Degree distribution of an inhomogeneous random intersection graph

Mindaugas Bloznelis, Julius Damarackas

Vilnius University, Faculty of mathematics and informatics,
Naugarduko 24, Vilnius, 03225, Lithuania

Abstract

We show the asymptotic degree distribution of the typical vertex of a sparse inhomogeneous random intersection graph.

key words: degree distribution, random graph, random intersection graph, power law

2000 Mathematics Subject Classifications: 05C80, 05C07, 05C82

1 Introduction

Let $X_1, \dots, X_m, Y_1, \dots, Y_n$ be independent non-negative random variables such that each X_i has the probability distribution P_1 and each Y_j has the probability distribution P_2. Given realized values $X = \{X_i\}_{i=1}^m$ and $Y = \{Y_j\}_{j=1}^n$ we define the random bipartite graph $H_{X,Y}$ with the bipartition $V = \{v_1, \dots, v_n\}$, $W = \{w_1, \dots, w_m\}$, where edges $\{w_i, v_j\}$ are inserted with probabilities $p_{ij} = \min\{1, X_i Y_j (nm)^{-1/2}\}$ independently for each $\{i, j\} \in [m] \times [n]$. The *inhomogeneous* random intersection graph $G(P_1, P_2, n, m)$ defines the adjacency relation on the vertex set V: vertices $v', v'' \in V$ are declared adjacent (denoted $v' \sim v''$) whenever v' and v'' have a common neighbour in $H_{X,Y}$.

The degree distribution of the typical vertex of the random graph $G(P_1, P_2, n, m)$ has been first considered by Shang [20]. The proof of the main result of [20] contains gaps and the result is incorrect in the regime where $m/n \to \beta \in (0, +\infty)$ as $m, n \to +\infty$. We remark that this regime is of particular importance, because it leads to inhomogeneous graphs with the clustering property: the clustering coefficient $\mathbf{P}(v_1 \sim v_2 | v_1 \sim v_3, v_2 \sim v_3)$ is bounded away from zero provided that $\mathbf{E}X_1^3 < \infty$ and $\mathbf{E}Y_1^2 < \infty$, see [8]. The aim of the present paper is to show the asymptotic degree distribution in the case where $m/n \to \beta$ for some $\beta \in (0, +\infty)$.

We consider a sequence of graphs $\{G_n = G(P_1, P_2, n, m)\}$, where $m = m_n \to +\infty$ as $n \to \infty$, and where P_1, P_2 do not depend on n. We denote $a_i = \mathbf{E}X_1^i$, $b_i = \mathbf{E}Y_1^i$. By $d(v_j) = d_{G_n}(v_j)$ we denote the degree of a vertex v_j in G_n (the number of vertices adjacent to v_j in G_n). We remark that for every n the random variables $d_{G_n}(v_1), \dots, d_{G_n}(v_n)$ are identically distributed. In Theorem 1 below we show the asymptotic distribution of $d(v_1)$.

Theorem 1. *Let $m, n \to \infty$.*

(i) Assume that $m/n \to 0$. Suppose that that $\mathbf{E}X_1 < \infty$. Then $\mathbf{P}(d(v_1) = 0) = 1 - o(1)$.
(ii) Assume that $m/n \to \beta$ for some $\beta \in (0, +\infty)$. Suppose that $\mathbf{E}X_1^2 < \infty$ and $\mathbf{E}Y_1 < \infty$. Then $d(v_1)$ converges in distribution to the random variable

$$d_* = \sum_{j=1}^{\Lambda_1} \tau_j, \tag{1}$$

where $\tau_1, \tau_2, \dots$ are independent and identically distributed random variables independent of the random variable Λ_1. They are distributed as follows. For $r = 0, 1, 2, \dots$, we have

$$\mathbf{P}(\tau_1 = r) = \frac{r+1}{\mathbf{E}\Lambda_2}\mathbf{P}(\Lambda_2 = r+1) \qquad \text{and} \qquad \mathbf{P}(\Lambda_i = r) = \mathbf{E}\, e^{-\lambda_i}\frac{\lambda_i^r}{r!}, \qquad i = 1, 2. \tag{2}$$

Here $\lambda_1 = Y_1 a_1 \beta^{1/2}$ and $\lambda_2 = X_1 b_1 \beta^{-1/2}$.
(iii) Assume that $m/n \to +\infty$. Suppose that $\mathbf{E}X_1^2 < \infty$ and $\mathbf{E}Y_1 < \infty$. Then $d(v_1)$ converges in distribution to a random variable Λ_3 having the probability distribution

$$\mathbf{P}(\Lambda_3 = r) = \mathbf{E}e^{-\lambda_3}\frac{\lambda_3^r}{r!}, \qquad r = 0, 1, \dots. \tag{3}$$

Here $\lambda_3 = Y_1 a_2 b_1$.

Remark 1. The probability distributions P_{Λ_i} of Λ_i, $i = 1, 2, 3$, are Poisson mixtures. One way to sample from the distribution P_{Λ_i} is to generate random variable λ_i and then, given λ_i, to generate Poisson random variable with the parameter λ_i. The realized value of the Poisson random variable has the distribution P_{Λ_i}.
Remark 2. The asymptotic distributions (1) and (3) admit heavy tails. In the case (ii) we obtain a power law asymptotic degree distribution (1) provided that at least one of P_1 and P_2 has a power law and $P_1(0), P_2(0) < 1$. In the case (iii) we obtain a power law asymptotic degree distribution (3) provided that P_2 has a power law.
Remark 3. Since the second moment a_2 does not show up in (1), (2) we expect that in the case (ii) the second moment condition $\mathbf{E}X_1^2 < \infty$ is redundant and could perhaps be replaced by the weaker first moment condition $\mathbf{E}X_1 < \infty$.
Random intersection graphs have attracted considerable attention in the recent literature, see, e.g., [1], [2], [3], [9], [10], [12], [18]. Starting with the paper by Karoński et al [16], see also [21], where the case of degenerate distributions $P_1 = P_{1n}$, $P_2 = P_{2n}$ depending on n was considered (i.e., $\mathbf{P}_{1n}(c_n) = \mathbf{P}_{2n}(c_n) = 1$, for some $c_n > 0$), several more complex random intersection graph models were later introduced by Godehardt and Jaworski [13], Spirakis et al. [17], Shang [20]. The asymptotic degree distribution for various random intersection graph models was shown in [4], [5], [6], [7], [11], [14], [15], [19], [22].

