le r} \prod_{v_i\in C_j} p_i = o_p(1)$. For $0 \le a_j \le A$ $(1 \le j \le r)$, we obtain

$$\mathbb{P}(\cap_{j=1}^r\{M_j = a_j\}) = \binom{m}{a_0, \cdots, a_r} q_0^{a_0} \prod_{j=1}^r q_j^{a_j} = \frac{m!}{a_0!} q_0^{a_0} \prod_{j=1}^r \frac{q_j^{a_j}}{a_j!}$$

$$= \frac{m!}{(m - \sum_{j=1}^r a_j)!}\left(1 - \sum_{j=1}^r q_j\right)^{m-\sum_{j=1}^r a_j} \prod_{j=1}^r \frac{q_j^{a_j}}{a_j!}$$

$$
\begin{aligned}
&= m^{\sum_{j=1}^r a_j} \cdot e^{-(m-\sum_{j=1}^r a_j)\sum_{j=1}^r q_j - O\left((m-\sum_{j=1}^r a_j)\left(\sum_{j=1}^r q_j\right)^2\right)} \prod_{j=1}^r \frac{q_j^{a_j}}{a_j!} \\
&= m^{\sum_{j=1}^r a_j} e^{O\left(\frac{r^2A^2}{m}\right)} \cdot e^{-m\sum_{j=1}^r q_j + O\left(r^2 A \min_j \prod_{v_i\in C_j} p_i + mr^2(\min_j \prod_{v_i\in C_j} p_i)^2\right)} \\
&\quad \cdot \prod_{j=1}^r \frac{q_j^{a_j}}{a_j!} \\
&\sim e^{-m\sum_{j=1}^r q_j} \prod_{j=1}^r \frac{m^{a_j} q_j^{a_j}}{a_j!} \\
&= e^{-m\sum_{j=1}^r \left[\prod_{v_i\in C_j} p_i + O\left(m \min_j \{(\max_{v_i\in V(H)\setminus C_j} p_i)\prod_{v_i\in C_j} p_i\}\right)\right]} \\
&\quad \cdot \prod_{j=1}^r \left[\frac{\left(m\prod_{v_i\in C_j} p_i\right)^{a_j}}{a_j!} \left(\prod_{v_i\in V(H)\setminus C_j} (1-p_i)\right)^{a_j}\right] \\
&\sim \prod_{j=1}^r e^{-m\prod_{v_i\in C_j} p_i} \cdot \frac{\left(m\prod_{v_i\in C_j} p_i\right)^{a_j}}{a_j!} \cdot e^{O\left(hA\max_{v_i\in V(H)\setminus C_j} p_i\right)} \\
&\sim \prod_{j=1}^r e^{-m\prod_{v_i\in C_j} p_i} \cdot \frac{\left(m\prod_{v_i\in C_j} p_i\right)^{a_j}}{a_j!} = \prod_{j=1}^r \mathbb{P}(\tilde{M}_j = a_j),
\end{aligned}
$$

as n tends to infinity. □

Denote by $\pi(H,\mathcal{C})$ the probability that H is induced by the clique cover $\mathcal{C}$. Recall that $N = \binom{n}{h} h!/aut(H)$ and $\mathcal{H} = \{H_1,\cdots,H_N\}$ collects all N copies of H in the complete graph K_n. Since each copy $H_i \in \mathcal{H}$ $(1 \le i \le N)$ can be induced by at most one clique cover in $\mathcal{C}(H_i)$, we have $\mathbb{E}X_i = \sum_{\mathcal{C}\in\mathcal{C}(H_i)} \pi(H_i,\mathcal{C}) = \sum_{\mathcal{C}\in\mathcal{C}(H)} \pi(H,\mathcal{C})$. The following lemma gives an asymptotic estimate for the probability $\pi(H,\mathcal{C})$.

