J. Appl. Probab. **59,** 755–776 (2022)
*doi:*10.1017/jpr.2021.84

ON LOCAL WEAK LIMIT AND SUBGRAPH COUNTS FOR SPARSE RANDOM GRAPHS

VALENTAS KURAUSKAS,* *Faculty of Mathematics and Informatics, Vilnius University*

Abstract

We use an inequality of Sidorenko to show a general relation between local and global subgraph counts and degree moments for locally weakly convergent sequences of sparse random graphs. This yields an optimal criterion to check when the asymptotic behaviour of graph statistics, such as the clustering coefficient and assortativity, is determined by the local weak limit.

As an application we obtain new facts for several common models of sparse random intersection graphs where the local weak limit, as we see here, is a simple random clique tree corresponding to a certain two-type Galton–Watson branching process.

Keywords: Local weak limit; subgraph count; clique tree; random intersection graph

2020 Mathematics Subject Classification: Primary 60C05; 05C80
Secondary 05C82

1. Introduction

A *rooted graph* is a pair (H, v) where H is a graph and $v \in V(H)$ is a distinguished vertex called the *root*. We often use only the symbol H to denote (H, v); in this case we write $\mathrm{root}(H) = v$. For a graph G and its vertex v, let B_r be the function that maps (G, v) to the the rooted graph (H, v), where H is the subgraph induced on the vertices of G with distance from v at most r. We simplify $B_r(G, v) = B_r((G, v))$ for B_r and other functions on rooted graphs.

A graph is *locally finite* if the degree of each of its vertices is finite. Let $\cong$ denote the isomorphism relation between connected rooted graphs which preserves the root. Let $(\mathcal{G}_*, d_{\mathrm{loc}})$ be the space of rooted connected locally finite graphs with equivalence relation $\cong$ and distance

$$d_{\mathrm{loc}}(G_1, G_2) = 2^{-\sup\{r\,:\,B_r(G_1)\cong B_r(G_2)\}}.$$

Consider a sequence of finite graphs $\{G_n, n = 1, 2, \ldots\}$. In this paper we assume $|V(G_n)| \geq 1$ for $n \geq 1$. Let v_n^* be a uniformly random vertex from $V(G_n)$. The component of G_n containing v_n^* together with root v_n^* induces a Borel measure μ_n on $(\mathcal{G}_*, d_{\mathrm{loc}})$ for each n. Let μ^* be another Borel measure on $(\mathcal{G}_*, d_{\mathrm{loc}})$, and let G^* denote a random element with law μ^*. (Without loss of generality we assume that all random objects we define in the paper are random elements in a single probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the specified laws; the integration $\mathbb{E}$ is over Ω.) Following Benjamini and Schramm [5], Aldous, Lyons, and Steele, and other authors

Received 6 December 2020; revision received 18 September 2021.
* Postal address: Akademijos 4, LT-08412 Vilnius, Lithuania. Email address: valentas@gmail.com

[2, 36, 37], we say that G^* is the *local weak limit* of $\{(G_n, v_n^*)\}$ and write $(G_n, v_n^*) \xrightarrow{d} G^*$ if and only if the measures μ_n converge weakly to μ^*: for each continuous bounded function $f : (\mathcal{G}_*, d_{\text{loc}}) \to \mathbb{R}$,

$$\mathbb{E}f(G_n, v_n^*) \to \mathbb{E}f(G^*). \tag{1.1}$$

Here and below all limits are as $n \to \infty$, unless stated otherwise. Since $(\mathcal{G}_*, d_{\text{loc}})$ is separable and complete [1], a standard argument (e.g. Theorem 2.3 of [8]) shows that $(G_n, v_n^*) \xrightarrow{d} G^*$ if and only if, for each non-negative integer r and each rooted connected graph H,

$$\mathbb{P}(B_r(G_n, v_n^*) \cong H) \to \mathbb{P}(B_r(G^*) \cong H).$$

We focus on models of random graphs with bounded average degree. Among others, the inhomogeneous random graph model of Bollobás, Janson, and Riordan [22] and the preferential attachment model (see Berger, Borgs, Chayes, and Saberi [7]) have been shown to have a weak limit (in an explicit form). Recently such a limit was also shown to exist for random planar graphs [46]. The local weak limit, if it exists, yields a lot of information about the asymptotics of various graph parameters; see e.g. [6], [19], [22], [23], [36], and [44].

The present contribution consists of a general result, Theorem 2.1, relating the asymptotics of subgraph counts with the local weak limit, and its application in the area of random intersection graphs.

The structure of the paper is as follows. In Section 2 we present and prove Theorem 2.1. In a separate result, Theorem 3.1 of Section 3, we determine the (very simple) local weak limit of several popular random intersection graph models. Combining this with Theorem 2.1 and using the fact that many important graph parameters can be expressed in terms of small subgraph counts, we obtain a number of previous and some new results for this type of models; see Section 4. The same method works for any sparse random graph model where we have weak local convergence (see e.g. Section 4.2).

The first manuscript of this paper was completed and posted to arXiv in 2015 [40]. Since then there has appeared some work in a similar general direction, including, for example, [47], unaware of the very general Theorem 2.1. A recent book in preparation [33] also devotes a chapter to weak limits as a general technique to study real world networks. The present version of the paper fixes several minor errors and omissions and has an updated literature list.

2. Local weak limit and subgraph counts

In Section 7 of [19], Bollobás, Janson, and Riordan remark that the local weak limit does not always determine the global subgraph count asymptotics; see also Example 2.1 below. They propose an extra condition of ‘exponentially bounded tree counts’. Our main result is that a simple condition on the degree moment is sufficient and, in general, necessary.

A homomorphism from a graph H to a graph G is a mapping from $V(H)$ to $V(G)$ that maps adjacent vertices in H to adjacent vertices in G. Let $\text{emb}(H, G)$ denote the number of embeddings (injective homomorphisms) from H to G. For a rooted graph H' let $\text{emb}'(H', G, v)$ denote the number of embeddings from H' to G that map $\text{root}(H')$ to v. Let $\mathcal{R}(H)$ denote the set of all $|V(H)|$ possible rooted graphs obtained from a graph H. Finally, let $d_G(v)$ denote the degree of vertex v in G.

Theorem 2.1. *Let $h \geq 2$ be an integer, let $\{G_n, n = 1, 2, \ldots\}$ be a sequence of graphs, such that $n_1 = n_1(n) = |V(G_n)| \to \infty$ and $n_1 \geq 1$, let v_n^* be chosen uniformly at random from $V(G_n)$, and suppose $(G_n, v_n^*) \xrightarrow{d} G^*$. Write $d_n = d_{G_n}(v_n^*)$, $d^* = d_{G^*}(r^*)$, where $r^* = \mathrm{root}(G^*)$, and assume $\mathbb{E}(d^*)^{h-1} < \infty$. Then the following statements are equivalent:*

(i) $\mathbb{E}d_n^{h-1} \to \mathbb{E}(d^*)^{h-1}$,

(ii) d_n^{h-1} *is uniformly integrable,*

(iii) *for any connected graph H on h vertices and any $H' \in \mathcal{R}(H)$,*

$$n_1^{-1}\,\mathrm{emb}(H, G_n) \to \mathbb{E}\,\mathrm{emb}'(H', G^*, r^*).$$

The above theorem provides a sufficient condition for the continuous but not necessarily bounded function $f_H : (\mathcal{G}_*, d_{\mathrm{loc}}) \to \mathbb{R}$ defined by $f_H(G, v) = \mathrm{emb}'(H, G, v)$ to satisfy (1.1). It is easy to construct weakly convergent sequences for which (i)–(iii) fail to hold.

Example 2.1. Let (G_n, v_n^*) be as in Theorem 2.1 and assume $|V(G_n)| = n$. Let (G'_n, v_n^*) be obtained by merging edges of a clique on a subset S_n of G_n. If $|S_n| = \Omega(n^{1/h})$ and $|S_n| = \mathrm{o}(n)$ then $(G'_n, v_n^*) \xrightarrow{d} G^*$, but (i)–(iii) do not hold for G'_n.

The proof of our theorem follows from the next basic but not widely known result of Sidorenko [45].

Theorem 2.2. (Sidorenko, 1994.) *Let H be a connected graph on h vertices. Then, for any graph G,*

$$\mathrm{hom}(H, G) \leq \mathrm{hom}(K_{1,h-1}, G) = \sum_{v \in V(G)} d_G(v)^{h-1}.$$

Here $\mathrm{hom}(H, G)$ is the number of homomorphisms from H to G, and $K_{1,s}$ is the complete bipartite graph with part sizes 1 and s. A special case where H is a path has been rediscovered in [27]; see also [24].

Below, in Section 4, we restate Theorem 2.1 in a random setting and demonstrate how it implies general results on network statistics expressible through subgraph counts such as the clustering and assortativity coefficients.

Recall that a sequence of random variables $\{X_n, n = 1, 2, \ldots\}$ is uniformly integrable if $\sup_{a\to\infty} \sup_n \mathbb{E}|X_n|\mathbb{I}_{|X_n|>a} = 0$, and equivalently if $\mathbb{E}|X_n|\mathbb{I}_{|X_n|>\omega_n} \to 0$ for any $\omega_n \to \infty$. A basic fact (see e.g. [9, pp. 31–32]) is as follows.

Lemma 2.1. *Suppose random variables $X^*, X_n, n = 1, 2, \ldots$ are non-negative, integrable and X_n converges to X^* in distribution as $n \to \infty$. Then $\{X_n\}$ is uniformly integrable if and only if $\mathbb{E}X_n \to \mathbb{E}X^*$.*

Proof of Theorem 2.1. In the proof denote $G = G_n$ and $v^* = v_n^*$.

(i) ⇔ (ii) We see that d_n converges in distribution to d^* by (1.1). Thus d_n^{h-1} converges in distribution to $(d^*)^{h-1}$ and the proof follows by Lemma 2.1.