2 Proofs

Before the proof we introduce some notation and give two auxiliary lemmas.
The event that the edge $\{w_i, v_j\}$ is present in $H = H_{X,Y}$ is denoted $w_i \to v_j$. We denote

$$\mathbb{I}_{ij} = \mathbb{I}_{\{w_i \to v_j\}}, \qquad \mathbb{I}_i = \mathbb{I}_{i1}, \qquad u_i = \sum_{2 \le j \le n} \mathbb{I}_{ij}, \qquad L = \sum_{i=1}^{m} u_i \mathbb{I}_i.$$

We remark, that u_i counts all neighbours of w_i in H belonging to the set $V \setminus \{v_1\}$. Denote

$$\hat{a}_k = m^{-1} \sum_{1 \le i \le m} X_i^k, \qquad \hat{b}_k = n^{-1} \sum_{2 \le j \le n} Y_j^k,$$
$$\lambda_{ij} = \frac{X_i Y_j}{\sqrt{mn}}, \qquad S_{XY} = \sum_{i=1}^{m} \sum_{j=2}^{n} \frac{X_i}{\sqrt{nm}} \lambda_{ij} \min\{1, \lambda_{ij}\}, \tag{4}$$

and introduce the event $\mathcal{A}_1 = \{\lambda_{i1} < 1,\ 1 \le i \le m\}$. By $\mathbf{P}_1$ and $\mathbf{E}_1$ we denote the conditional probability and conditional expectation given Y_1. By $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{E}}$ we denote the conditional probability and conditional expectation given X, Y. By $d_{TV}(\zeta, \xi)$ we denote the total variantion distance between the probability distributions of random variables ζ and ξ. In the case where ζ, ξ and X, Y are defined on the same probability space we denote by $\tilde{d}_{TV}(\zeta, \xi)$ the total variation distance between the conditional distributions of ζ and ξ given X, Y.
In the proof below we use the following simple fact about the convergence of a sequence of random variables $\{\varkappa_n\}$:

$$\varkappa_n = o_P(1), \quad \mathbf{E} \sup_n \varkappa_n < \infty \quad \Rightarrow \quad \mathbf{E}\varkappa_n = o(1). \tag{5}$$

We remark that the condition $\mathbf{E} \sup_n \varkappa_n < \infty$ (which assumes implicitly that all $\varkappa_n$ are defined on the same probability space) can be replaced by the more resctrictive condition that there exists a constant $c > 0$ such that $|\varkappa_n| \le c$, for all n. In the latter case we do not need all $\varkappa_n$ to be defined on the same probability space. In particular, given a sequence of bivariate random vectors $\{(\eta_n, \theta_n)\}$ such that, for every n and $m = m_n$ random variables η_n, θ_n and $\{X_i\}_{i=1}^m$, $\{Y_j\}_{j=1}^n$ are defined on the same probability space, we have

$$\tilde{d}_{TV}(\eta_n, \theta_n) = o_P(1) \ \Rightarrow \ d_{TV}(\eta_n, \theta_n) \le \mathbf{E}\tilde{d}_{TV}(\eta_n, \theta_n) = o(1).$$

Lemma 1. *Assume that* $\mathbf{E}X_1^2 < \infty$ *and* $\mathbf{E}Y_1 < \infty$. *We have as* $n, m \to +\infty$

$$\mathbf{P}(d(v_1) \ne L) = o(1), \tag{6}$$
$$\mathbf{P}(\mathcal{A}_1) = 1 - o(1), \tag{7}$$
$$S_{XY} = o_P(1), \qquad \mathbf{E}S_{XY} = o(1). \tag{8}$$

Proof of Lemma 1. Proof of (6). We observe that the event $\mathcal{A} = \{d(v_1) \ne L\}$ occurs in the case where for some $2 \le j \le n$ and some distinct $i_1, i_2 \in [m]$ the event $\mathcal{A}_{i_1,i_2,j} = \{w_{i_1} \to v_1,\ w_{i_1} \to v_j,\ w_{i_2} \to v_1,\ w_{i_2} \to v_j\}$ occurs. From the union bound and the inequality $\tilde{\mathbf{P}}(\mathcal{A}_{i_1,i_2,j}) \le m^{-2} n^{-2} X_{i_1}^2 X_{i_2}^2 Y_1^2 Y_j^2$ we obtain

$$\tilde{\mathbf{P}}(\mathcal{A}) = \tilde{\mathbf{P}}\left(\bigcup_{\{i_1,i_2\} \subset [m]} \bigcup_{2 \le j \le n} \mathcal{A}_{i_1,i_2,j} \right) \le n^{-1} \hat{b}_2 Y_1^2 Q_X. \tag{9}$$

Here $Q_X = m^{-2} \sum_{\{i_1,i_2\} \subset [m]} X_{i_1}^2 X_{i_2}^2$. We note that Q_X and Y_1^2 are stochastically bounded and $n^{-1}\hat{b}_2 = o_P(1)$ as $n \to +\infty$. Therefore, $\tilde{\mathbf{P}}(\mathcal{A}) = o_P(1)$. Now (5) implies (6).

Proof of (7). Let $\overline{\mathcal{A}}_1$ denote the complement event to $\mathcal{A}_1$. By the union bound and Markov's inequality

$$\mathbf{P}_1(\overline{\mathcal{A}}_1) \le \sum_{i \in [m]} \mathbf{P}_1(\lambda_{i1} \ge 1) \le (nm)^{-1} Y_1^2 \sum_{i \in [m]} \mathbf{E}X_i^2 = n^{-1} a_2 Y_1^2.$$

Hence we obtain $\mathbf{P}_1(\overline{\mathcal{A}}_1) = o_P(1)$. Now (5) implies $\mathbf{P}(\overline{\mathcal{A}}_1) = \mathbf{E}\mathbf{P}_1(\overline{\mathcal{A}}_1) = o(1)$.