Lemma 2. *Let $\mathcal{C} = \{C_1,\cdots,C_r\}$ be a clique cover of H, $I_1 = \{1 \le j \le r : |C_j| = 1\}$ and $I_2 = \{1 \le j \le r : |C_j| \ge 2\}$. Assume*

$$m \cdot \min_{1\le j\le r} \left\{ (\max_{v_i\in V(H)\setminus C_j} p_i) \cdot (\max_{v_i\in C_j} p_i) \right\} = o_p(1).$$

Then

$$\pi(H,\mathcal{C}) \sim_p \prod_{v_i\in C_j, j\in I_1} (1-e^{-mp_i}) \prod_{j\in I_2} \left(m \prod_{v_i\in C_j} p_i \right).$$

Consequently,

$$\pi(H,\mathcal{C}) = \Theta_p \left(\min \left\{ m^{|\mathcal{C}|} \prod_{v_i\in C_j, C_j\in\mathcal{C}} p_i, m^{|\mathcal{C}'|} \prod_{v_i\in C_j, C_j\in\mathcal{C}'} p_i \right\} \right).$$

Proof. Given $\{\theta_i\}_{i=1}^n$, let $(C_{r+1}, C_{r+2}, \cdots, C_t)$ be the subsets of $V(H)$ not in $\mathcal{C}$ of order of at least 2. For $1 \le j \le r$, let M_j be the number of elements that are connected to all vertices from C_j and no vertex from $V(H)\backslash C_j$ in $G(n, m, F, f_n)$. Since H is induced by $\mathcal{C}$, we have $M_j = 0$ for $j = r+1, \cdots, t$. Clearly, $M_j \sim \text{Bin}(m, q_j)$, i.e. M_j has binomial distribution with parameters m and q_j, where q_j is defined as in Lemma 1. Let $\tilde{M}_j$ be a Poisson random variable $\text{Poi}(m\prod_{v_i\in C_j} p_i)$. It follows from $\pi(H,\mathcal{C}) = \mathbb{P}(\cap_{j=1}^r\{M_j \ge 1\} \cap \cap_{j=r+1}^t\{M_j = 0\})$ and the Chernoff bound [7] that for sufficiently large $c > 0$, $\mathbb{P}(M_j \ge A) = o(\pi(H,\mathcal{C}))$ and $\mathbb{P}(\tilde{M}_j \ge A) = o(\pi(H,\mathcal{C}))$, where $A = A(n) = c\cdot\max\{m\max_{v_i\in C_j} p_i, \ln n\}$. Therefore,

$$\begin{aligned}
\pi(H,\mathcal{C}) &= \mathbb{P}\left(\cap_{j=1}^r\{M_j \ge 1\} \cap \cap_{j=r+1}^t\{M_j = 0\}\right) \\
&= \sum_{1\le a_j\le A, 1\le j\le r} \mathbb{P}\left(\cap_{j=1}^r\{M_j = a_j\} \cap \cap_{j=r+1}^t\{M_j = 0\}\right) \\
&\quad + o(\pi(H,\mathcal{C})).
\end{aligned}$$

Note that $m\cdot\min_{1\le j\le r}\{(\max_{v_i\in V(H)\backslash C_j} p_i)\cdot(\prod_{v_i\in C_j} p_i)\} = m$ $\cdot\min_{1\le j\le r}\{(\max_{v_i\in V(H)\backslash C_j} p_i)\cdot(\max_{v_i\in C_j} p_i)\} = o_p(1)$, $A = o(\sqrt{m})$, and $A\cdot\max_{v_i\in V(H)\backslash C_j} p_i = o_p(1)$. Since $\{p_i\}_{i=1}^n$ are i.i.d., we have $Ap_1 = o_p(1)$. It follows from Lemma 1 and Chebyshev's inequality that

