

# LIMITS OF SPARSE CONFIGURATION MODELS AND BEYOND: GRAPHEXES AND MULTIGRAPHEXES

BY CHRISTIAN BORGS<sup>1</sup>, JENNIFER T. CHAYES<sup>2</sup>, SOUVIK DHARA<sup>3</sup> AND  
SUBHABRATA SEN<sup>4</sup>

<sup>1</sup>Department of Electrical Engineering and Computer Science, University of California, Berkeley, [borgs@berkeley.edu](mailto:borgs@berkeley.edu)

<sup>2</sup>Division of Computing, Data Science, and Society, University of California, Berkeley, [jchayes@berkeley.edu](mailto:jchayes@berkeley.edu)

<sup>3</sup>Department of Mathematics, Massachusetts Institute of Technology, [sdhara@mit.edu](mailto:sdhara@mit.edu)

<sup>4</sup>Department of Statistics, Harvard University, [subhabratasen@fas.harvard.edu](mailto:subhabratasen@fas.harvard.edu)

We investigate structural properties of large, sparse random graphs through the lens of *sampling convergence* (Borgs et al. (*Ann. Probab.* **47** (2019) 2754–2800)). Sampling convergence generalizes left convergence to sparse graphs, and describes the limit in terms of a *graphex*. We introduce a notion of sampling convergence for sequences of multigraphs, and establish the graphex limit for the configuration model, a preferential attachment model, the generalized random graph and a bipartite variant of the configuration model. The results for the configuration model, preferential attachment model and bipartite configuration model provide necessary and sufficient conditions for these random graph models to converge. The limit for the configuration model and the preferential attachment model is an augmented version of an exchangeable random graph model introduced by Caron and Fox (*J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** (2017) 1295–1366).

## 1. Introduction.

**1.1. Aims and informal overview.** The study of large networks, arising from applications in the social, physical and life sciences, has witnessed meteoric growth over the past two decades. It is widely believed that a thorough understanding of the typical structural properties of these large networks can often provide deep insights into the workings of many social, economic and biological systems of practical interest. Random graph models have been extensively used to study properties of these networks, with many recent models aimed at capturing specific properties of real world networks (we refer the interested reader to [24] and the references therein for an overview).

In this light, it is desirable to study the asymptotic structural properties of random graphs. A natural question here is to identify a deterministic structure that captures the typical behavior of these random graph models. This question is analogous to deriving strong law of large numbers, but now on the space of graphs. The first challenge is to figure out the topology needed on the space of graphs for such convergence results. In case of dense graphs, where the number of edges in the graphs scale quadratically with the number of vertices, the theory of graph limits [9, 16, 17, 23, 35–37] provides the relevant framework to obtain this asymptotic description, and an extensive line of work [8, 20, 41] is aimed at describing the asymptotic behavior of dense random graphs. However, this framework fails to provide non-trivial information about sparse graph sequences, and thus motivates a recent line of work to extend the theory of graph limits to the sparse setting [7, 10–14, 28].

---

Received July 2019; revised March 2020.

*MSC2020 subject classifications.* Primary 60C05, 05C80; secondary 60G09, 60G55, 60G57.

*Key words and phrases.* Graphexes, sparse random graphs, Caron–Fox model, configuration model, preferential attachment, generalized random graph.

In this paper, we derive limits of fundamental sparse random graphs with respect to the notion of *sampling convergence* introduced recently by Borgs et al. [12]. Empirical evidence suggests that typical real-world networks are composed of high degree vertices, called “hubs,” which form a skeleton of the network, and low degree vertices that constitute the body of the network [3, 39]. The limiting object under sampling convergence, called graphex, makes this distinction explicit; see Section 1.5 for a more detailed discussion.

Our principal contributions in this article are as follows.

*Convergence of multigraphs.* We introduce a notion of sampling convergence for multigraphs, generalizing the notion of sampling convergence introduced in [12], and identify the resulting limit object, which we call a *multigraphex*. We also formulate an equivalent notion of convergence in terms of certain two-dimensional point processes; it is this representation we use when establishing the limits of the various random models considered in this paper.

*Limits of random graphs.* We deduce the (multi)graphex limit of fundamental random graph models, under the sparse setting—the configuration model (Theorem 1.2), a preferential attachment model (Theorem 1.6), the generalized random graph (Theorem 1.7) and a bipartite variant of the configuration model (Theorem 1.10). The proof techniques here are completely disjoint from the previous results in the dense settings [8, 20, 41]. In the dense case, graph convergence is equivalent to convergence of subgraph densities, which are real-valued random variables. Such equivalence breaks down in the sparse setting. To this end, we make use of an idea, put forth in [12], that sampling convergence is equivalent to weak convergence of certain two-dimensional point processes (Proposition 1.1). The relevant point processes for the configuration model, the preferential attachment model and the bipartite configuration model all have a specific “rank-one” structure (see Remark 5 below), which in turn allows us to conclude that weak convergence is equivalent to the weak convergence of a one-dimensional Lévy process. This facilitates a precise characterization of the necessary and sufficient conditions for sampling convergence in these random graph models. To illustrate the “nonrank-one case,” we analyze the generalized random graph, and derive sufficient conditions for sampling convergence. Our analysis in this case provides a fairly general template, and may prove to be useful for establishing sampling convergence for other graph sequences of practical interest.

*New interpretation of the Caron–Fox model.* Finally, our results provide a novel, alternative perspective on the Caron and Fox [19] model, which has induced immense recent interest in theoretical statistics. Specifically, a corollary of our result (Theorem 1.2) establishes that Caron–Fox graphs can be interpreted as the limit of samples from a configuration model or preferential attachment random graphs.

1.2. *Notation and terminologies.* Before we progress further, we introduce some notation used throughout our subsequent discussion. We use the standard notation of  $\xrightarrow{\mathbb{P}}$ , and  $\xrightarrow{d}$  to denote convergence in probability and in distribution, respectively. We use the Bachmann–Landau notation  $O(\cdot)$ ,  $o(\cdot)$ ,  $\Theta(\cdot)$ ,  $\Omega(\cdot)$ ,  $\omega(\cdot)$  for asymptotics of real numbers.  $\mathbb{R}_+ = [0, \infty)$ ,  $\overline{\mathbb{R}}_+ = \mathbb{R}_+ \cup \{\infty\}$ ,  $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$  and  $\otimes$ : = product of measures.

Given a multigraph  $G$ , we use the generic notation  $V(G)$ ,  $E(G)$  to denote the set of vertices and edges, respectively, and set  $v(G) = |V(G)|$ . Further, we denote the number of nonloop edges as  $e(G)$ . Let  $\mathfrak{G}_f$  denote the set of all multigraphs with finite number of vertices and edges. Thus  $\mathfrak{G}_f$  is countable and we equip this space with discrete topology.

For any topological space  $X$ ,  $\mathcal{B}(X)$  will denote the Borel sigma-algebra of  $X$ . We define a measure on a metric space to be locally finite if it assigns finite measure to all bounded Borel sets. Let  $\mathcal{M}(\mathbb{R}_+)$  and  $\mathcal{M}(\mathbb{R}_+^2)$  denote the space of locally finite Borel measures on  $\mathbb{R}_+$

and  $\mathbb{R}_+^2$ , respectively, equipped with the vague topology.  $\mathcal{N}(\mathbb{R}_+) \subset \mathcal{M}(\mathbb{R}_+)$  and  $\mathcal{N}(\mathbb{R}_+^2) \subset \mathcal{M}(\mathbb{R}_+^2)$  will denote the subspaces of counting measures, equipped with the vague topology. For a Polish space  $\mathcal{S}$ , let  $\mathcal{P}(\mathcal{S})$  denote the space of all probability measures on  $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ , equipped with the topology for weak convergence of probability measures. For an  $\mathcal{S}$ -valued random variable  $X$ , let  $\mathcal{L}(X)$  denote the law of  $X$ , which is an element of  $\mathcal{P}(\mathcal{S})$ .

We will also need the notion of completely random measures, defined as random measures  $\mu \in \mathcal{M}(\mathbb{R}_+)$  that obey the condition that for all finite families of bounded disjoint sets  $(A_i)_{i \leq k}$  in  $\mathcal{B}(\mathbb{R}_+)$ ,  $(\mu(A_i))_{i \leq k}$  is an independent collection of random variables. We will in particular be interested in completely random measures that are stationary, that is, completely random measures  $\mu$  such that the distribution of  $\mu([t, t + s])$  depends only on  $s$  for any  $t, s \in \mathbb{R}_+$ . These can be represented in the form

$$(1.1) \quad \mu = a\lambda + \sum_{i \geq 1} w_i \delta_{\theta_i},$$

where  $\{(w_i, \theta_i)\}_{i \geq 1}$  is a Poisson point process on  $(0, \infty) \times \mathbb{R}_+$  with rate measure  $\rho(dw)\lambda(d\theta)$ , with  $\rho$  obeying certain integrability conditions; see Appendix B for details. We use the notation  $\text{CRM}(a\lambda, \rho \times \lambda)$  for the law of a completely random measure of the form (1.1).

Finally, given a topological space  $X$  and an interval  $I \subseteq \mathbb{R}$ , we use  $\mathbb{D}(I, X)$  to denote the set of càdlàg functions  $f : I \rightarrow X$ , that is, the set of functions  $f$  that are right continuous, and have limits from the left.

### 1.3. Sampling convergence and multigraphexes.

*Sampling convergence.* The following two definitions are straight forward generalizations of the notion of sampling and sampling convergence from the simple graph setting of [12] to our multigraph setting.

**DEFINITION 1** (*p*-sampling). For a multigraph  $G$ , the *p*-sampled multigraph (denoted by  $\text{Smpl}(G, p)$ ) is an unlabeled random graph obtained by keeping each vertex independently with probability  $p$ , taking the induced edges on the kept vertices and deleting the isolated vertices.

**DEFINITION 2** (Sampling convergence). Let  $(G_n)_{n \geq 1}$  be a sequence of (multi)graphs.  $(G_n)_{n \geq 1}$  is said to be sampling convergent if for all  $t > 0$ ,  $\text{Smpl}(G_n, t/\sqrt{2e(G_n)})$  converges in distribution in  $\mathfrak{G}_f$ .

The distribution of the sampled graph is characterized by the subgraph frequencies, and thus relates to left convergence. In order for the sampled subgraph to be informative,  $p$  is to be chosen such that the sampled subgraph is nonempty. The choice of  $p \sim 1/\sqrt{e(G_n)}$  ensures that the sampled subgraph has  $\Theta(1)$  number of edges in expectation, and hence also an expected number of vertices that is  $\Theta(1)$ . Note that in the sparse setting, this holds since we removed the isolated vertices—without this, the expected number of sampled vertices,  $tv(G_n)/\sqrt{2e(G_n)}$ , would diverge.

*Graphexes and adjacency measures.* Next, we formally introduce the limiting objects for sampling convergence.

**DEFINITION 3** (Random adjacency measure). An adjacency measure is a measure  $\xi \in \mathcal{N}(\mathbb{R}_+^2)$  such that  $\xi(A \times B) = \xi(B \times A)$  for all  $A, B \in \mathcal{B}(\mathbb{R}_+)$ . A random adjacency measure is a  $\mathcal{N}(\mathbb{R}_+^2)$  valued random variable that is almost surely an adjacency measure. It is called *exchangeable* if  $\xi(\phi^{-1}(A) \times \phi^{-1}(B)) \stackrel{d}{=} \xi(A \times B)$  for every measure preserving map  $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ .

Expressing a random adjacency measure  $\xi$  as

$$\xi = \sum_{ij} \beta_{ij} \delta_{(\alpha_i, \alpha_j)},$$

where  $\beta_{ij} \in \mathbb{N}_0$ , one can naturally associate an unlabeled graph  $\mathcal{G}(\xi)$  as follows: Consider a countable vertex set where vertex  $i$  is labeled by  $\alpha_i$ .  $\mathcal{G}(\xi)$  is obtained by having  $\beta_{ij}$  many edges between the vertices labeled  $\alpha_i$  and  $\alpha_j$ , deleting the isolated vertices and finally erasing the labels of the vertices.

The limits of sampling convergence will be related to a  $\mathfrak{G}_f$  valued stochastic process obtained from specific random adjacency measures in a way we make more precise below. To define the random adjacency measures under consideration, we first define a multigraphex. We denote the sequence space  $\ell_1 = \{(x_i)_{i \geq 1} : x_i \in \mathbb{R}_+ \forall i, \sum_{i=1}^\infty x_i < \infty\}$ .

**DEFINITION 4 (Multigraphex).** A multigraphex is a triple  $\mathcal{W} = (W, S, I)$  such that  $I \in \ell_1$ ,  $S : \mathbb{R}_+ \mapsto \ell_1$  is a measurable function, and  $W : \mathbb{R}_+^2 \times \mathbb{N}_0 \mapsto \mathbb{R}_+$  is a measurable function satisfying  $W(x, y, k) = W(y, x, k)$ ,  $\sum_{k=0}^\infty W(x, y, k) = 1$ , for any  $x, y \in \mathbb{R}_+$  and  $k \in \mathbb{N}_0$ . We will assume throughout that  $\min\{\sum_{k \geq 1} S(\cdot, k), 1\}$  is integrable. Further, setting  $\mu_W(\cdot) = \int (1 - W(\cdot, y, 0)) dy$ , we assume that:

- (a)  $\Lambda(\{x : \mu_W(x) = \infty\}) = 0$  and  $\Lambda(\{x : \mu_W(x) > 1\}) < \infty$ ,
- (b)  $\int (1 - W(x, y, 0)) \mathbb{1}\{\mu_W(x) \leq 1\} \mathbb{1}\{\mu_W(y) \leq 1\} dy dx < \infty$ ,
- (c)  $\int (1 - W(x, x, 0)) dx < \infty$ .

$W$  is called a multigraphon,  $S$  is called a multistar function and  $I$  is called an isolated edge sequence. A graphon  $W$  is a multigraphon with  $W(x, y, k) = 0$  for  $k \geq 2$ . In this case, we describe the graphon as a function  $W : \mathbb{R}_+^2 \rightarrow [0, 1]$  and set  $W(x, y) = W(x, y, 1)$ . Similarly, a simple star function  $S : \mathbb{R}_+ \mapsto \ell_1$  satisfies  $S(x, k) = 0$  for  $k \geq 2$  and all  $x \in \mathbb{R}_+$ . In this case, we describe the star function as  $S : \mathbb{R}_+ \mapsto \mathbb{R}_+$ , and set  $S(x) = S(x, 1)$ . Finally, a simple isolated edge constant  $I$  corresponds to the case where  $I(k) = 0$  for  $k \geq 2$ . In this case, we represent  $I = I(1)$  as a constant. A graphex is obtained by replacing the multigraphon, multi-star function and multiedge sequence in Definition 4 by their simple analogues, with the isolated edge constant sometimes referred to as the dust component of the graphex [11, 28].

In this paper, the case where  $W$  is a multigraphon, but  $I$  and  $S$  are simple plays an important role; henceforth, whenever we specify a star function  $S : \mathbb{R}_+ \mapsto \mathbb{R}_+$  or an isolated edge constant  $I$ , we assume that these describe a star function or an edge sequence with  $S(x, k)$  or  $I(k) = 0$  for  $k \geq 2$  without explicitly mentioning it in every case.

**DEFINITION 5 (Adjacency measure of a multigraphex).** Given any multigraphex  $\mathcal{W} = (W, S, I)$ , define  $\xi_{\mathcal{W}}$ , the random adjacency measure generated by  $\mathcal{W}$  as follows:

$$\begin{aligned} \xi_{\mathcal{W}} &= \sum_{i \neq j} \zeta_{ij} \delta_{(\theta_i, \theta_j)} + \sum_i \zeta_{ii} \delta_{(\theta_i, \theta_i)} + \sum_{j,k} g(\theta_j, \chi_{jk}) (\delta_{(\theta_j, \sigma_{jk})} + \delta_{(\sigma_{jk}, \theta_j)}) \\ &\quad + \sum_k h(\eta_k'') (\delta_{(\eta_k, \eta_k')} + \delta_{(\eta_k', \eta_k)}), \end{aligned}$$

$$(1.2) \quad \zeta_{ij} = r \quad \text{if} \quad \sum_{l=0}^{r-1} W(v_i, v_j, l) \leq U_{\{i,j\}} \leq \sum_{l=0}^r W(v_i, v_j, l),$$

$$g(\theta_j, \chi_{jk}) = r \quad \text{if} \quad \sum_{l=0}^{r-1} S(v_j, l) \leq \chi_{jk} \leq \sum_{l=0}^r S(v_j, l),$$

$$h(\eta''_k) = r \quad \text{if} \quad \sum_{l=0}^{r-1} I(l) \leq \eta''_k \leq \sum_{l=0}^r I(l),$$

where  $(U_{\{i,j\}})_{i,j \geq 1}$  is a collection of independent uniform[0,1] random variables,  $\{(\theta_j, v_j)\}_{j \geq 1}$ ,  $\{(\chi_{jk}, \sigma_{jk})\}_{k \geq 1}$  for all  $j \geq 1$  are unit rate Poisson point processes on  $\mathbb{R}_+^2$ , and  $(\eta_k, \eta'_k, \eta''_k)_{k \geq 1}$  is a unit rate Poisson point processes on  $\mathbb{R}_+^3$ , where all the above Poisson point processes are independent of each other and  $(U_{\{i,j\}})_{i,j \geq 1}$ .

REMARK 1. It is not too hard to check that the measure  $\xi_{\mathcal{W}}$  introduced above is a.s. locally finite and defines an exchangeable random adjacency measure. This raises the question whether every exchangeable random adjacency measure necessarily corresponds to some (possibly random) multigraphex. For graphexes, this result was established in [42], Theorem 4.7. Their proof crucially uses a local finiteness criterion from Kallenberg [31], Proposition 9.25, which however turns out to not be quite correct. Indeed, the conditions for local finiteness in this proposition need to be supplemented by an extra condition, which was implicitly assumed by Kallenberg in his proof (as well as in the proof of [42], Theorem 4.7). We discuss this issue in [15], where we state and prove the corrected proposition, and then apply it to prove the analogue of [42], Theorem 4.7, for multigraphexes, stating that for any exchangeable random adjacency measure  $\xi$ , there exists a random multigraphex  $\mathcal{W}$  such that  $\xi \stackrel{d}{=} \xi_{\mathcal{W}}$ .

The adjacency measure  $\xi_{\mathcal{W}}$  associated with the multigraphex  $\mathcal{W} = (W, S, I)$  naturally defines a  $\mathfrak{G}_f$  valued stochastic process, as we define next. For a point process  $\xi$ , let us denote by  $\xi|_A$  the measure  $\xi$  restricted to  $A$ .

DEFINITION 6 (Multigraphex process). For any given multigraphex  $\mathcal{W}$ , we define the *multigraphex process generated by  $\mathcal{W}$*  as the  $\mathfrak{G}_f$ -valued stochastic process  $(GP_t(\mathcal{W}))_{t \geq 0}$  where  $GP_t(\mathcal{W}) = \mathcal{G}(\xi_{\mathcal{W}}|_{[0,t]^2})$ .

REMARK 2. There is an equivalent, albeit operationally slightly simpler description of the distribution of  $GP_t(\mathcal{W})$ . Indeed, it can be obtained by considering a single Poisson process  $\{v_j\}_{j \geq 1}$  of rate  $t$  on  $\mathbb{R}_+$ , and then adding edges according to the following procedure:

- ▷ for  $i \neq j$ , connect  $v_i$  and  $v_j$  with  $n_{ij}$  edges, where  $\mathbb{P}(n_{ij} = r) = W(v_i, v_j, r)$ ;
- ▷ for each  $j$ , add  $n_j$  self-loops to  $v_j$ , where  $\mathbb{P}(n_j = r) = W(v_j, v_j, r)$ ;
- ▷ for each  $j$  add a multistar to  $v_j$  by adding edges of multiplicity  $r$  at a rate  $tS(v_j, r)$ ;
- ▷ add isolated edges of multiplicity  $r$  with rate  $t^2I(r)$ .

Discard all isolated vertices (as well as all labels), and output the resulting unlabeled graph.

Finally, we define sampling convergence of a sequence of multigraphs to a multigraphex.

DEFINITION 7 (Convergence to multigraphex). A sequence  $(G_n)_{n \geq 1}$  of (multi)graphs is said to converge to a (multi)graphex  $\mathcal{W}$  if for all  $t > 0$ ,  $(\text{Smpl}(G_n, t/\sqrt{2e(G_n)}))_{n \geq 1}$  converges to  $GP_t(\mathcal{W})$  in distribution.

Of course, it is not clear whether sampling convergence in the sense of Definition 2 implies convergence to a multigraphex. To address this question, we first introduce an equivalent characterization of sampling convergence for multigraphs.

**DEFINITION 8 (Random labeling).** A labeling of a multigraph  $G$  into  $[0, s]$ , denoted by  $\text{Lbl}_s(G)$ , is a point process generated as follows: For a collection  $(U_v)_{v \in V(G)}$  of independent and identically distributed uniform $[0, s]$ , associate weight  $U_v$  to vertex  $v$ . Then  $\text{Lbl}_s(G) := \sum_{v,w \in V(G)} n_{vw} \delta_{(U_v, U_w)}$ , where  $n_{vw}$  denotes the number of edges between vertices  $v$  and  $w$ . The canonical labeling of  $G$ , denoted by  $\text{Lbl}(G)$ , refers to the case  $s = \sqrt{2e(G)}$ .

**PROPOSITION 1.1.** Consider a sequence of multigraphs  $(G_n)_{n \geq 1}$  with  $e(G_n) < \infty$  for all  $n \geq 1$  and  $\lim_{n \rightarrow \infty} e(G_n) = \infty$ . Then the following are equivalent:

- (i)  $(G_n)_{n \geq 1}$  is sampling convergent.
- (ii)  $(\text{Lbl}(G_n))_{n \geq 1}$  converges in distribution as random variables in  $\mathcal{N}(\mathbb{R}_+^2)$ .

Moreover, if the distributional limits of  $(\text{Smpl}(G_n, r/\sqrt{2e(G_n)}))_{n \geq 1}$  and  $(\text{Lbl}(G_n))_{n \geq 1}$  are given by  $H_r$  and  $\xi$ , then  $\text{Lbl}_r(H_r) \stackrel{d}{=} \xi|_{[0,r]^2}$ . Further,  $\xi$  is extremal. Therefore, there exists a multigraphex  $\mathcal{W}$  (nonrandom) such that  $\xi \stackrel{d}{=} \xi_{\mathcal{W}}$ , and  $(G_n)_{n \geq 1}$  is sampling convergent to  $\mathcal{W}$ .

The above proposition says that the limit of any sampling convergent sequence of graphs must be a multigraphex. For simple graphs, an analogue of Proposition 1.1 was proved in [12], Section 3, relying in turn on [42], Theorem 4.7. Modulo the issues mentioned in Remark 1, the proof in the multigraph setting is very similar. We provide an outline in Appendix A.

### 1.4. Limits of random graphs.

*Configuration model.* The configuration model is the canonical model for generating a random multigraph with a prescribed degree sequence. This model was introduced by Bollobás [5] to choose a uniform simple  $d$ -regular graph on  $n$  vertices, when  $dn$  is even. The idea was later generalized for general degree sequences  $\mathbf{d}$  by Molloy and Reed [38] and others (see [24]). Consider a sequence  $\mathbf{d} = (d_1, d_2, \dots, d_n)$  such that  $\ell_n = \sum_{i \in [n]} d_i$  is even, where  $[n] = \{1, \dots, n\}$ . Equip vertex  $j$  with  $d_j$  stubs or half-edges. Two half-edges create an edge once they are paired. Therefore, initially there are  $\ell_n = \sum_{i \in [n]} d_i$  half-edges. Pick any one half-edge and pair it with a uniformly chosen half-edge from the remaining unpaired half-edges. Keep repeating the above procedure until all the unpaired half-edges are exhausted. The random graph constructed in this way is called the configuration model, and will henceforth be denoted by  $\text{CM}_n(\mathbf{d})$ . Note that the graph constructed by the above procedure may contain self-loops and multiple edges. We define erased configuration model to be the graph obtained by collapsing all the multiple edges to single edges. We denote this graph by  $\text{ECM}_n(\mathbf{d})$ . This is slightly different compared to the erased configuration model in [24] since instead of deleting the loops, we merge multiple loops into single loops.

We will want to study the sampling limit of a sequence of configuration models, given in terms of a sequence of sequences of  $(\mathbf{d}_n)_{n \geq 1}$ , but for notational convenience, we suppress the index  $n$  of the sequence  $\mathbf{d}_n$ , and just speak of the limit of a sequence  $\text{CM}_n(\mathbf{d})$  random graphs. Since isolated vertices are removed in the process of sampling, we will assume without loss of generality that  $d_i > 0$  for all  $i$ . The following quantities determine this limit:

$$\rho_n := \frac{1}{\sqrt{\ell_n}} \sum_{i \in [n]} \delta_{\frac{d_i}{\sqrt{\ell_n}}}, \quad b_n = \int_0^\infty (x \wedge 1) \rho_n(dx),$$

with  $\rho_n$  considered as a measure on  $(0, \infty)$ . Throughout this discussion, we will assume that  $\max_{1 \leq i \leq n} d_i = o(\ell_n)$ . Note that the assumption  $d_i > 0$  trivially implies that  $\ell_n \geq n$ . This restriction is purely technical, and might possibly be relaxed. We do not pursue this in this paper.