Proof of (8). Since the first bound of (8) follows from the second one, we only prove the latter. Denote $\hat{X}_1 = \max\{X_1, 1\}$ and $\hat{Y}_1 = \max\{Y_1, 1\}$. We observe that $\mathbf{E}X_1^2 < \infty$, $\mathbf{E}Y_1 < \infty$ implies

$$\lim_{t\to+\infty} \mathbf{E}\hat{X}_1^2 \mathbb{I}_{\{\hat{X}_1 > t\}} = 0, \qquad \lim_{t\to+\infty} \mathbf{E}\hat{Y}_1 \mathbb{I}_{\{\hat{Y}_1 > t\}} = 0.$$

Hence one can find a strictly increasing function $\varphi : [1, +\infty) \to [0, +\infty)$ with $\lim_{t\to+\infty} \varphi(t) = +\infty$ such that

$$\mathbf{E}\hat{X}_1^2 \varphi(\hat{X}_1) < \infty, \qquad \mathbf{E}\hat{Y}_1 \varphi(\hat{Y}_1) < \infty. \tag{10}$$

In addition, we can choose φ satisfying

$$\varphi(t) < t \qquad \text{and} \qquad \varphi(st) \le \varphi(t)\varphi(s), \qquad \forall s, t \ge 1. \tag{11}$$

For this purpose we take a sufficiently slowly growing concave function $\psi : [0, +\infty) \to [0, +\infty)$ with $\psi(0) = 0$ and define $\varphi(t) = e^{\psi(\ln(t))}$. We note that the second inequality of (11) follows from the concavity property of ψ. We remark, that (11) implies

$$t/(st) = s^{-1} \le 1/\varphi(s) \le \varphi(t)/\varphi(st), \qquad s, t \ge 1. \tag{12}$$

Let $\hat{S}_{XY}$ be defined as in (4) above, but with λ_{ij} replaced by $\hat{\lambda}_{ij} = \hat{X}_i\hat{Y}_j/\sqrt{mn}$. We note, that $S_{XY} \le \hat{S}_{XY}$. Furthermore, from the inequalities

$$\min\{1, \hat{\lambda}_{ij}\} \le \min\Big\{1, \frac{\varphi(\hat{X}_i\hat{Y}_j)}{\varphi(\sqrt{mn})}\Big\} \le \frac{\varphi(\hat{X}_i\hat{Y}_j)}{\varphi(\sqrt{mn})} \le \frac{\varphi(\hat{X}_i)\varphi(\hat{Y}_j)}{\varphi(\sqrt{mn})} \tag{13}$$

we obtain $S_{XY} \le \hat{S}_{XY} \le S^*_{XY}/\varphi(\sqrt{nm})$, where

$$S^*_{XY} = (mn)^{-1} \Bigg(\sum_{1\le i\le m} \hat{X}_i^2 \varphi(\hat{X}_i)\Bigg) \Bigg(\sum_{2\le j\le n} \hat{Y}_j \varphi(\hat{Y}_j)\Bigg).$$

We remark that (11) and (12) and imply the third and the first inequality of (13), respectively. Finally, the bound $\mathbf{E}S_{XY} = o(1)$ follows from the inequality $S_{XY} \le S^*_{XY}/\varphi(\sqrt{nm})$ and the fact that $\mathbf{E}S^*_{XY}$ remains bounded as $n, m \to +\infty$, see (10). □

In the proof of Theorem 1 we use the following inequality refered to as LeCam's lemma, see e.g., [23].

Lemma 2. *Let* $S = \mathbb{I}_1 + \mathbb{I}_2 + \cdots + \mathbb{I}_n$ *be the sum of independent random indicators with probabilities* $\mathbf{P}(\mathbb{I}_i = 1) = p_i$. *Let* Λ *be Poisson random variable with mean* $p_1 + \cdots + p_n$. *The total variation distance between the distributions* P_S *of* P_Λ *of* S *and* Λ

$$\sup_{A\subset\{0,1,2...\}} |\mathbf{P}(S \in A) - \mathbf{P}(\Lambda \in A)| = \frac{1}{2}\sum_{k\ge0} |\mathbf{P}(S = k) - \mathbf{P}(\Lambda = k)| \le \sum_i p_i^2. \tag{14}$$

Proof of Theorem 1. In view of (6) the random variables $d(v_1)$ and L have the same asymptotic distribution (if any). We shall show the convergence in distribution of L.

The case (i). Here we prove that $\mathbf{P}(L > \varepsilon) = o(1)$, for any $\varepsilon \in (0, 1)$. In view of the identity $\mathbf{P}(L > \varepsilon) = \mathbf{E}\mathbf{P}_1(L > \varepsilon)$ and (5) it suffices to show that $\mathbf{P}_1(L > \varepsilon) = o_P(1)$. For this purpose we write, by the union bound and Markov's inequality,

$$\mathbf{P}_1(L > \varepsilon) \le \sum_{1\le k\le m} \mathbf{P}_1(\mathbb{I}_k = 1) \le \mathbf{E}_1 \sum_{1\le k\le m} \lambda_{k1} = \sqrt{m/n}\, Y_1 \mathbf{E}X_1 = o_P(1).$$

The case (ii). Here we prove that L converges in distribution to (1). We first approximate L by the random variable $L_3 = \sum_{k=1}^{m} \eta_k \xi_{3k}$. Then we show that L_3 converges in distribution to (1). Here $\eta_1, \dots, \eta_m$, $\xi_{31}, \dots, \xi_{3m}$ are conditionally independent (given X, Y) Poisson random variables with $\tilde{\mathbf{E}}\eta_k = \lambda_{k1}$ and $\tilde{\mathbf{E}}\xi_{3k} = X_k (n/m)^{1/2} b_1$. We assume, in addition, that given X, Y, the sequences $\{\mathbb{I}_k\}_{k=1}^m$ and $\{\xi_{3k}\}_{k=1}^m$ are conditionally independent.
Given X, Y, we generate independent Poisson random variables $\xi_{11}, \dots, \xi_{1m}$, $\Delta_{11}, \dots, \Delta_{1m}$, with the conditional mean values

$$\tilde{\mathbf{E}}\xi_{1k} = \sum_{2 \le j \le n} p_{kj}, \qquad \tilde{\mathbf{E}}\Delta_{1k} = \sum_{2 \le j \le n} (\lambda_{kj} - p_{kj}), \qquad 1 \le k \le m.$$