$$\begin{aligned}
\pi(H,\mathcal{C}) &\sim_p \sum_{1\le a_j\le A, 1\le j\le r} \prod_{j=1}^r \mathbb{P}(\tilde{M}_j = a_j) \prod_{j=r+1}^t \mathbb{P}(\tilde{M}_j = 0) \\
&= \prod_{j=1}^r \mathbb{P}(1 \le \tilde{M}_j \le A) \prod_{j=r+1}^t \mathbb{P}(\tilde{M}_j = 0) \\
&\sim_p \prod_{j=1}^r \mathbb{P}(\tilde{M}_j \ge 1) \prod_{j=r+1}^t \mathbb{P}(\tilde{M}_j = 0) \\
&= \prod_{j=1}^r \left(1 - e^{-m\prod_{v_i\in C_j} p_i}\right) \prod_{j=r+1}^t e^{-m\prod_{v_i\in C_j} p_i} \\
&\sim_p \prod_{j=1}^r \left(1 - e^{-m\prod_{v_i\in C_j} p_i}\right) \\
&\sim_p \prod_{v_i\in C_j, j\in I_1} \left(1 - e^{-mp_i}\right) \prod_{j\in I_2} \left(m \prod_{v_i\in C_j} p_i\right), \qquad (3)
\end{aligned}$$

as n tends to infinity.

For $v_i \in C_j$ $(j \in I_1)$, if $mp_i \ge 1$ then $1 - e^{-mp_i} = \Theta_p(1)$; if $mp_i < 1$ then $1 - e^{-mp_i} = \Theta_p(mp_i)$. Hence, $1 - e^{-mp_i} = \Theta_p(\min\{mp_i, 1\})$. Combining this with (3) completes the proof. □

Recall the definition (1) and define $\omega(H,\mathcal{C}) = \min_{\emptyset\neq S\subset V(H)} \psi(H,\mathcal{C},S)$. The next result provides another necessary ingredient regarding second moments for the proof of Theorem 1.

Lemma 3. *For any clique cover $\mathcal{C} = \{C_1,\cdots,C_r\} \in \mathcal{C}(H)$, suppose that $m\cdot\min_{1\le j\le r}\ \{(\max_{v_i\in V(H)\setminus C_j} p_i)\cdot(\max_{v_i\in C_j} p_i)\} = o_p(1)$. Let G_1 and G_2 be two subgraphs of K_n with $|V(G_1)\cap V(G_2)| = \ell$ and $G_1\cap G_2$ is an induced subgraph of both G_1 and G_2. Let $\mathcal{C}_i$ be a proper clique cover of G_i for $i = 1,2$. Denote by $X(G_i,\mathcal{C}_i)$ the indicator random variable of the event that G_i is induced by C_i in $G(n,m,F,f_n)$. Then*

$$\omega(G_2,\mathcal{C}_2) = O_p\left(\frac{n^\ell\cdot\mathbb{E}X(G_1,\mathcal{C}_1)\cdot\mathbb{E}X(G_2,\mathcal{C}_2)}{\mathbb{E}(X(G_1,\mathcal{C}_1)X(G_2,\mathcal{C}_2))}\right).$$

Proof. Let $\mathcal{C}_1 = \{C_{11},\cdots,C_{1r_1}\}$ and $\mathcal{C}_2 = \{C_{21},\cdots,C_{2r_2}\}$ be two proper clique covers of G_1 and G_2, respectively. We write $\mathcal{C}_1+\mathcal{C}_2$ for the set of clique covers on $V(G_1\cup G_2)$ satisfying $\mathcal{C}[V(G_1)] = \mathcal{C}_1$ and $\mathcal{C}[V(G_2)] = \mathcal{C}_2$ for each $\mathcal{C}\in\mathcal{C}_1+\mathcal{C}_2$. It is clear that each clique cover in $\mathcal{C}_1+\mathcal{C}_2$ is proper. Moreover, if G_i is induced by $\mathcal{C}_i$ on $V(G_i)$ for $i = 1,2$, then $G_1\cup G_2$ is induced by a unique element of $\mathcal{C}_1+\mathcal{C}_2$ and hence

$$\mathbb{E}(X(G_1,\mathcal{C}_1)X(G_2,\mathcal{C}_2)) = \sum_{\mathcal{C}\in\mathcal{C}_1+\mathcal{C}_2}\mathbb{E}X(G_1\cup G_2,\mathcal{C}). \tag{4}$$