Next, we introduce the limiting graphex for a sequence of configuration models. Given any measure  $\rho$  on  $(0, \infty)$  satisfying  $\int_0^\infty (x \wedge 1) d\rho(x) < \infty$ , we use  $\bar{\rho}$  to denote the tail Lévy intensity

$$\bar{\rho}(x) = \int_x^\infty \rho(dy).$$

Defining  $\bar{\rho}^{-1}$  as its inverse,  $\bar{\rho}^{-1}(y) := \inf\{x \in \mathbb{R}_+ : \bar{\rho}(x) \leq y\}$  we note that  $\bar{\rho}^{-1}$  is a càdlàg function from  $(0, \infty)$  to  $\mathbb{R}_+$ . It will be convenient to extend  $\bar{\rho}^{-1}$  to a function defined on  $\mathbb{R}_+$  by setting  $\bar{\rho}^{-1}(0) = 0$ . Finally, let  $p(k; \lambda)$  be the probability that a Poisson  $\lambda$  random variable takes the value  $k$ ,  $p(k; \lambda) = e^{-\lambda} \lambda^k / k!$ . For any  $a \in \mathbb{R}_+$  and any measure  $\rho$  on  $(0, \infty)$  satisfying  $\int_0^\infty (x \wedge 1) d\rho(x) < \infty$ , we define

$$(1.3) \quad W_{\text{CM}}(x, y, k) = \begin{cases} p(k; \bar{\rho}^{-1}(x)\bar{\rho}^{-1}(y)), & x \neq y, \\ p(k; \bar{\rho}^{-1}(x)^2/2), & x = y, \end{cases}$$

$$S_{\text{CM}}(x) = a\bar{\rho}^{-1}(x), \quad I_{\text{CM}} = \frac{a^2}{2}.$$

It is easy to see, by direct computation, that the graphex  $\mathcal{W}_{\text{CM}} = (W_{\text{CM}}, S_{\text{CM}}, I_{\text{CM}})$  satisfies the integrability criteria in Definition 4, and thus the associated random adjacency measure is locally finite. The multigraphex defined by (1.3) is closely related to the directed multigraph in Caron and Fox (2017); see [19], (12).

The following result derives necessary and sufficient conditions for the sampling convergence of  $\text{CM}_n(\mathbf{d})$  random graphs, and characterizes the limiting objects. To state our theorem, we introduce two random objects defined in terms of the sequence  $\mathbf{d} = (d_1, \dots, d_n)$  and the associated càdlàg function  $\bar{\rho}_n^{-1}$ : a càdlàg process  $(Y_n(t))_{t \geq 0}$  with

$$Y_n(t) = \frac{1}{\sqrt{\ell_n}} \sum_{i \in [n]} d_i \mathbb{1}\{U_i \leq t\},$$

where  $(U_i)_{i \in [n]}$  is an i.i.d. sequence of random variables  $U_i \sim \text{uniform}[0, \sqrt{\ell_n}]$ , and a completely random measure

$$(1.4) \quad \mu_n = \sum_{i \geq 1} \bar{\rho}_n^{-1}(v_i) \delta_{\theta_i},$$

where  $\{(v_i, \theta_i)\}_{i \geq 1}$  is a unit rate Poisson point process on  $\mathbb{R}_+^2$ .

We write  $\mathbb{P}_{\text{CM}}$  to denote the product measure  $\otimes_{n \geq 1} \mathbb{P}_n$ . The probability measure  $\mathbb{P}_{\text{ECM}}$  is defined analogously.

**THEOREM 1.2.** *The following are equivalent:*

- (i)  $\text{CM}_n(\mathbf{d})$  is sampling convergent a.s.  $\mathbb{P}_{\text{CM}}$ .
- (ii) There exists a random measure  $\mu$  such that  $\mathcal{L}(\mu_n)$  converges to  $\mathcal{L}(\mu)$  in  $\mathcal{P}(\mathcal{M}(\mathbb{R}_+))$ .
- (iii) There exists  $b$  and  $\rho$  such that  $b_n \rightarrow b$  and  $\rho_n \rightarrow \rho$  vaguely.
- (iv)  $(Y_n(t))_{t \geq 0}$  converges in distribution in  $\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+)$ .

Moreover, if (iii) holds, then  $\int_0^\infty (x \wedge 1) \rho(dx) \leq b \leq 1$ , and  $\mu$  is of the form (1.1) with  $a = b - \int (x \wedge 1) \rho(dx)$ . In this case,  $(Y_n(t))_{t \geq 0}$  converges in distribution to  $(Y(t))_{t \geq 0}$ , where  $Y(t) = \mu([0, t])$ , and  $\text{CM}_n(\mathbf{d})$  is sampling convergent to the multigraphex  $\mathcal{W}_{\text{CM}} = (W_{\text{CM}}, S_{\text{CM}}, I_{\text{CM}})$ , almost surely  $\mathbb{P}_{\text{CM}}$ , where  $\mathcal{W}_{\text{CM}}$  is defined in (1.3).

**REMARK 3.** The graphon limit of dense configuration models was derived in [41]. It is easy to see that Theorem 1.2 recovers the dense graph limit in case  $\ell_n = \Theta(n^2)$ . Ráth and Szakács [41] establish left convergence by explicitly computing the expected subgraph densities. In the sparse regime, this proof technique is no longer applicable.

REMARK 4. When proving the equivalence of (ii) and (iii), we will at the same time prove that under the condition (iii) from the above theorem,

$$(1.5) \quad a = \lim_{\varepsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \int_0^\varepsilon x \rho_n(dx) = \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \int_0^\varepsilon x \rho_n(dx)$$

(in fact, we will show that the condition  $b_n \rightarrow b$  in (iii) could be equivalently be replaced by the condition that the second equality in (1.5) holds). Since  $\int_0^\varepsilon x \rho_n(dx) = \frac{1}{\ell_n} \sum_i d_i 1_{d_i \leq \varepsilon \sqrt{\ell_n}}$ , the constant  $a$  therefore represents the limiting fraction of half-edges with degrees  $d_i = o(\sqrt{\ell_n})$ , and the condition  $b_n \rightarrow b$  is the condition that this limiting fraction exists.

REMARK 5. Let  $\xi_{\text{CM}}$  be the random adjacency measure associated to the multigraphex  $\mathcal{W}_{\text{CM}}$ , and let  $\xi_{\text{CM}}^* := \xi_{\text{CM}}|_{(x,y):y \leq x}$ . Then  $\xi_{\text{CM}}$  has the following ‘‘rank-one’’ structure: For any disjoint set  $A, B \in \mathcal{B}(\mathbb{R}_+)$  of finite measure, the distribution of  $\xi_{\text{CM}}^*(A \times A)$  is Poisson( $\mu(A)^2/2$ ) and that of  $\xi_{\text{CM}}(A \times B)$  is Poisson( $\mu(A)\mu(B)$ ). See Appendix C for the precise statement and proof.

We obtain the following corollaries from Theorem 1.2. Define the graphon  $W_{\text{ECM}}$  by

$$W_{\text{ECM}}(x, y) = \begin{cases} 1 - e^{-\bar{\rho}^{-1}(x)\bar{\rho}^{-1}(y)}, & x \neq y, \\ 1 - e^{-\bar{\rho}^{-1}(x)^2/2}, & x = y, \end{cases}$$

as well as a rescaled graphex  $\mathcal{W}_{\text{ECM}}^c := (W_{\text{ECM}}(\sqrt{c}\cdot, \sqrt{c}\cdot), \frac{1}{\sqrt{c}}S_{\text{CM}}(\sqrt{c}\cdot), \frac{1}{c}I_{\text{CM}})$ . Recall from [12] that any sequence of simple graphs with  $\#\text{loops} = O(\sqrt{\#\text{edges}})$  has a convergent subsequence. The following corollary characterizes possible limit points for  $\text{ECM}_n(\mathbf{d})$  under mild regularity conditions on the degree distribution.

COROLLARY 1.3. *Suppose that  $\rho_n \rightarrow \rho$  vaguely,  $b_n \rightarrow b \in \mathbb{R}_+$ . Further, assume that*

$$(1.6) \quad \int_0^\infty \int_0^\infty (1 - e^{-xy})\rho_n(dx)\rho_n(dy) \rightarrow c \quad \text{as } n \rightarrow \infty$$

for some  $0 < c < \infty$ . Then, as  $n \rightarrow \infty$ ,  $\text{ECM}_n(\mathbf{d})$  is sampling convergent to the graphex  $\mathcal{W}_{\text{ECM}}^c$  a.s.  $\mathbb{P}_{\text{ECM}}$ . Moreover, if the left-hand side of (1.6) is bounded away from zero, then the limit of any a.s.  $\mathbb{P}_{\text{ECM}}$  convergent subsequence of  $\text{ECM}_n(\mathbf{d})$  is of the form  $\mathcal{W}_{\text{ECM}}^C$ , for some constant  $C > 0$ .

As a further consequence of Theorem 1.2, we study when the limit is a pure graphon or purely isolated edges. To this end, we define the uniform tail regularity for a sequence of multigraphs.

DEFINITION 9. For a vertex  $G$ , let  $d_v(G)$  denote the degree of vertex  $v$ . A sequence of (multi)graphs  $(G_n)_{n \geq 1}$  is uniformly tail regular if for any  $\varepsilon > 0$ , there exists  $\delta > 0$  such that for all  $n \geq 1$ ,

$$\frac{1}{e(G_n)} \sum_{v: d_v \leq \delta \sqrt{e(G_n)}} d_v(G_n) < \varepsilon.$$

Note that Definition 9 is equivalent to [10], Definition 13, for simple graphs (see [10], Remark 14, and [11], Lemma 9.3). Also, recall the definition of stretched cut metric from [10].



COROLLARY 1.4. Assume that  $\rho_n \rightarrow \rho, b_n \rightarrow b \in \mathbb{R}_+$ . Then  $I_{CM} = 0$  if and only if  $CM_n(\mathbf{d})$  is uniformly tail regular a.s.  $\mathbb{P}_{CM}$ . Moreover, if  $I_{CM} = 0$  and (1.6) holds, then  $ECM_n(\mathbf{d})$  is uniformly tail regular a.s.  $\mathbb{P}_{ECM}$ . In this case,  $ECM_n(\mathbf{d})$  converges to  $W_{ECM}^c$  in the stretched cut metric a.s.  $\mathbb{P}_{ECM}$ .

COROLLARY 1.5. As  $n \rightarrow \infty, CM_n(\mathbf{d})$  is sampling convergent a.s.  $\mathbb{P}_{CM}$  to  $(0, 0, I_{CM})$  if and only if  $b_n \rightarrow b, \rho_n \rightarrow \rho$  vaguely, with  $\rho = 0$ .

*Preferential attachment model.* We consider a generalization of the preferential attachment model. This model was first introduced by Pittel [40], and the graph limit in the dense counter part of this model was studied in [8, 41]. Let  $\kappa = (\kappa_i)_{i \in [n]}$  be a sequence of nonnegative real numbers, and let  $\ell_{n,\kappa} = \sum_{i \in [n]} \kappa_i$ . Initially,  $PA_n(\kappa, 0)$  is an empty graph on vertex set  $[n]$ . Let  $d_i(l)$  denote the degree of vertex  $i$  in  $PA_n(\kappa, l)$ . Given the graph  $PA_n(\kappa, l)$  at time  $l$ ,  $PA_n(\kappa, l + 1)$  is created by adding one edge to the graph with the end points being chosen with probability proportional to  $d_i(l) + \kappa_i$ . More precisely, the edge  $(i, j)$  is added at step  $l$  with probability

$$\begin{cases} \frac{2(d_i(l) + \kappa_i)(d_j(l) + \kappa_j)}{(\ell_{n,\kappa} + 2l)^2} & \text{for } i \neq j, \\ \frac{(d_i(l) + \kappa_i)^2}{(\ell_{n,\kappa} + 2l)^2} & \text{for } i = j. \end{cases}$$

The above process iterates  $m_n$  times to yield  $PA_n(\kappa, m_n)$ . Throughout, we will assume that  $\max_i \kappa_i = o(\ell_{n,\kappa})$  and  $\min\{\ell_{n,\kappa}, m_n\} = \omega(\log n)$ .

When  $(\kappa_i)_{i \in [n]}$  are integers, the above process can simply be described by an urn scheme where we start with an urn with  $\kappa_i$  balls of color  $i$ , for  $i \in [n]$ . At step  $l$ , we select two balls with replacement from the urn. If the colors of the chosen balls are  $i$  and  $j$ , we add an edge between vertices  $i$  and  $j$ . We also add one additional copy of the balls in the urn with color  $i$  and  $j$  before the next iteration.

REMARK 6. The model in [8, 41] is slightly different in the sense that the  $l$ th edge is formed by first drawing the  $(2l - 1)$ th ball from the urn, replacing that ball in the urn and then drawing the  $2l$ th, instead of drawing two balls together at step  $l$ . However, this does not change the limiting result for the preferential attachment model.

The next theorem states that the limit for preferential attachment model. As we will see, as long as  $m_n = o(\ell_{n,\kappa}^2)$ , the model behaves very much like a configuration model with degree sequence  $(\bar{d}_i)_{i \in [n]}$  where  $\bar{d}_i$  is the expected degree of  $i$  at time  $m_n$ ,

$$\bar{d}_i = \mathbb{E}[d_i(m_n)] = \frac{2m_n}{\ell_{n,\kappa}} \kappa_i.$$

To formalize this, we set

$$\rho_{n,\kappa} := \frac{1}{\sqrt{2m_n}} \sum_{i \in [n]} \delta_{\frac{\bar{d}_i}{\sqrt{2m_n}}}, \quad b_{n,\kappa} = \int_0^\infty (x \wedge 1) \rho_{n,\kappa}(dx),$$

define  $\mu_{n,\kappa}$  in terms of  $\bar{\rho}_{n,\kappa}^{-1}$  instead of  $\bar{\rho}_n^{-1}$  similarly as in (1.4), and define

$$Y_{n,\kappa}(t) = \frac{1}{\sqrt{2m_n}} \sum_{i \in [n]} \bar{d}_i \mathbb{1}\{U_i \leq t\} = \frac{\sqrt{2m_n}}{\ell_{n,\kappa}} \sum_{i \in [n]} \kappa_i \mathbb{1}\{U_i \leq t\},$$

where  $(U_i)_{i \in [n]}$  is an i.i.d. sequence of uniform random variables in  $[0, \sqrt{2m_n}]$ . We will use sequences  $(\kappa^n)_{n \geq 1}$  and define the measure  $\mathbb{P}_{PAM}$  analogous to  $\mathbb{P}_{CM}$ .

**THEOREM 1.6.** *Assume that  $m_n = o(\ell_{n,\kappa}^2)$ . Then  $\text{PA}_n(\kappa, m_n)$ ,  $\mu_{n,\kappa}$ ,  $b_{n,\kappa}$ ,  $\rho_{n,\kappa}$ , and  $(Y_{n,\kappa}(t))_{t \geq 0}$  satisfies the the class of equivalent statements (i)–(iv) in Theorem 1.2 a.s.  $\mathbb{P}_{\text{PAM}}$ . Moreover, if  $\rho_{n,\kappa} \rightarrow \rho$  and  $a = \lim_{n \rightarrow \infty} \int (x \wedge 1) \rho_{n,\kappa}(\text{d}x) - \int (x \wedge 1) \rho(\text{d}x)$ , then  $\text{PA}_n(\kappa, m_n)$  is  $\mathbb{P}_{\text{PAM}}$  almost surely sampling convergent to the graphex  $\mathcal{W}_{\text{CM}}$  is defined in (1.3).*

**REMARK 7.** Consider the case where  $(\kappa_i)_{i \in [n]}$  is a collection of positive integers with  $\ell_{n,\kappa}$  being even. Then, if  $\ell_{n,\kappa}/2m_n \rightarrow 1$ , the sampling limits of  $\text{PA}_n(\kappa, m_n)$  and  $\text{CM}_n(\kappa)$  are identical. Further, if  $\kappa_i = 1$  for all  $i \in [n]$ , then  $\rho$  is the zero measure and  $b = 1$ , corresponding to the limiting graphex  $\mathcal{W}_{\text{CM}} = (0, 0, 1/2)$  and a sampling limit consisting of just isolated edges.

**REMARK 8.** Borgs et al. [8] and Ráth and Szakács [41] derived the graphon limit for this model in the setting  $m_n = \Theta(\ell_{n,\kappa}^2)$ , and  $\kappa_i = \kappa$  for all  $i \in [n]$ , where  $\kappa$  is a universal constant independent of  $n$ . A comparison between their results and Theorem 1.6 shows that the limiting graphons are very different. In particular, a naive extrapolation based on Theorem 1.6 turns out to be incorrect. Intuitively, this discrepancy is explained by nontrivial fluctuations of the vertex degrees around their expectations. As a result, the measure  $\rho_{n,\kappa}$  introduced above, does not adequately capture the degree characteristics in the dense setting. As a consequence, this establishes that our assumption  $m_n = o(\ell_{n,\kappa}^2)$  in Theorem 1.6 is, in fact, optimal.

*Generalized random graph.* Given a weight sequence  $(w_i)_{i \in [n]}$ , the generalized random graph model, denoted by  $\text{GRG}_n(\mathbf{w})$ , is obtained by connecting vertices  $i$  and  $j$  independently with probability

$$(1.7) \quad p_{ij} = \frac{w_i w_j}{L_n + w_i w_j},$$

where  $L_n = \sum_{i \in [n]} w_i$ . Throughout, we will assume  $L_n = \omega(\log n)$ . This model has been of considerable theoretical interest since, conditionally on the degree sequence  $\mathbf{d}$ , this gives a uniformly chosen simple graph with degree sequence  $\mathbf{d}$  [18, 24]. This is also related to the  $\beta$ -model, studied in [20]. The generalized random graph model falls into the larger class of inhomogeneous random graphs, studied in the seminal work of Bollobás et al. [6], where the phase transition in the connectivity structure was characterized for the finite average degree regime.

**ASSUMPTION 1.**

- (i)  $\rho_{n,w} = \frac{1}{\sqrt{L_n}} \sum_{i \in [n]} \delta_{\frac{w_i}{\sqrt{L_n}}}$  converges vaguely to some measure  $\rho_w$ .
- (ii)  $\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \int_0^\varepsilon x \rho_{n,w}(\text{d}x) = \lim_{\varepsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \int_0^\varepsilon x \rho_{n,w}(\text{d}x) = a$  for some constant  $a$ .
- (iii)  $\liminf_{n \rightarrow \infty} \int_0^\infty \int_0^\infty \frac{xy}{1+xy} \rho_n(\text{d}x) \rho_n(\text{d}y) > 0$ .

Define the graphex  $\mathcal{W}_{\text{GRG}}^C = (W_{\text{GRG}}^C, S_{\text{GRG}}^C, I_{\text{GRG}}^C)$ , given by

$$W_{\text{GRG}}^C(x, y) = \begin{cases} \frac{\bar{\rho}_w^{-1}(\sqrt{C}x) \bar{\rho}_w^{-1}(\sqrt{C}y)}{1 + \bar{\rho}_w^{-1}(\sqrt{C}x) \bar{\rho}_w^{-1}(\sqrt{C}y)} & \text{if } x \neq y, \\ 0 & \text{o.w.,} \end{cases}$$

$$S_{\text{GRG}}^C(x) = \frac{a}{\sqrt{C}} \bar{\rho}_w^{-1}(\sqrt{C}x), \quad I_{\text{GRG}}^C = \frac{a^2}{2C}.$$

We will use sequences  $(\mathbf{w}_n)_{n \geq 1}$  and suppress the dependence on  $n$  for notational convenience. Further, we use  $\mathbb{P}_{\text{GRG}}$  to denote the joint distribution of sequences of  $\text{GRG}_n(\mathbf{w})$  random graphs, where the graphs are sampled independently for each  $n$ .

THEOREM 1.7. *Suppose that  $\text{GRG}_n(\mathbf{w})$  satisfies Assumption 1, and*

$$(1.8) \quad \lim_{n \rightarrow \infty} \int_0^\infty \int_0^\infty \frac{xy}{1+xy} \rho_n(dx) \rho_n(dy) = c > 0.$$

*Then, as  $n \rightarrow \infty$ ,  $\text{GRG}_n(\mathbf{w})$  is sampling convergent to  $\mathcal{W}_{\text{GRG}}^c$  a.s.  $\mathbb{P}_{\text{GRG}}$ .*

COROLLARY 1.8. *If  $\text{GRG}_n(\mathbf{w})$  satisfies Assumption 1, then the limit of any a.s.  $\mathbb{P}_{\text{GRG}}$  convergent subsequence of  $\text{GRG}_n(\mathbf{w})$  is of the form  $\mathcal{W}_{\text{GRG}}^C$ , for some constant  $C > 0$ .*

COROLLARY 1.9. *For  $a = 0$ ,  $\text{GRG}_n(\mathbf{w})$  converges to  $\mathcal{W}_{\text{GRG}}^c$  in the stretched cut metric a.s.  $\mathbb{P}_{\text{GRG}}$ .*

*Bipartite configuration model.* In this section, we describe the sampling limit of bipartite configuration models. Let us introduce the model first. Consider two sets of vertices  $V_1$  and  $V_2$  with associated degree sequences  $\mathbf{d} = (d_{ij})_{i \in V_j, j=1,2}$  such that  $\sum_{i \in V_1} d_{i1} = \sum_{i \in V_2} d_{i2} := \ell_n/2$ . Equip the  $i$ th vertex in  $V_j$  with  $d_{ij}$  half-edges. A bipartite configuration model is generated by sequentially selecting unpaired half-edges one-by-one from  $V_1$ , and pairing it with a uniformly chosen unpaired half-edge from  $V_2$ . Replacing the paired half-edges by edges, one gets a bipartite random graph, which we denote by  $\text{BipCM}_n(\mathbf{d})$ . The probability measure  $\mathbb{P}_{\text{BCM}}$  is defined analogous to  $\mathbb{P}_{\text{CM}}$ .

Given a degree sequence  $\mathbf{d}$ , the following quantities determine the limit of  $\text{BipCM}_n(\mathbf{d})$ : For  $j = 1, 2$ , define

$$\rho_{nj} := \frac{1}{\sqrt{\ell_n}} \sum_{i \in V_j} \delta_{\frac{d_{ij}}{\sqrt{\ell_n}}}, \quad b_{nj} := \int_0^\infty (x \wedge 1) \rho_{nj}(dx).$$

Throughout, we will assume that  $\max_{1 \leq i \leq n, j=1,2} d_{ij} = o(\ell_n)$ . We do not try to relax this restriction here.

Next, we introduce the limiting graphex for  $\text{BipCM}_n(\mathbf{d})$ . In this case, we consider the feature space  $\Omega = \mathbb{R}_+ \times \{0, 1\}$ . We equip  $\{0, 1\}$  with the counting measure, and the feature space  $\Omega$  naturally inherits the product measure on the two component spaces. For measures  $\rho_j$  on  $\mathbb{R}_+$  satisfying  $\int_0^\infty (x \wedge 1) d\rho_j(x) < \infty$ , we define

$$(1.9) \quad W_{\text{BCM}} \left( \binom{x}{c_1}, \binom{y}{c_2}, k \right) = \begin{cases} p(k; \bar{\rho}_1^{-1}(x) \bar{\rho}_2^{-1}(y)), & c_1 \neq c_2, \\ 0, & c_1 = c_2, \end{cases}$$

$$S_{\text{BCM}} \left( \binom{x}{c} \right) = \begin{cases} a_2 \bar{\rho}_1^{-1}(x), & c = 0, \\ a_1 \bar{\rho}_2^{-1}(x), & c = 1, \end{cases} \quad I_{\text{BCM}} = a_1 a_2.$$

The extra coordinate  $\{0, 1\}$  in the feature space encodes the partition of a sampled vertex. Two vertices in the same partition cannot share an edge, and thus the graphon is zero whenever  $c_1 = c_2$ . Note that the graphex in (1.9) is closely related to the bipartite graph model in [19], Section 3.4.

The following result derives necessary and sufficient conditions for the sampling convergence of  $\text{BipCM}_n(\mathbf{d})$  random graphs, and characterizes the limiting objects. We note that while the proof of this result is related to that of Theorem 1.2 and the associated corollaries, these results help provide further intuition into sampling convergence, and provide interesting examples of possible limits that may be obtained under this notion of convergence.

THEOREM 1.10. *The following are equivalent:*

- (i)  $\text{BipCM}_n(\mathbf{d})$  is sampling convergent a.s.  $\mathbb{P}_{\text{BCM}}$ .

(ii) For  $j = 1, 2$ , there exists  $b_j$  and  $\rho_j$  such that  $b_{n_j} \rightarrow b_j$  and  $\rho_{n_j} \rightarrow \rho_j$  vaguely.

If (ii) holds, then  $\int_0^\infty (x \wedge 1)\rho_j(dx) \leq b_j$  and  $\text{BipCM}_n(\mathbf{d})$  is sampling convergent to the multigraphex  $\mathcal{W}_{\text{BCM}}$ , almost surely  $\mathbb{P}_{\text{BCM}}$ , with  $a_j = b_j - \int (x \wedge 1)\rho_j(dx)$ .