We assume that, given X, Y, these Poisson random variables are conditionally independent of the sequence $\{\eta_k\}_{k=1}^m$. We suppose, in addition, that $\{\eta_k\}_{k=1}^m$ is conditionally independent (given X, Y_1) of the set of edges of H that are not incident to v_1. We define $\xi_{2k} = \xi_{1k} + \Delta_{1k}$ and observe that ξ_{2k} has conditional (given X, Y) Poisson distribution with the conditional mean value $\tilde{\mathbf{E}}\xi_{2k} = \sum_{2 \le j \le n} \lambda_{kj}$. Introduce the random variables

$$L_0 = \sum_{1 \le k \le m} \eta_k u_k, \qquad L_1 = \sum_{1 \le k \le m} \eta_k \xi_{1k}, \qquad L_2 = \sum_{1 \le k \le m} \eta_k \xi_{2k}.$$

In order to show that L and L_3 have the same asymptotic probability distribution (if any) we prove that

$$d_{TV}(L, L_0) = o(1), \qquad d_{TV}(L_0, L_1) = o(1), \tag{15}$$
$$\mathbf{E}|L_1 - L_2| = o(1), \qquad \tilde{L}_2 - \tilde{L}_3 = o_P(1). \tag{16}$$

Here $\tilde{L}_2$ and $\tilde{L}_3$ are marginals of the random vector $(\tilde{L}_2, \tilde{L}_3)$ constructed below which has the property that $\tilde{L}_2$ has the same distribution as L_2 and $\tilde{L}_3$ has the same distribution as L_3.
Let us prove the first bound of (15). In view of (5) it suffices to show that $\tilde{d}_{TV}(L_0, L) = o_P(1)$. In order to prove the latter bound we apply the inequality

$$\tilde{d}_{TV}(L_0, L)\mathbb{I}_{\mathcal{A}_1} \le n^{-1} Y_1^2 \hat{a}_2 \tag{17}$$

shown below. We remark that (17) implies

$$\tilde{d}_{TV}(L_0, L) \le \tilde{d}_{TV}(L_0, L)\mathbb{I}_{\mathcal{A}_1} + \mathbb{I}_{\overline{\mathcal{A}}_1} \le n^{-1} Y_1^2 \hat{a}_2 + \mathbb{I}_{\overline{\mathcal{A}}_1} = o_P(1).$$

Here $n^{-1} Y_1^2 \hat{a}_2 = o_P(1)$, because $Y_1^2 \hat{a}_2$ is stochastically bounded. Furthermore, the bound $\mathbb{I}_{\overline{\mathcal{A}}_1} = o_P(1)$ follows from (7).
It remains to prove (17). We denote $L'_k = \sum_{i=1}^{k} \mathbb{I}_i u_i + \sum_{i=k+1}^{m} \eta_i u_i$ and write, by the triangle inequality,

$$\tilde{d}_{TV}(L_0, L) \le \sum_{k=1}^{m} \tilde{d}_{TV}(L'_{k-1}, L'_k).$$

Then we estimate $\tilde{d}_{TV}(L'_{k-1}, L'_k) \le \tilde{d}_{TV}(\eta_k, \mathbb{I}_k) \le (nm)^{-1} Y_1^2 X_k^2$. Here the first inequality follows from the properties of the total variation distance. The second inequality follows from Lemma 2 and the fact that on the event $\mathcal{A}_1$ we have $p_{k1} = \lambda_{k1}$.

Let us prove the second bound of (15). In view of (5) it suffices to show that $\tilde{d}_{TV}(L_0, L_1) = o_P(1)$. We denote $L^*_k = \sum_{i=1}^{k} \eta_i u_i + \sum_{i=k+1}^{m} \eta_i \xi_{1i}$ and write, by the triangle inequality,

$$\tilde{d}_{TV}(L_0, L_1) \le \sum_{k=1}^{m} \tilde{d}_{TV}(L^*_{k-1}, L^*_k). \tag{18}$$

Here

$$\tilde{d}_{TV}(L^*_{k-1}, L^*_k) \le \tilde{d}_{TV}(\eta_k u_k, \eta_k \xi_{1k}) \le \tilde{\mathbf{P}}(\eta_k \neq 0)\tilde{d}_{TV}(u_k, \xi_{1k}). \qquad (19)$$

Now, invoking the inequalities $\tilde{\mathbf{P}}(\eta_k \neq 0) = 1 - e^{-\lambda_{k1}} \le \lambda_{k1}$ and $\tilde{d}_{TV}(u_k, \xi_{1k}) \le \sum_{j=2}^n p_{kj}^2$, we obtain from (18), (19) and (8) that

$$\tilde{d}_{TV}(L_0, L_1) \le \sum_{k=1}^m \lambda_{k1} \sum_{j=2}^n p_{kj}^2 \le Y_1 S_{XY} = o_P(1).$$

Let us prove the first bound of (16). We observe that

$$|L_2 - L_1| = L_2 - L_1 = \sum_{1\le k\le m} \eta_k \Delta_{1k}$$

and

$$\tilde{\mathbf{E}} \sum_{1\le k\le m} \eta_k \Delta_{1k} = \sum_{1\le k\le m} \lambda_{k1} \sum_{2\le j\le n} (\lambda_{kj} - 1)\mathbb{I}_{\{\lambda_{kj}>1\}} \le Y_1 S_{XY}.$$

We obtain $\mathbf{E}|L_2 - L_1| \le \mathbf{E}Y_1 \mathbf{E}S_{XY} = o(1)$, see (8).