For each $\mathcal{C}\in\mathcal{C}_1+\mathcal{C}_2$ we define $J_1 = J_1(\mathcal{C}) = \{1\le j\le r_1 : \exists C\in\mathcal{C}$ such that $C = C_{1j}$ for some $1\le j\le r_1$ but $C\neq C_{2j}$ for any $1\le j\le r_2\}$, $J_2 = J_2(\mathcal{C}) = \{1\le j\le r_2 : \exists C\in\mathcal{C}$ such that $C = C_{2j}$ for some $1\le j\le r_2$ but $C\neq C_{1j}$ for any $1\le j\le r_1\}$, and $J_2 = J_3(\mathcal{C}) = \{(k,l), 1\le k\le r_1, 1\le l\le r_2 : \exists C\in\mathcal{C}$ such that $C = C_{1k}\cup C_{2l}$ for some $1\le k\le r_1$ and $1\le l\le r_2\}$. Furthermore, let $J_4 = \{1\le j\le r_1 : \exists l\in\{1,\cdots,r_2\}$ such that $(j,l)\in J_3\}$ and $J_5 = \{1\le j\le r_2 : \exists k\in\{1,\cdots,r_1\}$ such that $(k,j)\in J_3\}$, which readily imply that $J_1\cup J_4 = \{1,\cdots,r_1\}$ and $J_2\cup J_5 = \{1,\cdots,r_2\}$. For any $\mathcal{C}\in\mathcal{C}_1+\mathcal{C}_2$, it follows from Lemma 2 that

$$\begin{aligned}
&\mathbb{E}X(G_1\cup G_2,\mathcal{C})\\
\sim_p\ &\prod_{\substack{j\in J_1\\|C_{1j}|>1}}\left(m\prod_{v_i\in C_{1j}}p_i\right)\prod_{\substack{v_i\in C_{1j},j\in J_1\\|C_{1j}|=1}}\left(1-e^{-mp_i}\right)\\
&\cdot\prod_{\substack{j\in J_2\\|C_{2j}|>1}}\left(m\prod_{v_i\in C_{2j}}p_i\right)\prod_{\substack{v_i\in C_{2j},j\in J_2\\|C_{2j}|=1}}\left(1-e^{-mp_i}\right)\\
&\cdot\prod_{\substack{(k,l)\in J_3\\|C_{1k}\cup C_{2l}|>1}}\left(m\prod_{v_i\in C_{1k}\cup C_{2l}}p_i\right)\prod_{\substack{v_i\in C_{1k}\cup C_{2l},(k,l)\in J_3\\|C_{1k}\cup C_{2l}|=1}}\left(1-e^{-mp_i}\right).
\end{aligned}\tag{5}$$

If $mp_i \le 1$, then $mp_i = \Theta_p(1 - e^{-mp_i})$. By using (5) we obtain

$$\begin{aligned}
&\mathbb{E}X(G_1 \cup G_2, \mathcal{C})\\
=&O_p\left(\prod_{j\in J_1}\left(m\prod_{v_i\in C_{1j}} p_i\right)\prod_{j\in J_2}\left(m\prod_{v_i\in C_{2j}} p_i\right)\prod_{(k,l)\in J_3}\left(m\prod_{v_i\in C_{1k}\cup C_{2l}} p_i\right)\right)\\
=&O_p\left(\prod_{j\in J_1}\left(m\prod_{v_i\in C_{1j}} p_i\right)\prod_{j\in J_2}\left(m\prod_{v_i\in C_{2j}} p_i\right)\right.\\
&\left.\cdot\prod_{(k,l)\in J_3}\frac{\left(m\prod_{v_i\in C_{1k}} p_i\right)\left(m\prod_{v_i\in C_{2l}} p_i\right)}{m\prod_{v_i\in C_{1k}\cap C_{2l}} p_i}\right)\\
=&O_p\left(\prod_{1\le j\le r_1}\left(m\prod_{v_i\in C_{1j}} p_i\right)\prod_{1\le j\le r_2}\left(m\prod_{v_i\in C_{2j}} p_i\right)\prod_{(k,l)\in J_3}\frac{1}{m\prod_{v_i\in C_{1k}\cap C_{2l}} p_i}\right).
\end{aligned}$$