We observe here that the bipartite structure allows for interesting sampling limits as described below. A comparison between Theorem 1.2 and Theorem 1.10 shows that in the nonbipartite model,  $S_{\text{CM}} \neq 0$  implies that both  $W_{\text{CM}} \neq 0$  and  $I_{\text{CM}} \neq 0$ , while this is not necessarily true in the bipartite case.

REMARK 9. In the special case  $\rho_2 = 0$ ,  $\rho_1 \neq 0$ ,  $a_1 = 0$  and  $a_2 \neq 0$ , the corresponding limit  $\mathcal{W}_{\text{BCM}} = (0, S_{\text{BCM}}, 0)$ . This further illustrates that a configuration model type construction might also yield graphexes with pure star part.

REMARK 10. For degree sequences with  $\rho_1 = 0$ ,  $\rho_2 \neq 0$ ,  $a_1, a_2 \neq 0$ , we have a sampling limit with  $W_{\text{BCM}} = 0$  while  $S_{\text{BCM}} \neq 0$  and  $I_{\text{BCM}} \neq 0$ . Finally, if  $a_1 = 0$ ,  $a_2 \neq 0$  and  $\rho_1, \rho_2 \neq 0$ , the limiting graphex is of the form  $\mathcal{W}_{\text{BCM}} = (W_{\text{BCM}}, S_{\text{BCM}}, 0)$ .

### 1.5. Discussion.

*Background.* Diaconis and Janson [23], and Austin [2] identified a beautiful connection between the theory of graph limits, and convergence of exchangeable random arrays. For dense graphs, the notion of Left convergence is characterized by the convergence of all subgraph densities. Equivalently, one may permute the vertex labels of a graph  $G_n$  uniformly at random, and study the properties of the resulting permuted adjacency matrix. In the limit, these permuted matrices converge weakly to infinite exchangeable random arrays, and their laws are characterized by the celebrated Aldous–Hoover theorem [1, 26]. Further, the limiting law of the array has a one-to-one correspondence with the limiting graphon for the dense graph sequence. However, in contrast, for sparse graph sequences, these matrices converge to the zero array, and this framework fails to provide nontrivial information about the graph sequence. Identifying the exchangeable structures that characterize the limits for sparse graphs remained an open question for a decade.

Caron and Fox [19] introduced a family of random graph models based on a completely random measure, and introduced a notion of exchangeability for dynamically growing random graphs, via the exchangeability of their adjacency measure on  $\mathbb{R}_+^2$ . Extending this idea, and using the Kallenberg representation theorem [30, 31] for exchangeable point processes on  $\mathbb{R}_+^2$  as a conceptual cornerstone, [42, 43] (see also [28] for a review and some extensions of the results of [43]) introduced a very general class of exchangeable random graph models. They further examine structural properties of these graphs, and address questions related to statistical inference under these generative models. In parallel, [10] generalizes graph limit theory by introducing the notion of convergence in *stretched cut metric* for a sequence of sparse graphs. Finally, [12] formalized the relation between convergence of sparse graphs, and the convergence of corresponding limiting adjacency measure by introducing the notion of *sampling convergence*, a generalization of Left convergence for sparse graph sequences, and established that the limiting adjacency measure correspond exactly to graphex processes in [42, 43]. Further, they establish that under the assumption of uniform tail regularity sampling convergence is equivalent to convergence under the stretched cut metric from [10].

In this paper, we utilize these recent advances to study structural properties of random graphs, while simultaneously establishing the usefulness of this nascent theory.

REMARK 11. In a recent paper, Borgs et al. [11] proposed and studied the weak kernel metric on graphexes. This metric generalizes the cut metric for graphons and metrizes sampling convergence without additional regularity conditions. Further, two graphexes at zero distance in this metric lead to identically distributed graphex processes, and graphexes are equivalent in this sense if and only if they can be related by measure preserving transformations. It would be interesting to provide an analogous metric for multigraphexes, but this is beyond the scope of this paper, and also somewhat orthogonal to our purpose here.

*Insights on the graph structure.* Recall the definition of a multigraphex (Definition 4). We take this opportunity to provide further intuition for the components of a multigraphex, and what they imply for the multigraph sequence converging to this multigraphex. A sequence  $(G_n)_{n \geq 1}$  of multigraphs with  $m_n = e(G_n) = o(n^2)$ , is composed of three main parts:

(1) A dense core where the vertices have degree  $\Theta(\sqrt{m_n})$ . If the dense part contributes a positive proportion of edges (i.e., there are  $\Theta(\sqrt{m_n})$  many vertices of degree  $\Theta(\sqrt{m_n})$ ) then this part gives rise to the graphon, and thus gives the leading contribution to the subgraph densities for subgraphs that are more complex than isolated edges or stars.

(2) A sparse part where the vertices have degree  $o(\sqrt{m_n})$ . For the purpose of this discussion, assume that the edges out of these vertices are simple. Then the probability that after sampling, the degree of fixed vertex  $i$  is two or larger can be upper bounded by  $p^3 d_i^2$ , where  $d_i$  is the degree of the vertex  $i$  before sampling; as a consequence, the expected number of low degree vertices which after sampling have degree at least two is bounded by  $\sum_i p^3 d_i^2 = o(\sqrt{m_n}) \sum_i p^3 d_i = o(\sqrt{m_n}) p^3 m_n = o(1)$ . This shows that after sampling the low degree vertices will either have degree one or become isolated. Therefore, edges within the sparse part will appear as isolated edges in  $\text{Smpl}(G_n, t/\sqrt{2m_n})$ , contributing to the isolated edge constant  $I$ .

(3) Connections between dense and sparse part. Since the surviving vertices in the sparse part have degree one after sampling, these edges contribute to the edge and star densities and thus they appear as stars or isolated edges in  $\text{Smpl}(G_n, t/\sqrt{2m_n})$ .

Note that the vertices of degree  $\Omega(\sqrt{m_n})$  do not contribute anything to the graphex limit due to the fact that the probability of such a vertex being observed in the sampling is  $o(1)$ . (Note that in general, when the edges out of the low degree vertices have nontrivial multiplicities, we could also get isolated multiedges as well as stars with edges that have multiplicity bigger than one).

A visualization of  $\text{Smpl}(G_n, t/\sqrt{2m_n})$  is given by Figure 1. The asymptotic structure of  $\text{Smpl}(G_n, t/\sqrt{2m_n})$  constitutes a network between the dense part described by  $W$ , stars centered at the high degree vertices described by  $S$  representing the edges between the dense and sparse parts, and isolated edges described by  $I$  arising from the sparse part.

*Heuristic derivation of sampling limit for the configuration model.* We start by noting that  $p$ -sampling with  $p = t/\sqrt{2e(G)}$  is asymptotically equivalent to what one might want to call Poisson sampling, defined by first choosing  $k$  according to a Poisson random variable with expectation  $t|V(G)|/\sqrt{2e(G)}$ , and then choosing  $k$  vertices from  $V(G)$ , uniformly at random with replacement, which in turn is equivalent to considering a Poisson process of rate  $p = t/\sqrt{2e(G)}$  on  $V(G)$ . In the simple graph setting, this follows from Lemma 5.4 in [12], but an inspection of the proof shows that the lemma holds in the multigraph setting considered here as well. We also note that for the configuration model,  $2e(G_n) = \ell_n(1 + o(1))$ ; see Lemma 2.1 below for the precise statement. Finally, we couple the Poisson process on  $V(G)$  to a Poisson process as the one in Remark 2, that is, a Poisson process  $(x_i)$  of rate  $t$  on  $\mathbb{R}_+$  by assigning adjacent intervals of width  $1/\sqrt{\ell_n}$  to each vertex. The degree of the vertex  $v$  corresponding

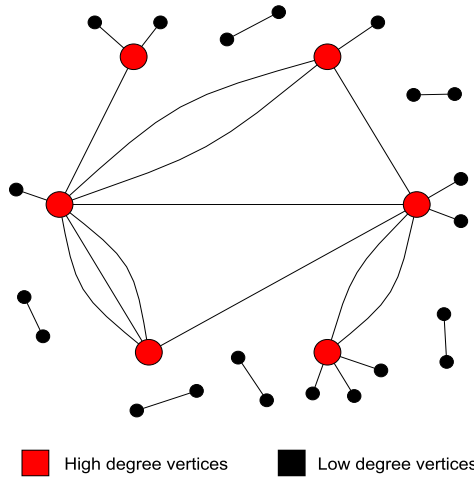


FIG. 1. Typical structure of  $\text{Smpl}(\text{CM}_n(\mathbf{d}), t/\sqrt{\ell_n})$ .

to  $x_i$  can then easily be seen to be equal to  $d_v = \bar{\rho}_n^{-1}(x_i)\sqrt{\ell_n}$ . Note also that  $\bar{\rho}_n^{-1}(x_i)\sqrt{\ell_n} = 0$  if  $x_i$  does not correspond to any vertex  $v \in [n]$ , that is, if  $x_i \notin [0, n/\sqrt{\ell_n}]$ .

Linking back to the above insights on the graph structure of the sampled graph, we next note that  $\int x\rho_n(dx) = \frac{1}{\ell_n} \sum_i d_i = 1$  and that

$$\int x\rho(dx) \leq \liminf_{n \rightarrow \infty} \int x\rho_n(dx) = 1$$

by Fatou’s lemma. We therefore interpret  $\int x\rho(dx)$  as the limiting fraction of (half)-edges whose endpoints have degrees of order  $\Theta(\sqrt{\ell_n})$ . Edges between vertices in this part therefore contributed to the graphon part of the limiting graphex. To “derive” the concrete form of this limiting graphon, we need one more fact, established in Proposition 2.2 below. It states that in the configuration model  $\text{CM}_n(\mathbf{d})$ , the number of edges created between two disjoint sets of half-edges  $S$  and  $S'$  of size  $O(\sqrt{\ell_n})$  is approximately distributed according to  $\text{Poisson}(|S||S'|/\ell_n)$ , and the number of edges connecting such a set  $S$  to itself is approximately distributed according to  $\text{Poisson}(|S|^2/2\ell_n)$ . Applied two Poisson points  $x_i, x_j$  such that the degrees of the corresponding vertices  $v, v'$  are of order  $\sqrt{\ell_n}$ , we then expect to see  $\text{Poisson}(d_v d_{v'}/\ell_n) = \text{Poisson}(\bar{\rho}_n^{-1}(x_i)\bar{\rho}_n^{-1}(x_j))$  many edges between  $i$  and  $j$ , and a loop of multiplicity  $\text{Poisson}(d_v^2/2\ell_n) = \text{Poisson}((\bar{\rho}_n^{-1}(x_i))^2)$  at the vertex  $i$ , explaining the form of the limiting graphon.

Next, observing that  $\int_M^\infty \rho(dx) \leq \frac{1}{M}$  and  $\int_M^\infty \rho_n(dx) \leq \frac{1}{M}$  by Markov’s inequality, we see that the high degree vertices don’t contribute to  $b_n$  or  $b$ , showing that  $a = \lim_{n \rightarrow \infty} b_n - b$  is the limiting fraction of half-edges belonging to low degree vertices. Considering the set of half-edges  $S_L$  attached to some low degree vertex, let  $S_L^p$  be the set of half-edges surviving after sampling. Then  $S_L$  contains approximately  $a\ell_n$  many half-edges, showing that  $|S_L^p|$  is approximately equal to  $pa\ell_n = ta\sqrt{\ell_n}$ . The number of full edges formed between these is then approximately equal to  $\text{Poisson}((ta\sqrt{\ell_n})^2/2\ell_n) = \text{Poisson}((ta)^2/2)$ , motivating the appearance of the term  $a^2/2$  in (1.3).

Finally, to derive the form of the star intensity  $S$  in (1.3), we consider the edges between the high and low degree vertices. A vertex  $v$  corresponding to a Poisson point  $x_i$  such that  $d_v = \bar{\rho}_n^{-1}(x_i)\sqrt{\ell_n}$  is of order  $\sqrt{\ell_n}$  then approximately has degree  $\text{Poisson}(d_v|S_L^p|/\ell_n) \approx \text{Poisson}(ta\bar{\rho}_n^{-1}(x_i))$  into  $S_L^p$ , explaining the appearance of the term  $S$  in (1.3).

To relate the results for the configuration model to those of the erased configuration model we use that a Poisson random variable with rate  $w$  is nonzero with probability  $1 - e^{-w}$ . This

in turn implies that asymptotically, the number of nonloop edges in the erased configuration model is by a factor  $c$  smaller than the number of nonloop edges in the original configuration model, with  $c$  given by (1.6). Since sampling convergence of a sequence  $G_n$  involves a random coin flip with probabilities  $p = t/\sqrt{2e(G_n)}$ , we have to rescale time by a factor  $\sqrt{c}$  when translating our results for the configuration model to that of the erased configuration model. This leads to the graphex  $\mathcal{W}_{\text{ECM}}^c$  in Corollary 1.3.

The limit for the bipartite configuration model can be motivated using analogous heuristics.

*Heuristic derivation of sampling limit for the preferential attachment model.* It turns out that the preferential attachment model behaves very much like a configuration model with degree sequence equal to the expected degrees at time  $m_n$ ,  $(\bar{d}_i)_{i \in [n]}$ . The proof details are different, with Proposition 2.2 replaced by Proposition 3.2 below as well as other differences in the details, but the essence will again be that we control the dependence of the number of edges between different sets of vertices and approximate them by suitable Poisson random variable, eventually giving the same limiting graphex as the configuration model with degree sequence  $(\bar{d}_i)_{i \in [n]}$ ; see also Remark 12 in Section 3 below.

*Heuristic derivation of the sampling limit for the generalized random graph.* It will be convenient to sample vertices with probability  $p' = t/\sqrt{L_n}$  rather than with probability  $p = 1/\sqrt{2m_n}$  where  $m_n$  is the number of nonloop edges in  $\text{GRG}_n(\mathbf{w})$ . It turns out that, asymptotically, this just corresponds to rescaling of time by a factor  $\sqrt{c}$ , a fact which follows from the observation that

$$\frac{1}{L_n} \mathbb{E}[2m_n] = \frac{1}{L_n} \sum_{i,j} \frac{w_i w_j}{L_n + w_i w_j} = \int_0^\infty \int_0^\infty \frac{xy}{1 + xy} \rho_n(dx) \rho_n(dy) = c + o(1)$$

(plus a concentration argument). This explains the rescaling by  $\sqrt{c}$  in  $\mathcal{W}_{\text{GRG}}^c$ , but obviously, not yet the particular form of the limiting graphex.

To derive the latter, we proceed very similar to our heuristic derivation for the configuration model, except that we now consider a core of vertices defined by the *weights* of the vertices. Specifically, we consider a core of vertices with weights  $w_i = \Theta(\sqrt{L_n})$  and a set of low-weight vertices with weights  $w_i = o(\sqrt{L_n})$ . It is then again not hard to argue that the low weight vertices will have degree at most 1 after sampling, and it is also clear that asymptotically, the sum of the weights of all low weight vertices is  $aL_n$  with  $a$  as in Assumption 1.

Furthermore, following the steps in our heuristic derivation of the limiting graphex for the configuration model, replacing the Poisson number of edges  $\text{Poisson}(d_v d_{v'} / \ell_n)$  between two vertices of degree  $d_i, d_j$  by  $\text{Bern}(p_{ij})$  (with  $p_{ij}$  given in (1.7)), the reader can now easily “derive” the form of the limiting graphon for  $\text{GRG}_n(\mathbf{w})$ . To obtain the other two parts of the limiting graphex, we approximate the probability (1.7) for an edge between two low weight vertices (or a low and high weight vertex) by  $w_i w_j / L_n$  and approximate the sum of independent Bernoulli random variables by a Poisson random variable; using these approximations, the “derivation” of the limiting graphex is now very similar to that for the limiting graphex for the configuration model.

*Relation to Caron–Fox graph process.* Corollary 1.4 establishes that the sampling limit of certain  $\text{ECM}_n(\mathbf{d})$  random graphs is given by the random graph model introduced by Caron and Fox [19] (see [19], Section 3). Thus our result gives a new perspective on the Caron–Fox random graph. Indeed, certain Caron–Fox graphs may be looked upon as sampling limits of suitable  $\text{ECM}_n(\mathbf{d})$  random graphs. [12] characterizes graphex processes as the limits of sampling convergent graph sequences, and thus conceptually clarifies the innate importance of these processes. Our result has a similar conceptual interpretation, in that it identifies a

prominent graphex process, that is, the Caron–Fox process, as the sampling limit of a natural sequence of random graphs. Put differently, rather than obtaining the model by first postulating exchangeability of a rather abstract random measure on  $\mathbb{R}_+^2$ , then invoking Kallenberg’s representation theorem and finally making further simplifications to arrive at the final model, our results *derive* the Caron–Fox graph as a subsample of an underlying latent configuration model. In turn, this further reinforces the importance of the Caron–Fox model, and provides some practical insights into its suitability as a model in real applications.

*Outline.* The rest of the paper is structured as follows. We prove Theorem 1.2 and the associated corollaries in Section 2, Theorem 1.6 in Section 3, Theorem 1.7 in Section 4, and Theorem 1.10 in Section 5. For completeness, we collect some properties of completely random measures in Appendix B. In Appendix C, we compute some functionals of specific random adjacency measures arising in the proofs of Theorems 1.2 and 1.7, respectively. Finally, Appendix D establishes some facts about random adjacency measures under rescaling.

**2. Proof for configuration model results.** Our proofs rely on one lemma and three propositions. For any (multi)graph  $G$ , let  $e(G)$  denote the number of nonloop edges in  $G$ .

LEMMA 2.1 (Nonloop edges in  $\text{CM}_n(\mathbf{d})$ ). *As  $n \rightarrow \infty$ ,  $\frac{e(\text{CM}_n(\mathbf{d}))}{\ell_n} \rightarrow 1/2$  a.s.  $\mathbb{P}_{\text{CM}}$ . Further, as  $n \rightarrow \infty$ , for all  $\varepsilon > 0$ ,*

$$(2.1) \quad \frac{2\mathbb{E}[e(\text{ECM}_n(\mathbf{d}))]}{\ell_n} - \int_0^\infty \int_0^\infty (1 - e^{-xy})\rho_n(dx)\rho_n(dy) \rightarrow 0,$$

$$(2.2) \quad \mathbb{P}_{\text{ECM}}(|e(\text{ECM}_n(\mathbf{d})) - \mathbb{E}[e(\text{ECM}_n(\mathbf{d}))]| > \varepsilon\ell_n) \leq 2\exp(-C_0\varepsilon\ell_n)$$

for some universal constant  $C_0 > 0$ .

PROPOSITION 2.2. *Let  $\mathcal{E}_n(S, S')$  denote the number of edges created between the set of half-edges  $S$  and  $S'$  in the construction of  $\text{CM}_n(\mathbf{d})$ . Consider  $k$  disjoint subsets of half-edges  $(S_j)_{j \in [k]}$  such that  $|S_j| = s_j = O(\sqrt{\ell_n})$  for all  $j \in [k]$ . Let  $\mathcal{E}_n = (\mathcal{E}_n(S_i, S_j))_{1 \leq i \leq j \leq k}$ ,  $\mathcal{E} := (\mathcal{E}_{ij})_{1 \leq i \leq j \leq k}$ , where  $\mathcal{E}$  is an independent collection and  $\mathcal{E}_{ij} \sim \text{Poisson}(s_i s_j / \ell_n)$  for  $i \neq j$ ,  $\mathcal{E}_{ii} \sim \text{Poisson}(s_i^2 / 2\ell_n)$ . Then, as  $n \rightarrow \infty$ ,*

$$d_{\text{TV}}(\mathcal{E}_n, \mathcal{E}) \rightarrow 0,$$

where  $d_{\text{TV}}(\cdot, \cdot)$  denotes the total variation distance. Moreover, if  $S_j$ ’s are random disjoint subsets chosen independently of  $\text{CM}_n(\mathbf{d})$  and satisfying  $\mathbb{E}[s_j] = O(\sqrt{\ell_n})$  for all  $j \in [k]$ , then  $\lim_{n \rightarrow \infty} d_{\text{TV}}(\mathcal{E}_n, \mathcal{E}) = 0$ , where both  $\mathcal{E}_n$  and  $\mathcal{E}$  refer to the joint distribution, including in particular the randomness stemming from the random sets  $S_j$ ’s.

To state the next proposition, we recall the definition of  $Y_n(t)$  from Theorem 1.2. For  $A \in \mathcal{B}(\mathbb{R}_+)$ , let  $V_n(A)$  be the set of vertices obtained by labeling the vertices in  $[n]$  uniformly from  $[0, \sqrt{\ell_n}]$  and then retaining the vertices with labels in  $A$ . This induces a random measure  $\bar{S}_n$  on  $\mathbb{R}_+$  via  $\bar{S}_n(A) = \frac{1}{\sqrt{\ell_n}} \sum_{i \in V_n(A)} d_i$  that is related to  $Y_n(t)$  via  $Y_n(t) = \bar{S}_n([0, t])$ . As we will see, the next proposition immediately implies that distributional convergence of the measure  $\mu_n$  defined in Theorem 1.2 is equivalent to the convergence of the finite-dimensional distributions of  $(Y_n(t))_{t \geq 0}$ .

PROPOSITION 2.3. *For any disjoint collection of sets  $(A_j)_{j \in [k]}$  from  $\mathcal{B}(\mathbb{R}_+)$ , and  $\alpha = (\alpha_j)_{j \in [k]} \in \mathbb{R}^k$ , define  $\Phi_n(A_1, \dots, A_k) := \mathbb{E}[e^{i \sum_{j \in [k]} \alpha_j \bar{S}_n(A_j)}]$ . If  $\max_i \delta_i = o(\ell_n)$ , then*

$$\Phi_n(A_1, \dots, A_k) = \exp\left(\sum_{j \in [k]} \Lambda(A_j) \int (e^{i\alpha_j x} - 1)\rho_n(dx) + o(1)\right).$$



PROPOSITION 2.4. *Let  $\xi_n$  denote the point process  $\text{Lbl}_{\sqrt{\ell_n}}(\text{CM}_n(\mathbf{d}))$  on  $\mathbb{R}_+^2$ . For any  $A, B \in \mathcal{B}(\mathbb{R}_+)$ ,  $l \in \mathbb{N}^*$  and  $\delta > 0$ ,*

$$(2.3) \quad \begin{aligned} & \mathbb{P}(|\mathbb{P}(\xi_n(A \times B) = l | \text{CM}_n(\mathbf{d})) - \mathbb{P}(\xi_n(A \times B) = l)| > \delta) \\ & \leq 2 \exp\left(-\frac{\delta^2 \ell_n}{(12\Lambda(A)\Lambda(B))^2}\right). \end{aligned}$$

Consequently,  $\mathbb{P}_{\text{CM}}$  a.s.,  $d_L(\mathcal{L}(\xi_n | \text{CM}_n(\mathbf{d})), \mathcal{L}(\xi_n)) \rightarrow 0$  as  $n \rightarrow \infty$ , where  $d_L(\cdot, \cdot)$  denotes the Lévy–Prohorov metric on  $\mathcal{P}(\mathcal{M}(\mathbb{R}_+^2))$ .

We first establish Theorem 1.2 and its corollaries given Lemma 2.1, Propositions 2.2, 2.3 and 2.4, and defer the proofs of the lemma and propositions to the end of the section.

*Proofs of Theorem 1.2, Corollaries 1.3, 1.4 and 1.5.*

PROOF OF THEOREM 1.2. (ii)  $\Leftrightarrow$  (iii). If  $\{(v_i, \theta_i)\}_{i \geq 1}$  is a unit rate Poisson point process on  $\mathbb{R}_+^2$ , then  $\{(\bar{\rho}_n^{-1}(v_i), \theta_i)\}_{i \geq 1}$  is a Poisson process with intensity measure  $\rho_n \times \lambda$ , showing that  $\mu_n \sim \text{CRM}(0, \rho_n \times \lambda)$ . Let  $X_n(t) = \mu_n([0, t])$ . Then  $(X_n(t))_{t \geq 0}$  is a Lévy process (see Appendix B for the definition and some important properties of Lévy processes) with characteristic function

$$(2.4) \quad \begin{aligned} \mathbb{E}[e^{i\theta X_n(t)}] &= \mathbb{E}[e^{i\theta \mu_n([0, t])}] = \exp\left(t \int (e^{i\theta x} - 1) \rho_n(dx)\right) \\ &= \exp\left(t \left(i\theta b_n + t \int (e^{i\theta x} - 1 - i\theta(x \wedge 1)) \rho_n(dx)\right)\right), \end{aligned}$$

where the third step follows using Lemma B.1. Using the standard terminology for the theory of Lévy processes (see Appendix B)  $(X_n(t))_{t \geq 0}$  is a Lévy process with characteristics  $(b_n, 0, \rho_n)$ . By [22], Lemma 11.1.XI, statement (ii) is equivalent to assuming that  $(X_n(t))_{t \geq 0}$  converges as a stochastic process in  $\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+)$ . Therefore, [27], Chapter VII, Corollary 3.6 (restated as Lemma B.2 in Appendix B for the special case of Lévy processes) implies that (ii) is equivalent to the following two conditions:  $b_n \rightarrow b$  and  $\int f d\rho_n \rightarrow \int f d\rho$  for all bounded continuous functions  $f : \mathbb{R}_+ \rightarrow \mathbb{R}$  vanishing near zero, which by Lemma B.3 is equivalent to (iii). Note that Lemma B.3 also proves Remark 4.