Let us prove the second bound of (16). Given X, Y, generate independent Poisson random variables $\xi'_{31}, \dots, \xi'_{3m}, \Delta_{21}, \dots, \Delta_{2m}, \Delta_{31}, \dots, \Delta_{3m}$ which are conditionally independent of the sequence $\{\eta_k\}_{k=1}^m$ and have the conditional mean values

$$\tilde{\mathbf{E}}\xi'_{3k} = X_k (n/m)^{1/2} b, \qquad \tilde{\mathbf{E}}\Delta_{2k} = X_k (n/m)^{1/2}\delta_2, \qquad \tilde{\mathbf{E}}\Delta_{3k} = X_k (n/m)^{1/2}\delta_3.$$

Here $b = \min\{\hat{b}_1, b_1\}$, $\delta_2 = \hat{b}_1 - b$, $\delta_3 = b_1 - b$. We note that $\delta_2, \delta_3 \ge 0$ and observe that the random vector

$$(L'_2, L'_3), \qquad L'_2 = \sum_{1\le k\le m} \eta_k(\xi'_{3k} + \Delta_{2k}), \qquad L'_3 = \sum_{1\le k\le m} \eta_k(\xi'_{3k} + \Delta_{3k})$$

has the marginal distributions of (L_2, L_3). In addition, we have

$$\tilde{\mathbf{E}}|L'_2 - L'_3| = (\delta_2 + \delta_3) Y_1 \hat{a}_2 = |\hat{b}_1 - b_1| Y_1 \hat{a}_2 = o_P(1). \qquad (20)$$

In the last step we used the fact that $Y_1\hat{a}_2 = O_P(1)$ and $\hat{b}_1 - b_1 = o_P(1)$, by the law of large numbers. Finally, we show that (20) implies the bound $|L'_2 - L'_3| = o_P(1)$. Denoting, for short, $H = |L'_2 - L'_3|$ and $h = \tilde{\mathbf{E}}\mathbb{I}_{\{H\ge\varepsilon\}}$ we write, for $\varepsilon \in (0,1)$,

$$\mathbf{P}(H \ge \varepsilon) = \mathbf{E}h = \mathbf{E}h(\mathbb{I}_{\{\tilde{\mathbf{E}}H \ge \varepsilon^2\}} + \mathbb{I}_{\{\tilde{\mathbf{E}}H < \varepsilon^2\}}). \qquad (21)$$

Using the simple inequality $h \le 1$ and the inequality, $h \le \varepsilon^{-1}\tilde{\mathbf{E}}H$, which follows from Markov's inequality, we obtain

$$\begin{aligned} \mathbf{E}h\mathbb{I}_{\{\tilde{\mathbf{E}}H\ge\varepsilon^2\}} &\le \mathbf{E}\mathbb{I}_{\{\tilde{\mathbf{E}}H\ge\varepsilon^2\}} = \mathbf{P}(\tilde{\mathbf{E}}H \ge \varepsilon^2) = o(1), \\ \mathbf{E}h\mathbb{I}_{\{\tilde{\mathbf{E}}H<\varepsilon^2\}} &\le \mathbf{E}(\varepsilon^{-1}\tilde{\mathbf{E}}H)\mathbb{I}_{\{\tilde{\mathbf{E}}H<\varepsilon^2\}} < \varepsilon. \end{aligned}$$

Invoking these inequalities in (21) we obtain $\mathbf{P}(H \ge \varepsilon) < \varepsilon + o(1)$. Hence $H = o_P(1)$.

Now we prove that L_3 converges in distribution to (1). Introduce the random variable $\bar{L} = \sum_{1\le k\le m} \eta_k \bar{\xi}_k$, where, given X, Y, the random variables $\bar{\xi}_1, \dots, \bar{\xi}_m$ are conditionally independent of $\{\eta_k\}_{k=1}^m$ and have the conditional mean values $\tilde{\mathbf{E}}\bar{\xi}_k = X_k \beta^{-1/2} b_1$. Proceeding as in the proof of the second bound of (16) above, we construct a random vector $(L_3'', \bar{L}')$ with the same marginals as $(L_3, \bar{L})$ and such that

$$\tilde{\mathbf{E}}|L_3'' - \bar{L}'| \le |1 - (m/n)^{1/2}\beta^{-1/2}|Y_1 b_1 \hat{a}_2 = o_P(1). \tag{22}$$

In the last step we used the fact that $Y_1\hat{a}_2 = O_P(1)$ and $m/n \to \beta$. Now, (22) implies $L_3'' - \bar{L}' = o_P(1)$. We conclude that L_3 and $\bar{L}$ have the same asymptotic distribution (if any).
Next we prove that $\bar{L}$ converges in distribution to (1). For this purpose we show that $\mathbf{E}e^{it\bar{L}} \to \mathbf{E}e^{itd_*}$, for each $t \in (-\infty, +\infty)$. Denote $\Delta(t) = e^{it\bar{L}} - e^{itd_*}$. We shall show below that, for any real t and any realized value Y_1 there exists a positive constant $c = c(t, Y_1)$ such that for every $0 < \varepsilon < 0.5$ we have

$$\limsup_{n,m\to+\infty} |\mathbf{E}(\Delta(t)|Y_1)| < c\varepsilon. \tag{23}$$

Clearly, (23) implies $\mathbf{E}(\Delta(t)|Y_1) = o(1)$. This fact together with the simple inequality $|\Delta(t)| \le 2$ yields $\mathbf{E}\Delta(t) = o(1)$, by Lebesgue's dominated convergence theorem. Observing that $\mathbf{E}\Delta(t) = \mathbf{E}e^{it\bar{L}} - \mathbf{E}e^{itd_*}$ we conclude that $\mathbf{E}e^{it\bar{L}} \to \mathbf{E}e^{itd_*}$.

We fix $0 < \varepsilon < 0.5$ and prove (23). Before the proof we introduce some notation. Denote

$$f_\tau(t) = \mathbf{E}e^{it\tau_1}, \qquad \bar{f}_\tau(t) = \sum_{r\ge0} e^{itr}\bar{p}_r, \qquad \bar{p}_r = \bar{\lambda}^{-1}\sum_{1\le k\le m}\lambda_{k1}\mathbb{I}_{\{\bar{\xi}_k = r\}}, \qquad \bar{\lambda} = \sum_{k=1}^m \lambda_{k1},$$
$$\delta = (\bar{f}_\tau(t) - 1)\bar{\lambda} - (f_\tau(t) - 1)\lambda_1, \qquad f(t) = \mathbf{E}_1 e^{itd_*}, \qquad \bar{f}(t) = \bar{\mathbf{E}}e^{it\bar{L}}.$$