(i) ⇒ (iii) Suppose (i) holds. Fix any connected graph H with $|V(H)| = h$. Let r be the diameter of H. Write $b_j(G, v) = |B_j(G, v)|$ and $b_j^* = |B_j(G^*)|$. Note that $\mathbb{P}(b_j^* = \infty) = 0$ for $j = 0, 1, \ldots$ since G^* is locally finite. For a rooted graph H' denote

$$X(H') = \mathrm{emb}'(H', G, v^*) \quad \text{and} \quad X^*(H') = \mathrm{emb}'(H', G^*, r^*).$$

Using Sidorenko's theorem, Theorem 2.2, for any $H' \in \mathcal{R}(H)$,

$$\begin{aligned} \mathbb{E}X(H') &= n_1^{-1}\mathrm{emb}(H, G) \le n_1^{-1}\,\mathrm{hom}(H, G) \\ &\le n_1^{-1}\,\mathrm{hom}(K_{1,h-1}, G) = \mathbb{E}d_n^{h-1} \to \mathbb{E}(d^*)^{h-1} < \infty. \end{aligned} \tag{2.1}$$

Next, we have $X(H') \to X^*(H)$ in distribution for each $H' \in \mathcal{R}(H)$ (apply (1.1) to the continuous and bounded function $f(G, v) = \mathbb{I}_{\mathrm{emb}'(G,H',v)=k}$). Therefore, by Fatou's lemma and (2.1),

$$\mathbb{E}X^*(H') \le \liminf \mathbb{E}X(H') < \infty.$$

Let $\epsilon \in (0, 1)$. Since $\mathbb{E}(d^*)^{h-1}$ and $\mathbb{E}X^*(H')$ are finite, we can find a $t > 0$ such that

$$\begin{aligned} &\mathbb{E}(d^*)^{h-1}\mathbb{I}_{d^*>t} \le \mathbb{E}(d^*)^{h-1}\mathbb{I}_{b^*_{r+1}>t} < \epsilon \quad \text{and} \\ &\mathbb{E}X^*(H')\mathbb{I}_{b^*_{r+1}>t} < \epsilon \quad \text{for each } H' \in \mathcal{R}(H). \end{aligned}$$

Pick $s \ge t$ large enough that $\mathbb{P}(b^*_{r+1} > s) \le 0.5\epsilon t^{-(h+r-1)}$. By Lemma 2.1 and (1.1), for each $H' \in \mathcal{R}(H)$,

$$\mathbb{E}d_n^{h-1}\mathbb{I}_{d_n\le t} \to \mathbb{E}(d^*)^{h-1}\mathbb{I}_{d^*\le t}, \tag{2.2}$$

$$\mathbb{E}X(H')\mathbb{I}_{b_{r+1}(G,v^*)\le s} \to \mathbb{E}X^*(H')\mathbb{I}_{b^*_{r+1}\le s} \ge \mathbb{E}X^*(H') - \epsilon, \tag{2.3}$$

$$\mathbb{E}X(H')\mathbb{I}_{b_{r+1}(G,v^*)\in(t,s]} \to \mathbb{E}X^*(H')\mathbb{I}_{b^*_{r+1}\in(t,s]} \le \epsilon, \tag{2.4}$$

$$\mathbb{P}(b_{r+1}(G, v^*) > s) \to \mathbb{P}(b^*_{r+1} > s) \le 0.5\epsilon t^{-(h+r-1)}. \tag{2.5}$$

Define subsets of $V(G)$:

$$R_1 := \{v : d_G(v) > t\}, \quad R_2 := \{v : b_{r+1}(G, v) > s\}.$$

We call an embedding σ of H into G *bad* if its image shares a vertex with $R_1 \cup R_2$. Denote the set of all bad embeddings by $\mathcal{X}_{\mathrm{bad}}$. Note that for $H' \in \mathcal{R}(H)$

$$0 \le \mathrm{emb}(H, G) - n_1\mathbb{E}X(H')\mathbb{I}_{b_{r+1}(G,v^*)\le s} \le |\mathcal{X}_{\mathrm{bad}}|. \tag{2.6}$$

Let $\mathcal{X}_1$ be the set of all embeddings σ whose image intersects both R_1 and $V \setminus (R_1 \cup R_2)$. Let $\mathcal{X}_2 = \mathcal{X}_{\mathrm{bad}} \setminus \mathcal{X}_1$. By Theorem 2.2, the number of bad embeddings which have the image entirely contained in R_1 is

$$\begin{aligned} \mathrm{emb}(H, G[R_1]) &\le \mathrm{hom}(H, G[R_1]) \\ &\le \sum_{v\in R_1} d_G(v)^{h-1} \\ &= n_1\left(\mathbb{E}d_n^{h-1} - \mathbb{E}d_n^{h-1}\mathbb{I}_{d_n\le t}\right) \\ &\le n_1\mathbb{E}d_n^{h-1} - n_1\mathbb{E}(d^*)^{h-1} + \epsilon n_1 + \mathrm{o}(n_1) \\ &\le \epsilon n_1 + \mathrm{o}(n_1). \end{aligned} \tag{2.7}$$

Here the last two inequalities follow by (i) and (2.2). Let $v \in V \setminus (R_1 \cup R_2)$ be a vertex in the image of an embedding in $\mathcal{X}_1$. By the definition of R_1 and R_2, $b_{r+1}(G, v) \in (t, s]$. So using (2.4),

$$|\mathcal{X}_1| \le n_1 \sum_{H'\in\mathcal{R}(H)} \mathbb{E}X(H')\mathbb{I}_{b_{r+1}(G,v^*)\in(t,s]} \le h\epsilon n_1 + \mathrm{o}(n_1).$$

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

Now consider a subgraph H_σ of G, $H_\sigma \cong H$ corresponding to an embedding $\sigma \in \mathcal{X}_2$. H_σ cannot have an edge in $E_1 = \{xy \in G : x \in R_1, y \in V(G) \setminus (R_1 \cup R_2)\}$, otherwise σ would be an element of $\mathcal{X}_1 = \mathcal{X}_{\text{bad}} \setminus \mathcal{X}_2$. So $V(H_\sigma)$ is contained in $R_1 \cup Q$, where

$$Q = \bigcup_{v \in R_2} V(B_r(G - E_1, v)) \setminus R_1.$$

Note that since each vertex in Q has degree at most t, $|Q| \le 2|R_2|t^r$. By Theorem 2.2,

$$|\mathcal{X}_2| \le \sum_{v \in R_1} d_G(v)^{h-1} + \sum_{v \in Q} d_G(v)^{h-1}. \tag{2.8}$$

For the second term we have by (2.5)

$$\sum_{v \in Q} d_G(v)^{h-1} \le |Q|t^{h-1} \le 2|R_2|t^r t^{h-1} \le \epsilon n_1 + \mathrm{o}(n_1). \tag{2.9}$$

Combining (2.7), (2.8), and (2.9), we obtain $|\mathcal{X}_2| \le 2\epsilon n_1 + \mathrm{o}(n_1)$. We have proved

$$|\mathcal{X}_{\text{bad}}| = |\mathcal{X}_1| + |\mathcal{X}_2| \le (h+2)\epsilon n_1 + \mathrm{o}(n_1).$$

Since the proof holds for arbitrarily small ϵ, we see that $n_1^{-1}|\mathcal{X}_{\text{bad}}| \to 0$. Thus (iii) follows using (2.3) and (2.6).

(iii) $\Rightarrow$ (i) Write $(x)_p = x(x-1)\cdots(x-p+1)$. Then (iii) applied to $H = K_{1,h-1}$ yields $\mathbb{E}(d_n)_{h-1} \to \mathbb{E}(d^*)_{h-1}$, while $(G, v^*) \xrightarrow{d} G^*$ shows that $(d_n)_{h-1} \to (d^*)_{h-1}$ in distribution. Thus $(d_n)_{h-1}$ is uniformly integrable by Lemma 2.1. This implies that for any $j = 1, 2, \ldots, h-1$ $(d_n)_j \le (d_n)_{h-1}$ is uniformly integrable, so by Lemma 2.1 again $\mathbb{E}(d_n)_j \to \mathbb{E}(d^*)_j$. Using $S(h-1, j)$ to denote Stirling numbers of the second kind,

$$\mathbb{E}(d_n)^{h-1} = \sum_{j=1}^{h-1} S(h-1, j)\mathbb{E}(d_n)_j \to \sum_{j=1}^{h-1} S(h-1, j)\mathbb{E}(d^*)_j = \mathbb{E}(d^*)^{h-1}.$$ □

The next fact is simple and known (see Lemma 9.3 of [22]), but we include a proof for completeness.

Lemma 2.2. *Suppose $(G_n, v_n^*) \xrightarrow{d} G^*$ and the degree d_n of a uniformly random vertex v_n^* from $V(G_n)$ is uniformly integrable. Write $n_1 = n_1(n) = |V(G_n)|$. Let G_n' be obtained from G_n by adding or removing edges incident to a set $S_n \subseteq V(G_n)$ of size $\mathrm{o}(n_1)$. Then $(G_n', v_n^*) \xrightarrow{d} G^*$.*

Proof of Example 2.1. It is straightforward that the uniform integrability condition (ii) fails for G_n', so the other two conditions also fail by Theorem 2.1. The fact that G_n and G_n' have the same local weak limit G^* follows from Lemma 2.2. □

Proof of Lemma 2.2. Let N_n denote the set of vertices in $V(G_n) \setminus S_n$ which have a neighbour in S_n.

Claim 2.3. *For any $\epsilon \in (0, 1)$ there are $\delta > 0$, $n_0 > 0$ such that if $n \ge n_0$ and $0 < |S_n| < \delta n_1$ then $|N_n| \le \epsilon n_1$.*

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

Proof. Let δ and n_0 be such that $\delta < \epsilon$, $\mathbb{E}d_n\mathbb{I}_{d_n>0.5\epsilon\delta^{-1}} < 0.5\epsilon$ for all $n \geq n_0$. Assume that $|N_n| > \epsilon n_1$ for some $n \geq n_0$. Write $d(v) = d_{G_n}(v)$. We have

$$\begin{aligned} n_1\mathbb{E}d_n\mathbb{I}_{d_n>0.5\epsilon\delta^{-1}} &\geq \sum_{v\in S_n} d(v)\mathbb{I}_{d(v)>0.5\epsilon\delta^{-1}} \\ &\geq \sum_{v\in S_n}\big(d(v) - 0.5\epsilon\delta^{-1}\big)\mathbb{I}_{d(v)>0.5\epsilon\delta^{-1}} \\ &\geq |S_n|\left(|S_n|^{-1}\sum_{v\in S_n} d(v) - 0.5\epsilon\delta^{-1}\right) \\ &\geq \epsilon n_1 - 0.5\epsilon n_1 \\ &\geq 0.5\epsilon n_1, \end{aligned}$$

which is a contradiction. Here we used Jensen's inequality and the assumption $\sum_{v\in S} d(v) \geq |N_n| > \epsilon n_1$. □

Now fix any positive integer r. As is done in [22], we apply the above claim r times to get that the set $N_n^{(r)}$ of vertices at distance at most r from S_n in G_n has size $o(n_1)$. Now $(G_n', v_n^*) \xrightarrow{d} G^*$ follows since

$$\mathbb{P}(B_r(G_n, v_n^*) \cong B_r(G_n', v_n^*)) \geq \mathbb{P}\big(v_n^* \notin N_n^{(r)}\big) = 1 - o(1). \qquad \square$$

3. Uncorrelated random clique trees

Random intersection graphs were introduced in [38] and received some attention as a potential model for large empirical networks with clustering; see e.g. the survey papers [16] and [17]. We show that in the regime that yields sparse graphs with a positive clustering coefficient in such models the weak limit is very specific, namely it is an uncorrelated random clique tree, defined formally below.

Let $H = (V^1, V^2, E)$ be a bipartite graph. The *intersection graph* $G = G(H)$ of H is the graph on the vertex set $V(G) = V^1$ with edges

$$E(G) = \big\{uv : \exists w \in V^2 \text{ such that } uw, wv \in H\big\},$$

where $e \in H$ is shorthand for $e \in E(H)$. An intersection graph of a random bipartite graph H is called a *random intersection graph*. It will be convenient to assume that V^i consists of the first n_i elements of a countable set $\mathcal{V}^i$, where $\mathcal{V}^1 \cap \mathcal{V}^2 = \emptyset$. The set V^2 is often called the set of *attributes*. (The names V and W are often used in the literature for V^1 and V^2.) We will call the elements of $\mathcal{V}^i$ *vertices of type i*. For $v \in V^i$ we denote $S_v = \Gamma_H(v)$, $X_v = |S_v|$, where $\Gamma_H(x)$ is the set of neighbours of v in the graph H. Sometimes we will want to stress the type of v in the notation. Since $V^i = \big\{v_1^{(i)}, v_2^{(i)}, \ldots\big\}$ consists of the first n_i vertices of $\mathcal{V}_i$, for $v = v_j^{(i)}$ we will set $X_v^{(i)} := X_v$ and $S_v^{(i)} := S_v$. We will let $X \sim Y$ denote the fact that X and Y have the same distribution.