Noting that $V(G_1)\cap V(G_2)\cap C_{2l} = C_{1k}\cap C_{2l}$, we have

$$\prod_{(k,l)\in J_3}\left(m\prod_{v_i\in C_{1k}\cap C_{2l}} p_i\right) \ge \prod_{C\in\mathcal{C}_2[V(G_1)\cap V(G_2)]}\left(m\prod_{v_i\in C} p_i\right) \ge \frac{\omega(G_2,\mathcal{C}_2)}{n^\ell},$$

which completes the proof in this case employing Lemma 2.

On the other hand, if $mp_i > 1$, $1 = \Theta_p(1 - e^{-mp_i})$. It follows from (5) that

$$\begin{aligned}
\mathbb{E}X(G_1 \cup G_2, \mathcal{C}) =&O_p\left(\prod_{\substack{j\in J_1\\ |C_{1j}|>1}}\left(m\prod_{v_i\in C_{1j}} p_i\right)\prod_{\substack{j\in J_2\\ |C_{2j}|>1}}\left(m\prod_{v_i\in C_{2j}} p_i\right)\right.\\
&\left.\cdot\prod_{\substack{(k,l)\in J_3\\ |C_{1k}\cup C_{2l}|>1}}\left(m\prod_{v_i\in C_{1k}\cup C_{2l}} p_i\right)\right).
\end{aligned}$$

When $|C_{1k}| = |C_{2l}| = 1$, $m\prod_{v_i\in C_{1k}\cup C_{2l}} p_i = o_p(1)$; when $|C_{1k}| = 1$ and $|C_{2l}| > 1$, $m\prod_{v_i\in C_{1k}\cup C_{2l}} p_i \le m\prod_{v_i\in C_{2l}} p_i$; when $|C_{1k}| > 1$ and $|C_{2l}| = 1$, $m\prod_{v_i\in C_{1k}\cup C_{2l}} p_i \le m\prod_{v_i\in C_{1k}} p_i$; when $|C_{1k}| > 1$ and $|C_{2l}| > 1$, $m\prod_{v_i\in C_{1k}\cup C_{2l}} p_i = (m\prod_{v_i\in C_{1k}} p_i)(m\prod_{v_i\in C_{2l}} p_i)\ /(m\prod_{v_i\in C_{1k}\cap C_{2l}} p_i)$. Arguing similarly as above, we derive

$$\begin{aligned}
&\mathbb{E}X(G_1 \cup G_2, \mathcal{C})\\
=&O_p\left(\prod_{\substack{1\le j\le r_1\\ |C_{1j}|>1}}\left(m\prod_{v_i\in C_{1j}} p_i\right)\prod_{\substack{1\le j\le r_2\\ |C_{2j}|>1}}\left(m\prod_{v_i\in C_{2j}} p_i\right)\right.
\end{aligned}$$

$$\cdot \prod_{\substack{(k,l)\in J_3 \\ |C_{1k}|>1, |C_{2l}|>1}} \frac{1}{m \prod_{v_i \in C_{1k}\cap C_{2l}} p_i} \Bigg)$$

$$= O_p \left(\prod_{\substack{1\le j\le r_1 \\ |C_{1j}|>1}} \left(m \prod_{v_i\in C_{1j}} p_i \right) \prod_{\substack{1\le j\le r_2 \\ |C_{2j}|>1}} \left(m \prod_{v_i\in C_{2j}} p_i \right) \right.$$