Here $\bar{\mathbf{E}}$ denotes the conditional expectation given X, Y and $\bar{\xi}_1, \dots, \bar{\xi}_m$. Introduce the event $\mathcal{D} = \{|\hat{a}_1 - a_1| < \varepsilon \min\{1, a_1\}\}$ and let $\overline{\mathcal{D}}$ denote the complement event. Furthermore, select the number $T > 1/\varepsilon$ such that $\mathbf{P}(\tau_1 \ge T) < \varepsilon$. By $c_1, c_2, \dots$ we denote positive numbers which do not depend on n, m.
We observe that, given Y_1, the conditional distribution of d_* is the compound Poisson distribution with the characteristic function $f(t) = e^{\lambda_1(f_\tau(t)-1)}$. Similarly, given X, Y and $\bar{\xi}_1, \dots, \bar{\xi}_m$, the conditional distribution of $\bar{L}$ is the compound Poisson distribution with the characteristic function $\bar{f}(t) = e^{\bar{\lambda}(\bar{f}_\tau(t)-1)}$. In the proof of (23) we exploit the convergence $\bar{\lambda} \to \lambda_1$ and $\bar{f}_\tau(t) \to f_\tau(t)$. In what follows we assume that m, n are so large that $\beta \le 2m/n \le 4\beta$.
Let us prove (23). We write

$$\mathbf{E}_1\Delta(t) = I_1 + I_2, \qquad I_1 = \mathbf{E}_1\Delta(t)\mathbb{I}_{\mathcal{D}}, \qquad I_2 = \mathbf{E}_1\Delta(t)\mathbb{I}_{\overline{\mathcal{D}}}.$$

Here $|I_2| \le 2\mathbf{P}_1(\overline{\mathcal{D}}) = 2\mathbf{P}(\overline{\mathcal{D}}) = o(1)$, by the law of large numbers. Next we estimate I_1. Combining the identity $\mathbf{E}_1\Delta(t) = \mathbf{E}_1 f(t)(e^\delta - 1)$ with the inequalities $|f(t)| \le 1$ and $|e^s - 1| \le |s|e^{|s|}$, we obtain

$$|I_1| \le \mathbf{E}_1|\delta|e^{|\delta|}\mathbb{I}_{\mathcal{D}} \le c_1\mathbf{E}_1|\delta|\mathbb{I}_{\mathcal{D}}. \tag{24}$$

Here we estimated $e^{|\delta|} \le e^{8\Lambda_1} =: c_1$ using the inequalities

$$|\delta| \le 2\bar{\lambda} + 2\lambda_1, \qquad \bar{\lambda} = Y_1(m/n)^{1/2}\hat{a}_1 \le 3\lambda_1.$$

We remark that the last inequality holds for $m/n \le 2\beta$ provided that the event $\mathcal{D}$ occurs. Finally, we show that $\mathbf{E}_1|\delta|\mathbb{I}_{\mathcal{D}} \le (c_2 + \lambda_1 c_3 + \lambda_1 c_4)\varepsilon + o(1)$. We first write

$$\delta = (\bar{f}_\tau(t) - 1)(\bar{\lambda} - \lambda_1) + (\bar{f}_\tau(t) - f_\tau(t))\lambda_1,$$

and estimate $|\delta| \leq 2|\bar{\lambda}-\lambda_1|+\lambda_1|\bar{f}_\tau(t)-f_\tau(t)|$. From the inequalities $|\hat{a}_1-a_1|<\varepsilon$ and $m/n \leq 2\beta$ we obtain

$$|\bar{\lambda}-\lambda_1| \leq Y_1|\hat{a}_1-a|(m/n)^{1/2}+Y_1a_1|(m/n)^{1/2}-\beta^{1/2}| \leq 2Y_1\beta^{1/2}\varepsilon+o(1).$$

Hence $\mathbf{E}_1|\bar{\lambda}-\lambda_1|\mathbb{I}_{\mathcal{D}} \leq c_2\varepsilon+o(1)$, where $c_2=2Y_1\beta^{1/2}$. We secondly show that

$$\mathbf{E}_1|\bar{f}_\tau(t)-f_\tau(t)|\mathbb{I}_{\mathcal{D}} \leq (c_3+c_4)\varepsilon+o(1).$$

To this aim we split

$$\bar{f}_\tau(t)-f_\tau(t)=\sum_{r\geq 0} e^{itr}(\bar{p}_r-p_r)=R_1-R_2+R_3,$$

and estimate separately the terms

$$R_1=\sum_{r\geq T} e^{itr}\bar{p}_r, \qquad R_2=\sum_{r\geq T} e^{itr}p_r, \qquad R_3=\sum_{0\leq r<T} e^{itr}(\bar{p}_r-p_r).$$

Here we denote $p_r=\mathbf{P}(\tau_1=r)$. The upper bound for R_2 follows by the choice of T

$$|R_2| \leq \sum_{r\geq T} p_r=\mathbf{P}(\tau_1\geq T)<\varepsilon.$$

Next, combining the identity $\bar{p}_r=(\hat{a}_1m)^{-1}\sum_{1\leq k\leq m}X_k\mathbb{I}_{\{\bar{\xi}_k=r\}}$ with the inequalities

$$|R_1| \leq (\hat{a}_1m)^{-1}\sum_{r\geq T}\sum_{1\leq k\leq m} X_k\mathbb{I}_{\{\bar{\xi}_k=r\}}=(\hat{a}_1m)^{-1}\sum_{1\leq k\leq m} X_k\mathbb{I}_{\{\bar{\xi}_k\geq T\}} \tag{25}$$

and $\hat{a}_1^{-1}\mathbb{I}_{\mathcal{D}} \leq 2a_1^{-1}$, we estimate

$$\mathbf{E}_1|R_1|\mathbb{I}_{\mathcal{D}} \leq 2a_1^{-1}\mathbf{E}_1(X_1\mathbb{I}_{\{\bar{\xi}_1\geq T\}}) \leq 2a_1^{-1}T^{-1}\mathbf{E}_1(X_1\bar{\xi}_1)=2a_1^{-1}a_2b_1\beta^{-1/2}\varepsilon.$$

Hence $\mathbf{E}_1|R_1|\mathbb{I}_{\mathcal{D}} \leq c_4\varepsilon$, where $c_4=2a_1^{-1}a_2b_1\beta^{-1/2}$.
Now we estimate R_3. We denote $p'_r=(\hat{a}_1/a_1)\bar{p}_r$ and observe that the inequality $|\hat{a}_1-a| \leq \varepsilon a_1$ implies $|\hat{a}_1a_1^{-1}-1| \leq \varepsilon$ and

$$|\sum_{0\leq r\leq T} e^{itr}(\bar{p}_r-p'_r)| \leq \varepsilon\sum_{0\leq r\leq T}\bar{p}_r \leq \varepsilon.$$