Many different variants of the random bipartite graph H have been studied; see e.g. the survey papers [16] and [17].

- The *active* random intersection graph: each $v \in V^1$ independently chooses $X_v^{(1)}$ from a distribution P on $\{0, \ldots, n_2\}$, then draws a uniformly random subset S_v^1 of size $X_v^{(1)}$ of

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

its neighbours from V^2 (independently of other vertices). A special case is the *binomial* random intersection graph.

- The *passive* random intersection graph: each $v \in V^2$ independently chooses $X_v^{(2)}$ from a distribution P on $\{0, \ldots, n_1\}$, then draws a uniformly random subset S_v^2 of size $X_v^{(2)}$ of its neighbours from V^1 (independently of other vertices).
- The *inhomogeneous* random intersection graph $G^{\text{inhomog}}(n_1, n_2, \xi^{(1)}, \xi^{(2)})$: the vertices $v \in V^i$ are independently assigned random non-negative weights $\xi_v^{(i)} \sim \xi^{(i)}$. Given the weights, edges vw appear in H independently with probability

$$\min\left(\frac{\xi_v^{(1)}\xi_w^{(2)}}{\sqrt{n_1 n_2}}, 1\right).$$

- We will also consider random intersection graphs $G^{\text{conf}}(d_1, d_2)$ based on the *configuration model*: see [32], [34], and [48]. Let $d_1 = \{d_{1,u}, u \in V^1\}$ and $d_2 = \{d_{2,v}, v \in V^2\}$ be sequences of non-negative integers indexed by V_1 and V_2 respectively such that $\sum_u d_{1,u} = \sum_v d_{2,v}$. The random bipartite multigraph $H^{\text{conf}}(d_1, d_2)$ with parts (V^1, V^2) of sizes n_1 and n_2 is obtained as follows. Distribute the total number of $2\sum d_{1,u}$ half-edges among the vertices of $V^1 \cup V^2$ so that the jth vertex of part i, $v = v_j^{(i)}$, receives $d_{i,v}$ half-edges. Pick a uniformly random perfect matching between the half-edges of parts V^1 and V^2. In the bipartite graph, add an edge between u and v whenever a half-edge from u is matched with a half-edge from v (we allow multi-edges).

Usually (see e.g. [12]) the above models yield random graphs with a linear number of edges and a clustering coefficient bounded away from zero only if $n_2/n_1 = \Theta(1)$. Therefore we will assume $n_2/n_1 = \Theta(1)$ in this paper.

Let μ be the distribution of a random variable Z on $[0, \infty)$ with $0 < \mathbb{E}Z < \infty$. We let Z^* denote a random variable with the *size-biased* distribution

$$\mu^*(A) = (\mathbb{E}Z)^{-1} \int_A t \, \mathrm{d}\mu(t)$$

for any Borel set A. If Z is integer-valued, then $\mathbb{P}(Z^* = k) = (\mathbb{E}Z)^{-1} k \mathbb{P}(Z = k)$. (We follow the star notation of other authors; see e.g. [4] and [36]. We also use symbols such as G^*, d^*, v^* to denote objects unrelated to size-biased random variables; the actual meaning should be clear from the context.)

Given two random variables D_1, D_2 on $\{0, 1, 2, \ldots\}$ with $\mathbb{E}D_1, \mathbb{E}D_2 \in (0, \infty)$, define a multi-type Galton–Watson process $\mathcal{T} = \mathcal{T}(D_1, D_2)$ as follows. $S(0)$ consists of a single root node $r = \text{root}(\mathcal{T})$. The root r has a set $S(1)$ of offspring, where $|S(1)| \sim D_1$. For each $k \geq 1$, $S(k+1)$ consists of the offspring of the nodes in $S(k)$. Given $|S(k)|$, the number of offspring of each node in $S(k)$ is independent and distributed as $D_{i(k)}^* - 1$. Here $i(k) = 2$ if k is odd and $i(k) = 1$ otherwise. We call $S(k)$, the set of vertices at distance k from the root, the *generation* k of $\mathcal{T}$. A corresponding random tree, also denoted by $\mathcal{T}$, is a graph on the vertex set $\cup_k S(k)$ with edges $\{uv : v \text{ is an offspring of } u\}$ and root r. Consider $\mathcal{T}$ as a bipartite graph with parts (V^1, V^2), where V^1 and V^2 consists of all nodes in generations $0, 2, \ldots$ and $1, 3, \ldots$ respectively. We define the *uncorrelated random clique tree* $G_{\mathcal{T}}$ to be the intersection graph of $\mathcal{T}$ rooted at r.

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

For a finite (random) sequence A, we write $X \in_u A$ to denote the fact that X is chosen uniformly at random from all the elements of A (given A). For random variables $Z, Z_1, Z_2, \ldots$, we let $Z_n \xrightarrow{d} Z$ denote the fact that Z_n converges in distribution to Z.

Let H be a rooted connected graph. For a (multi-)graph G of size $n_1 \geq 1$, write $p_r(G, H) = n_1^{-1}|\{v \in V(G) : B_r(G, v) \cong H\}|$. Let $\{G_n, n = 1, 2, \ldots\}$ be a sequence of finite random graphs with $|V(G_n)| \geq 1$, let $v_n^* \in_u V(G_n)$, and let G^* be a random graph on $(\mathcal{G}_*, d_{\mathrm{loc}})$. (For multigraphs we define $G_1 \cong G_2$ if and only if there are bijections $\phi_1 : V(G_1) \to V(G_2)$ and $\phi_2 : E(G_1) \to E(G_2)$ such that ϕ_1 maps the endpoints of e to the endpoints of $\phi_2(e)$ for each edge $e \in G_1$, and $\phi_1(\mathrm{root}(G_1)) = \mathrm{root}(G_2)$.) We write $\mathcal{L}((G_n, v_n^*) \mid G_n) \xrightarrow{p} \mathcal{L}(G^*)$ as $n \to \infty$ if, for each non-negative integer r and each rooted connected graph H,

$$p_r(G_n, H) \xrightarrow{p} \mathbb{P}(B_r(G^*) \cong H). \tag{3.1}$$

As observed in the recent literature [46], this is equivalent to the convergence of the conditional random measures $\mathcal{L}((G_n, v_n^*) \mid G_n)$ to the fixed measure $\mathcal{L}(G^*)$ in probability, also known as *quenched* convergence. That is, consider the space of Borel measures on $(\mathcal{G}_*, d_{\mathrm{loc}})$ with the Lévy–Prokhorov metric π. Then $\pi(\mathcal{L}((G_n, v_n^*) \mid G_n), \mathcal{L}(G^*)) \xrightarrow{p} 0$ if and only if (3.1) holds for each r and each H as above. This can be seen using an argument similar to (iv) on page 72 of Billingsley [9]. While this equivalence reduces many questions related to local weak limits to the classical theory for separable metric spaces, in this paper it is only used to justify our notation.

Theorem 3.1. *Let $\{G_n\}$ be a sequence of random intersection graphs where the underlying bipartite graphs are $H_n = (V^1, V^2, F)$ with $V^1 = V^1(n)$, $V^2 = V^2(n)$ and $F = F(n)$. For $i = 1, 2$ write $v_i^* = v_i^*(n)$, where $v_i^*(n) \in_u V^i$, $n_i = n_i(n) = |V^i|$ and $X^{(i)} = X^{(i)}(n) = X_{v_i^*}$.*

Suppose $\{n_1\}$, $\{n_2\}$ are sequences of positive integers, such that $n_1, n_2 \to \infty$, $n_2/n_1 \to \beta \in (0, \infty)$ and

(i) *either G_n, $n = 1, 2, \ldots$ is an active random intersection graph and there is a random variable D_1 with $\mathbb{E}D_1 \in (0, \infty)$ such that $\mathbb{E}X^{(1)} \to \mathbb{E}D_1$ and*

$$X^{(1)} \xrightarrow{d} D_1; \tag{3.2}$$

(ii) *or G_n, $n = 1, 2, \ldots$ is a passive random intersection graph and there is a random variable D_2 with $\mathbb{E}D_2 \in (0, \infty)$ such that $\mathbb{E}X^{(2)} \to \mathbb{E}D_2$ and*

$$X^{(2)} \xrightarrow{d} D_2; \tag{3.3}$$

(iii) *or $G_n = G^{\mathrm{inhomog}}(n_1, n_2, \xi^{(1)}, \xi^{(2)})$, $n = 1, 2, \ldots$, such that for $i = 1, 2$ $0 < \mathbb{E}\xi^{(i)} < \infty$ and $\xi^{(i)}$ does not depend on n;*

(iv) *or $G_n = G^{\mathrm{conf}}(d_1, d_2)$, $n = 1, 2, \ldots$, where $d_1 = d_1(n)$, $d_2 = d_2(n)$ are non-random and for $i = 1, 2$ we have $\mathbb{E}d_{1, v_i^*} \to \mathbb{E}D_i$ and $d_{1, v_i^*} \xrightarrow{d} D_i$.*

Then both (3.2) and (3.3) hold and $\mathcal{L}((G_n, v_1^*) \mid G_n) \xrightarrow{p} \mathcal{L}(G_{\mathcal{T}})$ with $\mathcal{T} = \mathcal{T}(D_1, D_2)$ where any D_i that is not defined here is defined in Remark 3.1.

The proof is available in the arXiv version of this paper [40, Appendix A]. (In case (iv) a previous version (v2) of [40] stated an analogous result for random sequences d_1, d_2.

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

We simplified the condition to match, for example, [3] and [34]. The previous result follows by a simple technical argument.)

Recall that given a non-negative random variable X, a mixed Poisson random variable with parameter X attains value k with probability $\mathbb{E}\, e^{-X}X^k(k!)^{-1}$ for $k = 0, 1, \ldots$. We denote this distribution by $\mathrm{Po}\,(X)$.

Remark 3.1. (See also [11] and [12].) In case (i) we have $D_2 \sim \mathrm{Po}\,(\beta^{-1}\mathbb{E}D_1)$, in case (ii) we have $D_1 \sim \mathrm{Po}\,(\beta\mathbb{E}D_2)$, and in case (iii) we have $D_1 \sim \mathrm{Po}\big(\beta^{1/2}\xi^{(1)}\mathbb{E}\xi^{(2)}\big)$, $D_2 \sim \mathrm{Po}\big(\beta^{-1/2}\xi^{(2)}\mathbb{E}\xi^{(1)}\big)$. Thus in (i)–(iv) $\beta\mathbb{E}D_2 = \mathbb{E}D_1$.

Remark 3.2. For arbitrary random variables D_1', D_2' on $\{0, 1, 2, \ldots\}$ with positive means, there is a sequence of random configuration intersection graphs as in (iv) for which $D_1 = D_1'$, $D_2 = D_2'$.

Thus in active, passive, and inhomogeneous models, either D_1, or D_2, or both, has a (mixed) Poisson distribution. The configuration model generalises these models in terms of local weak limits. For example, both D_1 and D_2 can be power-law.

The fact that $\mathcal{T}(D_1, D_2)$ is a limit for many sparse bipartite graph sequences is intuitive and in some physics literature has been assumed implicitly [39, 41]. A result similar to Theorem 3.1(iv) can be found in [23] and [43]. A nice proof for almost sure convergence in random configuration graphs (which could possibly be extended to bipartite graphs) can be found in [25]. For completeness, we provide our own formal proof in the Appendix of the arXiv version [40]. We do not use the second moment condition and work under slightly weaker assumptions (convergence in probability). We are not aware of prior literature on weak limits in cases (i)–(iii).