(ii)  $\Leftrightarrow$  (iv). By Proposition 2.3 and (2.4), convergence of the finite dimensional distributions of  $(Y_n(t))_{t \geq 0}$  is equivalent to convergence of the characteristic functions of  $X_n(t)$  for all  $t$ . Since  $X_n(t)$  is a Lévy process, this in turn is equivalent to the convergence of this process in law, which is equivalent to (ii).

All that remains to show is therefore tightness of  $(Y_n(t))_{t \geq 0}$  in  $\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+)$ . To this end, we note that for  $t > u > s$ ,

$$(2.5) \quad \begin{aligned} \mathbb{E}[(Y_n(t) - Y_n(u))(Y_n(u) - Y_n(s))] &= \frac{1}{\ell_n^2} \sum_{i \neq j} d_i d_j (t - u)(u - s) \\ &\leq \frac{1}{4}(t - s)^2. \end{aligned}$$

Tightness of  $(Y_n(t))_{t \geq 0}$  follows using [27], Chapter VI, Theorem 4.1.

(i)  $\Leftrightarrow$  (iv). By Proposition 1.1, statement (i) is equivalent to  $\mathbb{P}_{\text{CM}}$  almost sure convergence of  $(\text{Lbl}(\text{CM}_n(\mathbf{d})))_{n \geq 1}$  in distribution. To compare this to  $\mathbb{P}_{\text{CM}}$  almost sure convergence of  $\xi_n$  in distribution, we will use the fact that by Lemma 2.1,  $2e(\text{CM}_n(\mathbf{d}))/\ell_n \rightarrow 1$   $\mathbb{P}_{\text{CM}}$  almost surely. This in turn implies that  $\mathbb{P}_{\text{CM}}$  almost sure convergence of  $(\text{Lbl}(\text{CM}_n(\mathbf{d})))_{n \geq 1}$  in distribution

is equivalent to  $\mathbb{P}_{\text{CM}}$  almost sure convergence of  $\xi_n$  in distribution (the formal argument requires a technical lemma, Lemma D.1 from Appendix D). Combining these facts with Proposition 2.4 and the fact that  $\ell_n = \Omega(\log n)$  we concluded that (i) is equivalent to the statement that  $\mathcal{L}(\xi_n(A_i \times A_j))_{1 \leq i \leq j \leq k}$  converges for any  $k \geq 1$  and disjoint sets  $A_i \in \mathcal{B}(\mathbb{R}_+)$ ,  $i = 1, \dots, k$ .

Given a collection of labels  $(U_j)_{j=1, \dots, n}$  chosen i.i.d. uniformly at random in  $[0, \sqrt{\ell_n}]$ , let  $V_n(A_i)$  be the set of vertices with label in  $A_i$ , and let  $S_i$  be the set of half-edges whose endpoint is in  $V_n(A_i)$ . As before, let  $\bar{S}_n(A) = \frac{1}{\sqrt{\ell_n}} \sum_{j \in V_n(A)} d_j$ . Then  $|S_i| = \bar{S}_n(A_i) \sqrt{\ell_n}$  and  $\mathbb{E}[|S_i|] = \Lambda(A_i) \sqrt{\ell_n}$ , so by Proposition 2.2 convergence of  $\mathcal{L}(\xi_n(A_i \times A_j))_{1 \leq i \leq j \leq k}$  is equivalent to distributional convergence of  $\{\text{Poisson}(\bar{S}_n(A_i) \bar{S}_n(A_j))\}_{i, j \in [k]}$ , which in turn is equivalent to distributional convergence of  $\{\bar{S}_n(A_i) \bar{S}_n(A_j)\}_{i, j \in [k]}$ . The latter clearly implies convergence of the random vector  $\{(\bar{S}_n(A_i))^2\}_{i \in [k]}$ , and since  $\bar{S}_n(A_i) \geq 0$ , this in turn implies convergence of  $\{\bar{S}_n(A_i)\}_{i \in [k]}$ . Conversely, the latter clearly implies convergence of  $\{\bar{S}_n(A_i) \bar{S}_n(A_j)\}_{i, j \in [k]}$ , so we have shown that (i) is equivalent to convergence of the finite dimensional distributions of  $(Y_n(t))_{t \geq 0}$ . To show that this is equivalent to (iv), we use the tightness condition (2.5).

Finally, to obtain the required descriptions for the limiting objects, note that (ii) implies  $\mathcal{L}(\mu_n)$  converges weakly. Since  $\mu_n$  is completely random, it follows that there exists a completely random measure  $\mu$  such that  $\mathcal{L}(\mu_n) \rightarrow \mathcal{L}(\mu)$  in  $\mathcal{P}(\mathcal{N}(\mathbb{R}_+))$ . Thus,  $\mu$  admits a representation (1.1). Moreover, the convergence of the characteristics of the Lévy process  $(X_n(t))_{t \geq 0}$  yields that  $a = \lim_{n \rightarrow \infty} \int (x \wedge 1) \rho_n(dx) - \int (x \wedge 1) \rho(dx)$ . Finally, Proposition 2.3, (2.4) and convergence of the Lévy process  $(X_n(t))_{t \geq 0}$  to a Lévy process with characteristics  $(a, 0, \rho)$  gives that

$$\Phi_n(A_1, \dots, A_k) \rightarrow \exp\left(\sum_{j \in [k]} \Lambda(A_j) \left(a + \int (e^{ia_j x} - 1) \rho(dx)\right)\right),$$

for any disjoint collection of sets  $(A_j)_{j \in [k]}$  from  $\mathcal{B}(\mathbb{R}_+)$ . This shows that  $\bar{S}_n$  converges to the completely random measure  $\mu$ , and that  $(Y_n(t))_{t \geq 0}$  converges to the Lévy process  $(\mu([0, t]))_{t \geq 0}$ . Using the convergence of  $\bar{S}_n$  to  $\mu$  and following the argument from the proof of (i)  $\Leftrightarrow$  (iv) we then get that in distribution,  $(\xi_n(A_i \times A_j))_{1 \leq i \leq j \leq k}$  converges to  $\{\text{Poisson}(\mu(A_i) \mu(A_j))\}_{i, j \in [k]}$ . As established in Lemma C.1, this is equal in distribution to  $(\xi_{\mathcal{W}_{\text{CM}}}(A_i \times A_j))_{1 \leq i \leq j \leq k}$ , as required.  $\square$

**PROOF OF COROLLARY 1.3.** To establish this corollary, we first note that

$$0 \leq \int_0^\infty \int_0^\infty (1 - e^{-xy}) \rho_n(dx) \rho_n(dy) \leq 1,$$

using  $1 - e^{-x} \leq x$  for  $x \geq 0$ , and  $\int_0^\infty x \rho_n(dx) = 1$ . Thus this sequence is compact, and equivalently, every sequence has a convergent subsequence. Let us assume

$$\int_0^\infty \int_0^\infty (1 - e^{-xy}) \rho_n(dx) \rho_n(dy) \rightarrow c > 0,$$

along a subsequence. The proof of Theorem 1.2 implies that  $\text{Lbl}_{\sqrt{\ell_n}}(\text{ECM}_n(\mathbf{d}))$  converges weakly to the random adjacency measure corresponding to the graphex  $\mathcal{W}_{\text{ECM}}^1$ . The proof is now complete, once we use Lemma 2.1 and Lemma D.1.  $\square$

Next, we prove Corollary 1.4. It is easy to see that in this specific case, the result follows almost directly from [10], Remark 14. Below, we provide a more detailed proof from first principles. We feel that this proof is more intuitive, and also more generally applicable, as evidenced by its easy adaptation to establish Corollary 1.9 in Section 4.

PROOF OF COROLLARY 1.4. By Theorem 1.2,  $I_{CM} = 0$  if and only if  $a = 0$ , which by Remark 4 is equivalent to

$$(2.6) \quad \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \int_0^\varepsilon x \rho_n(dx) = 0,$$

which by Definition 9 is equivalent to uniform tail regularity of  $CM_n(\mathbf{d})$ .

We now establish that (2.6) implies that the sequence  $ECM_n(\mathbf{d})$  is uniform tail regular  $\mathbb{P}_{ECM}$  a.s. To this end, set  $V_{>\varepsilon} = \{i : d_i > \varepsilon \sqrt{\ell_n}\}$ ,  $V_{\leq\varepsilon} = V_{>\varepsilon}^c$  and denote by  $\mathcal{E}(V_{>\varepsilon})$  the number of edges in  $CM_n(\mathbf{d})$  with both end points in  $V_{>\varepsilon}$ . We have, by direct computation,

$$\mathbb{E}[\mathcal{E}(V_{>\varepsilon})] = \frac{1}{2}(1 + o(1))\ell_n \left(1 - \int_0^\varepsilon x \rho_n(dx)\right)^2 = \ell_n \left(\frac{1}{2} - \text{Err}(\varepsilon, n)\right),$$

where  $\text{Err}(\varepsilon, n)$  is an error term such that  $\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \text{Err}(\varepsilon, n) = 0$ . By Lemma 2.5 below, there exist a constant  $C_0 > 0$  such that for all  $\delta_n > 0$ ,

$$\mathbb{P}(|\mathcal{E}(V_{>\varepsilon}) - \mathbb{E}[\mathcal{E}(V_{>\varepsilon})]| \geq \delta_n \ell_n) \leq 2 \exp(-C_0 \ell_n \delta_n^2).$$

Since  $\ell_n = \omega(\log n)$ , we can choose  $\delta_n$  in such a way that this bound is summable and  $\delta_n = o(1)$ . Thus  $\mathbb{P}_{CM}$  a.s.,

$$(2.7) \quad e(CM_n(\mathbf{d})) - \mathcal{E}(V_{>\varepsilon}) \leq \text{Err}(\varepsilon, n)\ell_n + o(\ell_n).$$

Let  $\mathcal{E}_{ECM}(\varepsilon)$  denote the number of edges in  $ECM_n(\mathbf{d})$  with at least one end point in  $V_{\leq\varepsilon}$ . Since this is bounded by the number of edges in  $CM_n(\mathbf{d})$  with at least one end point in  $V_{\leq\varepsilon}$ , (2.7) implies that  $\mathbb{P}_{ECM}$  a.s.,  $\mathcal{E}_{ECM}(\varepsilon) \leq \ell_n \text{Err}(\varepsilon, n) + o(\ell_n)$ . Now, recall the definitions of stretched canonical graphon [10], Section 2.3, and uniformly tail regular graphs [10], Definition 13. Let  $W_n$  denote the stretched canonical graphon for  $ECM_n(\mathbf{d})$ . Set  $\varepsilon > 0$ , and consider  $U_n$  to be the set corresponding to the vertices in  $V_{>\varepsilon}$ . With this choice of  $U_n$ ,

$$(2.8) \quad \|W_n - W_n \mathbb{1}\{U_n \times U_n\}\|_1 \leq \text{Err}(\varepsilon, n) + o(1),$$

where  $\|\cdot\|_1$  denotes the  $L_1$  norm. Note here that for constructing the stretched canonical graphon, the space is scaled with  $\sqrt{e(ECM_n(\mathbf{d}))}$ , rather than  $\sqrt{\ell_n}$ , but that does not change the order in (2.8) due to Lemma 2.1. To derive the required claim, it suffices to show that the Lebesgue measure of the set corresponding to the vertices in  $U_n$  is uniformly bounded in  $n$ . Upon direct computation, we note that this measure is exactly  $\rho_n((\varepsilon, \infty)) \leq \frac{1}{\varepsilon} \int_0^\infty x \rho_n(dx) = \frac{1}{\varepsilon}$ . This establishes that  $ECM_n(\mathbf{d})$  is uniformly tail regular a.s.  $\mathbb{P}_{ECM}$ . Finally, note that convergence in stretched cut metric follows immediately from [12], Theorem 5.5.  $\square$

PROOF OF COROLLARY 1.5. The corollary is an immediate consequence of Theorem 1.2.  $\square$

*Proofs of Propositions 2.3 and 2.4 and Lemma 2.1.*

PROOF OF PROPOSITION 2.3. Let  $l_i$  denote the label for vertex  $i$  that is uniformly distributed over the interval  $[0, \sqrt{\ell_n}]$ , independently over  $i \in [n]$ . Note that

$$\begin{aligned} \Phi(A_1, \dots, A_k) &= \prod_{i \in [n]} \mathbb{E}\left[e^{\frac{id_i}{\sqrt{\ell_n}} \sum_{j \in [k]} \alpha_j \mathbb{1}\{l_i \in A_j\}}\right] \\ &= \prod_{i \in [n]} \left(1 - \frac{1}{\sqrt{\ell_n}} \sum_{j \in [k]} \Lambda(A_j) + \frac{1}{\sqrt{\ell_n}} \sum_{j \in [k]} \Lambda(A_j) e^{\frac{id_i \alpha_j}{\sqrt{\ell_n}}}\right) \end{aligned}$$

$$\begin{aligned}
 &= \prod_{i \in [n]} \exp\left(\frac{1}{\sqrt{\ell_n}} \sum_{j \in [k]} \Lambda(A_j) (e^{\frac{id_i \alpha_j}{\sqrt{\ell_n}}} - 1) + O(d_i^2 / \ell_n^2)\right) \\
 &= \exp\left(\sum_{j \in [k]} \Lambda(A_j) \int (e^{i \alpha_j x} - 1) \rho_n(dx) + o(1)\right),
 \end{aligned}$$

where in the last step we used that  $\sum_i d_i^2 \leq \max_i d_i \sum_i d_i = o(\ell_n^2)$ . This completes the proof.  $\square$

In the next two proofs, we will use the following simple switching lemma.

**LEMMA 2.5 (Switching lemma).** *Let  $P$  be a perfect matching of the half-edges, and  $P'$  be another perfect matching which can be obtained from  $P$  by one switch. Let  $X$  be a function on perfect matchings satisfying the Lipchitz condition  $|X(P) - X(P')| \leq c$ . Then, for any  $\varepsilon > 0$ ,*

$$\mathbb{P}(|X - E[X]| > \varepsilon) \leq 2e^{-\frac{\varepsilon^2}{\ell_n c^2}}.$$

**PROOF.** This follows using identical arguments as [44], Theorem 2.19. Note that [44], Theorem 2.19, was only stated for random regular graphs, but the same argument works for the general configuration model as well.  $\square$

**PROOF OF PROPOSITION 2.4.** Recall that an instance of  $CM_n(\mathbf{d})$  is generated by choosing a uniformly random matching  $P$  of the  $\ell_n$  half-edges corresponding to  $\mathbf{d}$ . Denoting the multi-graph corresponding to a matching  $P$  by  $G_n = G_n(P)$ , we apply Lemma 2.5 with  $X_n(P) = \mathbb{P}(\xi_{G_n}(P = l | G_n))$ . Consider two matchings  $P$  and  $P'$  that differ by at most one switch, and label the half-edges involved in this switch by  $1, \dots, 4$  in such a way that in  $P$ , 13 and 24 are matched, and in  $P'$ , 14 and 23 are matched. Then  $\xi_{G_n}(A \times B)$  remains unchanged between  $P$  and  $P'$  unless at least one of the half-edges  $1, \dots, 4$  has a label in  $A$ , and a second, different one, has a label in  $B$ . A union bound then easily shows that  $|X_n(P) - X_n(P')| \leq 12\Lambda(A)\Lambda(B)/\ell_n$ . This proves (2.3).

The proof of the final statement follows by observing that  $\mathcal{M}(\mathbb{R}_+^2)$ , equipped with vague topology is a Polish space, and that weak convergence on such spaces is determined by countable classes of sets.  $\square$

**PROOF OF LEMMA 2.1.** The expected number of loops in  $CM_n(\mathbf{d})$  is given by

$$\begin{aligned}
 (2.9) \quad &\frac{1}{2} \mathbb{E} \left[ \sum_{i \in [n]} \sum_{j \in [d_i]} \mathbb{1}\{j\text{th half-edge of the } i\text{th vertex creates a loop}\} \right] \\
 &= \frac{1}{2} \sum_{i \in [n]} \sum_{j \in [d_i]} \frac{d_i - 1}{\ell_n - 1} \leq \sum_{i \in [n]} \frac{d_i^2}{\ell_n} \leq \max_{i \in [n]} d_i = o(\ell_n).
 \end{aligned}$$

This immediately proves that  $\mathbb{E}[e(CM_n(\mathbf{d}))]/\ell_n \rightarrow 1/2$  as  $n \rightarrow \infty$ . To prove almost sure convergence, we use Lemma 2.5. Note that if  $P$  and  $P'$  are two perfect matchings differing by at most one switch,  $e(CM_n(\mathbf{d}))/\ell_n$  might change by at most  $2/\ell_n$ . This gives the required concentration.

Next, we compute the expected number of edges in  $ECM_n(\mathbf{d})$ . To this end, we compute first the expected number of multiple edges. Let  $X_{ij}$  denote the number of edges between  $i$  and  $j$ . Thus the total number of multiple edges is given by

$$(2.10) \quad \sum_{i < j} (X_{ij} - 1) \mathbb{1}\{X_{ij} \geq 2\}.$$

Now,

$$\begin{aligned}
 \mathbb{E}[(X_{ij} - 1)\mathbb{1}\{X_{ij} \geq 2\}] &= \mathbb{E}[X_{ij}\mathbb{1}\{X_{ij} \geq 2\}] - \mathbb{P}(X_{ij} \geq 2) \\
 &= \mathbb{E}[X_{ij}] - \mathbb{E}[X_{ij}\mathbb{1}\{X_{ij} \leq 1\}] - \mathbb{P}(X_{ij} \geq 2) \\
 (2.11) \quad &= \mathbb{E}[X_{ij}] - \mathbb{P}(X_{ij} \geq 1) \\
 &= \mathbb{E}[X_{ij}] - 1 + \mathbb{P}(X_{ij} = 0) \\
 &= \frac{d_i d_j}{\ell_n - 1} - 1 + e^{-\frac{d_i d_j}{\ell_n}} + O\left(\frac{d_i^2 d_j + d_j^2 d_i}{\ell_n^2}\right),
 \end{aligned}$$

where we have used that

$$0 \leq \mathbb{P}(X_{ij} = 0) - \prod_{t=0}^{d_i-1} \left(1 - \frac{d_j}{\ell_n - 1 - 2t}\right) \leq \frac{d_i^2 d_j}{(\ell_n - 2d_i)^2},$$

see, for example, [25], (4.9), together with the fact that

$$\left(1 - \frac{d_j}{\ell_n - 1 - 2t}\right) = \exp\left(-\frac{d_j}{\ell_n}\right) \left(1 + O\left(\frac{d_i d_j + d_j^2}{\ell_n^2}\right)\right).$$

Therefore, (2.10) and (2.11) together with (2.9) imply that the expected number of edges in  $\text{ECM}_n(\mathbf{d})$  is

$$\begin{aligned}
 \frac{\ell_n}{2} - \sum_{i < j} \left(\frac{d_i d_j}{\ell_n} - 1 + e^{-\frac{d_i d_j}{\ell_n}}\right) + o(\ell_n) &= \frac{1}{2} \sum_{i \neq j} (1 - e^{-\frac{d_i d_j}{\ell_n}}) + o(\ell_n) \\
 &= \frac{\ell_n}{2} \int_0^\infty \int_0^\infty (1 - e^{-xy}) \rho_n(dx) \rho_n(dy) + o(\ell_n),
 \end{aligned}$$

where we have again used the fact that  $\max_{i \in [n]} d_i = o(\ell_n)$ . This proves (2.1).

To prove (2.2), we again use Lemma 2.5, noting that the number of edges in  $\text{ECM}_n(\mathbf{d})$  becomes fixed once the uniform matching of half-edges has been fixed, and that a switch can alter this function by at most a bounded constant. This completes the proof.  $\square$

*2.1. Edge counts for configuration model. PROOF OF PROPOSITION 2.2.* We prove this for  $k = 2$  and the general case follows similarly. Thus, we need to show that

$$d_{\text{TV}}\left(\mathcal{L}(\mathcal{E}_n(S_1, S_1), \mathcal{E}_n(S_1, S_2), \mathcal{E}_n(S_2, S_2)), \text{Poi}\left(\frac{s_1^2}{2\ell_n}\right) \otimes \text{Poi}\left(\frac{s_1 s_2}{\ell_n}\right) \otimes \text{Poi}\left(\frac{s_2^2}{2\ell_n}\right)\right) \rightarrow 0.$$

For simplicity, we write  $S = S_1 \cup S_2$  and  $s = |S| = s_1 + s_2$ . Let us enumerate the half-edges in  $S_1$  arbitrarily by  $\{1, \dots, s_1\}$  and the half-edges in  $S_2$  by  $\{s_1 + 1, \dots, s\}$ . We first pair the half-edges of  $S_1$  and then the remaining unpaired half-edges of  $S_2$ . Consider sequential pairing of the half-edges of  $S$  and at step  $\alpha$ ,  $\alpha = 1, \dots, s$ , we take the half-edge labeled  $\alpha$  and if it is not already paired to some previous half-edge, we pair it with another unpaired half-edge chosen uniformly at random. Let  $I_\alpha$  denote the three-dimensional Bernoulli random vector where

$$(2.12) \quad I_\alpha = \begin{cases} (1, 0, 0) & \text{if } \alpha \in S_1, \text{ half-edge } \alpha \text{ is not paired previously} \\ & \text{and it is paired with another half-edge in } S_1, \\ (0, 1, 0) & \text{if } \alpha \in S_1, \text{ half-edge } \alpha \text{ is not paired previously} \\ & \text{and it is paired with another half-edge in } S_2, \\ (0, 0, 1) & \text{if } \alpha \in S_2, \text{ half-edge } \alpha \text{ is not paired previously} \\ & \text{and it is paired with another half-edge in } S_2, \\ (0, 0, 0) & \text{otherwise.} \end{cases}$$

Note that  $\sum_{\alpha=1}^s I_\alpha = (\mathcal{E}_n(S_1, S_1), \mathcal{E}_n(S_1, S_2), \mathcal{E}_n(S_2, S_2))$ . We first couple the  $I'_\alpha$ 's to independent multivariate Bernoulli random variables, and then use Stein's method to obtain multivariate Poisson approximation for  $(\mathcal{E}_n(S_1, S_1), \mathcal{E}_n(S_1, S_2), \mathcal{E}_n(S_2, S_2))$ .

*Coupling.* We approximate the collection  $(I_\alpha)_{\alpha=1}^s$  with a collection of independent random variables  $(\hat{I}_\alpha)_{\alpha=1}^s$ . To this end, we describe an algorithm that sequentially pairs the half-edges and keeps track of a special set of half-edges, called *bad half-edges*. Let  $\mathcal{B}_\alpha$  denote the set of bad half-edges at step  $\alpha$ . Initially, all the half-edges are *nonbad*, that is,  $\mathcal{B}_0 = \emptyset$ . As before, the half-edges of  $S_1$  take labels in  $\{1, \dots, s_1\}$  and the half-edges of  $S_2$  take labels in  $\{s_1 + 1, \dots, s\}$ . At stage  $\alpha$ , we pair the half-edge labeled  $\alpha$  (call it  $e_\alpha$ ) in  $S$  to a *uniformly chosen* half edge from  $\{e_1, \dots, e_\alpha\}^c$  (call it  $f_\alpha$ ). If  $e_\alpha \in \mathcal{B}_{\alpha-1}$ , we set  $\mathcal{B}_\alpha = \mathcal{B}_{\alpha-1}$ . If  $e_\alpha \in \mathcal{B}_{\alpha-1}^c$  and  $f_\alpha \in \mathcal{B}_{\alpha-1}^c$ , we set  $\mathcal{B}_\alpha = \mathcal{B}_{\alpha-1} \cup \{f_\alpha\}$  and  $g_\alpha = f_\alpha$ . Finally, if  $e_\alpha \in \mathcal{B}_{\alpha-1}^c$  and  $f_\alpha \in \mathcal{B}_{\alpha-1}$ , we choose  $g_\alpha$  uniformly at random from  $\mathcal{B}_{\alpha-1}^c \cap \{e_1, \dots, e_\alpha\}^c$  independently and set  $\mathcal{B}_\alpha = \mathcal{B}_{\alpha-1} \cup \{g_\alpha\}$ . Under this scheme, we define

$$(2.13) \quad \hat{I}_\alpha = \begin{cases} (1, 0, 0) & \text{if } e_\alpha \in S_1, f_\alpha \in S_1, \\ (0, 1, 0) & \text{if } e_\alpha \in S_1, f_\alpha \in S_2, \\ (0, 0, 1) & \text{if } e_\alpha \in S_2, f_\alpha \in S_2, \\ (0, 0, 0) & \text{otherwise,} \end{cases}$$

$$I_\alpha = \begin{cases} (1, 0, 0) & \text{if } e_\alpha \in S_1 \cap \mathcal{B}_{\alpha-1}^c, g_\alpha \in S_1, \\ (0, 1, 0) & \text{if } e_\alpha \in S_1 \cap \mathcal{B}_{\alpha-1}^c, g_\alpha \in S_2, \\ (0, 0, 1) & \text{if } e_\alpha \in S_2 \cap \mathcal{B}_{\alpha-1}^c, g_\alpha \in S_2, \\ (0, 0, 0) & \text{otherwise.} \end{cases}$$

Note that  $(\hat{I}_\alpha)_{\alpha=1}^s$  is an independent collection, and for  $\alpha = 1, \dots, s_1$ ,  $\hat{I}_\alpha = (1, 0, 0)$  with probability  $(s_1 - \alpha)/(\ell_n - \alpha)$ ,  $\hat{I}_\alpha = (0, 1, 0)$  with probability  $s_2/(\ell_n - \alpha)$ , and zero otherwise. Furthermore, for  $\alpha = s_1 + 1, \dots, s$ ,  $\hat{I}_\alpha = (0, 0, 1)$  with probability  $(s_2 - (\alpha - s_1))/(\ell_n - \alpha)$ , and zero otherwise. Moreover, the distribution of  $(I_\alpha)_{\alpha=1}^s$  is same as described in (2.12).