In the last inequality we use the fact that the probabilities $\{\bar{p}_r\}_{r\geq 0}$ sum up to 1. It follows now that

$$|R_3|\mathbb{I}_{\mathcal{D}} \leq \varepsilon+\sum_{0\leq r\leq T}|p'_r-p_r|.$$

Furthermore, observing that $\mathbf{E}_1p'_r=a^{-1}\mathbf{E}_1X_k\mathbb{I}_{\{\bar{\xi}_k=r\}}=p_r$, for $1\leq k\leq m$, we obtain

$$\mathbf{E}_1|p'_r-p_r|^2=m^{-1}\mathbf{E}_1|a^{-1}X_1\mathbb{I}_{\{\bar{\xi}_1=r\}}-p_r|^2 \leq m^{-1}a_1^{-2}\mathbf{E}X_1^2.$$

Hence, $\mathbf{E}_1|p'_r-p_r|=O(m^{-1/2})$. We conclude that

$$\mathbf{E}_1|R_3|\mathbb{I}_{\mathcal{D}} \leq \varepsilon+O(|T|m^{-1/2})=\varepsilon+o(1).$$

The case (iii). We start with introducing some notation. Denote $m/n = \beta_n$. Given $\varepsilon \in (0,1)$ introduce random variables

$$\gamma = \sum_{1 \le k \le m} \mathbb{I}'_k \lambda_{k1} \gamma_k, \qquad \gamma_k = X_k \beta_n^{-1/2} b_1 \mathbb{I}'_k, \qquad \mathbb{I}'_k = \mathbb{I}_{\{X_k \beta_n^{-1/2} b_1 < \varepsilon\}}.$$

Given X, Y, let $\tilde{\mathbb{I}}_1, \dots, \tilde{\mathbb{I}}_m$ be conditionally independent Bernoulli random variables with success probabilities

$$\tilde{\mathbf{P}}(\tilde{\mathbb{I}}_k = 1) = 1 - \tilde{\mathbf{P}}(\tilde{\mathbb{I}}_k = 0) = \gamma_k.$$

We assume that, given X, Y, the sequences $\{\mathbb{I}_k\}_{k=1}^m$, $\{\tilde{\mathbb{I}}_k\}_{k=1}^m$ and $\{\xi_{3k}\}_{k=1}^m$ are conditionally independent. Introduce random variables

$$L_4 = \sum_{1 \le k \le m} \mathbb{I}_k \xi_{3k}, \qquad L_5 = \sum_{1 \le k \le m} \mathbb{I}_k \mathbb{I}'_k \xi_{3k}, \qquad L_6 = \sum_{1 \le k \le m} \mathbb{I}_k \tilde{\mathbb{I}}_k.$$

Furthermore, we define the random variable L_7 as follows. We first generate X, Y. Then, given X, Y, we generate a Poisson random variable with the conditional mean value γ. The realized value of the Poisson random variable is denoted L_7. Thus, we have $\mathbf{P}(L_7 = r) = \mathbf{E} e^{-\gamma} \gamma^r / r!$, for $r = 0, 1, \dots$.

We note that L and L_3 have the same asymptotic distribution (if any), by (15), (16). Now we prove that L_3 converges in distribution to Λ_3. For this purpose we show that for any $\varepsilon \in (0,1)$

$$d_{TV}(L_3, L_4) = o(1), \qquad \mathbf{E}(L_4 - L_5) = o(1), \tag{26}$$

$$d_{TV}(L_5, L_6) \le a_2 b_1^2 \varepsilon, \qquad d_{TV}(L_6, L_7) = o(1), \tag{27}$$

$$\mathbf{E} e^{itL_7} - \mathbf{E} e^{it\Lambda_3} = o(1). \tag{28}$$

Let us prove (26), (27), (28). The first bound of (26) is obtained in the same way as the first bound of (15). To show the second bound of (26) we invoke the inequality

$$\tilde{\mathbf{E}}(L_4 - L_5) = \sum_{1 \le k \le m} (1 - \mathbb{I}'_k) \tilde{\mathbf{E}} \mathbb{I}_{k1} \tilde{\mathbf{E}} \xi_{3k} \le Y_1 b_1 m^{-1} \sum_{1 \le k \le m} (1 - \mathbb{I}'_k) X_k^2$$

and obtain

$$\mathbf{E}(L_4 - L_5) = \mathbf{E}\tilde{\mathbf{E}}(L_4 - L_5) \le b_1^2 \mathbf{E} X_1^2 \mathbb{I}_{\{X_1 \ge \varepsilon b_1^{-1} \beta_n^{1/2}\}} = o(1).$$

We note that the right hand side tends to zero since $\beta_n \to +\infty$.

Let us prove the first inequality of (27). Proceeding as in (18), (19) and using the identity $\tilde{\mathbb{I}}_k = \tilde{\mathbb{I}}_k \mathbb{I}'_k$ we write

$$\tilde{d}_{TV}(L_5, L_6) \le \sum_{1 \le k \le m} \mathbb{I}'_k \tilde{\mathbf{P}}(\mathbb{I}_k \ne 0) \tilde{d}_{TV}(\xi_{3k}, \tilde{\mathbb{I}}_k).$$

Next, we estimate $\mathbb{I}'_k \tilde{d}_{TV}(\xi_{3k}, \tilde{\mathbb{I}}_k) \le \gamma_k^2$, by LeCam's inequality (14), and invoke the inequality $\tilde{\mathbf{P}}(\mathbb{I}_k \ne 0) \le \lambda_{k1}$. We obtain

$$\tilde{d}_{TV}(L_5, L_6) \le \sum_{1 \le k \le m} \mathbb{I}'_k \lambda_{k1} \gamma_k^2 \le \varepsilon \sum_{1 \le k \le m} \mathbb{I}'_k \lambda_{k1} \gamma_k \le \varepsilon Y_1 b_1 \hat{a}_2.$$

Here we estimated $\gamma_k^2 \le \varepsilon \gamma_k$. Now the inequalities $d_{TV}(L_5, L_6) \le \mathbf{E}\tilde{d}_{TV}(L_5, L_6) \le a_2 b_1^2 \varepsilon$ imply the first relation of (27).