4. Applications

The proofs of the results in this section are given in Section 4.4. Let $\{G_n\}$ be a sequence of finite random graphs, and let G^* be a random element on $(\mathcal{G}_*, d_{\mathrm{loc}})$. Assume $|V(G_n)| \geq 1$ for all n and let v_n^* be chosen uniformly at random from $V(G_n)$ (given G_n).

4.1. Subgraph counts in random graphs

To apply Theorem 2.1 in a random setting we need some easy technical facts. Due to the equivalence mentioned after (3.1), the Lévy–Prokhorov metric and Skorokhod's representation theorem could be used to show these or stronger properties (see e.g. [23], [46]), but we derive them from more basic arguments.

Lemma 4.1. *Suppose* $\mathcal{L}((G_n, v_n^*)\,|\,G_n) \xrightarrow{p} \mathcal{L}(G^*)$ *and* $\{(G_n, v_n^*)\}$ *are defined on the same probability space. Then there is a random set A of positive integers such that*

(a) $\mathbb{P}(n \in A) \to 1$ *as* $n \to \infty$ *and*

(b) *almost surely* $|A| = \infty$ *and* $(G_n, v_n^*) \xrightarrow{d} G^*$ *as* $n \to \infty$, $n \in A$.

Lemma 4.2. $\mathcal{L}((G_n, v_n^*)\,|\,G_n) \xrightarrow{p} \mathcal{L}(G^*)$ *if and only if, for each bounded continuous function* $f : (\mathcal{G}_*, d_{\mathrm{loc}}) \to \mathbb{R}$, *we have* $\mathbb{E}(f(G_n, v_n^*)\,|\,G_n) \xrightarrow{p} \mathbb{E}f(G^*)$.

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

We now restate Theorem 2.1 for sequences of random graphs.

Lemma 4.3. *Let $h \ge 2$ be an integer, suppose $\mathcal{L}((G_n, v_n^*) \mid G_n) \xrightarrow{p} \mathcal{L}(G^*)$ and assume $\inf |V(G_n)| \to \infty$. As before, denote $d^* = d_{G^*}(r^*)$, $r^* = \mathrm{root}(G^*)$, $n_1 = n_1(n) = |V(G_n)|$ and assume $\mathbb{E}(d^*)^{h-1} < \infty$. Let d_n denote the degree of a uniformly random vertex in G_n. Then the following statements are equivalent:*

(i) $\mathbb{E}d_n^{h-1} \to \mathbb{E}(d^*)^{h-1}$,

(ii) d_n^{h-1} *is uniformly integrable,*

(iii) *for any connected graph H on h vertices and any $H' \in \mathcal{R}(H)$,*

$$n_1^{-1}\mathbb{E}\,\mathrm{emb}(H, G_n) \to \mathbb{E}\,\mathrm{emb}'(H', G^*, r^*).$$

Each of the above statements implies that for any connected graph H on h vertices and any $H' \in \mathcal{R}(H)$,

$$n_1^{-1}\mathrm{emb}(H, G_n) \xrightarrow{p} \mathbb{E}\,\mathrm{emb}'(H', G^*, r^*). \tag{4.1}$$

4.2. General weakly convergent sequences

In this section we assume that $\mathcal{L}((G_n, v_n^*) \mid G_n) \xrightarrow{p} \mathcal{L}(G^*)$, $n_1 = n_1(n) = |V(G_n)| \ge 3$ is non-random and $n_1 \to \infty$. As before, d_n is the degree of $v^* = v_n^*$ in G_n and d^* is the degree of the root r^* of G^*. Lemma 4.3 yields convergence of ${n_1}^{-1}\mathrm{emb}(H, G_n)$ provided that the $(|V(H)| - 1)$th degree moment of G_n converges. This allows us to determine the limit behaviour of statistics based on subgraph counts.

The *clustering coefficient* of a graph G is defined as

$$\alpha(G) := \frac{\mathrm{emb}(K_3, G)}{\mathrm{emb}(P_3, G)},$$

where K_3 is the clique on three vertices and P_t is the path on t vertices. (Set $\alpha(G) := 0$ when the denominator is zero.) For a rooted graph H' let $\mathrm{hom}'(H', G, v)$ denote the number of homomorphisms from H' to G that map $\mathrm{root}(H')$ to v. For $t \ge 2$, let K_t' be K_t rooted at any vertex and let $K_{1,t}'$ be the bipartite graph $K_{1,t}$ rooted at the vertex of degree t.

Corollary 4.1. *Suppose $\mathcal{L}((G_n, v_n^*) \mid G_n) \xrightarrow{p} \mathcal{L}(G^*)$ and $\mathbb{E}d_n^2 \to \mathbb{E}(d^*)^2 \in (0, \infty)$. Then*

$$\alpha(G_n) \xrightarrow{p} \alpha^* := \frac{\mathbb{E}\,\mathrm{emb}'(K_3', G^*, r^*)}{\mathbb{E}\,\mathrm{emb}'(K_{1,2}', G^*, r^*)} = \frac{\mathbb{E}\,\mathrm{emb}'(K_3', G^*, r^*)}{\mathbb{E}(d^*)_2}.$$

The *assortativity coefficient* (see e.g. [18], [35]) for a graph G is defined as Pearson's correlation of the degrees over the neighbouring vertices,

$$r(G) := \frac{g(G) - b(G)^2}{b'(G) - b(G)^2},$$

where

$$g(G) := (2e(G))^{-1} \sum d_G(u)d_G(v), \quad b(G) := (2e(G))^{-1} \sum d_G(u),$$
$$b'(G) := (2e(G))^{-1} \sum d_G(u)^2, \quad e(G) := |E(G)|,$$

and the sums are over all $2e(G)$ ordered pairs (u, v) of adjacent vertices in G. We define $r(G) := 0$ when either $e(G)$ or $b'(G) - b(G)^2$ is zero (i.e. G is regular). The above quantities can be easily expressed in terms of subgraph count statistics; see e.g. [22] and Section 4.4. Let P'_4 denote the graph P_4 rooted at one of its internal vertices.

Corollary 4.2. *Suppose* $\mathcal{L}((G_n, v_n^*) \mid G_n) \xrightarrow{p} \mathcal{L}(G^*)$, $\mathbb{E}d_n^3 \to \mathbb{E}(d^*)^3 < \infty$ *and* $\operatorname{Var}(d^*) > 0$. *Then*

$$\mathbb{E}r(G_n) \xrightarrow{p} \rho^* := \frac{\mathbb{E}d^*\mathbb{E}\,\hom'(P'_4, G^*, r^*) - (\mathbb{E}(d^*)^2)^2}{\mathbb{E}d^*\mathbb{E}(d^*)^3 - (\mathbb{E}(d^*)^2)^2}. \tag{4.2}$$

Corollaries 4.1 and 4.2 easily follow from Lemma 4.3; see Section 4.4. Notice that since $\alpha(G), r(G) \in [0, 1]$, convergence in probability in these corollaries implies convergence of means.

Statistics that can be expressed in terms of integrals of bounded functions, such as the limit degree distribution, are obtained directly from the local weak limit. Hence no degree moment conditions are necessary. Let $\pi_k(G)$ be the fraction of vertices of degree k in G. By Lemma 4.2,

$$\pi_k(G_n) = \mathbb{E}\big(\mathbb{I}_{d_{G_n}(v_n^*)=k} \mid G_n\big) \xrightarrow{p} \mathbb{P}(d^* = k). \tag{4.3}$$

Given a graph G and an integer $k \geq 2$, let (u_1^*, u_2^*, u_3^*) be a uniformly random triple of distinct vertices from $V(G)$. The *conditional clustering coefficient* is

$$\alpha_k(G) := \mathbb{P}\big(u_1^*u_3^* \in G \mid u_1^*u_2^*, u_2^*u_3^* \in G, d(u_2^*) = k\big),$$

and set $\alpha_k(G) := 0$ if the event in the condition has probability zero. Lemma 4.2 implies that if $\mathbb{P}(d^* = k) > 0$ then

$$\alpha_k(G_n) \xrightarrow{p} \alpha_k^* = \frac{\mathbb{E}\mathbb{I}_{d^*=k}\,\operatorname{emb}'(K'_3, G^*, r^*)}{k(k-1)\mathbb{P}(d^* = k)}. \tag{4.4}$$

The *conditional assortativity* (see [18]) is defined as

$$r_k(G) := \mathbb{E}\big(d_G(u_2^*) \mid u_1^*u_2^* \in G, d(u_1^*) = k\big),$$

and set $r_k(G) := 0$ if the event in the condition has probability zero. Let P'_3 be P_3 rooted at one of the endpoints.

Corollary 4.3. *Suppose* $\mathcal{L}((G_n, v_n^*) \mid G_n) \xrightarrow{p} \mathcal{L}(G^*)$, $\mathbb{E}d_n^2 \to \mathbb{E}(d^*)^2 < \infty$ *and* $\mathbb{P}(d^* = k) > 0$. *Then*

$$r_k(G_n) \xrightarrow{p} r_k^* = \frac{\mathbb{E}\mathbb{I}_{d^*=k}\,\hom'(P'_3, G^*, r^*)}{k\mathbb{P}(d^* = k)} = 1 + \frac{\mathbb{E}\mathbb{I}_{d^*=k}\,\operatorname{emb}'(P'_3, G^*, r^*)}{k\mathbb{P}(d^* = k)}.$$

In a similar way we can study the bivariate degree distribution [13] and many other functionals.

4.3. The case of random intersection graphs

Here we apply the above general results in the case where the limit is the uncorrelated clique tree of Section 3. We stress that Theorem 2.1 and its corollaries are applicable to a much broader class of sequences, including the inhomogeneous sparse random graph and the preferential attachment model [7, 22], general random configuration graphs with their many potential applications (see e.g. [34]), and random graphs from certain minor-closed classes, including random planar graphs [29, 42, 46].

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

Theorem 3.1 yields the first main condition (convergence to a local weak limit) for Lemma 4.3. For the other condition (convergence of a degree moment) we prove the following result.

Lemma 4.4. *Let $\{G_n\}$ be a sequence as in Theorem 3.1 and let k be a positive integer. Suppose an additional condition for each of cases* (i)–(iv) *of Theorem 3.1 holds:*

(i) $\mathbb{E}(X^{(1)})^k \to \mathbb{E}D_1^k < \infty$,

(ii) $\mathbb{E}(X^{(2)})^{k+1} \to \mathbb{E}D_2^{k+1} < \infty$,

(iii) $\mathbb{E}(\xi^{(1)})^k < \infty$ *and* $\mathbb{E}(\xi^{(2)})^{k+1} < \infty$,

(iv) $\mathbb{E}d_{1,v_1^*}^k \to \mathbb{E}D_1^k < \infty$ *and* $\mathbb{E}d_{2,v_2^*}^{k+1} \to \mathbb{E}D_2^{k+1} < \infty$.

Then $\mathbb{E}(d^*)^k < \infty$ *and* $\mathbb{E}d_n^k \to \mathbb{E}(d^*)^k$.

(We simplified (iv) of a previous version of [40] to fixed sequences and dropped a redundant assumption. To extend it to random d_1, d_2, use arguments similar to those of the proof of Lemma 4.3.) The special case of (i) where $k \le 2$ was shown in [15]. Here we use a different argument based on Theorem 3.1; see Section 4.4.