Next, we investigate the probability that the two random variables in (2.13) are unequal. First, if  $e_\alpha, f_\alpha \in \mathcal{B}_{\alpha-1}^c$ , then  $\hat{I}_\alpha = I_\alpha$ . Next, at stage  $\alpha$ ,  $e_\alpha \in \mathcal{B}_{\alpha-1}$  if it was already paired previously to a nonbad half-edge. Now, there were at most  $s$  previous steps performed and at each of those steps, the probability of pairing with the half-edge labeled  $\alpha$  is at most  $1/(\ell_n - s)$ , so that  $\mathbb{P}(e_\alpha \in \mathcal{B}_{\alpha-1}) \leq s/(\ell_n - s)$ . Moreover, conditional on the fact that  $e_\alpha \in \mathcal{B}_{\alpha-1}$ ,  $\hat{I}_\alpha \neq I_\alpha$  if and only if  $\hat{I}_\alpha \neq (0, 0, 0)$  and  $\mathbb{P}(\hat{I}_\alpha \neq (0, 0, 0) | e_\alpha \in \mathcal{B}_{\alpha-1}) = \mathbb{P}(\hat{I}_\alpha \neq (0, 0, 0)) \leq s/(\ell_n - s)$ . Thus,

$$(2.14) \quad \mathbb{P}(e_\alpha \in \mathcal{B}_{\alpha-1} \text{ and } I_\alpha \neq \hat{I}_\alpha) \leq s^2/(\ell_n - s)^2.$$

On the other hand, if  $f_\alpha \in \mathcal{B}_{\alpha-1}$  and  $e_\alpha \in \mathcal{B}_{\alpha-1}^c$  (we call this event  $F_\alpha$ ), then the event  $f_\alpha \in S^c$  implies  $g_\alpha \in S$ . Therefore,

$$(2.15) \quad \begin{aligned} \mathbb{P}(F_\alpha \text{ and } I_\alpha \neq \hat{I}_\alpha) &\leq \mathbb{P}(F_\alpha, f_\alpha \in S) + \mathbb{P}(F_\alpha, f_\alpha \in S^c, g_\alpha \in S) \\ &= \text{(I)} + \text{(II)}. \end{aligned}$$

First, we obtain an upper bound on (II). Noting that  $\mathbb{P}(F_\alpha) \leq \mathbb{P}(f_\alpha \in \mathcal{B}_{\alpha-1}) \leq s/(\ell_n - s)$  and  $\mathbb{P}(g_\alpha \in S | F_\alpha) \leq s/(\ell_n - s)$  due to the choice of  $g_\alpha$ , it follows that

$$(2.16) \quad \text{(II)} \leq s^2/(\ell_n - s)^2.$$

To derive an upper bound on (I), we use the fact that, conditioned on  $|\mathcal{B}_{\alpha-1} \cap S| \leq \ell_n^{1/3}$ , the probability that  $f_\alpha \in \mathcal{B}_{\alpha-1} \cap S$  is at most  $\ell_n^{1/3}/(\ell_n - s)$ . Thus,

$$\begin{aligned} \text{(I)} &\leq \mathbb{P}(f_\alpha \in \mathcal{B}_{\alpha-1} \cap S) \leq \frac{\ell_n^{1/3}}{\ell_n - s} + \mathbb{P}(|\mathcal{B}_{\alpha-1} \cap S| > \ell_n^{1/3}) \\ &\leq \frac{\ell_n^{1/3}}{\ell_n - s} + \mathbb{P}(\mathcal{E}_n(S, S) > \ell_n^{1/3}), \end{aligned}$$

where the last inequality follows using the fact that an element is added to  $\mathcal{B}_{\alpha-1} \cap S$  if and only if either  $e_\alpha, f_\alpha \in \mathcal{B}_{\alpha-1}^c \cap S$  or  $e_\alpha, g_\alpha \in \mathcal{B}_{\alpha-1}^c \cap S$  but  $f_\alpha \in \mathcal{B}_{\alpha-1}$ , and in both cases  $\mathcal{E}_n(S, S)$  increases by 1. Now, in the sequential pairing scheme for creating the configuration model in (2.12), let  $\mathcal{F}_\alpha$  denote the sigma-algebra with respect to which the pairing obtained up to time  $\alpha$  is measurable. Let  $X_\alpha = \mathbb{E}[\mathcal{E}_n(S, S)|\mathcal{F}_\alpha]$ , so that  $X_s = \mathcal{E}_n(S, S)$  and  $X_0 = \mathbb{E}[\mathcal{E}_n(S, S)]$ . Thus  $(X_\alpha)_{\alpha=1}^s$  is the Doob-martingale for  $\mathcal{E}_n(S, S)$  and  $|X_\alpha - X_{\alpha-1}| \leq 1$ . Using Azuma–Hoeffding inequality [29], Theorem 2.25, and the fact that  $\mathbb{E}[\mathcal{E}_n(S, S)] \leq s^2/\ell_n = O(1)$ , it follows that for all sufficiently large  $n$

$$\mathbb{P}(\mathcal{E}_n(S, S) > \ell_n^{1/3}) \leq \mathbb{P}\left(\mathcal{E}_n(S, S) - \mathbb{E}[\mathcal{E}_n(S, S)] > \frac{1}{2}\ell_n^{1/3}\right) \leq e^{-\ell_n^{2/3}/4s},$$

and, therefore,

$$(2.17) \quad \text{(I)} \leq \frac{\ell_n^{1/3}}{\ell_n - s} + e^{-c\ell_n^{1/6}}$$

for some constant  $c > 0$ .

Combining (2.15), (2.16) and (2.17), we get that for all sufficiently large  $n$ ,

$$\mathbb{P}(F_\alpha \text{ and } I_\alpha \neq \hat{I}_\alpha) \leq \frac{\ell_n^{1/3}}{\ell_n - s} + e^{-c\ell_n^{1/6}} + \frac{s^2}{(\ell_n - s)^2}.$$

With the help of (2.14) this implies that for all sufficiently large  $n$ ,

$$\begin{aligned} (2.18) \quad \mathbb{P}\left(\sum_{\alpha=1}^s \hat{I}_\alpha \neq \sum_{\alpha=1}^s I_\alpha\right) &\leq \sum_{\alpha=1}^s \mathbb{P}(\hat{I}_\alpha \neq I_\alpha) \\ &\leq \frac{s^3}{(\ell_n - s)^2} + \frac{s\ell_n^{1/3}}{\ell_n - s} + se^{-c\ell_n^{1/6}} =: \text{Err}_1. \end{aligned}$$

Since  $s = O(\sqrt{\ell_n})$ ,  $\text{Err}_1 \rightarrow 0$ .

*Multivariate Stein’s method.* We will use the result for multivariate Poisson approximation [4], Theorem 1. Let

$$\begin{aligned} \lambda_{11} &= \sum_{\alpha=1}^s \mathbb{P}(\hat{I}_\alpha = (1, 0, 0)) = \sum_{\alpha=1}^{s_1} \frac{s_1 - \alpha}{\ell_n - \alpha}, \\ \lambda_{12} &= \sum_{\alpha=1}^s \mathbb{P}(\hat{I}_\alpha = (0, 1, 0)) = \sum_{\alpha=1}^{s_1} \frac{s_2}{\ell_n - \alpha}, \\ \lambda_{22} &= \sum_{\alpha=1}^s \mathbb{P}(\hat{I}_\alpha = (0, 0, 1)) = \sum_{\alpha=s_1+1}^s \frac{s_2 - (\alpha - s_1)}{\ell_n - \alpha}, \end{aligned}$$

$$c(\lambda) = c(\lambda_{11}, \lambda_{12}, \lambda_{22}) = \frac{1}{2} + \max\{0, \log(2(\lambda_{11} + \lambda_{12} + \lambda_{22}))\}.$$

[4], Theorem 1, together with an easy calculation using the fact that  $\lambda_{11} \geq s_1(s_1 - 1)/2\ell_n$ ,  $\lambda_{12} \geq s_1s_2/\ell_n$  and  $\lambda_{22} \geq s_2(s_2 - 1)/2\ell_n$  then yields

$$\begin{aligned}
 & d_{\text{TV}}\left(\sum_{\alpha=1}^s \hat{I}_\alpha, \text{Poi}(\lambda_{11}) \otimes \text{Poi}(\lambda_{12}) \otimes \text{Poi}(\lambda_{22})\right) \\
 (2.19) \quad & \leq \frac{1}{(\ell_n - s)^2} \min\{s^3, c(\lambda)(2\ell_n(2s_1 - 1) + \ell_ns_2 + 2\ell_n(2s_2 - 1))\} \\
 & \leq \frac{1}{(\ell_n - s)^2} \min\{s^3, 5c(\lambda)s\ell_n\} =: \text{Err}_2.
 \end{aligned}$$

Combining (2.18) and (2.19) with the bound

$$\left| \lambda_{11} - \frac{s_1^2}{2\ell_n} \right| + \left| \lambda_{12} - \frac{s_1s_2}{\ell_n} \right| + \left| \lambda_{22} - \frac{s_2^2}{2\ell_n} \right| \leq \max\left\{ \frac{s}{\ell_n}, \frac{s^3}{2\ell_n(\ell_n - s)} \right\} =: \text{Err}_3,$$

we obtain that

$$\begin{aligned}
 & d_{\text{TV}}\left(\sum_{\alpha=1}^s I_\alpha, \text{Poi}\left(\frac{s_1^2}{2\ell_n}\right) \otimes \text{Poi}\left(\frac{s_1s_2}{\ell_n}\right) \otimes \text{Poi}\left(\frac{s_2^2}{2\ell_n}\right)\right) \\
 (2.20) \quad & \leq \text{Err}_1 + \text{Err}_2 + (1 - e^{-\text{Err}_3}) \\
 & \leq \text{Err}_1 + \text{Err}_2 + \text{Err}_3.
 \end{aligned}$$

When  $S_1$  and  $S_2$  are fixed subsets with  $s_1, s_2 = O(\sqrt{\ell_n})$ , the proof of Proposition 2.2 now follows from (2.20).

Let us now consider the case where  $S_j$ 's are random sets. Observe that that for  $\varepsilon$  sufficiently small,  $\text{Err}_i = o(1)$  if  $s \leq \ell_n^{1/2+\varepsilon}$ . If we condition on the sets  $(S_j)_{j \in [k]}$  and assume that  $s \leq \ell_n^{1/2+\varepsilon}$ , we can therefore couple  $\mathcal{E}_n$  and  $\mathcal{E}$  in such a way that  $\mathcal{E}_n \neq \mathcal{E}$  with probability  $o(1)$ ; if  $s > \ell_n^{1/2+\varepsilon}$ , we couple them arbitrarily. Since  $\mathbb{P}(s > \ell_n^{1/2+\varepsilon}) \leq \ell_n^{-1/2-\varepsilon} \mathbb{E}[s] = O(\ell_n^{-\varepsilon})$  by Markov's inequality and our assumptions on the expectations of  $s_1$  and  $s_2$ , we see that the resulting coupling is such that  $\mathcal{E}_n = \mathcal{E}$  with probability  $1 - o(1)$ , showing that  $\lim_{n \rightarrow \infty} d_{\text{TV}}(\mathcal{E}_n, \mathcal{E}) = 0$ , as required.

The proof of Proposition 2.2 is now complete.  $\square$

**3. Proof of results on preferential attachment model.** In this section, we prove Theorem 1.6. To avoid notational overhead, we recycle some notation and denote the random point process  $\text{Lbl}_{\sqrt{2m_n}}(\text{PA}_n(\kappa, m_n))$  for the graph  $\text{PA}_n(\kappa, m_n)$  by  $\xi_n$ . For a subset  $A \in \mathcal{B}(\mathbb{R}_+)$ , let  $V_n(A)$  denote the set of vertices obtained by labeling the vertices uniformly from  $[0, \sqrt{2m_n}]$  independently and retaining the vertices with labels in  $A$ . For any  $V \subset [n]$ , let  $S_n(l; V) = \sum_{i \in V} (\kappa_i + d_i(l))$ . We will also define a random measure  $\bar{S}_{n,\kappa}$  by  $\bar{S}_{n,\kappa}(A) = \frac{\sqrt{2m_n}}{\ell_{n,\kappa}} S_n(0, V_n(A))$ .

The main ingredients of the proof can be decomposed into the following lemma and propositions; in all of them, we assume that  $\min\{\ell_{n,\kappa}, m_n\} = \omega(\log n)$  and  $\max_i \kappa_i = o(\ell_{n,\kappa})$ , stating any additional assumption explicitly.

LEMMA 3.1. *If  $\log m_n = o(\ell_{n,\kappa})$ , then  $e(\text{PA}_n(\kappa, m_n)) = m_n(1 + o(1))$  a.s.  $\text{PA}_n(\kappa, m_n)$ .*

PROPOSITION 3.2. *Let  $(V_i)_{i \in [k]}$  be a disjoint collection of vertex subsets such that, for all  $i \in [k]$ ,  $S_n(0, V_i) = O(\ell_{n,\kappa}/\sqrt{m_n})$ , and for any  $\varepsilon > 0$ ,*

$$(3.1) \quad \lim_{n \rightarrow \infty} \mathbb{P}\left(\sup_{l \leq m_n} \left| \frac{\sqrt{2m_n} S_n(l, V_i)}{\ell_{n,\kappa} + 2l} - \frac{\sqrt{2m_n} S_n(0, V_i)}{\ell_{n,\kappa}} \right| > \varepsilon\right) = 0.$$



Let  $\mathcal{E}_n := (\mathcal{E}(V_i, V_j))_{1 \leq i \leq j \leq k}$ ,  $\mathcal{E} := (\mathcal{E}_{ij})_{1 \leq i \leq j \leq k}$ , where  $\mathcal{E}(V_i, V_j)$  denotes the number of edges in  $\text{PA}_n(\kappa, m_n)$  with one end-point in  $V_i$  and the other end-point in  $V_j$ , and  $\mathcal{E}$  is a collection of independent random variables with  $\mathcal{E}_{ij} \sim \text{Poisson}(2m_n S_n(0, V_i) S_n(0, V_j) / \ell_{n,\kappa}^2)$ ,  $i \neq j$ , while  $\mathcal{E}_{ii} \sim \text{Poisson}(m_n S_n(0, V_i)^2 / \ell_{n,\kappa}^2)$ . Then

$$\lim_{n \rightarrow \infty} d_{\text{TV}}(\mathcal{E}_n, \mathcal{E}) \rightarrow 0.$$

Further, suppose that  $V_i$ 's are random subsets chosen independently of  $\text{PA}_n(\kappa, m_n)$  such that  $\mathbb{E}[S_n(0, V_i)] = O(\ell_{n,\kappa} / \sqrt{m_n})$  and such for any  $\varepsilon > 0$ , (3.1) holds, where the probabilities and expectations under consideration are to be taken over the joint distribution of  $V_i$ 's and  $d_i(l)$ 's. Then  $\lim_{n \rightarrow \infty} d_{\text{TV}}(\mathcal{E}_n, \mathcal{E}) = 0$ .

We will want to apply this proposition to random sets of the form  $V_i = V_n(A_i)$ , where  $A_i \in \mathcal{B}(\mathbb{R}_+)$ ,  $i = 1, \dots, k$  are pairwise disjoint. To do this, we use the following proposition, and the fact that  $\mathbb{E}[S_n(0, V_n(A_i))] = \Lambda(A_i) \ell_{n,\kappa} / \sqrt{m_n} = O(\ell_{n,\kappa} / \sqrt{m_n})$ .

REMARK 12. Before stating the next proposition, let us apply the statements of the previous one to the sets  $V_i = \{i\}$  to relate the proposition to the heuristic arguments given for the preferential attachment model in the Introduction. Since our actual proof will not use this argument, let us not worry about verifying the condition (3.1). Then the statement of the proposition say that the number of edges between  $i$  and  $j$  is a Poisson random variable with parameter  $2m_n \kappa_i \kappa_j / \ell_{n,\kappa}^2 = \bar{d}_i \bar{d}_j / 2m_n$ , and the number of loops at  $i$  is a Poisson random variable with parameter  $\bar{d}_i^2 / 4m_n$ , showing at least heuristically that the preferential attachment model behaves like a configuration model with degree sequence  $(\bar{d}_i)_{i \in [n]}$ .

PROPOSITION 3.3. Fix  $A \in \mathcal{B}(\mathbb{R}_+)$ , and  $\varepsilon > 0$ . If  $m_n = o(\ell_{n,\kappa}^2)$ , then

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \sup_{l \leq m_n} \left| \frac{\sqrt{2m_n} S_n(l, V_n(A))}{\ell_{n,\kappa} + 2l} - \frac{\sqrt{2m_n} S_n(0, V_n(A))}{\ell_{n,\kappa}} \right| > \varepsilon \right) = 0.$$

Finally, we need the following analogue of Proposition 2.4.

PROPOSITION 3.4. For any  $A, B \in \mathcal{B}(\mathbb{R}_+)$ ,  $l \in \mathbb{N}$  and  $\delta > 0$ ,

$$\begin{aligned} & \mathbb{P}(|\mathbb{P}(\xi_n(A \times B) = l | \text{PA}_n(\kappa, m_n)) - \mathbb{P}(\xi_n(A \times B) = l)| > \delta) \\ & \leq 2 \exp\left(-\frac{\delta^2 m_n}{8\Lambda(A)^2 \Lambda(B)^2}\right). \end{aligned}$$

First, we complete the proof of Theorem 1.6, given Lemma 3.1, and Propositions 3.3, 3.2 and 3.4, and defer their proofs to the end of the section.

PROOF OF THEOREM 1.6. The equivalence of (ii), (iii) and (iv) can be read of Theorem 1.2 applied to the sequence  $(\bar{d}_i)_{i \in [n]}$ . All we need to observe is that  $\sum_i \bar{d}_i = 2m_n$ , and that our assumptions on  $\max_i \kappa_i$ ,  $\ell_{n,\kappa}$  and  $m_n$  imply that  $\max_i \bar{d}_i = o(2m_n)$  and  $m_n = \omega(\log n)$ .

To show equivalence of (i) and (iv), we first note that the assumptions of Theorem 1.6 implies those of Lemma 3.1 and Proposition 3.3. We then apply Proposition 3.2 to the sets  $V_n(A_1), \dots, V_n(A_k)$ , where  $A_i \in \mathcal{B}(\mathbb{R}_+)$ ,  $i = 1, \dots, k$ , and rewrite the statement of the proposition for this case in terms of the random variables  $\bar{S}_{n,\kappa}(A_i)$ ,  $i = 1, \dots, k$ . The proof is then identical to the proof of the equivalence of (i) and (iv) in Theorem 1.2, once we replace Lemma 2.1 by Lemma 3.1, Proposition 2.4 by Proposition 3.4 and Proposition 2.2 by

Proposition 3.2 and Proposition 3.3. The proof that the sampling limit is given by the graphex  $\mathcal{W}_{\text{CM}}$  is again the same, once we observe that Proposition 2.3 and (2.4) hold for  $S_{n,\kappa}$ ,  $\rho_{n,\kappa}$  and  $(X_{n,\kappa}(t))_{t \geq 0}$  where  $X_{n,\kappa}(t) = \mu_{n,\kappa}([0, t])$ .  $\square$

Next, we prove Propositions 3.3, Lemma 3.1, and Propositions 3.4 and 3.2, in that order. We let  $(\mathcal{F}_l)_{l=1}^{m_n}$  denote the canonical filtration associated with the graph process  $(\text{PA}_n(\kappa, l))_{l=1}^{m_n}$ , and let  $(\mathcal{F}'_l)_{l=1}^{m_n}$  be a filtration, where  $\mathcal{F}'_l$  is the minimal sigma algebra containing the information about  $V_n(A)$  and  $\text{PA}_n(\kappa, l)$ . Note that  $\mathcal{F}'_0$  is the sigma algebra containing the information about  $V_n(A)$  only, while both  $\text{PA}_n(\kappa, m_n)$  and  $V_n(A)$  are measurable with respect to the filtration  $\mathcal{F}'_{m_n}$ .

PROOF OF PROPOSITION 3.3. To avoid cumbersome notation, we write  $S_n(l)$  for  $S_n(l, V_n(A))$  in this proof. Clearly,  $(S_n(l))_{l \in [0, m_n]}$  is a Markov chain with  $S_n(0) = \sum_{i \in V_n(A)} \kappa_i$  and conditionally on  $\mathcal{F}'_l$ ,

$$(3.2) \quad S_n(l + 1) = \begin{cases} S_n(l) & \text{with probability } \frac{(\ell_{n,\kappa} + 2l - S_n(l))^2}{(\ell_{n,\kappa} + 2l)^2}, \\ S_n(l) + 1 & \text{with probability } \frac{2S_n(l)(\ell_{n,\kappa} + 2l - S_n(l))}{(\ell_{n,\kappa} + 2l)^2}, \\ S_n(l) + 2 & \text{with probability } \frac{S_n(l)^2}{(\ell_{n,\kappa} + 2l)^2}. \end{cases}$$

Note that for any  $l \geq 0$ ,

$$\mathbb{E}[S_n(l + 1) - S_n(l) | \mathcal{F}'_l] = \frac{2S_n(l)}{\ell_{n,\kappa} + 2l} \implies \mathbb{E}\left[\frac{S_n(l + 1)}{\ell_{n,\kappa} + 2(l + 1)} \middle| \mathcal{F}'_l\right] = \frac{S_n(l)}{\ell_{n,\kappa} + 2l},$$

and, therefore,  $(\frac{\sqrt{2m_n}S_n(l)}{\ell_{n,\kappa} + 2l})_{l=0}^{m_n}$  is a martingale with respect to  $(\mathcal{F}'_l)_{l=0}^{m_n}$ . Let QV denote the quadratic variation of this martingale. To compute QV, note that

$$(3.3) \quad \begin{aligned} & \mathbb{E}\left[\left(\frac{\sqrt{2m_n}S_n(l + 1)}{\ell_{n,\kappa} + 2l + 2} - \frac{\sqrt{2m_n}S_n(l)}{\ell_{n,\kappa} + 2l}\right)^2 \middle| \mathcal{F}'_l\right] \\ & \leq \frac{4m_n}{(\ell_{n,\kappa} + 2l)^2(\ell_{n,\kappa} + 2l + 2)^2} \\ & \quad \times ((\ell_{n,\kappa} + 2l)^2 \mathbb{E}[(S_n(l + 1) - S_n(l))^2 | \mathcal{F}'_l] + 4S_n^2(l)) \\ & \leq \frac{4m_n}{(\ell_{n,\kappa} + 2l)^4} (2(\ell_{n,\kappa} + 2l)^2 \mathbb{E}[(S_n(l + 1) - S_n(l)) | \mathcal{F}'_l] + 4S_n^2(l)) \\ & = \frac{16m_n S_n(l)}{(\ell_{n,\kappa} + 2l)^3} + \frac{16m_n S_n^2(l)}{(\ell_{n,\kappa} + 2l)^4}. \end{aligned}$$

We will need the following fact, whose proof is given immediately after completing the proof of this proposition.

FACT 3.1.  $\mathbb{E}[S_n(l)] = O(\frac{\ell_{n,\kappa} + 2l}{\sqrt{m_n}})$  and  $\mathbb{E}[S_n(l)^2] = o(\frac{(\ell_{n,\kappa} + 2l)^2}{\sqrt{m_n}})$ .