Let us prove the second relation of (27). In view of (5) it suffices to show that $\tilde{d}_{TV}(L_6, L_7) = o_P(1)$. For this purpose we write

$$\tilde{d}_{TV}(L_6, L_7) \leq \mathbb{I}_{\mathcal{A}_1} \tilde{d}_{TV}(L_6, L_7) + \mathbb{I}_{\overline{\mathcal{A}}_1},$$

where $\mathbb{I}_{\overline{\mathcal{A}}_1} = o_P(1)$, see (7), and estimate using LeCam's inequality (14)

$$\mathbb{I}_{\mathcal{A}_1} \tilde{d}_{TV}(L_6, L_7) \leq \mathbb{I}_{\mathcal{A}_1} \sum_{1 \leq k \leq m} \tilde{\mathbf{P}}^2(\mathbb{I}_k \tilde{\mathbb{I}}_k = 1)\mathbb{I}'_k \leq b_1^2 Y_1^2 m^{-1} \hat{a}_4 = o_P(1).$$

Here we used the fact that $\mathbf{E}X_1^2 < \infty$ implies $m^{-1}\hat{a}_4 = o_P(1)$.
Finally, we show (28). We write $\tilde{\mathbf{E}} e^{itL_7} = e^{\gamma(e^{it}-1)}$ and observe that

$$Y_1 b_1 a_2 - \gamma = o_P(1). \tag{29}$$

Furthermore, since for any real t the function $z \to e^{z(e^{it}-1)}$ is bounded and uniformly continuous for $z \geq 0$, we conclude that (29) implies the convergence

$$\mathbf{E}e^{itL_7} = \mathbf{E}e^{\gamma(e^{it}-1)} \to \mathbf{E}e^{Y_1 b_1 a_2 (e^{it}-1)} = \mathbf{E}e^{it\Lambda_3}.$$

It remains to prove (29). We write $Y_1 b_1 a_2 - \gamma = Y_1 b_1 (a_2 - \hat{a}_2) + Y_1 b_1 \hat{a}_2 - \gamma$ and note that $a_2 - \hat{a}_2 = o_P(1)$, by the law of large numbers, and

$$0 \leq \mathbf{E}(Y_1 b_1 \hat{a}_2 - \gamma) = b_1^2 \mathbf{E} X_1^2 \mathbb{I}_{\{X_1 \geq \varepsilon b_1^{-1} \beta_n^{1/2}\}} = o(1).$$

□

Acknowledgement. Research of M. Bloznelis was supported by the Research Council of Lithuania grant MIP-053/2011.

References

[1] A. D. Barbour and G. Reinert, The shortest distance in random multi-type intersection graphs, Random Structures and Algorithms 39 (2011), 179–209.

[2] M. Behrisch, Component evolution in random intersection graphs, The Electronical Journal of Combinatorics 14 (2007), #R17.

[3] S. Blackburn and S. Gerke, Connectivity of the uniform random intersection graph, Discrete Mathematics 309 (2009), 5130-5140.

[4] M. Bloznelis, Degree distribution of a typical vertex in a general random intersection graph, Lithuanian Mathematical Journal 48 (2008), 38–45.

[5] M. Bloznelis, A random intersection digraph: Indegree and outdegree distributions, Discrete Mathematics 310 (2010), 2560–2566.

[6] M. Bloznelis, The largest component in an inhomogeneous random intersection graph with clustering, The Electronic Journal of Combinatorics 17(1) (2010), R110.

[7] M. Bloznelis, Degree and clustering coefficient in sparse random intersection graphs, to appear in The Annals of Applied Probability.

[8] M. Bloznelis, V. Kurauskas, Clustering function: a measure of social influence, http://arxiv.org/abs/1207.4941

[9] M. Bradonjic, A. Hagberg, N. W. Hengartner, A. G. Percus, Component Evolution in General Random Intersection Graphs, The 7th Workshop on Algorithms and Models for the Web Graph, WAW2010. Lecture Notes in Computer Science (Springer-Verlag, Berlin, 2010), Vol. 6516, pp. 36-49.

[10] T. Britton, M. Deijfen, M. Lindholm, and N. A. Lageras, Epidemics on random graphs with tunable clustering. J. Appl. Prob. 45 (2008), 743–756.

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

[12] L. Eschenauer and V. D. Gligor, A key-management scheme for distributed sensor networks, in: Proceedings of the 9th ACM Conference on Computer and Communications Security (2002), 41–47.

[13] E. Godehardt and J. Jaworski, Two models of random intersection graphs and their applications, *Electronic Notes in Discrete Mathematics* **10** (2001), 129–132.

[14] J. Jaworski, M. Karoński, and D. Stark, The degree of a typical vertex in generalized random intersection graph models, Discrete Mathematics 306 (2006), 2152–2165.

[15] J. Jaworski and D. Stark, The vertex degree distribution of passive random intersection graph models, *Combinatorics, Probability and Computing* **17** (2008), 549–558.

[16] M. Karoński, E. R. Scheinerman, and K. B. Singer-Cohen, On random intersection graphs: The subgraph problem, Combinatorics, Probability and Computing 8 (1999), 131–159.

[17] S. Nikoletseas, C. Raptopoulos, and P. G. Spirakis, The existence and efficient construction of large independent sets in general random intersection graphs, in ICALP (2004), J. Daz, J. Karhumki, A. Lepist, and D. Sannella, Eds., vol. 3142 of Lecture Notes in Computer Science, Springer, pp. 10291040.

[18] K. Rybarczyk, Diameter, connectivity, and phase transition of the uniform random intersection graph, Discrete Mathematics 311 (2011), 1998–2019.

[19] K. Rybarczyk, The degree distribution in random intersection graphs, in: Studies in Classification, Data Analysis and Knowledge Organization, Springer, Berlin–Heidelberg–New York, (2012), 291–299.

[20] Y. Shang, Degree distributions in general random intersection graphs, The Electronical Journal of Combinatorics 17 (2010), #R23.

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

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

[23] J. M. Steele, Le Cam's inequality and Poisson approximations, The American Mathematical Monthly 101 (1994), 48–54.