Using the same notation as in Section 4.2, assume that $\mathcal{L}((G_n, v_n^*) \mid G_n) \xrightarrow{p} \mathcal{L}(G^*) = \mathcal{L}(G_{\mathcal{T}})$, where $\mathcal{T} = \mathcal{T}(D_1, D_2)$, $\mathbb{E}D_1 > 0$ and $\mathbb{E}D_2 > 0$. Let $Z_1, Z_2, \ldots \sim D_2^* - 1$ be independent and independent of D_1. By (4.3) we have

$$\pi_k(G_n) \xrightarrow{p} \mathbb{P}(d^* = k) = \mathbb{P}\left(\sum_{i=1}^{D_1} Z_i = k\right).$$

For sequences of graphs as in Theorem 3.1(i)–(iii), the corresponding convergence of means has been shown in [10] and [14]; see also Remark 3.1. We also notice that the second moment condition required in [14] for the inhomogeneous model is not necessary.

By simple calculations we get

$$\begin{aligned}
&\mathbb{E}Z_1^k = \frac{\mathbb{E}(D_2-1)^k D_2}{\mathbb{E}D_2}, \quad \mathbb{E}(Z_1)_k = \mathbb{E}(D_2)_{k+1}(\mathbb{E}D_2)^{-1}, \quad k = 1, 2, \ldots,\\
&\mathbb{E}d^* = \mathbb{E}(Z_1 + \cdots + Z_{D_1}) = \mathbb{E}D_1 \mathbb{E}Z_1,\\
&\mathbb{E}(d^*)^2 = \mathbb{E}D_1 \mathbb{E}Z_1^2 + \mathbb{E}(D_1)_2(\mathbb{E}Z_1)^2,\\
&\mathbb{E}(d^*)^3 = \mathbb{E}D_1 \mathbb{E}Z_1^3 + 3\mathbb{E}(D_1)_2 \mathbb{E}Z_1 \mathbb{E}Z_1^2 + \mathbb{E}(D_1)_3(\mathbb{E}Z_1)^3,\\
&\mathbb{E}\,\mathrm{emb}'(K_3', G^*, r^*) = \mathbb{E}D_1 \mathbb{E}(Z_1)_2 = \mathbb{E}D_1(\mathbb{E}Z_1^2 - \mathbb{E}Z_1)
\end{aligned}$$

and

$$\mathbb{E}\,\mathrm{hom}'(P_4', G^*, r^*) = \mathbb{E}D_1 \mathbb{E}Z_1^3 + \mathbb{E}(D_1)_2 \mathbb{E}Z_1 \mathbb{E}Z_1^2 + \frac{\mathbb{E}(D_1)_2}{\mathbb{E}D_1}\mathbb{E}(d^*)^2 \mathbb{E}Z_1. \tag{4.5}$$

(The above estimates also hold in the case when either side is infinite.)

When $\mathbb{E}D_2^3 < \infty$ and $\mathbb{E}D_1^2 < \infty$, we have

$$\alpha^* = \frac{\mathbb{E}D_1 \mathbb{E}D_2 \mathbb{E}(D_2)_3}{\mathbb{E}D_1 \mathbb{E}D_2 \mathbb{E}(D_2)_3 + \mathbb{E}(D_1)_2(\mathbb{E}(D_2)_2)^2}$$

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

in Corollary 4.1. Using Remark 3.1, this simplifies to $\alpha^* = \mathbb{E}D_1/\mathbb{E}D_1^2$ for active random intersection graphs and to $\alpha^* = \mathbb{E}(D_2)_3(\mathbb{E}(D_2)_3 + \beta(\mathbb{E}(D_2)_2)^{-2}$ for passive random intersection graphs. This is equal to a related estimate $\hat{\alpha} = \lim \mathbb{E}\,\mathrm{emb}(K_3, G_n)(\mathbb{E}\,\mathrm{emb}(P_3, G_n))^{-1}$ obtained by Bloznelis [12] and the estimates of Godehardt, Jaworski, and Rybarczyk [30] for these particular models.

Similarly, if $\mathbb{E}D_1^2 < \infty$ and $\mathbb{E}D_2^4 < \infty$ then ρ^* in Corollary 4.2 is a rational function of $\mathbb{E}D_1$, $\mathbb{E}D_1^2$, and $\mathbb{E}D_2^k$, $k = 1, 2, 3, 4$ obtained by using (4.5) and the above expressions for $\mathbb{E}(d^*)^j$ in (4.2). One can check by simple algebra that ρ^* is equal to

$$\hat{\rho} = \lim\big(\mathbb{E}g(G_n) - \mathbb{E}b(G_n)^2\big)\big(\mathbb{E}b'(G_n) - \mathbb{E}b(G_n)^2\big)^{-1}$$

computed in [18] for sparse passive and active random intersection graphs.

Assuming only that $\mathbb{P}(d^* = k) > 0$, we get in (4.4)

$$\alpha_k^* = \frac{\mathbb{E}\big(\sum_{j=1}^{D_1} Z_i(Z_i - 1) \mid d^* = k\big)}{k(k-1)}.$$

If $D_2 \sim \mathrm{Po}\,(\lambda)$, as is the case for the active random intersection graph of Theorem 3.1(i), for example, then as in [18] (but without a second moment assumption)

$$\alpha_k^* = \frac{\lambda \mathbb{P}(d^* = k-1)}{k\mathbb{P}(d^* = k)}.$$

Finally, if $\mathbb{E}D_1^2 < \infty$, $\mathbb{E}D_2^2 < \infty$ in Corollary 4.3, we have

$$r_k^* = k^{-1}\mathbb{E}\left(\sum_{i=1}^{D_1} Z_i^2 \,\middle|\, d^* = k\right) + \frac{\mathbb{E}(D_1)_2\mathbb{E}(D_2)_2}{\mathbb{E}D_1\mathbb{E}D_2}.$$

This agrees with a related estimate obtained in [18] for active and passive random intersection graphs.

Thus Corollaries 4.1–4.3 generalise several previous results for particular random intersection graph models to arbitrary sequences of graphs with the uncorrelated clique tree as a limit. Applying them together with Lemma 4.4 with an appropriate k yields slightly stronger versions (i.e. convergence in probability and optimal moment conditions) of these results for the active and passive random intersection graphs with bounded expected degree. We are not aware of similar prior results for the inhomogeneous and configuration models.

4.4. Proofs

The *radius* of a connected rooted graph is the maximum distance from any vertex of the graph to the root.

Proof of Lemma 4.1. Let $H_1, H_2, \ldots$ be an enumeration of finite graphs in $\mathcal{G}_*$. For positive integers i, n, define the event

$$B(i, n) = \big\{\exists j \le i : |p_{r_j}(G_n, H_j) - \mathbb{P}(B_{r_j}(G^*) \cong H_j)| > i^{-1}\big\},$$

where r_j is the radius of H_j. By the assumption of the lemma, $\mathbb{P}(B(i, n)) \to 0$ for each $i = 1, 2, \ldots$. Define $N_1 = 1$, N_i, $i = 2, 3, \ldots$ by taking

$$N_i = 1 + \sup\big\{n > N_{i-1} : \mathbb{P}(B(i, n)) > i^{-1}\big\}.$$

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

Let $i(n) = \max\{i : N_i \le n\}$. Now let $A = \left\{n : \overline{B(i(n), n)}\right\}$. We have $\mathbb{P}(n \notin A) = \mathbb{P}(B(i(n), n)) \le i(n)^{-1} \to 0$. For any sequence of events $\{A_n, n \ge 1\}$ on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we have

$$\mathbb{P}(\cap_{m\ge 1} \cup_{n\ge m} A_n) \ge \limsup \mathbb{P}(A_n). \tag{4.6}$$

Thus

$$\mathbb{P}(|A| = \infty) \ge \limsup \mathbb{P}\left(\overline{B(i(n), n)}\right) = 1.$$

Now, by the definition of $B(i, n)$ on the event $|A| = \infty$, we have $(G_n, v_n^*) \xrightarrow{d} G^*$ as $n \to \infty$, $n \in A$. □

Proof of Lemma 4.2. ($\Leftarrow$) The function $(\mathcal{G}_*, d_{\mathrm{loc}}) \to \mathbb{R}$ that maps (G, v) to $\mathbb{I}_{B_r(G,v)\cong H}$ is bounded and continuous for each $r \ge 0$ and connected rooted graph H.

($\Rightarrow$) Without loss of generality we may assume $\{(G_n, v_n^*)\}$ are defined on a single probability space. Let A be a random set guaranteed by Lemma 4.1. Suppose there is some $\epsilon > 0$, a bounded continuous function f, and an infinite subset of positive integers B, such that $\mathbb{P}(|\mathbb{E}(f(G_n, v_n^*) \mid G_n) - \mathbb{E}f(G^*)| > \epsilon) > \epsilon$ for all $n \in B$. Define a random set $C = \{n \in B : |\mathbb{E}(f(G_n, v_n^*) \mid G_n) - \mathbb{E}f(G^*)| > \epsilon\}$. Since for $n \in A \cap B$ we have $\mathbb{P}(n \in A \cap C) \ge \epsilon - o(1)$, by (4.6) $\mathbb{P}(|A \cap C| = \infty) \ge \epsilon$. However, $(G_n, v_n^*) \xrightarrow{d} G^*$ when $n \to \infty$, $n \in A$, by (1.1), which contradicts our assumption. □

Proof of Lemma 4.3. We can assume $n_1 \ge 1$.

(i) $\Leftrightarrow$ (ii) This follows by Lemma 2.1, since Lemma 4.2 implies convergence in distribution of d_n^{h-1} to $(d^*)^{h-1}$.

(i) $\Rightarrow$ (4.1), (iii) Assume (i). Then there is a positive sequence $a_n \to 0$, such that $|\mathbb{E}d_n^{h-1} - \mathbb{E}(d^*)^{h-1}| \le a_n$. The empirical $(h-1)$th moment of the degree of G_n is $D_n = n_1^{-1} \hom(K_{1,h-1}, G_n)$. Also, $D_n = \sum_{H\in\mathcal{S}} d_H(\mathrm{root}(H))^{h-1} p_r(G_n, H)$, where $\mathcal{S}$ consists of graphs in $\mathcal{G}_*$ of radius 1. Since $\mathbb{E}D_n = \mathbb{E}d_n^{h-1}$, it follows by (i) and (3.1) that $D_n \xrightarrow{p} \mathbb{E}(d^*)^{h-1}$. So there is a positive sequence $\epsilon_n \to 0$, such that for all n,

$$\mathbb{P}\left(|D_n - \mathbb{E}(d^*)^{h-1}| > \epsilon_n\right) \le \epsilon_n.$$

We may assume that $\epsilon_n \ge a_n$. Let the random set C consist of those n for which $|D_n - \mathbb{E}(d^*)^{h-1}| \le \epsilon_n$. It follows by (4.6) that $\mathbb{P}(|C| = \infty) = 1$, $\mathbb{P}(n \in C) \to 1$ and on the event $|C| = \infty$, $D_n \to \mathbb{E}(d^*)^{h-1}$ as $n \to \infty$, $n \in C$.

Let A be the random set guaranteed by Lemma 4.1. On the event $|A \cap C| = \infty$, the subsequence of graphs $\{G_n, n \in A \cap C\}$ satisfies the conditions of Theorem 2.1.