$$\left. \cdot \prod_{C\in\mathcal{C}_2'[V(G_1)\cap V(G_2)]} \frac{1}{m\prod_{v_i\in C} p_i} \right)$$

$$= O_p \left(\prod_{\substack{1\le j\le r_1 \\ |C_{1j}|>1}} \left(m \prod_{v_i\in C_{1j}} p_i \right) \prod_{\substack{1\le j\le r_2 \\ |C_{2j}|>1}} \left(m \prod_{v_i\in C_{2j}} p_i \right) \cdot \frac{n^\ell}{\omega(G_2,\mathcal{C}_2)} \right),$$

which combining with (4) and Lemma 2 proves the lemma. □

The dependency graph is an essential concept in applying stein's method [3]. A dependency graph D of a family of random variables $\{Z_i\}_{i\in V(D)}$ is a graph with vertex set $V(D)$ and edge set $E(D)$ such that if A and B are two disjoint subsets of $V(D)$ with no edges running between them, then the families $\{Z_i\}_{i\in A}$ and $\{Z_i\}_{i\in B}$ are mutually independent.

Lemma 4. ([7, p.154]) *Suppose that $Z = \sum_{i\in V(D)} Z_i$, where the Z_i are random indicator variables with a dependency graph D. Then, with $\pi_i = \mathbb{E}Z_i$ and $\lambda = \mathbb{E}Z = \sum_{i\in V(D)} \pi_i$ (and with summation over ordered pairs (i,j)),*

$$d_{TV}(Z, \mathrm{Poi}(\lambda)) \le \min\{\lambda^{-1}, 1\} \left(\sum_{i\in V(D)} \pi_i^2 + \sum_{(i,j)\in E(D)} (\pi_i\pi_j + \mathbb{E}(Z_iZ_j)) \right).$$

Proof of Theorem 1. Given $\{\theta_i\}_{i=1}^n$, for each $1\le i\le N$ we write $X(H_i,\mathcal{C})$ for the indicator random variable that the clique cover in $\mathcal{C}(H_i)$ corresponding (via the isomorphism between H and H_i) to $\mathcal{C}\in\mathcal{C}(H)$ induces H_i. Hence, we decompose $X(H)$ as $X(H) = \sum_{i=1}^N X_i = Y_0 + Y_1$, where
$Y_0 = \sum_{i=1}^N \sum_{\mathcal{C}\in\mathcal{C}_0(H)} X(H_i,\mathcal{C})$ and $Y_1 = \sum_{i=1}^N \sum_{\mathcal{C}\in\mathcal{C}(H)\setminus\mathcal{C}_0(H)} X(H_i,\mathcal{C})$.

Capitalizing Lemma 2 and noting the fact that $\eta_0(H) = \eta_1(H,\mathcal{C})$ $= \eta_2(H,\mathcal{C},V(H))$ for $\mathcal{C}\in\mathcal{C}_0(H)$, we deduce

$$\mathbb{E}Y_0 \sim \frac{n^h}{aut(H)} \sum_{\mathcal{C}\in\mathcal{C}_0(H)} \pi(H,\mathcal{C}) \sim \frac{1}{aut(H)} \sum_{\mathcal{C}\in\mathcal{C}_0(H)} \left(\beta^{|\mathcal{C}|}\gamma^{\Sigma\,\mathcal{C}} \prod_{v_i\in C\in\mathcal{C}} \theta_i \right) = \lambda,$$

as n tends to infinity. Define $\phi = \phi(H) = \min_{\mathcal{C}\in\mathcal{C}_0(H)} \omega(H,\mathcal{C})$. Since H is strictly α-balanced, for each $\mathcal{C}\in\mathcal{C}_0(H)$ and $\emptyset \ne S\subset V(H)$, $\psi(H,\mathcal{C},S)\to\infty$ as $n\to\infty$. This implies that $\lim_{n\to\infty}\phi = \infty$.