Using Fact 3.1, (3.3) now shows that for all  $t \in [0, 1]$ , we have

$$\begin{aligned} \mathbb{E}[\text{QV}(tm_n)] &= \sum_{l < tm_n} \mathbb{E}\left[\left(\frac{\sqrt{2m_n}S_n(l + 1)}{\ell_{n,\kappa} + 2l + 2} - \frac{\sqrt{2m_n}S_n(l)}{\ell_{n,\kappa} + 2l}\right)^2 \middle| \mathcal{F}'_l\right] \\ &\leq 8m_n \sum_{l=0}^{\infty} \left(\frac{\mathbb{E}[S_n(l)]}{(\ell_{n,\kappa} + 2l)^3} + \frac{\mathbb{E}[S_n(l)^2]}{(\ell_{n,\kappa} + 2l)^4}\right) = O\left(\frac{\sqrt{m_n}}{\ell_{n,\kappa}}\right) = o(1). \end{aligned}$$

An application of Doob’s inequality [34], Chapter 1, Section 9, Theorem 1(3), now completes the proof.  $\square$

**PROOF OF FACT 3.1.** Recall that  $(\frac{\sqrt{2m_n}S_n(l)}{\ell_{n,\kappa}+2l})_{l=1}^{m_n}$  is a martingale with respect to  $(\mathcal{F}_l)_{l=1}^{m_n}$ . Thus,

$$\mathbb{E}[S_n(l)] = \frac{\ell_{n,\kappa} + 2l}{\ell_{n,\kappa}} \mathbb{E}[S_n(0)] = O\left(\frac{\ell_{n,\kappa} + 2l}{\sqrt{m_n}}\right).$$

For bound on the second moment, note that (3.2) implies

$$\begin{aligned} \mathbb{E}[S_n(l+1)^2] &= \mathbb{E}[(S_n(l))^2] \left(1 + \frac{4}{(\ell_{n,\kappa} + 2l)} + \frac{2}{(\ell_{n,\kappa} + 2l)^2}\right) + \frac{2\mathbb{E}[S_n(l)]}{\ell_{n,\kappa} + 2l} \\ &= \mathbb{E}[(S_n(l))^2] \left(1 + \frac{2}{(\ell_{n,\kappa} + 2l)}\right)^2 + O\left(\frac{1}{\sqrt{m_n}}\right), \end{aligned}$$

which in turn shows that

$$\begin{aligned} \frac{\mathbb{E}[(S_n(l))^2]}{(\ell_{n,\kappa} + 2l)^2} &\leq \frac{\mathbb{E}[S_n(0)^2]}{\ell_{n,\kappa}^2} + O\left(\frac{1}{\sqrt{m_n}}\right) \sum_{k=0}^l \frac{1}{(\ell_{n,\kappa} + 2k + 1)^2} \\ &= \frac{\mathbb{E}[S_n(0)^2]}{\ell_{n,\kappa}^2} + O\left(\frac{1}{\ell_{n,\kappa}\sqrt{m_n}}\right) \end{aligned}$$

Combined with the bound

$$\mathbb{E}[S_n(0)^2] = \text{Var}(S_n(0)) + \mathbb{E}^2[S_n(0)] = O\left(\frac{\ell_{n,\kappa}}{\sqrt{m_n}} \max_{i \in [n]} \kappa_i + \frac{\ell_{n,\kappa}^2}{m_n}\right) = O\left(\frac{\ell_{n,\kappa}^2}{\sqrt{m_n}}\right),$$

(where the last step follows using our assumption that  $\max_{i \in [n]} \kappa_i = o(\ell_{n,\kappa})$ ), the claim now follows.  $\square$

**PROOF OF LEMMA 3.1.** Recall that, according to the definition preceding Lemma 3.1, we have  $S_n(l, \{i\}) := d_i(l) + \kappa_i$ . Let  $L_l$  denote the number of loops in  $\text{PA}_n(\kappa, l)$ , and let  $P_l = \frac{\sum_{i \in [n]} (S_n(l, \{i\}))^2}{(\ell_{n,\kappa} + 2l)^2}$ . Conditionally on  $\mathcal{F}_l$ , we then have that

$$L_{l+1} = \begin{cases} L_l + 1 & \text{with probability } P_l, \\ L_l & \text{otherwise.} \end{cases}$$

Setting  $L'_l = L_l - \sum_{k=0}^{l-1} P_k$ , we see that  $(L'_l)_{l=0}^{m_n}$  is a Martingale with respect to the filtration  $(\mathcal{F}_l)_{l=0}^{m_n}$ , and that  $L'_l - P_l \leq L'_{l+1} \leq L'_l + 1$ . The Azuma–Hoeffding inequality and the fact that  $L_0 = 0$  then implies that

$$(3.4) \quad \mathbb{P}\left(L_{m_n} \geq \sum_{l=0}^{m_n-1} P_l + \lambda m_n\right) = \mathbb{P}(L'_{m_n} \geq \lambda m_n) \leq e^{-\frac{\lambda^2 m_n}{2}}.$$

Next, define  $Q_l := \sum_{i \in [n]} (S_n(l, \{i\}))^2$ . By a simple calculation using the analogue of (3.2) for  $S_n(l, \{i\})$ , we get that

$$\mathbb{E}[Q_{l+1} | \mathcal{F}_l] = Q_l \left(1 + \frac{4}{(\ell_{n,\kappa} + 2l)} + \frac{2}{(\ell_{n,\kappa} + 2l)^2}\right) + 2 \leq Q_l \left(\frac{\ell_{n,\kappa} + 2l + 2}{\ell_{n,\kappa} + 2l}\right)^2 + 2,$$

which we rewrite as

$$\mathbb{E}[P_{l+1} | \mathcal{F}_l] \leq P_l + \frac{2}{(\ell_{n,\kappa} + 2l + 2)^2}$$

to conclude that  $P'_l = P_l - \sum_{k=1}^l \frac{2}{(\ell_{n,\kappa} + 2k)^2}$  is a supermartingale. Next, we note that upon the addition of an edge at step  $l + 1$ , either one of the terms in  $Q_l$  change from  $(S_n(l, \{i\}))^2$  to  $(S_n(l, \{i\}) + 2)^2$  or two of them change from  $(S_n(l, \{i\}))^2$  to  $(S_n(l, \{i\}) + 1)^2$ . In either case, the total change is at most  $4(\ell_{n,\kappa} + 2\ell) + 4$ . A straightforward calculation using the fact that  $Q_l \leq (\ell_{n,\kappa} + 2\ell)^2$  then shows that

$$-\frac{4}{\ell_{n,\kappa} + 2l + 2} \leq P'_{l+1} - P'_l \leq \frac{4}{\ell_{n,\kappa} + 2l + 2}.$$

The Azuma–Hoeffding inequality combined with the facts that  $P'_0 = P_0$  then implies that

$$(3.5) \quad \mathbb{P}\{\exists l \leq m_n : P'_l \geq P_0 + \lambda\} \leq m_n e^{-\frac{\lambda^2 \ell_{n,\kappa}}{32}} \leq e^{-\frac{\lambda^2 \ell_{n,\kappa}}{64}}$$

provided  $\log m_n = o(\ell_{n,\kappa})$  and  $n$  large enough. Combining (3.4) and (3.5) then shows that with probability at least  $1 - e^{-\lambda^2 m_n / 2} - e^{-\lambda^2 \ell_{n,\kappa} / 64}$ , we have that

$$L_{m_n} < \sum_{l=0}^{m_n-1} P_l + \lambda m_n \leq \sum_{l=0}^{m_n-1} P'_l + \frac{m_n}{\ell_{n,\kappa}^2} + \lambda m_n \leq m_n P_0 + \frac{m_n}{\ell_{n,\kappa}^2} + 2\lambda m_n = (2\lambda + o(1))m_n,$$

where in the last step we used that  $P_0 = \frac{1}{\ell_{n,\kappa}^2} \sum_{i \in [n]} \kappa_i^2 = o(\ell_{n,\kappa}) \frac{1}{\ell_{n,\kappa}^2} \sum_{i \in [n]} \kappa_i = o(1)$ . By our assumption that  $\min\{m_n, \ell_{n,\kappa}\} = \omega(\log n)$ , the error probability is summable for all fixed  $\lambda > 0$ . Since  $\lambda$  was arbitrary, this proves that  $L_{m_n} / m_n \rightarrow 0$  with probability 1.  $\square$

**PROOF OF PROPOSITION 3.4.** Let  $X := \mathbb{P}(\xi_n(A \times B) = l | \mathcal{P}_n(\kappa, m_n))$  and let  $(Y_l := \mathbb{E}[X | \mathcal{F}_l])_{l=0}^{m_n}$  denote the Doob martingale with respect to the filtration  $(\mathcal{F}_l)_{l=0}^{m_n}$ . Therefore,  $Y_0 = \mathbb{E}[X]$ , while  $Y_{m_n} = X$ . Moreover,  $X$  can change by changing the status of the edge  $(i, j)$  only if  $(i, j) \in V_n(A) \times V_n(B)$  or  $(j, i) \in V_n(A) \times V_n(B)$ . Thus, the martingale difference  $|Y_l - Y_{l-1}|$  can be bounded as

$$|Y_l - Y_{l-1}| \leq \frac{2\Lambda(A)\Lambda(B)}{m_n}.$$

An application of the Azuma–Hoeffding inequality now completes the proof.  $\square$

**PROOF OF PROPOSITION 3.2.** We prove the proposition for  $k = 2$ , the general case follows similarly. For  $l \in [m_n]$ , we denote the added edge at time  $l$  by  $\{v_1(l), v_2(l)\}$ , and let

$$I_{l+1} = \begin{cases} (1, 0, 0) & \text{if } v_1(l), v_2(l) \in V_1, \\ (0, 0, 1) & \text{if } v_1(l), v_2(l) \in V_2, \\ (0, 1, 0) & \text{if } v_1(l) \in V_1, \text{ and } v_2(l) \in V_2 \text{ or vice versa,} \\ (0, 0, 0) & \text{otherwise.} \end{cases}$$

Note that  $(\mathcal{E}(V_1, V_1), \mathcal{E}(V_1, V_2), \mathcal{E}(V_2, V_2)) = \sum_{l=1}^{m_n} I_l$ . To approximate this sum in total variation distance, we first couple the vectors  $I_l$  to independent random vectors, and then we couple the sum of independent indicators to independent Poisson random variables.

For the first step, we will use an explicit coupling. Let  $((U_1(l), U_2(l)))_{l=1}^{m_n}$  be i.i.d. random variables, with  $(U_1(1), U_2(1))$  being uniformly distributed on the unit square  $[0, 1]^2$ . For  $i = 1, 2$ , we denote the events

$$\begin{aligned} G_{i1}(l) &= \left\{ U_i(l) \leq \frac{S_n(l, V_i)}{\ell_{n,\kappa} + 2l} \right\}, & G_{i2}(l) &= \left\{ \frac{S_n(l, V_i)}{\ell_{n,\kappa} + 2l} < U_i(l) \leq \frac{S_n(l, V_i) + S_n(l, V_2)}{\ell_{n,\kappa} + 2l} \right\}, \\ \hat{G}_{i1}(l) &= \left\{ U_i(l) \leq \frac{S_n(0, V_i)}{\ell_{n,\kappa}} \right\}, & \hat{G}_{i2}(l) &= \left\{ \frac{S_n(0, V_i)}{\ell_{n,\kappa}} < U_i(l) \leq \frac{S_n(0, V_i) + S_n(0, V_2)}{\ell_{n,\kappa}} \right\}. \end{aligned}$$

We then generate the processes  $(S_n(l, V_i))_{l=0}^{m_n}, i = 1, 2, (I_l)_{l=0}^{m_n}$  and an independent collection  $(\hat{I}_l)_{l=0}^{m_n}$  jointly as follows:

$$(3.6) \quad S_n(l + 1, V_i) = \begin{cases} S_n(l, V_i) + 2 & \text{on } G_{1i}(l) \cap G_{2i}(l), \\ S_n(l, V_i) + 1 & \text{on } (G_{1i}(l) \cap G_{2i}(l)^c) \cup (G_{1i}(l)^c \cap G_{2i}(l)), \\ S_n(l, V_i) & \text{otherwise,} \end{cases}$$

$$I_{l+1} = \begin{cases} (1, 0, 0) & \text{on } G_{11}(l) \cap G_{21}(l), \\ (0, 0, 1) & \text{on } G_{12}(l) \cap G_{22}(l), \\ (0, 1, 0) & \text{on } (G_{11}(l) \cap G_{22}(l)) \cup (G_{12}(l) \cap G_{21}(l)), \\ (0, 0, 0) & \text{otherwise.} \end{cases}$$

The definition of  $\hat{I}_{l+1}$  is identical to (3.6) by replacing  $G_{ij}$  by  $\hat{G}_{ij}, i, j = 1, 2$ . It is easy to see that the above give the processes  $(S_n(l, V_i))_{l=0}^{m_n}, i = 1, 2, (I_l)_{l=0}^{m_n}$  defined before and  $(\hat{I}_l)_{l=0}^{m_n}$  is an independent collection. Next, fix  $\varepsilon > 0$  and let

$$\mathcal{A}_l = \left\{ \sup_{l' \leq l} \left| \frac{\sqrt{2m_n} S_n(l', V_i)}{\ell_{n,\kappa} + 2l'} - \frac{\sqrt{2m_n} S_n(0, V_i)}{\ell_{n,\kappa}} \right| \leq \varepsilon \right\}, \quad \mathcal{A} = \bigcup_{l=1}^{m_n} \mathcal{A}_l.$$

By the assumption (3.1),  $\mathbb{P}(\mathcal{A}^c) = o(1)$  and

$$\begin{aligned} & \mathbb{P}(I_{l+1} \neq I'_{l+1} | \mathcal{F}'_l) \\ & \leq 2 \left[ \left| \frac{S_n(l, V_1)}{\ell_{n,\kappa} + 2l} - \frac{S_n(0, V_1)}{\ell_{n,\kappa}} \right| + \left| \frac{S_n(l, V_1 \cup V_2)}{\ell_{n,\kappa} + 2l} - \frac{S_n(0, V_1 \cup V_2)}{\ell_{n,\kappa}} \right| \right] \\ & \quad \times \left[ \frac{S_n(l, V_1 \cup V_2)}{\ell_{n,\kappa} + 2l} + \frac{S_n(0, V_1 \cup V_2)}{\ell_{n,\kappa}} \right], \end{aligned}$$

so that  $\mathbb{P}(I_{l+1} \neq I'_{l+1} | \mathcal{F}'_l) \mathbb{1}_{\mathcal{A}_l} \leq C\varepsilon/m_n$ , for some constant  $C > 0$ . Thus,

$$\begin{aligned} & \mathbb{P}(\exists l : I_{l+1} \neq \hat{I}_{l+1}) \\ & = \mathbb{P}(\exists l : I_{l+1} \neq \hat{I}_{l+1}, \mathcal{A}) + o(1) \leq \mathbb{P}(\exists l : I_{l+1} \neq \hat{I}_{l+1}, \mathcal{A}_l) + o(1) \\ & \leq \sum_{l=1}^{m_n} \mathbb{E}[\mathbb{P}(I_{l+1} \neq I'_{l+1} | \mathcal{F}'_l) \mathbb{1}_{\mathcal{A}_l}] + o(1) \leq C\varepsilon + o(1). \end{aligned}$$

Since  $\varepsilon > 0$  is arbitrary,  $\lim_{n \rightarrow \infty} \mathbb{P}(\exists l : I_{l+1} \neq \hat{I}_{l+1}) = 0$ . Now, we can use multivariate Stein’s method to approximate  $\sum_{l=1}^{m_n} \hat{I}_l$  in an identical manner as at the end of Section 2.1, this time with

$$\begin{aligned} \lambda_{11} &= \sum_{\alpha=1}^{m_n} \mathbb{P}(\hat{I}_\alpha = (1, 0, 0)) = \frac{m_n}{\ell_{n,\kappa}^2} S_n(0, V_1)^2, \\ \lambda_{12} &= \sum_{\alpha=1}^{m_n} \mathbb{P}(\hat{I}_\alpha = (0, 1, 0)) = \frac{2m_n}{\ell_{n,\kappa}^2} S_n(0, V_1) S_n(0, V_2), \\ \lambda_{22} &= \sum_{\alpha=1}^{m_n} \mathbb{P}(\hat{I}_\alpha = (0, 0, 1)) = \frac{m_n}{\ell_{n,\kappa}^2} S_n(0, V_2)^2. \end{aligned}$$

□

**4. Proof of results on generalized random graph.** We first establish that the number of edges in  $\text{GRG}_n(\mathbf{w})$  is concentrated around a deterministic value. Recall that for any graph  $G$ , we use  $e(G)$  to denote the number of nonloop edges in  $G$ , and that  $L_n$  denotes the  $\ell_1$  norm of the weight vector,  $L_n = \sum_{i \in [n]} w_i$ .

LEMMA 4.1. For  $0 < \varepsilon \leq 1$ ,

$$(4.1) \quad \begin{aligned} \mathbb{P}(|e(\text{GRG}_n(\mathbf{w})) - \mathbb{E}[e(\text{GRG}_n(\mathbf{w}))]| > \varepsilon L_n) &\leq 2 \exp\left(-\frac{\varepsilon^2 L_n}{3}\right), \\ \frac{1}{L_n} \mathbb{E}[e(\text{GRG}_n(\mathbf{w}))] &= \frac{1}{2} \int_0^\infty \int_0^\infty \frac{xy}{1+xy} \rho_n(dx) \rho_n(dy) + o(1). \end{aligned}$$

We denote the the random point process  $\text{Lbl}_{\sqrt{L_n}}(\text{GRG}_n(\mathbf{w}))$  for the graph  $\text{GRG}_n(\mathbf{w})$  by  $\xi_n$ . Let  $\xi$  denote the random adjacency measure corresponding to the graphex  $\mathcal{W}_{\text{GRG}}^1$ .

PROPOSITION 4.2. Under Assumption 1, as  $n \rightarrow \infty$ ,

$$\mathbb{P}(\xi_n(A) = 0) \rightarrow \mathbb{P}(\xi(A) = 0),$$

for any  $A$  that is a union of disjoint rectangles in  $\mathbb{R}_+^2$ .

PROPOSITION 4.3. Suppose that Assumption 1 holds. For  $0 < \varepsilon \leq 1$  and  $A \subset \mathcal{B}(\mathbb{R}_+^2)$  that is a union of disjoint rectangles, there exists a constant  $C = C(\varepsilon, A) > 0$  such that

$$\mathbb{P}(|\mathbb{P}(\xi_n(A) = 0 | \text{GRG}_n(\mathbf{w})) - \mathbb{P}(\xi_n(A) = 0)| > \varepsilon) \leq e^{-CL_n}.$$

We first prove Theorem 1.7, given these results.

PROOF OF THEOREM 1.7. Lemma 4.1 implies that for any  $\delta_n \rightarrow 0$ ,

$$\mathbb{P}(|e(\text{GRG}_n(\mathbf{w})) - \mathbb{E}[e(\text{GRG}_n(\mathbf{w}))]| > \delta_n L_n) \leq 2 \exp\left(-\frac{\delta_n^2 L_n}{3}\right).$$

As  $L_n = \omega(\log n)$ , we can choose  $\delta_n \rightarrow 0$  such that the above probabilities are summable. As a result, Lemma 4.1 implies  $\mathbb{P}_{\text{GRG}}$  a.s.,

$$\frac{1}{L_n} e(\text{GRG}_n(\mathbf{w})) - \frac{1}{L_n} \mathbb{E}[e(\text{GRG}_n(\mathbf{w}))] \rightarrow 0.$$

By Lemma D.1, it is therefore enough to show that

$$\mathcal{L}(\xi_n | \text{GRG}_n(\mathbf{w})) \rightarrow \mathcal{L}(\xi), \quad \mathbb{P}_{\text{GRG}} \text{ a.s.}$$

To this end, we use [33], Theorem A.1. To apply this theorem, we need to show that for every union  $A$  is a union of disjoint rectangles in  $\mathbb{R}_+^2$ , we have

$$(4.2) \quad \mathbb{P}(\xi_n(A) = 0 | \text{GRG}_n(\mathbf{w})) \rightarrow \mathbb{P}(\xi(A) = 0), \quad \mathbb{P}_{\text{GRG}} \text{ a.s.},$$

$$(4.3) \quad \mathbb{E}[\xi_n(A) | \text{GRG}_n(\mathbf{w})] \rightarrow \mathbb{E}[\xi(A)], \quad \mathbb{P}_{\text{GRG}} \text{ a.s.}$$

Propositions 4.2, and 4.3 together directly imply (4.2). We will verify (4.3) only for  $A = [0, t]^2$ , leaving the general case to the reader. Note that

$$\mathbb{E}[\xi_n(A) | \text{GRG}_n(\mathbf{w})] = (1 + o(1)) \frac{t^2}{L_n} e(\text{GRG}_n(\mathbf{w})) \rightarrow \frac{t^2}{2} c, \quad \mathbb{P}_{\text{GRG}} \text{ a.s.},$$

where the last step follows from Lemma 4.1. This concludes the proof.  $\square$

PROOF OF COROLLARY 1.8. Consider a sampling convergent subsequence of  $\text{GRG}_n(\mathbf{w})$ . Observing that

$$0 \leq \int_0^\infty \int_0^\infty \frac{xy}{1+xy} \rho_n(dx) \rho_n(dy) \leq \left( \int_0^\infty x \rho_n(dx) \right) \left( \int_0^\infty y \rho_n(dy) \right) = 1.$$

We may therefore choose a further subsequence such that (1.8) holds for some  $c > 0$ . By Theorem 1.7, this subsequence is a.s. sampling convergent to  $\mathcal{W}_{\text{GRG}}^c$ . Since a sequence cannot converge to  $\mathcal{W}_{\text{GRG}}^c$  and  $\mathcal{W}_{\text{GRG}}^{c'}$  for  $c \neq c'$ , this completes the proof.  $\square$

PROOF OF COROLLARY 1.9. The proof is similar to that of Corollary 1.4 in Section 2.1, and thus we only sketch the main ideas. First, by our assumption that  $a = 0$ ,

$$\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \int_0^\varepsilon x \rho_n(dx) = 0.$$

As in the proof of Corollary 1.4, we set  $V_{>\varepsilon} = \{i : w_i > \varepsilon \sqrt{L_n}\}$  and set  $\mathcal{E}_{\text{GRG}}(\varepsilon)$  to denote the number of edges with at least one end in  $V_{\leq \varepsilon}$ . This implies

$$\mathbb{E}[\mathcal{E}_{\text{GRG}}(\varepsilon)] = \sum_{i,j \in V_{\leq \varepsilon}: i < j} p_{ij} + \sum_{i \in V_{\leq \varepsilon}} \sum_{j \notin V_{\leq \varepsilon}} p_{ij} \leq \sum_{i \in V_{< \varepsilon}} \sum_{j \in [n]} \frac{w_i w_j}{L_n} = L_n \int_0^\varepsilon x \rho_n(dx),$$

where in the second to last step we used that  $p_{ij} \leq w_i w_j / L_n$ . Next, concentration for sum of independent Bernoulli variables [29], (2.5) and (2.6), Theorem 2.8, and the fact that  $\mathbb{E}[\mathcal{E}_{\text{GRG}}(\varepsilon)] \leq L_n$  immediately implies that as long as  $0 < \delta_n \leq 1$ ,

$$\begin{aligned} \mathbb{P}(|\mathcal{E}_{\text{GRG}}(\varepsilon) - \mathbb{E}[\mathcal{E}_{\text{GRG}}(\varepsilon)]| > \delta_n L_n) &\leq 2 \exp\left(-\frac{L_n^2 \delta_n^2}{2(\mathbb{E}[\mathcal{E}_{\text{GRG}}(\varepsilon)] + L_n \delta_n / 3)}\right) \\ &\leq 2e^{-\frac{L_n \delta_n^2}{3}}. \end{aligned}$$

Choosing  $\delta_n = o(1)$  in such a way that the error bound is summable (which is possible by our assumption that  $L_n = \omega(\log n)$ ), we conclude that  $\mathbb{P}_{\text{GRG}}$  a.s.,  $\mathcal{E}_{\text{GRG}}(\varepsilon) \leq L_n \text{Err}(\varepsilon, n) + o(L_n)$ , where we recall the notation  $\text{Err}(\varepsilon, n)$  from the proof of Corollary 1.4. The rest of the proof follows exactly as Corollary 1.4, upon setting  $U_n$  to be the set of vertices corresponding to  $V_{\leq \varepsilon}$ .  $\square$

It remains to establish Propositions 4.2–4.3 and Lemma 4.1. We prove Lemma 4.1 first and defer the proof of Proposition 4.2 to Section 4.1 and that of Proposition 4.3 to Section 4.2.

PROOF OF LEMMA 4.1. Note that

$$\mathbb{E}[e(\text{GRG}_n(\mathbf{w}))] = \frac{1}{2} \sum_{i \neq j} \frac{w_i w_j}{L_n + w_i w_j} = \frac{1}{2} \sum_{i,j \in [n]} \frac{w_i w_j}{L_n + w_i w_j} - \frac{1}{2} \sum_{i \in [n]} \frac{w_i^2}{L_n + w_i^2}.$$

Since

$$\sum_{i \in [n]} \frac{w_i^2}{L_n + w_i^2} \leq \sum_{i:w_i \leq \sqrt{L_n}} \frac{w_i^2}{L_n} + \sum_{i:w_i > \sqrt{L_n}} 1 = O(\sqrt{L_n}),$$

it follows that

$$\frac{1}{L_n} \mathbb{E}[e(\text{GRG}_n(\mathbf{w}))] = \frac{1}{2} \int_0^\infty \int_0^\infty \frac{xy}{1+xy} \rho_n(dx) \rho_n(dy) + o(1).$$

(4.1) follows by standard application of martingale concentration inequalities for sums of independent Bernoulli random variables such as [29], (2.5) and (2.6), Theorem 2.8, together with the observation that  $\mathbb{E}[e(\text{GRG}_n(\mathbf{w}))] \leq L_n \int \int \frac{xy}{1+xy} \rho_n(dx) \rho_n(dy) \leq L_n$ .  $\square$

4.1. *Proof of Proposition 4.2.* Fix any  $k \geq 1$ , let  $(B_i)_{i \in [k]}$  be a collection of disjoint intervals, and let  $\mathcal{E}_n(B_i, B_j)$  denote the number of edges between vertices with labels in  $B_i$  and  $B_j$ , respectively. We will want to prove that

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{E}_n(B_i \times B_j) = 0, \forall 1 \leq i \leq j \leq k) = \mathbb{P}(\xi(B_i \times B_j) = 0, \forall 1 \leq i \leq j \leq k).$$

Let  $N = (w_i, \theta_i)_{i \geq 1}$  denote the Poisson point process on  $\mathbb{R}_+^2$  with intensity  $\rho(dw) \otimes d\theta$ . Further let  $N_{i\varepsilon} = N([\varepsilon, \infty) \times B_i)$ . Throughout the proof,  $\text{Err}(\varepsilon, n)$  is a generic notation for some function  $f(\varepsilon, n)$  s.t.  $\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} f(\varepsilon, n) = 0$ . Similarly,

$$\lim_{\varepsilon \rightarrow 0} \limsup_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} \text{Err}(\varepsilon, K, n) = 0$$

and

$$\lim_{K \rightarrow \infty} \limsup_{\varepsilon \rightarrow 0} \limsup_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \text{Err}(K, \varepsilon, M, n) = 0.$$

Fix  $\varepsilon > 0$  and let  $V_{>\varepsilon} = \{i \in [n] : w_i > \varepsilon\sqrt{L_n}\}$  and  $V_{\leq\varepsilon}^c = V_{>\varepsilon}$ . Recalling that we assigned a random label in  $[0, \sqrt{L_n}]$  to each vertex in  $[n]$ , let  $V_i$  be the set of vertices with labels in  $B_i$ . We set  $V_i^> = V_i \cap V_{>\varepsilon}$ ,  $V_i^{\leq} = V_i \cap V_{\leq\varepsilon}$  and  $T_i = \sum_{u \in V_i^{\leq}} \bar{w}_u$ , where  $\bar{w}_u = w_u / \sqrt{L_n}$ .