Assume (4.1) does not hold. Then there is $H' \in \mathcal{R}(H)$, $\epsilon > 0$ and a deterministic infinite set D of positive integers such that for all $n \in D$,

$$\mathbb{P}(|X_n - \mathbb{E}X^*| > \epsilon) > \epsilon.$$

Here $X_n = n_1^{-1} \mathrm{emb}(H, G_n)$ and $X^* = \mathrm{emb}'(H', G^*, r^*)$. Let $D_1 \subseteq D$ consist of those n in D for which $|X_n - \mathbb{E}X^*| > \epsilon$. Again, by (4.6), we get that $\mathbb{P}(|A \cap C \cap D_1| = \infty) \ge \epsilon$. On this event $X_n \not\to \mathbb{E}X^*$ as $n \to \infty$, $n \in A \cap C$. This is a contradiction to Theorem 2.1(iii).

It remains to show (iii). Write $Y_n = n_1^{-1} \hom(H, G_n)$. Trivially, $X_n \le Y_n$, and by Lemma 2.2 $Y_n \le D_n$. So, for any $t > 0$,

$$\mathbb{E}X_n \mathbb{I}_{X_n > t} \le \mathbb{E}Y_n \mathbb{I}_{Y_n > t} \le \mathbb{E}D_n \mathbb{I}_{D_n > t}.$$

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

Since $\mathbb{E}D_n \to \mathbb{E}(d^*)^{h-1}$ and $D_n \xrightarrow{p} \mathbb{E}(d^*)^{h-1}$, D_n is uniformly integrable, and so is X_n. Since X_n also converges in probability by (4.1), (iii) follows by Lemma 2.1.

(iii) $\Rightarrow$ (i) The proof is identical to that of the corresponding implication of Theorem 2.1. □

Proof of Corollary 4.1. Apply Lemma 4.3. □

Proof of Corollary 4.2. For non-empty G, we have

$$g(G) = \frac{\hom(P_4, G)}{\mathrm{emb}(K_2, G)}, \quad b(G) = \frac{\hom(K_{1,2}, G)}{\mathrm{emb}(K_2, G)}, \quad b'(G) = \frac{\hom(K_{1,3}, G)}{\mathrm{emb}(K_2, G)}.$$

Let $S(t, j)$ denote Stirling numbers of the second kind. Using Lemma 4.3,

$$\begin{aligned}
&\frac{1}{n}\,\mathrm{emb}(K_2, G_n) \xrightarrow{p} \mathbb{E}d^*,\\
&\frac{1}{n}\hom(P_4, G_n)\\
&\quad = \frac{1}{n}\big(\mathrm{emb}(P_4, G_n) + \mathrm{emb}(K_3, G_n) + 2\,\mathrm{emb}(P_3, G_n) + \mathrm{emb}(K_2, G_n)\big)\\
&\quad \xrightarrow{p} \mathbb{E}\big(\mathrm{emb}'(P_4', G^*) + \mathrm{emb}'(K_3', G^*) + \mathrm{emb}(K_{1,2}', G^*) + \mathrm{emb}(P_3', G^*) + \mathrm{emb}'(K_2', G^*)\big)\\
&\quad = \mathbb{E}\hom'(P_4', G^*, r^*),\\
&\frac{1}{n}\hom(K_{1,t}, G_n) = \frac{1}{n}\sum_{j=1}^{t} S(t, j)\mathrm{emb}(K_{1,j}, G_n) \xrightarrow{p} \mathbb{E}\hom'(K_{1,t}', G^*, r^*) = \mathbb{E}(d^*)^t
\end{aligned}$$

for $t = 2, 3$. The claim follows by the definition of $r(G)$. □

Proof of Corollary 4.3. Note that $\pi_k(G_n) \xrightarrow{p} \mathbb{P}(d^* = k) > 0$ and when $\pi_k(G) > 0$ we have

$$\begin{aligned}
r_k(G) &= \frac{\mathbb{E}\big(d_G(u_2^*)\mathbb{I}_{d_{G_n}(u_1^*)=k}\mathbb{I}_{u_1^*u_2^*\in G} \mid G_n = G\big)}{\mathbb{P}(d_{G_n}(u_1^*) = k,\, u_1^*u_2^* \in G \mid G_n = G)}\\
&= \frac{(n_1)_2^{-1} H(G)}{k(n_1 - 1)^{-1}\pi_k(G)}\\
&= n_1^{-1} H(G)(k\pi_k(G))^{-1},
\end{aligned} \tag{4.7}$$

where $H(G)$ is the number of homomorphisms from $P_3 = xyz$ to G so that x is mapped to a vertex of degree k. Let $H_t(G)$ denote the number of such homomorphisms where additionally y is mapped to a vertex of degree at most t, and let $\bar{H}_t(G) = H(G) - H_t(G)$.

Fix $\delta > 0$. We will show that for any $\epsilon > 0$ and all n large enough,

$$\mathbb{P}\big(|n_1^{-1} H(G_n) - \mathbb{E}\mathbb{I}_{d^*=k}\hom'(P_3', G^*, r^*)| > \delta\big) \le \epsilon, \tag{4.8}$$

i.e. $n_1^{-1} H(G_n) \xrightarrow{p} \mathbb{E}\mathbb{I}_{d^*=k}\hom'(P_3', G^*, r^*)$.

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

By Lemma 4.2,

$$n_1^{-1}H_t(G_n) = \mathbb{E}\left(\mathbb{I}_{d_{G_n}(u_1^*)=k} \sum_{u\,:\,uu_1^*\in G_n} d_{G_n}(u)\mathbb{I}_{d_{G_n}(u)\le t} \,\middle|\, G_n\right)$$
$$\xrightarrow{p} h_t^* = \mathbb{E}\left(\mathbb{I}_{d^*=k} \sum_{u\,:\,ur^*\in G^*} d_{G^*}(u)\mathbb{I}_{d_{G^*}(u)\le t}\right).$$

Also, $h_t^* \to h^* = \mathbb{E}\mathbb{I}_{d^*=k}\,\mathrm{hom}'(P_3', G^*, r^*)$ as $t\to\infty$ since $\mathcal{L}((G_n, v_n^*) \mid G_n) \xrightarrow{p} \mathcal{L}(G^*)$ and $h^* \le \mathbb{E}\,\mathrm{hom}'(P_3', G^*, r^*) < \infty$ by Lemma 4.3. Therefore we can pick t_1 such that for $t \ge t_1$ and all n large enough,

$$\mathbb{P}\left(|n_1^{-1}H_t(G_n) - h_t^*| > \frac{\delta}{4}\right) \le \frac{\epsilon}{4}, \quad |h_t^* - h^*| \le \frac{\delta}{4}$$

and so

$$\mathbb{P}\left(|n_1^{-1}H_t(G_n) - h^*| > \frac{\delta}{2}\right) \le \frac{\epsilon}{2}. \tag{4.9}$$

Next, note that

$$\bar{H}_t(G_n) \le \sum_{v\in V(G_n)} d_{G_n}(v)^2\mathbb{I}_{d_{G_n}(v)>t}.$$

So, by Markov's inequality,

$$\mathbb{P}\big(n_1^{-1}\bar{H}_t(G_n) > \delta/2\big) \le 2\delta^{-1}n_1^{-1}\mathbb{E}\bar{H}_t(G_n) \le 2\delta^{-1}\mathbb{E}d_n^2\mathbb{I}_{d_n>t}.$$

Now d_n^2 is uniformly integrable by Lemma 4.3, so there is t_2 such that for all $t \ge t_2$ and all large enough n,

$$\mathbb{P}\big(n_1^{-1}\bar{H}_t(G_n) > \delta/2\big) \le \frac{\epsilon}{2}. \tag{4.10}$$

Now (4.8) follows by setting $t = \max(t_1, t_2)$ and combining (4.7), (4.9), and (4.10). □

Proof of Lemma 4.4. Let $Z_1, Z_2, \ldots \sim D_2^* - 1$ and D_1 be independent. For a random variable X with $\mathbb{E}X \in (0, \infty)$ and its size-biased version X^*, we have

$$\mathbb{E}(X^*-1)_j = \sum_{m\ge 1}(m-1)_j\frac{m\mathbb{P}(X=m)}{\mathbb{E}X} = \frac{\mathbb{E}(X)_{j+1}}{\mathbb{E}X}, \quad j = 1, 2, \ldots.$$

Thus, using the assumptions and Remark 3.1,

$$\mathbb{E}(Z_1)_j = \mathbb{E}(D_2)_{j+1}(\mathbb{E}D_2)^{-1} < \infty \quad \text{for } j = 1, \ldots, k. \tag{4.11}$$

Similarly $\mathbb{E}Z_1^k = (\mathbb{E}D_2)^{-1}\mathbb{E}D_2^{k+1} < \infty$. Write

$$\binom{k}{k_1, \ldots, k_j} = \frac{k!}{k_1!\cdots k_j!}.$$

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

Conditioning on D_1, using linearity of expectation and symmetry, we get

$$\mathbb{E}(d^*)^k = \mathbb{E}\left(\sum_{i=1}^{D_1} Z_i\right)^k$$
$$= \sum \binom{k}{k_1, \ldots, k_j} \mathbb{E}\binom{D_1}{j} \mathbb{E}Z_1^{k_1} \times \cdots \times \mathbb{E}Z_j^{k_j} \in (0, \infty). \tag{4.12}$$

Here the sum is over all j and all tuples of positive integers $(k_1, \ldots, k_j)$ such that $k_1 + \cdots + k_j = k$. Write $d_n = d_{G_n}(v_n^*)$ and recall that v_n^* is a uniformly random vertex from $V(G_n)$. By Theorem 3.1 $\mathcal{L}((G_n, v_n^*) \mid G_n) \xrightarrow{p} \mathcal{L}(G^*)$, so $d_n^k \xrightarrow{d} (d^*)^k$. By Fatou's lemma,

$$\mathbb{E}(d^*)^k \le \liminf \mathbb{E}d_n^k.$$

We assume without loss of generality that in case (iv) the sequences $d_1(n)$ and $d_2(n)$ (not to be confused with the random variable d_n) are symmetric random permutations of two fixed-degree sequences (each permutation of a particular sequence is equally likely). So in all cases (i)–(iv) by symmetry $\mathbb{E}d_{G_n}(v_1)^k = \mathbb{E}d_n^k$, where v_1 is a fixed vertex in $V(G_n)$. For each of the random intersection graph models we will show

$$\mathbb{E}d_{G_n}(v_1)^k \le \mathbb{E}(d^*)^k + o(1). \tag{4.13}$$

Let $\mathcal{T} \sim \mathcal{T}(D_1, D_2)$. Assume that $D_1 = d_{\mathcal{T}}(\text{root}(\mathcal{T}))$, $x_1, \ldots, x_{D_1}$ are the children of root($\mathcal{T}$) and Z_i is the number of children of x_i. For each n define a bipartite graph (a tree) $\tilde{H}_n$ as follows. On the event $D_1 > n_2$, let $\tilde{H}_n$ be a tree consisting of just the root $\tilde{v}_1$. On the event $D_1 \le n_2$, let $\tilde{H}_n$ be the subtree induced by generations 0, 1 and 2 of $\mathcal{T}$, but take only the first $Z_i' = Z_i \mathbb{I}_{Z_i \le n_1 - 1}$ children for the node x_i, $i = 1, \ldots, D_1$. Label the root v_1. Given $D_1, Z_1, \ldots, Z_{D_1}$, draw labels for $x_1, \ldots, x_{D_1}$ from V^2 uniformly at random without replacement and draw Z_i' distinct labels from $V^1 \setminus \{v_1\}$ for the children of x_i $i = 1, \ldots, D_1$, for each i (conditionally) independently. Here $V^i = V_i(H_n)$ is the set of first n_i vertices of the fixed ground set $\mathcal{V}^i$ as in Theorem 3.1.