Next, we define a dependency graph D as in [13] with the vertex set $V(D) = \{1, \cdots, N\} \times \mathcal{C}_0(H)$ such that $\{(i, \mathcal{C}_1), (j, \mathcal{C}_2)\} \in E(D)$ if and only if $V(H_i) \cap V(H_j) \neq \emptyset$. Recall that $\mathbb{E}X(H_i, \mathcal{C}) = \pi(H_i, \mathcal{C})$, we obtain from Lemma 4 that

$$d_{TV}(Y_0, \mathrm{Poi}(\mathbb{E}Y_0)) \leq \min\{(\mathbb{E}Y_0)^{-1}, 1\}\Bigg(\sum_{(i,\mathcal{C})} \pi^2(H_i, \mathcal{C})$$

$$+ \sum_{\{(i,\mathcal{C}_1),(j,\mathcal{C}_2)\} \in E(D)} \Big(\pi(H_i, \mathcal{C}_1)\pi(H_j, \mathcal{C}_2) + \mathbb{E}(X(H_i, \mathcal{C}_1)X(H_j, \mathcal{C}_2))\Big)\Bigg). \tag{6}$$

We now estimate each term on the righthand side of (6) as follows. By Lemma 2, $N = O(n^h)$, and the definition of dependency graph D we obtain

$$\sum_{(i,\mathcal{C}) \in V(D)} \pi^2(H_i, \mathcal{C}) = O\left(\sum_{\mathcal{C} \in \mathcal{C}_0} n^h m^{2|\mathcal{C}|}\Big(\prod_{v_i \in C \in \mathcal{C}} p_i\Big)^2\right) = O(n^{-h})$$

and

$$\begin{aligned}
&\sum_{\{(i,\mathcal{C}_1),(j,\mathcal{C}_2)\} \in E(D)} \pi(H_i, \mathcal{C}_1)\pi(H_j, \mathcal{C}_2) \\
=&O\left(n^{2h-1} m^{2|\mathcal{C}|}\Big(\prod_{v_i \in C \in \mathcal{C}_1} p_i\Big)\Big(\prod_{v_i \in C \in \mathcal{C}_2} p_i\Big)\right) \\
=&O(n^{-1}).
\end{aligned}$$

For any two different proper clique covers $\mathcal{C}_1, \mathcal{C}_2 \in \mathcal{C}_0(H_i)$, we have $\mathbb{E}(X(H_i, \mathcal{C}_1)X(H_i, \mathcal{C}_2)) = 0$ $(1 \leq i \leq N)$. Furthermore, using Lemma 2 and Lemma 3, we obtain

$$\begin{aligned}
&\sum_{\{(i,\mathcal{C}_1),(j,\mathcal{C}_2)\} \in E(D)} \mathbb{E}(X(H_i, \mathcal{C}_1)X(H_j, \mathcal{C}_2)) \\
=&\sum_{\substack{1 \leq i,j \leq N \\ V(H_i) \cap V(H_j) \neq \emptyset}} \sum_{\substack{\mathcal{C}_1 \in \mathcal{C}_0(H_i) \\ \mathcal{C}_2 \in \mathcal{C}_0(H_j)}} \mathbb{E}(X(H_i, \mathcal{C}_1)X(H_j, \mathcal{C}_2)) \\
=&O\left(\sum_{\mathcal{C}_1, \mathcal{C}_2 \in \mathcal{C}_0(H)} n^h \sum_{\ell=1}^{h-1} \frac{n^{h-\ell}}{\phi} \pi(H, \mathcal{C}_1)\pi(H, \mathcal{C}_2) n^{\ell}\right) \\
=&O(\phi^{-1}) = o(1),
\end{aligned}$$

where $\ell = |V(H_i) \cap V(H_j)|$. Applying the above estimates to (6), we obtain $d_{TV}(Y_0, \mathrm{Poi}(\mathbb{E}Y_0)) = o(1)$. Consequently,

$$\begin{aligned}
d_{TV}(Y_0, \mathrm{Poi}(\lambda)) &\leq d_{TV}(Y_0, \mathrm{Poi}(\mathbb{E}Y_0)) + d_{TV}(\mathrm{Poi}(\mathbb{E}Y_0), \mathrm{Poi}(\lambda)) \\
&= o(1) + O(|\mathbb{E}Y_0 - \lambda|) = o(1).
\end{aligned} \tag{7}$$