Also let  $I_{ij} := \mathbb{1}\{i \text{ and } j \text{ create an edge}\}$ . Thus,  $(I_{ij})_{1 \leq i < j \leq n}$  is an independent collection of Bernoulli random variables with  $\mathbb{P}(I_{ij} = 1) = p_{ij} = \bar{w}_i \bar{w}_j / (1 + \bar{w}_i \bar{w}_j)$ . Defining

$$\begin{aligned} F_{ii}^n(\varepsilon) &= \prod_{\substack{u < v \\ u, v \in V_i^>}} (1 - I_{uv}) \prod_{\substack{u \in V_i^> \\ v \in V_i^{\leq}}} (1 - I_{uv}) \prod_{\substack{u < v \\ u, v \in V_i^{\leq}}} (1 - I_{uv}) \\ &= F_{ii}^n(\varepsilon, 1) F_{ii}^n(\varepsilon, 2) F_{ii}^n(\varepsilon, 3), \\ (4.4) \quad F_{ij}^n(\varepsilon) &= \prod_{\substack{u \in V_i^> \\ v \in V_j^>}} (1 - I_{uv}) \prod_{\substack{u \in V_i^{\leq} \\ v \in V_j^>}} (1 - I_{uv}) \prod_{\substack{u \in V_i^> \\ v \in V_j^{\leq}}} (1 - I_{uv}) \prod_{\substack{u \in V_i^{\leq} \\ v \in V_j^{\leq}}} (1 - I_{uv}) \\ &= F_{ij}^n(\varepsilon, 1) F_{ij}^n(\varepsilon, 2) F_{ji}^n(\varepsilon, 2) F_{ij}^n(\varepsilon, 3), \end{aligned}$$

we note that

$$(4.5) \quad \mathbb{P}(\mathcal{E}_n(B_i \times B_j) = 0, \forall 1 \leq i \leq j \leq k) = \mathbb{E} \left[ \prod_{1 \leq i \leq j \leq k} F_{ij}^n(\varepsilon) \right]$$

We first state a lemma which identifies a ‘‘good’’ event.

LEMMA 4.4. *Define the events*

$$\begin{aligned} \mathcal{A}_{1i} &:= \left\{ |T_i - \mathbb{E}[T_i]| \leq \varepsilon^{1/4}, \sum_{u \in V_i^{\leq}} w_u^2 \leq \varepsilon^{1/2} L_n \right\}, \\ \mathcal{A}_{2i} &:= \left\{ \sum_{u \in V_i} \bar{w}_u \leq K \right\}, \quad \mathcal{A}_{3i} := \{|V_i^>| \leq M\}. \end{aligned}$$

Then, for  $\mathcal{A} = \bigcap_{i \in [k]} (\mathcal{A}_{1i} \cap \mathcal{A}_{2i} \cap \mathcal{A}_{3i})$ ,  $\mathbb{P}(\mathcal{A}^c) = \text{Err}(K, \varepsilon, M, n)$ .

Let  $\mathbb{P}_{[k]}$  (resp.,  $\mathbb{E}_{[k]}$ ) denote the conditional probability measure (resp., expectation) conditional on the choices of the random sets  $(V_i^r)_{i \in [k], r = >, \leq}$ . The next lemma characterizes the asymptotic behavior of  $F_{ii}^n, F_{ij}^n$ .



LEMMA 4.5. *On the set  $\mathcal{A}$ , for all  $1 \leq i \leq j \leq k$ ,*

$$(4.6) \quad \mathbb{E}_{[k]}[F_{ij}^n(\varepsilon, 3)] = \begin{cases} e^{-(\mathbb{E}[T_i])^2/2} + \text{Err}(\varepsilon, n) & \text{for } i = j, \\ e^{-\mathbb{E}[T_i]\mathbb{E}[T_j]} + \text{Err}(\varepsilon, n) & \text{for } i \neq j, \end{cases}$$

$$(4.7) \quad \mathbb{E}_{[k]}[F_{ij}^n(\varepsilon, 2)] = e^{-\mathbb{E}[T_i]\sum_{v \in V_j^>} \bar{w}_v} (1 + \text{Err}(K, \varepsilon, n)).$$

Moreover, for all  $i \in [k]$ ,

$$(4.8) \quad d_{\text{TV}}(\mathcal{L}(|V_i^>|), \text{Poi}(\Lambda(B_i)\rho([\varepsilon, \infty)))) = \text{Err}(\varepsilon, M, n).$$

Next, we prove Proposition 4.2, deferring the proof of the lemmas to the later part of this section.

PROOF OF PROPOSITION 4.2. To prove the proposition, we use the explicit expression for  $\mathbb{P}(\xi(B_i \times B_j) = 0, \forall 1 \leq i \leq j \leq k)$  given in Lemma C.2, the expression (4.5) for  $\mathbb{P}(\mathcal{E}_n(B_i \times B_j) = 0, \forall 1 \leq i \leq j \leq k)$ , and Lemmas 4.4 and 4.5. To avoid cumbersome notation, we prove this result for  $k = 1$ , and  $B_1 = [0, t]$ . The generalization to  $k \geq 1$  and arbitrary  $B_i$ 's is identical except the notational overhead, and thus we will sketch the general proof after proving the  $k = 1$  case. Let  $V_t$  denote the set of vertices with labels in  $[0, t]$ , and  $V_t^> = V_t \cap V_{>\varepsilon}$  and  $V_t^{\leq} = V_t \cap V_{\leq\varepsilon}$ . Also recall that  $T_1 = \sum_{u \in V_t^{\leq}} \bar{w}_u$ . Define the quantities

$$(4.9) \quad \begin{aligned} F_1 &:= \prod_{i < j, i, j \in V_t^>} (1 - I_{ij}), & F_2 &:= \prod_{i \in V_t^>, j \in V_t^{\leq}} (1 - I_{ij}), \\ F_3 &:= \prod_{i < j, i, j \in V_t^{\leq}} (1 - I_{ij}). \end{aligned}$$

Thus,  $\mathbb{P}(\xi_n([0, t]^2) = 0) = \mathbb{E}[F_1 F_2 F_3]$ . Notice that  $F_1, F_2, F_3 \leq 1$  almost surely, which we will use throughout the proof.

Next, by Assumption 1,

$$(4.10) \quad \mathbb{E}[T_1] = \frac{t}{\sqrt{L_n}} \sum_{i \in V_{\leq\varepsilon}} \frac{w_i}{\sqrt{L_n}} = t \int_0^\varepsilon x \rho_n(dx) = at + \text{Err}(\varepsilon, n).$$

Using Lemma 4.5, (4.6) and (4.7) together with (4.10) and Lemma 4.4,

$$(4.11) \quad \begin{aligned} \mathbb{E}[F_1 F_2 F_3] &= \mathbb{E}[F_1 F_2 F_3 \mathbb{1}_{\mathcal{A}}] + \text{Err}(K, \varepsilon, M, n) \\ &= e^{-\frac{a^2 t^2}{2}} \mathbb{E}\left[ e^{-at \sum_{i \in V_t^>} \bar{w}_i} \prod_{i < j, i, j \in V_t^>} \frac{1}{1 + \bar{w}_i \bar{w}_j} \mathbb{1}_{\mathcal{A}} \right] + \text{Err}(K, \varepsilon, M, n) \\ &= e^{-\frac{a^2 t^2}{2}} \mathbb{E}[f(n, |V_t^>|)] + \text{Err}(K, \varepsilon, M, n), \end{aligned}$$

where we set

$$f(n, k) := \int_{[\varepsilon, \infty)^k} e^{-at \sum_{i=1}^k w_i} \prod_{1 \leq i < j \leq k} \frac{1}{1 + w_i w_j} \prod_{i=1}^k \frac{\rho_n(dw_i)}{\rho_n([\varepsilon, \infty))}.$$

Indeed, if we take the expectation in the second line in (4.11) and condition on  $|V_t^>| = k$ , the elements of  $V_t^>$  are a sequence of  $k$  numbers chosen without replacement from  $\{i \in [n] : \bar{w}_i > \varepsilon\}$ . On the event  $\mathcal{A}$ , where  $k \leq K$ , we can replace the without replacement sampling by sampling with replacement at the cost of an error  $\text{Err}(K, n)$ , at which point we get a  $k$  independent samples from  $\rho_n(dw)$  conditioned on  $w > \varepsilon$ . This proves the last identity in (4.11).

Now, using the vague convergence of  $\rho_n$  from Assumption 1, we get that for any  $k \geq 1$  and any  $\varepsilon > 0$  such that  $\rho$  has no atom at  $\varepsilon$ ,

$$\lim_{n \rightarrow \infty} f(n, k) = \int_{[\varepsilon, \infty)^k} e^{-at \sum_{i=1}^k w_i} \prod_{1 \leq i < j \leq k} \frac{1}{1 + w_i w_j} \prod_{i=1}^k \frac{\rho(dw_i)}{\rho([\varepsilon, \infty))} =: f(k).$$

Therefore, Lemma 4.5, (4.8) implies that, for continuity point  $\varepsilon > 0$  of  $\rho$ , and any  $R \geq 1$ ,

$$(4.12) \quad \lim_{n \rightarrow \infty} \sum_{k=1}^R |f(n, |V_t^>|) \mathbb{P}(|V_t^>| = k) - f(k) \mathbb{P}(\text{Poi}(t\rho([\varepsilon, \infty))) = k)| = 0.$$

Now, notice that  $\max\{f(n, k), f(k)\} \leq 1$ . Thus,

$$(4.13) \quad \begin{aligned} \sum_{k > R} \mathbb{E}[f(n, k) | |V_t^>| = k] \mathbb{P}(|V_t^>| = k) &\leq \sum_{k > R} \binom{|V_{>\varepsilon}|}{k} \left(\frac{t}{\sqrt{L_n}}\right)^k \\ &\leq \sum_{k > R} \frac{1}{k!} (t\rho([\varepsilon, \infty)))^k, \end{aligned}$$

which goes to zero as  $R \rightarrow \infty$ . Thus (4.11), (4.12) and (4.13) together with Lemma C.2 imply that

$$\mathbb{E}[F_1 F_2 F_3 \mathbb{1}_{\mathcal{A}}] = \mathbb{P}(\xi([0, t]^2) = 0) + \text{Err}(K, \varepsilon, M, n),$$

and the proof follows using Lemma 4.4.

Let us now sketch the proof for the general  $k$  case briefly. For simplicity, let us consider  $B_i = [t_{i-1}, t_{i-1} + t_i]$ , where  $t_0 = 0$  and  $t_i > 0$  for  $i \in [k]$ . Recall the notations in (4.4). From the identity (4.5), we can use identical computations as in (4.11) that yields

$$\begin{aligned} \mathbb{P}(\mathcal{E}_n(B_i \times B_j) = 0, \forall 1 \leq i \leq j \leq k) \\ = \mathbb{E} \left[ \prod_{1 \leq i \leq j \leq k} F_{ij}^n(\varepsilon) \mathbb{1}_{\mathcal{A}} \right] + \text{Err}(K, \varepsilon, M, n) \\ = e^{-\frac{a^2}{2} \sum_{i=1}^k t_i^2 - a^2 \sum_{i < j} t_i t_j} \mathbb{E}[\tilde{f}(n, (|V_t^>|)_{i=1}^k)] + \text{Err}(K, \varepsilon, M, n), \end{aligned}$$

where we set

$$\begin{aligned} \tilde{f}(n, (r_i)_{i=1}^k) &:= \int_{[\varepsilon, \infty)^{\sum_{i=1}^k r_i}} e^{-a \sum_{i \neq j} t_i \sum_{l=1}^{r_j} w_{jl}} \prod_{1 \leq i < j \leq k} \prod_{\substack{1 \leq l_1 \leq r_i \\ 1 \leq l_2 \leq r_j}} \frac{1}{1 + w_{il_1} w_{jl_2}} \\ &\quad \times \prod_{i=1}^k \prod_{1 \leq l_1 < l_2 \leq r_i} \frac{1}{1 + w_{il_1} w_{il_2}} \prod_{i=1}^k \prod_{l=1}^{r_i} \frac{\rho_n(dw_{il})}{\rho_n([\varepsilon, \infty))}. \end{aligned}$$

The rest of the proof is identical to the case  $k = 1$ .  $\square$

Finally, we prove Lemma 4.4 and Lemma 4.5. To avoid notational overhead, we again prove these for the special case  $k = 1$ , and  $B_1 = [0, t]$  and the generalization to  $k > 1$  and general  $B_i$ 's follow using identical arguments. Recall the notations  $V_t, V_t^>, V_t^{\leq}$  defined above (4.9), which will be used throughout the proof.

PROOF OF LEMMA 4.4. First, note that

$$(4.14) \quad \mathbb{E} \left[ \sum_{i \in V_t^{\leq}} w_i^2 \right] \leq \varepsilon \sqrt{L_n} \mathbb{E} \left[ \sum_{i \in V_t} w_i \right] = t\varepsilon \sum_{i \in [n]} w_i = t\varepsilon L_n,$$

where the first inequality follows using  $\max_{i \in V_t^\leq} w_i \leq \varepsilon \sqrt{L_n}$ . Further,

$$\text{Var}(T_1) = \sum_{i \in V_{\leq \varepsilon}} \frac{w_i^2}{L_n} \frac{t}{\sqrt{L_n}} \left(1 - \frac{t}{\sqrt{L_n}}\right) \leq \frac{t\varepsilon}{L_n} \sum_{i \in V_{> \varepsilon}^c} w_i \leq t\varepsilon.$$

Thus, by Chebyshev’s inequality, together with (4.14) and Markov’s inequality yields  $\mathbb{P}(\mathcal{A}_1^c) = \text{Err}(\varepsilon, n)$ . Next, again by Markov’s inequality,

$$\mathbb{P}(\mathcal{A}_2^c) \leq \frac{1}{K} \mathbb{E} \left[ \sum_{i \in V_t} \bar{w}_i \right] = \frac{1}{K} \frac{t}{L_n} \sum_{i \in [n]} w_i = \frac{t}{K} = \text{Err}(1/K, n).$$

Finally,  $|V_t^>| \sim \text{Bin}(|V_{> \varepsilon}|, t/\sqrt{L_n})$ . Thus, another application of Markov’s inequality yields

$$\mathbb{P}(\mathcal{A}_3^c) \leq \frac{t|V_{> \varepsilon}|}{\sqrt{L_n}M} = \frac{t\rho_n([\varepsilon, \infty))}{M} = \text{Err}(\varepsilon, M, n). \quad \square$$

**PROOF OF LEMMA 4.5.** First, observe that, on the event  $\mathcal{A}$  defined in Lemma 4.4,

$$\begin{aligned} \mathbb{E}[F_2|V_t] &= \prod_{i \in V_t^>, j \in V_t^\leq} \frac{1}{1 + \bar{w}_i \bar{w}_j} = e^{-T_1 \sum_{i \in V_t^>} \bar{w}_i + O(\sum_{i \in V_t^>} \bar{w}_i^2 \sum_{j \in V_t^\leq} \bar{w}_j^2)} \\ &= e^{-\mathbb{E}[T_1] \sum_{i \in V_t^>} \bar{w}_i} (1 + \text{Err}(K, \varepsilon, n)), \end{aligned}$$

where we used that  $\sum_{i \in V_t^>} \bar{w}_i^2 \leq (\sum_{i \in V_t} \bar{w}_i)^2$  to obtain the final error bound, proving (4.7).

Next, let  $\lambda = \sum_{i, j \in V_t^\leq} p_{ij}$ . Using standard bounds for coupling sums of independent Bernoulli random variables to Poisson random variables [24], Theorem 2.10,

$$d_{\text{TV}} \left( \mathcal{L} \left( \sum_{i < j, i, j \in V_t^\leq} I_{ij} | V_t^>, V_t^\leq \right), \text{Poi}(\lambda) \right) \leq \sum_{i < j, i, j \in V_t^\leq} p_{ij}^2 \leq \frac{(\sum_{i \in V_t^\leq} w_i^2)^2}{L_n^2} \leq \varepsilon,$$

where the last inequality holds on  $\mathcal{A}_1$ . Moreover, on  $\mathcal{A}_1$ ,

$$\sum_{i < j, i, j \in V_t^\leq} p_{ij} \leq \frac{(\sum_{i \in V_t^\leq} w_i)^2}{2L_n^2} = \frac{T_1^2}{2} = \frac{(\mathbb{E}[T_1])^2}{2} + \frac{\sqrt[4]{\varepsilon}}{2} \left( \frac{\sqrt[4]{\varepsilon}}{2} + \mathbb{E}[T_1] \right) = O(\sqrt[4]{\varepsilon}),$$

$$\sum_{i < j, i, j \in V_t^\leq} p_{ij} \geq \frac{(\sum_{i \in V_t^\leq} w_i)^2}{2L_n^2} - \frac{1}{L_n^2} \sum_{i \in V_t^\leq} w_i^2 = \frac{(\mathbb{E}[T_1])^2}{2} + O(\sqrt[4]{\varepsilon})$$

and, therefore,

$$\sum_{i < j, i, j \in V_t^\leq} p_{ij} = \frac{(\mathbb{E}[T_1])^2}{2} + O(\sqrt[4]{\varepsilon}).$$

Thus, on  $\mathcal{A}$ ,

$$d_{\text{TV}} \left( \mathcal{L} \left( \sum_{i < j, i, j \in V_t^\leq} I_{ij} | V_t^>, V_t^\leq \right), \text{Poi}((\mathbb{E}[T_1])^2/2) \right) = \text{Err}(\varepsilon, n),$$

and (4.6) follows immediately.

Finally, we prove (4.8). Note that  $|V_t^>| \sim \text{Bin}(|V_{> \varepsilon}|, t/\sqrt{L_n})$ . The proof follows from standard inequalities for distance between Binomial and Poisson random variables [24], Theorem 2.10, which implies that the left-hand side of (4.8) is bounded by  $|V_{> \varepsilon}|t^2/L_n \leq Mt^2/L_n = \text{Err}(M, n)$ . Since  $|V_{> \varepsilon}|t/\sqrt{L_n} = t\rho_n([\varepsilon, \infty)) = t\rho([\varepsilon, \infty)) + \text{Err}(n)$ , provided  $\rho$  does not have an atom at  $\varepsilon$ ; since the limit  $\varepsilon \rightarrow 0$  can be taken through the continuity points, the proof follows.  $\square$

4.2. Concentration.

PROOF OF PROPOSITION 4.3. We only give a proof for  $A = [0, t] \times [0, s]$  leaving the general case to the reader. Let  $R = \binom{n}{2}$  and let  $(p_{i_r j_r})_{r \in [R]}$  denote a nonincreasing ordering of the  $p_{ij}$ 's. Let  $I_r$  denote the indicator that an edge has been created between  $i_r$  and  $j_r$ ; thus  $I_r \sim \text{Ber}(p_{i_r j_r})$ , independently over  $r \in [R]$ . To simplify notation, let  $X = \mathbb{P}(\xi_n(A) = 0 | G_n)$ . Further, for  $r = 1, \dots, R$ , let  $\mathcal{F}_r = \sigma(I_i : i \in [r])$  (where we used the notation  $[0] = \emptyset$ ) and define  $X_r = \mathbb{E}[X | \mathcal{F}_r]$ . Thus,  $(X_r)_{r=0}^R$  is a martingale with respect to the filtration  $(\mathcal{F}_r)$  satisfying  $X_0 = \mathbb{E}[X]$  and  $X_R = X$ . We will apply a concentration inequality from [21], Theorems 18, 22. Thus, if we can show that

$$(4.15) \quad \text{Var}(X_r | \mathcal{F}_{r-1}) \leq \sigma_r^2, \quad |X_r - X_{r-1}| \leq M,$$

then

$$(4.16) \quad \mathbb{P}(|X - \mathbb{E}[X]| > \varepsilon) \leq 2 \exp\left(-\frac{\varepsilon^2}{2(\sum_{r=1}^R \sigma_r^2 + M\varepsilon/3)}\right).$$

Thus, we need to obtain the correct  $M$  and  $\sigma_r^2$  such that (4.15) holds. Note that

$$(4.17) \quad \begin{aligned} \mathbb{E}[X | \mathcal{F}_{r-1}] &= p_{i_r j_r} (\mathbb{E}[X | \mathcal{F}_{r-1}, I_r = 1] - \mathbb{E}[X | \mathcal{F}_{r-1}, I_r = 0]) \\ &\quad + \mathbb{E}[X | (\mathcal{F}_{r-1}, I_r = 0)], \\ \mathbb{E}[X | \mathcal{F}_r] &= I_{i_r j_r} (\mathbb{E}[X | \mathcal{F}_{r-1}, I_r = 1] - \mathbb{E}[X | \mathcal{F}_{r-1}, I_r = 0]) \\ &\quad + \mathbb{E}[X | \mathcal{F}_{r-1}, I_r = 0]. \end{aligned}$$

Moreover,  $X$  can change by changing the status of the edge  $(i_r, j_r)$  only if both  $i_r, j_r \in V_s \cup V_t$ , and thus

$$(4.18) \quad |\mathbb{E}[X | \mathcal{F}_{r-1}, I_r = 1] - \mathbb{E}[X | \mathcal{F}_{r-1}, I_r = 0]| \leq \frac{(s+t)^2}{L_n}.$$

Combining (4.17) and (4.18),

$$(4.19) \quad |X_r - X_{r-1}| \leq |I_r - p_{i_r j_r}| \frac{(s+t)^2}{L_n}.$$

Therefore,

$$\text{Var}(X_r | \mathcal{F}_{r-1}) = \mathbb{E}[(X_r - X_{r-1})^2 | \mathcal{F}_{r-1}] \leq \frac{(s+t)^4}{L_n^2} \mathbb{E}[(I_r - p_{i_r j_r})^2] \leq \frac{(s+t)^4}{L_n^2} p_{i_r j_r},$$

where the second step follows from  $\mathbb{E}[X_r | \mathcal{F}_{r-1}] = X_{r-1}$  and the third step follows from (4.19). Thus, we can apply (4.16) with  $\sigma_r^2 = (s+t)^4 p_{i_r j_r} / L_n^2$ , and  $M = (s+t)^2 / L_n$ . Now, the proof of Proposition 4.3 follows by using the fact that  $\sum_{i \neq j} p_{ij} \leq \sum_{i \neq j} w_i w_j / L_n \leq L_n$ .  $\square$

**5. Proofs of results on bipartite configuration model.** The proof of Theorem 1.10 is very similar to that of Theorem 1.2 for the configuration model and again relies on three key propositions, whose proofs are also similar to those of the corresponding key propositions from the proof of Theorem 1.2. We will outline this proof strategy by stating the key propositions, but we will leave both the reduction of Theorem 1.10 to these propositions, and the proofs of the propositions themselves to the reader.

PROPOSITION 5.1. *Let  $\mathcal{E}_n(S, S')$  denote the number of edges created between the set of half-edges  $S$  and  $S'$  in the construction of  $\text{BipCM}_n(\mathbf{d})$ . Consider  $k$  disjoint subsets of half-edges  $(S_i)_{i \in [k]}$  such that  $|S_i| = O(\sqrt{\ell_n})$  for all  $i \in [k]$ . Let  $S_{ij} = S_i \cap V_j$ ,  $i \in [k]$ ,  $j = 1, 2$ . Let  $\mathcal{E}_n = (\mathcal{E}_n(S_i, S_j))_{1 \leq i \leq j \leq k}$ ,  $\mathcal{E} := (\mathcal{E}_{ij})_{1 \leq i \leq j \leq k}$ , where  $\mathcal{E}_{ij} \sim \text{Poisson}((|S_{i1}||S_{j2}| + |S_{i2}||S_{j1}|)/\ell_n)$  for  $i \neq j$ ,  $\mathcal{E}_{ii} \sim \text{Poisson}(|S_{i1}||S_{i2}|/\ell_n)$ , and  $\mathcal{E}$  is an independent collection. Then, as  $n \rightarrow \infty$ ,*

$$d_{\text{TV}}(\mathcal{E}_n, \mathcal{E}) \rightarrow 0.$$

Moreover, if  $S_j$ 's are random disjoint subsets chosen independently of  $\text{BipCM}_n(\mathbf{d})$  and satisfying  $\mathbb{E}[|S_j|] = O(\sqrt{\ell_n})$ , then  $\lim_{n \rightarrow \infty} d_{\text{TV}}(\mathcal{E}_n, \mathcal{E}) = 0$ , where both  $\mathcal{E}_n$  and  $\mathcal{E}$  refer to the joint distribution, including in particular the randomness stemming from the random sets  $S_j$ 's.