Write $\tilde{d}_n = \mathbb{I}_{D_1 \le n_2} \sum_{i=1}^{D_1} Z_i'$ and notice that $\tilde{d}_n$ is an upper bound on the degree of v_1 in the resulting intersection graph. We have, as in (4.12),

$$\mathbb{E}\tilde{d}_n^k = \sum \binom{k}{k_1, \ldots, k_j} \mathbb{E}\binom{D_1}{j} \mathbb{I}_{D_1 \le n_2} \mathbb{E}(Z_1')^{k_1} \times \cdots \times \mathbb{E}(Z_t')^{k_t}$$
$$= \mathbb{E}(d^*)^k - o(1). \tag{4.14}$$

Here we used (4.11), (4.12) and bounds

$$\mathbb{E}(D_1)_j \mathbb{I}_{D_1 \le n_2} = \mathbb{E}(D_1)_j - o(1), \quad \mathbb{E}(Z_1')^j = \mathbb{E}Z_1^j - \mathbb{E}Z_1^j \mathbb{I}_{Z_1 > n_1 - 1} = \mathbb{E}Z_1^j - o(1),$$

valid for any $j \le k$ by Lemma 2.1. Thus it suffices to prove that $\mathbb{E}d_{G_n}(v_1)^k \le \mathbb{E}\tilde{d}_n^k + o(1)$. Recall that H_n is the bipartite graph underlying the intersection graph G_n. Call a path xyz *good* if $x = v_1$, $y \in \mathcal{V}_2$ and $z \in \mathcal{V}_1 \setminus \{v_1\}$. We have

$$d_{G_n}(v_1) \le \sum \mathbb{I}(v_1 w v, H_n) \quad \text{and} \quad \tilde{d}_n = \sum \mathbb{I}(v_1 w v, \tilde{H}_n)$$

where the sum is over all good paths $v_1 w v$ and $\mathbb{I}(F, H)$ is the indicator of the event that $F \subseteq E(H)$.

For any graph H we denote $v(H)=|V(H)|$ and $e(H)=|E(H)|$. If H is bipartite (more precisely, 2-coloured), $v_j(H)$, $j=1,2$ denotes the size of jth part V_j of H. Define an equivalence relation between bipartite graphs $H'=(V_1', V_2', E')$, $H''=(V_1'', V_2'', E'')$: $H'\sim H''$ if and only if there is an isomorphism from H' to H'' that maps V_j' to V_j'', $j=1,2$. Let $\mathcal{F}_k$ consist of one member for each equivalence class of all graphs formed from a union of k good paths (a not necessarily disjoint *union* of graphs $G_1=(V_1, E_1), \ldots, G_k=(V_k, E_k)$ is a graph $(V_1\cup\cdots\cup V_k, E_1\cup\cdots\cup E_k)$). For $H'\in\mathcal{F}_k$, let $N(H')$ be the number of distinct tuples of k good paths whose union is a bipartite graph H'' with parts $V_1''\subseteq V^1$ and $V_2''\subseteq V^2$ such that $H''\sim H'$. It is easy to see that there are positive constants $c(H')$, $C(H')$ such that, for all n large enough,

$$N(H')=c(H')(n_1)_{v_1(H')-1}(n_2)_{v_2(H')}=C(H')n_1^{v(H')-1}(1+\mathrm{o}(1)). \tag{4.15}$$

By linearity of expectation,

$$\mathbb{E}d_{G_n}(v_1)^k\le\mathbb{E}\Big(\sum\mathbb{I}(v_1wv, H_n)\Big)^k=\sum_{H'\in\mathcal{F}_k}N(H')\mathbb{E}\mathbb{I}(H', H_n),$$

and similarly

$$\mathbb{E}\tilde{d}_n^k=\sum_{H'\in\mathcal{F}_k}N(H')\mathbb{E}\mathbb{I}(H', \tilde{H}_n).$$

Using (4.14), (4.15) and the fact that $\mathcal{F}_k$ is finite, in order to prove (4.13) it suffices to check that

$$\mathbb{E}\mathbb{I}(H', H_n)\le\mathbb{E}\mathbb{I}(H', \tilde{H}_n)+\mathrm{o}\big(n_1^{-v(H')+1}\big)\quad\text{for each } H'\in\mathcal{F}_k. \tag{4.16}$$

So fix any $H'\in\mathcal{F}_k$. Suppose $v_2(H')=t$ and the degrees of vertices in $V_2(H')$ are $b_1, \ldots, b_t$. Note that $b_j\le k+1$ for $j=1, \ldots, t$. Conditioning on D_1 and the positions of generation 1 nodes labelled $V_2(H')$ and using (4.11)

$$\begin{aligned}\mathbb{E}\mathbb{I}(H', \tilde{H}_n)&=\mathbb{E}\frac{(D_1)_t\mathbb{I}_{D_1\le n_1}}{(n_2)_t}\frac{(Z_1')_{b_1-1}}{(n_1-1)_{b_1-1}}\times\cdots\times\frac{(Z_t')_{b_t-1}}{(n_1-1)_{b_t-1}}\\&=n_2^{-t}n_1^{-e(H')+t}\mathbb{E}(D_1)_t\prod_{i=1}^{t}\mathbb{E}(Z_i')_{b_i-1}(1+\mathrm{o}(1))\\&=\beta^{-t}n_1^{-e(H')}\mathbb{E}(D_1)_t(\mathbb{E}D_2)^{-t}\prod_{i=1}^{t}\mathbb{E}(D_2)_{b_i}(1+\mathrm{o}(1)).\end{aligned} \tag{4.17}$$

Now if H' is a tree then $e(H')=v(H')-1$, and (4.16) follows if

$$\mathbb{E}\mathbb{I}(H', H_n)\le\mathbb{E}\mathbb{I}(H', \tilde{H}_n)(1+\mathrm{o}(1)). \tag{4.18}$$

Meanwhile, if H' has a cycle then $e(H')\ge v(H')$ and $\mathbb{E}\mathbb{I}(H', \tilde{H}_n)=\mathrm{O}\big(n_1^{-v(H')}\big)$, so (4.16) follows whenever

$$\mathbb{E}\mathbb{I}(H', H_n)=\mathrm{o}\big(n^{-v(H')+1}\big). \tag{4.19}$$

We now consider (4.16) for each model separately.

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

(i) *Active intersection graph.* Let $a_1, \ldots, a_s$ be the degrees of vertices in $V_1(H')$. We can assume $a_1 = d_{H'}(v_1) = t$. Of course, $a_j \le k$, $j = 1, \ldots, s$. Since the vertices in $V_1(H')$ choose their neighbours independently, using Lemma 2.1,

$$\mathbb{E}\mathbb{I}(H', H_n) = \mathbb{E}\prod_{i=1}^{s}\frac{(X_v)_{a_i}}{(n_2)_{a_i}} = n_2^{-e(H')}\prod_{i=1}^{s}\mathbb{E}(D_1)_{a_i}(1 + \mathrm{o}(1)).$$

If H' has a cycle then $e(H') \ge v(H')$ and $\mathbb{E}\mathbb{I}(H', H_n) = \mathrm{O}\big(n_1^{-e(H')}\big)$, so (4.19) holds.

By Remark 3.1, $D_2 \sim \mathrm{Po}\,(\beta^{-1}\mathbb{E}D_1)$. So $\mathbb{E}(D_2)_{b_i} = (\beta^{-1}\mathbb{E}D_1)^{b_i}$. Thus (4.17) reduces to

$$\mathbb{E}\mathbb{I}(H', \tilde{H}_n) = (\beta n_1)^{-e(H')}\mathbb{E}(D_1)_t(\mathbb{E}D_1)^{e(H')-t}(1 + \mathrm{o}(1))$$

If H' has no cycle, then $a_j = 1$ for all $j \ge 2$. Thus

$$\mathbb{E}\mathbb{I}(H', H_n) \le n_2^{-e(H')}\mathbb{E}(D_1)_t(\mathbb{E}D_1)^{e(H')-t}(1 + \mathrm{o}(1))$$

and (4.19) follows.

(ii) *Passive intersection graph.* Since $b_i \le k + 1$ for $i = 1, \ldots, t$ by assumption (ii) of the lemma,

$$\mathbb{E}\mathbb{I}(H', H_n) = \mathbb{E}\prod_{i=1}^{s}\frac{(X_v)_{b_i}}{(n_1)_{b_i}} = n_1^{-e(H')}\prod_{i=1}^{s}\mathbb{E}(D_2)_{b_i}(1 + \mathrm{o}(1)).$$

Using Remark 3.1, $D_1 \sim \mathrm{Po}\,(\beta\mathbb{E}D_2)$, so $\mathbb{E}(D_1)_t = \beta^t(\mathbb{E}D_2)^t$. Therefore (4.17) reduces to

$$\mathbb{E}\mathbb{I}(H', \tilde{H}_n) = n_1^{-e(H')}\prod_{i=1}^{s}\mathbb{E}(D_2)_{b_i}(1 + \mathrm{o}(1))$$

and (4.16) follows.

(iii) *Inhomogeneous random intersection graph.* Let $\{\xi_u : u \in V(H')\}$ be independent random variables such that $\xi_u \sim \xi^{(i)}$ for $u \in V_i(H')$, $i = 1, 2$. Write $a \wedge b = \min\,(a, b)$. Then

$$\begin{aligned}\mathbb{E}\mathbb{I}(H', H_n) &= \mathbb{E}\prod_{uv\in E(H')}\left(\frac{\xi_u\xi_v}{\sqrt{n_1 n_2}} \wedge 1\right)\\ &\le \beta^{-e(H')/2}n_1^{-e(H')}\prod_{u\in V(H')}\mathbb{E}\xi_u^{d_{H'}(u)}(1 + \mathrm{o}(1)).\end{aligned}$$

If H' contains a cycle, then by the assumption that $\mathbb{E}(\xi^{(1)})^k$ and $\mathbb{E}(\xi^{(2)})^{k+1}$ are finite, we get that $\mathbb{E}\mathbb{I}(H', H_n) = \mathrm{O}\big(n_1^{-v(H')}\big)$, so (4.19) holds. If H' is a tree then

$$\mathbb{E}\mathbb{I}(H', H_n) \le \beta^{-e(H')/2}n_1^{-e(H')}\mathbb{E}\big(\xi^{(1)}\big)^t\big(\mathbb{E}\xi^{(1)}\big)^{s-1}\prod_{j=1}^{t}\mathbb{E}\big(\xi^{(2)}\big)^{b_j}(1 + \mathrm{o}(1)). \tag{4.20}$$

Using Remark 3.1, we have $D_1 \sim \mathrm{Po}\big(\beta^{1/2}\xi^{(1)}\mathbb{E}\xi^{(2)}\big)$ and $D_2 \sim \mathrm{Po}\big(\beta^{-1/2}\xi^{(2)}\mathbb{E}\xi^{(1)}\big)$, so $\mathbb{E}(D_1)_t = \beta^{t/2}\mathbb{E}(\xi^{(1)})^t(\mathbb{E}\xi^{(2)})^t$ and $\mathbb{E}(D_2)_j = \beta^{-j/2}\mathbb{E}(\xi^{(2)})^j(\mathbb{E}\xi^{(1)})^j$ for $j \le k + 1$. Putting these

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

estimates into (4.17) and simplifying, we get the expression on the right-hand side of (4.20). Then (4.18) follows.