We claim that $\mathbb{P}(Y_1 > 0) = o(1)$. If it is true, then $d_{TV}(X, \mathrm{Poi}(\lambda)) \leq d_{TV}(X, Y_0) + d_{TV}(Y_0, \mathrm{Poi}(\lambda)) \leq \mathbb{P}(Y_1 > 0) + d_{TV}(Y_0, \mathrm{Poi}(\lambda)) = o(1)$ by (7). The proof of Theorem 1 is then complete.

What remains to show is the vanishing probability of the event $\{Y_1 > 0\}$. In fact, for $\mathcal{C} \in \mathcal{C}(H)\backslash\mathcal{C}_0(H)$, there exists some non-empty $S \subseteq V(H)$ satisfying $\eta_2(H, \mathcal{C}, S) < \eta_0(H)$. Let $H_i[S]$ be the induced subgraph of S. We have $\mathbb{P}\big(\sum_{i=1}^N X(H_i, \mathcal{C}) > 0\big) \le \mathbb{P}\big(\exists 1 \le i \le N, X(H_i[S], \mathcal{C}[S]) > 0\big) \le \psi(H, \mathcal{C}, S) = o(1)$. We arrive at $\mathbb{P}(Y_1 > 0) = o(1)$ due to the finiteness of the clique covers. □

Acknowledgment

The author would like to thank the anonymous reviewer for detailed constructive comments that helped improve the quality of the paper.

References

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

[2] 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, M. Böhmer (eds.) Data Science, Learning by Latent Structures, and Knowledge Discovery, pp. 69–78, Springer, Berlin, Heidelberg, 2015.

[3] M. Coulson, R. E. Gaunt, and G. Reinert, *Poisson approximation of subgraph counts in stochastic block models and a graphon model*, ESAIM: Prob. Stat. **20** (2016), 131–142.

[4] M. Deijfen and W. Kets, *Random intersection graphs with tunable degree distribution and clustering*, Probab. Engin. Inform. Sci. **23** (2009), 661–674.

[5] S. Janson, T. Luczak, and A. Ruciński, *Random Graphs*, Wiley, New York, 2000.

[6] T. Jeong, Y. M. Jung, and S. Yun, *Iterative reweighted algorithm for non-convex Poissonian image restoration model*, J. Korean Math. Soc. **55** (2018), 719–734.

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

[8] S. Nikoletseas, C. Raptopoulos, and P. Spirakis, *Large independent sets in general random intersection graphs*, Theor. Comput. Sci. **406** (2008), 215–224.

[9] K. Rybarczyk, *The coupling method for inhomogeneous random intersection graphs*, Electron. J. Combin. **24** (2017), #P2.10.

[10] K. Rybarczyk and D. Stark, *Poisson approximation of the number of cliques in random intersection graphs*, J. Appl. Probab. **47** (2010), 826–840.

[11] ———, *Poisson approximation of counts of induced subgraphs in random intersection graphs*, Discrete Math. **340** (2017), 2183–2193.

[12] Y. Shang, *Degree distributions in general random intersection graphs*, Electron. J. Combin. **17** (2010), #P23.

[13] ———, Isoperimetric numbers of randomly perturbed intersection graphs. *Symmetry*, 11(2019) article 452.

Yilun Shang
Department of Computer and Information Sciences
Faculty of Engineering and Environment
Northumbria University
Newcastle upon Tyne NE1 8ST, United Kingdom
E-mail address: `yilun.shang@northumbria.ac.uk`