(iv) *Random configuration graph.* For $i = 1, 2$, let

$$\tilde{d}_{i,m} = n_i^{-1} \sum_{v \in V^i} (d_{i,v})_m.$$

Recall that $N = \sum_{u \in V^1}^{n_1} d_{1,u}$ is the total number of half-edges in each of the parts. Since $n_1, n_2 \to \infty$ and by the assumption of the lemma $N n_1^{-1} \to \mathbb{E} D_1$, there exists $\omega_n \to \infty$ such that for all n

$$n_1, n_2, N \ge \omega_n, \quad \omega_n \ge k + 2. \tag{4.21}$$

The probability that H_n contains H' as a subgraph is at most

$$a(H') = \frac{1}{(N)_{e(H')}} \mathbb{E} \prod_{u \in V(H')} (d_{H_n}(u))_{d_{H'}(u)}.$$

Here the product counts the number of ways to choose particular half-edges forming H'. Let $(u_1^*, \ldots, u_s^*)$ and $(w_1^*, \ldots, w_t^*)$ be independent uniformly random tuples of distinct vertices from $V_1(H_n)$ and $V_2(H_n)$ respectively. Using symmetry, (4.21), and the assumption of the lemma,

$$\begin{aligned} a(H') &= \frac{1}{(N)_{e(H')}} \mathbb{E} \prod_{i=1}^{s} (d_{1,u_i^*})_{a_i} \prod_{j=1}^{t} (d_{2,w_j^*})_{b_j} \\ &\le \frac{1}{(N)_{e(H')}} \prod_{i=1}^{s} \tilde{d}_{1,a_i} \prod_{j=1}^{t} \tilde{d}_{2,b_j} (1 + \mathrm{o}(1)) \\ &= (\mathbb{E} D_1 n_1)^{-e(H')} \prod_{i=1}^{s} \mathbb{E}(D_1)_{a_i} \prod_{j=1}^{t} \mathbb{E}(D_2)_{b_j} (1 + \mathrm{o}(1)). \end{aligned}$$

Again, if H' has a cycle then (4.19) follows. Otherwise, if H' is a tree, then since $\mathbb{E} D_1 n_1 = \mathbb{E} D_2 n_2 (1 + \mathrm{o}(1))$ we have $\mathbb{E} D_1 = \beta \mathbb{E} D_2$ and

$$(\mathbb{E} D_1)^{-e(H')} \prod_{i=1}^{s} \mathbb{E}(D_1)_{a_i} = \mathbb{E}(D_1)_t \frac{(\mathbb{E} D_1)^{s-1}}{(\mathbb{E} D_1)^{s+t-1}} = \beta^{-t} \mathbb{E}(D_1)_t (\mathbb{E} D_2)^{-t}.$$

By comparing $a(H')$ with (4.17), we see that (4.18) holds. □

Acknowledgements

I would like to thank the referees for their helpful remarks, in particular for pointing out the important connections to the classical theory of convergence of measures in complete separable metric spaces.

Funding information

Part of this work was supported by the Research Council of Lithuania (MIP-067/2013).

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] ALDOUS, D. AND LYONS, R. (2007). Processes on unimodular random networks. *Electron. J. Prob.* **12**, 1454–1508.

[2] ALDOUS, D. AND STEELE, J. M. (2004). The objective method: probabilistic combinatorial optimization and local weak convergence, In *Probability on Discrete Structures* (Encyclopaedia Math. Sci. **110**), pp. 1–72. Springer, Berlin.

[3] ANDERSSON, H. (1998). Limit theorems for a random graph epidemic model. *Ann. Appl. Prob.* **8**, 1331–1349.

[4] ARRATIA, R., GOLDSTEIN, L. AND KOCHMAN, F. (2019). Size bias for one and all. *Prob. Surv.* **16**, 1–61.

[5] BENJAMINI, I. AND SCHRAMM, O. (2001). Recurrence of distributional limits of finite planar graphs. *Electron. J. Prob.* **6**, 1–13.

[6] BENJAMINI, I., LYONS, R. AND SCHRAMM, O. (2013). Unimodular random trees. *Ergodic Theory Dyn. Syst.* **35**, 1–15.

[7] BERGER, N., BORGS, C., CHAYES J. T. AND SABERI, A. (2014). Asymptotic behavior and distributional limits of preferential attachment graphs. *Ann. Prob.* **42**, 1–40.

[8] BILLINGSLEY, P. (1971). *Weak Convergence of Measures: Applications in Probability*. Society for Industrial and Applied Mathematics, Philadelphia.

[9] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.

[10] BLOZNELIS, M. (2008). Degree distribution of a typical vertex in a general random intersection graph. *Lith. Math. J.* **48**, 38–45.

[11] BLOZNELIS, M. (2010). The largest component in an inhomogeneous random intersection graph with clustering. *Electron. J. Combin.* **17**, R110.

[12] BLOZNELIS, M. (2013). Degree and clustering coefficient in sparse random intersection graphs. *Ann. Appl. Prob.* **23**, 1254–1289.

[13] BLOZNELIS, M. (2015). Degree-degree distribution in a power law random intersection graph with clustering. In *Algorithms and Models for the Web Graph: 12th International Workshop (WAW 2015)* (Lecture Notes Comput. Sci. **9479**), eds D. F. Gleich *et al.*, pp. 42–53. Springer.

[14] BLOZNELIS, M. AND DAMARACKAS, J. (2013). Degree distribution of an inhomogeneous random intersection graph. *Electron. J. Combin.* **20**, P3.

[15] BLOZNELIS, M. AND KURAUSKAS, V. (2017). Large cliques in sparse random intersection graphs. *Electron. J. Combin*. **24**, P2.5.

[16] BLOZNELIS, M., JAWORSKI, J., GODEHARDT, E., KURAUSKAS, V. AND RYBARCZYK, K. (2015). Recent progress in complex network analysis: models of random intersection graphs. In *Data Science, Learning by Latent Structures and Knowledge Discovery* (Studies in Classification, Data Analysis and Knowledge Organization), eds B. Lausen *et al.*, pp. 69–78. Springer, Berlin and Heidelberg.

[17] BLOZNELIS, M., JAWORSKI, J., GODEHARDT, E., KURAUSKAS, V. AND RYBARCZYK, K. (2015). Recent progress in complex network analysis: properties of random intersection graphs. In *Data Science, Learning by Latent Structures and Knowledge Discovery* (Studies in Classification, Data Analysis and Knowledge Organization), eds B. Lausen *et al.*, pp. 79–88. Springer, Berlin and Heidelberg.

[18] BLOZNELIS, M., JAWORSKI, J. AND KURAUSKAS, V. (2013). Assortativity and clustering in sparse random intersection graphs. *Electron. J. Prob.* **18**, 1–24.

[19] BOLLOBÁS, B. AND RIORDAN, O. (2011). Sparse graphs: metrics and random models. *Random Structures Algorithms* **39**, 1–38.

[20] BOLLOBÁS, B. AND RIORDAN, O. (2015). An old approach to the giant component problem. *J. Combinatorial Theory B* **113**, 236–260.

[21] BOLLOBÁS, B., JANSON, S. AND RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31**, 3–122.

[22] BOLLOBÁS, B., JANSON, S. AND RIORDAN, O. (2011). Sparse random graphs with clustering. *Random Structures Algorithms* **38** 269–323.

[23] BORDENAVE, C. AND LELARGE, M. (2010). Resolvent of large random graphs. *Random Structures Algorithms* **37**, 332–352.

[24] CSIKVÁRI, P. AND ZHICONG, L. (2014). Graph homomorphisms between trees. *Electron. J. Prob.* **21**, P4.9.

[25] DEMBO, A. AND MONTANARI, A. (2010). Gibbs measures and phase transitions on sparse random graphs. *Braz. J. Prob. Stat.* **24**, 137–211.

[26] DURRETT, R. (2010). *Probability: Theory and Examples*, 4th edn. Cambridge University Press.

https://doi.org/10.1017/jpr.2021.84 Published online by Cambridge University Press

[27] FIOL, M. A. AND GARRIGA, E. (2009). Number of walks and degree powers in a graph. *Discrete Math.* **309**, 2613–2614.
[28] GAMARNIK, D. AND MISRA, S. (2015). Giant component in multipartite graphs with given degree sequences. *Stoch. Syst.* **5**, 372–408.
[29] GEORGAKOPOULOS, A. AND WAGNER, S. (2016). Limits of subcritical random graphs and random graphs with excluded minors. Available at arXiv:1512.03572.
[30] GODEHARDT, E., JAWORSKI, J. AND RYBARCZYK, K. (2012). Clustering coefficients of random intersection graphs. In Challenges at the Interface of Data Analysis, *Computer Science and Optimization*, eds W. Gaul *et al*., pp. 243–253. Springer, Berlin.
[31] GRIMMETT, G. AND STIRZAKER, D. (2001). *Probability and Random Processes*. Oxford University Press.
[32] GUILLAUME J.-L. AND LATAPY, M. (2006). Bipartite graphs as models of complex networks. *Physica A* **371**, 795–813.
[33] VAN DER HOFSTAD, R. (2020+). *Random Graphs and Complex Networks*, Vol. II, preliminary version. https://www.win.tue.nl/~rhofstad/NotesRGCNII_colleagues_25_04_2022.pdf, accessed 2022-06-09.
[34] JANSON, S. AND LUCZAK, M. J. (2009). A new approach to the giant component problem. *Random Structures Algorithms* **34**, 197–216.
[35] LITVAK, N. AND VAN DER HOFSTAD, R. (2013). Uncovering disassortativity in large scale-free networks. *Phys. Rev. E* **87**, 022801.
[36] LOVÁSZ, L. (2012). *Large Networks And Graph Limits* (American Mathematical Society Colloquium Publications **60**). American Mathematical Society, Providence, RI.
[37] LYONS, R. (2005). Asymptotic enumeration of spanning trees. *Combinatorics Prob. Comput.* **14**, 491–522.
[38] KAROŃSKI, M., SCHEINERMAN, E. R. AND SINGER-COHEN, K. B. (1999). On random intersection graphs: the subgraph problem. *Combinatorics Prob. Comput.* **8**, 131–159.
[39] KARRER, B. AND NEWMAN, M. E. J. (2010). Random graphs containing arbitrary distributions of subgraphs. *Phys. Rev. E* **82**, 066118.
[40] KURAUSKAS, V. (2015). On local weak limit and subgraph counts for sparse random graphs. Available at arXiv:1504.08103.
[41] 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.
[42] PANAGIOTOU, K., STUFLER, B. AND WELLER, K. (2016). Scaling limits of random graphs from subcritical classes. *Ann. Prob.* **44**, 3291–3334.
[43] RICHARDSON, T. AND URBANKE, R. (2008). *Modern Coding Theory*. Cambridge University Press.
[44] SALEZ, J. (2011). Some implications of local weak convergence for sparse random graphs. Doctoral thesis, Université Pierre et Marie Curie – Paris VI, Ecole Normale Supérieure de Paris – ENS Paris.
[45] SIDORENKO, A. (1994). A partially ordered set of functionals corresponding to graphs. *Discrete Math.* **131**, 263–277.
[46] STUFLER, B. (2021). Local convergence of random planar graphs. In *Extended Abstracts EuroComb 2021* (Trends in Mathematics **14**), eds J. Nešetřil *et al.* Birkhäuser, Cham.
[47] VADON, V., KOMJÁTHY, J. AND VAN DER HOFSTAD, R. (2019). A new model for overlapping communities with arbitrary internal structure. *Applied Network Science* **4**, 1–19.
[48] WORMALD, N. C. (1999). Models of random regular graphs. In *Surveys in Combinatorics 1999 (Canterbury)* (London Math. Soc. Lecture Notes **267**), pp. 239–298. Cambridge University Press.