PROPOSITION 5.2. *Let  $V_n(A)$  denote the set of vertices obtained by labeling the vertices uniformly from  $[0, \sqrt{\ell_n}]$  and then retaining the vertices with labels in  $A$ . For a vertex set  $V$ , define  $\bar{S}_n(V) = \frac{1}{\sqrt{\ell_n}} \sum_{i \in V} d_i$ . For any disjoint collection of sets  $(A_i)_{i \in [k]}$  from  $\mathcal{B}(\mathbb{R}_+)$ , let  $V_{ij}$  denote the set of vertices in  $V_j$  with labels in  $A_i$ , and  $\alpha = (\alpha_{ij})_{i \in [k], j=1,2} \in \mathbb{R}^{2k}$ . Define  $\Phi((V_{ij})_{i \in [k], j=1,2}) := \mathbb{E}[e^{i \sum_{j=1}^2 \sum_{i \in [k]} \alpha_{ij} \bar{S}_n(A_{ij})}]$ . Then*

$$\Phi((V_{ij})_{i \in [k], j=1,2}) = \exp\left( (1 + o(1)) \sum_{j=1,2} \sum_{i \in [k]} \Lambda(A_i) \int (e^{i\alpha_{ij}x} - 1) \rho_{nj}(dx) \right).$$

PROPOSITION 5.3. *For any  $A, B \in \mathcal{B}(\mathbb{R}_+)$ ,  $l \in \mathbb{N}^*$  and  $\delta > 0$ ,*

$$\begin{aligned} & \mathbb{P}(|\mathbb{P}(\xi_n(A \times B) = l | \text{BipCM}_n(\mathbf{d})) - \mathbb{P}(\xi_n(A \times B) = l)| > \delta) \\ & \leq 2 \exp\left( -\frac{\delta^2 \ell_n}{12 \Lambda(A)^2 \Lambda(B)^2} \right). \end{aligned}$$

### APPENDIX A: SAMPLING CONVERGENCE FOR MULTIGRAPHS

In this section, we prove Proposition 1.1. The corresponding result for simple graphs was established in [12]. The extension to multigraphs is relatively straightforward, and thus we just sketch the proof.

PROOF OF PROPOSITION 1.1. Taking into account Remark 1, which implies that any exchangeable adjacency measure can be represented by a possibly random multigraphex, the proof of [12], Lemma 3.2, can be immediately adapted to the multigraph setting. The only crucial point to note is that [12] use [43], Lemma 4.11, which, in turn, depends on the discreteness of the space of finite graphs. In the case of multigraphs, that is again true because the sampled graph almost surely take values in the space of multigraphs with finite number of edges, on which the discrete topology can be similarly defined.  $\square$

### APPENDIX B: PROPERTIES OF COMPLETELY RANDOM MEASURES AND LÉVY PROCESSES

In the proof of Theorem 1.2, we require the notion of completely random measure, which we define here.

DEFINITION 10 (Completely random measure). A random measure  $\mu$  on  $\mathbb{R}_+$  is called a completely random measure if for all finite families of bounded disjoint sets  $(A_i)_{i \leq k}$  in  $\mathcal{B}(\mathbb{R}_+)$ ,  $(\mu(A_i))_{i \leq k}$  is an independent collection of random variables.

Any completely random measure  $\mu$  is a random element of  $\mathcal{M}(R_+)$  and admits a nice representation [22, 32], Theorem 10.1III. In the special case where  $\mu$  is stationary, that is, the distribution of  $\mu([t, t + s])$  depends only on  $s$  for any  $t, s \in \mathbb{R}_+$ , the representation takes the form (1.1) where the measure  $\rho$  satisfies the condition

$$\int_0^\infty (x \wedge 1)\rho(dx) < \infty;$$

see [22], Example 10.1 (a).

LEMMA B.1. *Let  $\mu$  be a completely random measure of the form (1.1), and let  $A \in \mathcal{B}(\mathbb{R}_+)$  with  $\mu(A) < \infty$ . Then the characteristic function of  $\mu(A)$  is given by*

$$\mathbb{E}[e^{i\theta\mu(A)}] = \exp\left(i\theta a\lambda(A) + \lambda(A) \int (e^{i\theta x} - 1)\rho(dx)\right).$$

PROOF. This is a straightforward calculation, very similar to the one in Exercise 10.1.2 in [22].  $\square$

We will also need the notion of a Lévy process. It is defined as a real valued càdlàg process  $X = (X(t))_{t \geq 0}$  such that  $X(0) = 0$ , the increments  $X(t_1) - X(0), X(t_2) - X(t_1), \dots, X(t_n) - X(t_{n-1})$  are independent whenever  $0 < t_1 < \dots < t_n$ , and such that  $X(t + s) - X(t)$  is equal in distribution to  $X(s)$  for all  $s, t > 0$ . It is well known that given any bounded function  $h : \mathbb{R} \rightarrow \mathbb{R}$  such that  $h(x) = x$  in a neighborhood of 0, the characteristic function,  $\chi_t(\theta) = \mathbb{E}(e^{i\theta X(t)})$ , can be written as  $e^{t\psi(\theta)}$  with

$$\psi(\theta) = ia\theta - \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R} \setminus \{0\}} d\rho(x)(e^{i\theta x} - 1 - i\theta h(x)),$$

where  $a \in \mathbb{R}, \sigma \geq 0$ , and  $\rho$  is a  $\sigma$ -finite measure on  $\mathbb{R}$  such that  $\int (x^2 \wedge 1) d\rho(x) < \infty$ . Following [27], we call the triple  $(a, \sigma, \rho)$  the characteristics associated with  $h$ , or simply the characteristics of  $X$  when  $h$  is clear from the context. While  $h$  is usually chosen as  $h(x) = x1_{|x| \leq 1}$ , here we follow the approach of [27] insisting that  $h$  is continuous (since this is more convenient when considering limits); specifically, we will choose  $h(x) = (|x| \wedge 1)\text{sign}(x)$ . We will need the following lemma, which is a special case of Corollary 3.6 in Chapter VII in [27].

LEMMA B.2 ([27]). *Let  $X_n = (X_n(t))_{t \geq 0}$  be a sequence of Lévy processes with characteristics  $(b_n, \sigma_n, \rho_n)$ . Then  $X_n$  converges to a Lévy process  $X$  with characteristics  $(b, \sigma, \rho)$  in law if and only if  $b_n \rightarrow b, \sigma_n \rightarrow \sigma$  and  $\int f d\rho_n \rightarrow \int f d\rho$  for all bounded continuous functions  $f$  vanishing in a neighborhood of zero.*

We will apply the lemma in the special case where  $\rho_n$  has support on  $\mathbb{R}_+, \int x\rho_n(dx)$  is bounded uniformly in  $n$  and  $b_n$  is given in terms of  $\rho_n$  as  $b_n = \int (|x| \wedge 1)\rho_n(dx)$ . To facilitate the application in this case, we prove the following, auxiliary lemma.

LEMMA B.3. *Let  $\rho_n$  be a sequence of measures on  $\mathbb{R}_+$  such that  $\limsup_{n \rightarrow \infty} \int x\rho_n(dx) < \infty$ , let  $b_n = \int (x \wedge 1)\rho_n(dx)$ , and let*

$$a_- = \lim_{\varepsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \int_0^\varepsilon x\rho_n(x) dx, \quad a_+ = \lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \int_0^\varepsilon x\rho_n(x) dx$$

*Then  $\int f d\rho_n \rightarrow \int f d\rho$  for all bounded continuous functions  $f$  vanishing in a neighborhood of zero if and only if  $\rho_n$  converges vaguely to  $\rho$ . Furthermore, if  $\rho_n \rightarrow \rho$  vaguely then  $b_n$  converges to some  $b$  if and only if  $a_- = a_+$ , in which case  $b = \int_0^\infty (x \wedge 1)\rho(dx) + a_+$ .*

PROOF. Restricting ourself to large enough  $n$ , we may w.l.o.g assume that  $\int x\rho_n(dx) \leq 2c$ . Combined with the fact that  $\int x\rho(dx) \leq \limsup_{n \rightarrow \infty} \int x\rho_n(dx) \leq c$ , we conclude that  $\rho_n([M, \infty)) \leq c/M$  and  $\rho([M, \infty)) \leq 2c/M$ . Thus convergence for all bounded, continuous functions is equivalent to vague convergence.

Let us now assume that  $\rho_n$  is vaguely convergent to  $\rho$ , and let  $0 < \varepsilon \leq 1$  be such that  $\rho$  has not atom at  $\varepsilon$ . Then  $\int_\varepsilon^\infty (x \wedge 1)\rho_n(dx)$  converges to  $\int_\varepsilon^\infty (x \wedge 1)\rho(dx)$ , showing that

$$b_n = \int_\varepsilon^\infty (x \wedge 1)\rho(dx) + \int_0^\varepsilon (x \wedge 1)\rho_n(dx) + o(1).$$

This implies that

$$\liminf_{n \rightarrow \infty} b_n = \int_0^\infty (x \wedge 1)\rho(dx) + a_- \quad \text{and} \quad \limsup_{n \rightarrow \infty} b_n = \int_0^\infty (x \wedge 1)\rho(dx) + a_+,$$

which completes the proof.  $\square$

### APPENDIX C: PROPERTIES OF LIMITING ADJACENCY MEASURES

In this Appendix, we calculate the finite dimensional distributions of random adjacency measures corresponding to the graphexes in Theorem 1.2, Theorem 1.7 and Theorem 1.10. These are used extensively in the respective proofs.

*Configuration model.* Let  $\xi_{\text{CM}}$  denote the random adjacency measure associated to the multigraphex  $\mathcal{W}_{\text{CM}} = (W_{\text{CM}}, S_{\text{CM}}, I_{\text{CM}})$  and  $\xi_{\text{CM}}^* := \xi_{\text{CM}}|_{(x,y):y \leq x}$ .

Then we have the following.

LEMMA C.1. *For any  $A, B \in \mathcal{B}(\mathbb{R}_+)$  with  $A \cap B = \emptyset$ , the conditional distribution of  $\xi_{\text{CM}}^*(A \times A)$ , conditional on  $\{(\theta_i, v_i)\}_{i \geq 1}$ , is  $\text{Poisson}(\mu(A)^2/2)$  and that of  $\xi_{\text{CM}}(A \times B)$  is  $\text{Poisson}(\mu(A)\mu(B))$ . Moreover, for a disjoint collection  $(B_i)_{i=1}^k$ , conditionally on  $\{(\theta_i, v_i)\}_{i \geq 1}$ ,  $(\xi_{\text{CM}}^*(B_i \times B_i))_{i \in [k]}$ ,  $(\xi_{\text{CM}}(B_i \times B_j))_{1 \leq i < j \leq k}$  is an independent collection.*

PROOF. Let  $\{(\theta_i, v_i)\}$  be a unit rate Poisson process on  $\mathbb{R}_+^2$  and set  $w_i := \bar{\rho}^{-1}(v_i)$ . Now, conditionally on  $\{(\theta_i, v_i)\}_{i \geq 1}$ ,

$$\begin{aligned} \xi_{\text{CM}}^*(A \times A) &= \sum_{i > j} \text{Poi}(w_i w_j) \mathbb{1}\{\theta_i \in A, \theta_j \in A\} + \sum_i \text{Poi}(w_i^2/2) \mathbb{1}\{\theta_i \in A\} \\ &\quad + \sum_{j,k} \mathbb{1}\{\chi_{jk} \leq a w_j\} \mathbb{1}\{\theta_j \in A, \sigma_{jk} \in A\} \\ &\quad + \sum_k \mathbb{1}\{\eta_k'' \leq a^2/2\} \mathbb{1}\{\eta_k \in A, \eta_k' \in A\} \\ &= \sum_{i > j} \text{Poi}(w_i w_j) \mathbb{1}\{\theta_i \in A, \theta_j \in A\} + \sum_i \text{Poi}(w_i^2/2) \mathbb{1}\{\theta_i \in A\} \\ &\quad + \sum_j \text{Poi}(a \Lambda(A) w_j) \mathbb{1}\{\theta_j \in A\} + \text{Poi}(a^2 \Lambda(A)^2/2), \end{aligned}$$

where, by construction, all the  $\text{Poi}(\cdot)$  random variables above are mutually independent. Therefore,

$$\begin{aligned} \xi_{\text{CM}}^*(A \times A) &= \text{Poi}\left(\frac{a^2 \Lambda(A)^2}{2} + a \Lambda(A) \sum_{i \geq 1} w_i \mathbb{1}\{\theta_i \in A\} + \frac{1}{2} \left(\sum_{i \geq 1} w_i \mathbb{1}\{\theta_i \in A\}\right)^2\right) \\ &= \text{Poi}(\mu(A)^2/2). \end{aligned}$$

Similarly, conditionally on  $(w_i, \theta_i)_{i \geq 1}$ ,

$$\begin{aligned} &\xi_{\text{CM}}(A \times B) \\ &= \sum_{i \neq j} \text{Poi}(w_i w_j) \mathbb{1}\{\theta_i \in A, \theta_j \in B\} \\ &\quad + \sum_{j,k} \mathbb{1}\{\chi_{jk} \leq a w_j\} \mathbb{1}\{\theta_j \in A, \sigma_{jk} \in B\} + \sum_{j,k} \mathbb{1}\{\chi_{jk} \leq a w_j\} \mathbb{1}\{\theta_j \in B, \sigma_{jk} \in A\} \\ &\quad + \sum_k \mathbb{1}\{\eta''_k \leq a^2/2\} \mathbb{1}\{\eta_k \in A, \eta'_k \in B\} + \sum_k \mathbb{1}\{\eta''_k \leq a^2/2\} \mathbb{1}\{\eta_k \in B, \eta'_k \in A\} \\ &= \sum_{i \neq j} \text{Poi}(w_i w_j) \mathbb{1}\{\theta_i \in A, \theta_j \in B\} + \sum_j \text{Poi}(a \Lambda(A) w_j) \mathbb{1}\{\theta_j \in B\} \\ &\quad + \sum_j \text{Poi}(a \Lambda(B) w_j) \mathbb{1}\{\theta_j \in A\} + \text{Poi}(a^2 \Lambda(A) \Lambda(B)), \end{aligned}$$

and thus

$$\begin{aligned} \xi_{\text{CM}}(A \times B) &= \text{Poi}\left(\left(a \Lambda(A) + \sum_{i \geq 1} w_i \mathbb{1}\{\theta_i \in B\}\right) \times \left(a \Lambda(B) + \sum_{i \geq 1} w_i \mathbb{1}\{\theta_i \in A\}\right)\right) \\ &= \text{Poi}(\mu(A) \mu(B)). \end{aligned}$$

The stated conditional independence follows by construction.  $\square$

*Generalized random graphs.* Let  $\xi_{\text{GRG}}$  denote the random adjacency measure associated to the graphex  $\mathcal{W}_{\text{GRG}}^1$  in Theorem 1.7, and  $\xi_{\text{GRG}}^* := \xi_{\text{GRG}}|_{(x,y):y \leq x}$ . We fix any  $k \geq 1$  and let  $(B_i)_{i \in [k]}$  be a collection of disjoint intervals such that  $B_{i+1}$  lies to the left of  $B_i$  on  $\mathbb{R}_+$ . Let  $N$  denote the Poisson point process on  $\mathbb{R}_+^2$  with intensity  $\rho(dw) \otimes d\theta$ . Further let  $N_{i\varepsilon} = N([\varepsilon, \infty) \times B_i)$ .

LEMMA C.2.

$$\begin{aligned} &\mathbb{P}(\xi_{\text{GRG}}^*(B_i \times B_j) = 0, 1 \leq i \leq j \leq k) = \lim_{\varepsilon \rightarrow 0} \mathbb{E}[G(N_{1\varepsilon}, \dots, N_{k\varepsilon})], \\ \text{(C.1)} \quad &G(l_1, \dots, l_k) = \int \prod_{1 \leq i \leq j \leq k} f_{ij}(\mathbf{w}) \prod_{i \in [k], \ell \in [l_i]} \mathbb{1}\{w_\ell^{(i)} \in [\varepsilon, \infty)\} \frac{\rho(dw_\ell^{(i)})}{\rho([\varepsilon, \infty))}, \end{aligned}$$

where  $\mathbf{w}$  is the collection of random variables  $(w_\ell^{(i)})_{i \in [k], \ell \in [l_i]}$  and

$$\begin{aligned} f_{ii}(\mathbf{w}) &= e^{-a^2 \Lambda(B_i)^2/2} \prod_{1 \leq u \leq v \leq l_i} \frac{1}{1 + w_u^{(i)} w_v^{(i)}} e^{-a \Lambda(B_i) \sum_{u \in [l_i]} w_u^{(i)}}, \\ f_{ij}(\mathbf{w}) &= e^{-a^2 \Lambda(B_i) \Lambda(B_j)} \prod_{u \in [l_i], v \in [l_j]} \frac{1}{1 + w_u^{(i)} w_v^{(j)}} e^{-a \Lambda(B_i) \sum_{u \in [l_j]} w_u^{(j)} - a \Lambda(B_j) \sum_{u \in [l_i]} w_u^{(i)}}. \end{aligned}$$

PROOF. Fix  $\varepsilon > 0$  and note that, conditional on  $N_{i\varepsilon} = k_i$  for all  $i \in [k]$ , the collection  $(w_j : \theta_j \in \bigcup_{i \in [k]} B_i)$  can be considered as i.i.d. samples from the normalized measure  $\rho|_{[\varepsilon, \infty)}$ . Fix  $1 \leq i < j \leq k$ . Given  $(\theta_u^{(i)}, w_u^{(i)})$  and  $(\theta_v^{(j)}, w_v^{(j)})$  with  $\theta_u^{(i)} \in B_i$ ,  $\theta_v^{(j)} \in B_j$ ,  $w_u^{(i)}, w_v^{(j)} \geq \varepsilon$ , the adjacency measure  $\xi_{\text{GRG}}$  has a point at  $(\theta_u^{(i)}, \theta_v^{(j)})$  with probability  $\frac{w_u^{(i)} w_v^{(j)}}{1 + w_u^{(i)} w_v^{(j)}}$ . Moreover, these points are independent given  $\{(\theta_u^{(i)}, w_u^{(i)}) : 1 \leq u \leq l_i\}$  and



$\{(\theta_v^{(j)}, w_v^{(j)}) : 1 \leq v \leq l_j\}$  variables. Further, (1.2) implies that the set  $B_i \times B_j$  has an independent  $\text{Poi}(a^2 \Lambda(B_i) \Lambda(B_j) + a \Lambda(B_i) \sum_{v \in [l_j]} w_v^{(j)} + a \Lambda(B_j) \sum_{u \in [l_i]} w_u^{(i)})$  points from points  $\{(\theta, w)\} \in N$  with  $w \geq \varepsilon$ . Using independence of the contributions, the probability of having zero points in the adjacency measure is precisely  $f_{ij}(\mathbf{w})$ . Further, we note that given  $N$ , the edges are all independent. This directly motivates the right-hand side of (C.1). Finally, we let  $\varepsilon \downarrow 0$  to get the desired equality. The argument for  $f_{ii}(\mathbf{w})$  is similar, and is therefore omitted.  $\square$

*Bipartite configuration model.* The limiting adjacency measure in case of Bipartite configuration model is given by  $\xi_{\mathcal{W}_{\text{BCM}}}$ , where  $\mathcal{W}_{\text{BCM}}$  is defined in (1.9). Let  $\{(\theta_i, v_i, c_i)\}_{i \geq 1}$  be a unit rate Poisson process on  $\mathbb{R}_+^2 \times \{0, 1\}$  and set  $w_i := \bar{\rho}_j^{-1}(v_i)$  if  $c_i = j$ . For  $r = 1, 2$ , define the completely random measure  $\mu_r := a_r \Lambda + \sum_{i \geq 1} \bar{\rho}_r^{-1}(v_i) \delta_{\theta_i} \mathbb{1}\{c_i = r\}$ . We will show the following.

LEMMA C.3. *For any Borel subsets  $A, B$  of  $\mathbb{R}$  with  $A \cap B = \emptyset$ , the conditional distribution of  $\xi_{\mathcal{W}_{\text{BCM}}}^*(A \times A)$ , conditional on  $\{(\theta_i, v_i, c_i)\}_{i \geq 1}$ , is  $\text{Poisson}(\mu_1(A) \mu_2(A))$  and that of  $\xi_{\mathcal{W}_{\text{BCM}}}(A \times B)$  is  $\text{Poisson}(\mu_1(A) \mu_2(B) + \mu_2(A) \mu_1(B))$ . Moreover, for a disjoint collection  $(B_i)_{i=1}^k$ , conditionally on  $\{(\theta_i, v_i)\}_{i \geq 1}$ ,  $(\xi_{\mathcal{W}_{\text{BCM}}}^*(B_i \times B_i))_{i \in [k]}$ ,  $(\xi_{\mathcal{W}_{\text{BCM}}}(B_i \times B_j))_{1 \leq i < j \leq k}$  is an independent collection.*

PROOF. Note that conditionally on  $\{(\theta_i, v_i, c_i)\}_{i \geq 1}$ ,

$$\begin{aligned} \xi_{\mathcal{W}_{\text{BCM}}}^*(A \times A) &= \sum_{i > j} \text{Poi}(w_i w_j) \mathbb{1}\{c_i \neq c_j\} \mathbb{1}\{\theta_i \in A, \theta_j \in A\} \\ &\quad + \sum_{j,k} \sum_{r=0,1} \mathbb{1}\{\chi_{jk} \leq a_r w_j\} \mathbb{1}\{c_j \neq r\} \mathbb{1}\{\theta_j \in A, \sigma_{jk} \in A\} \\ &\quad + \sum_k \mathbb{1}\{\eta_k'' \leq a_1 a_2\} \mathbb{1}\{\eta_k \in A, \eta_k' \in A\} \\ &= \sum_{i > j} \text{Poi}(w_i w_j) \mathbb{1}\{c_i \neq c_j\} \mathbb{1}\{\theta_i \in A, \theta_j \in A\} \\ &\quad + \sum_j \sum_{r=0,1} \text{Poi}(a_r \Lambda(A) w_j) \mathbb{1}\{c_j \neq r\} \mathbb{1}\{\theta_j \in A\} + \text{Poi}(a_1 a_2 \Lambda(A)^2), \end{aligned}$$

where, by construction, all the  $\text{Poi}(\cdot)$  random variables above are mutually independent. Therefore,

$$\begin{aligned} \xi_{\mathcal{W}_{\text{BCM}}}^*(A \times A) &= \text{Poi}\left(a_1 a_2 \Lambda(A)^2 + \sum_{r=0,1} a_r \Lambda(A) \sum_{i \geq 1} \mathbb{1}\{c_i \neq r\} w_i \mathbb{1}\{\theta_i \in A\}\right. \\ &\quad \left. + \sum_{i > j} w_i w_j \mathbb{1}\{c_i \neq c_j\} \mathbb{1}\{\theta_i \in A, \theta_j \in A\}\right) \\ &= \text{Poi}(\mu_1(A) \mu_2(A)). \end{aligned}$$

Similar argument can be carried out for  $A, B \in \mathcal{B}(\mathbb{R}_+)$  with  $A \cap B = \emptyset$  to conclude that

$$\xi_{\mathcal{W}_{\text{BCM}}}(A \times B) = \text{Poi}(\mu_1(A) \mu_2(B) + \mu_2(A) \mu_1(B)).$$

The stated conditional independence follows by construction.  $\square$

APPENDIX D: RESCALING OF A GRAPHON PROCESS

LEMMA D.1 (Rescaling lemma). *Given a sequence of multigraphs  $(G_n)_{n \geq 1}$  and real numbers  $(\ell_n)_{n \geq 1}$ , suppose that  $\lim_{n \rightarrow \infty} \frac{2e(G_n)}{\ell_n} = c > 0$ . Further, let  $\text{Lbl}_{\sqrt{\ell_n}}(G_n) \xrightarrow{d} \xi_{\mathcal{W}}$  for some multigraphex  $\mathcal{W} = (W, S, I)$ . Then  $\text{Lbl}(G_n) \xrightarrow{d} \xi_{\mathcal{W}^c}$ ,  $\mathcal{W}^c = (W^c, S^c, I^c)$  with*

$$W^c(x, y, \cdot) = W(\sqrt{c}x, \sqrt{c}y, \cdot), \quad S^c(x, \cdot) = \frac{1}{\sqrt{c}}S(\sqrt{c}x, \cdot) \quad \text{and} \quad I^c(\cdot) = \frac{I(\cdot)}{c}.$$

PROOF. Define the point process  $\xi'$  by  $\xi'([0, s] \times [0, t]) = \xi_{\mathcal{W}}([0, c^{-1/2}s] \times [0, c^{-1/2}t])$ , for any  $0 < s, t < \infty$ . First, let us show that, as  $n \rightarrow \infty$ ,

$$(D.1) \quad \text{Lbl}(G_n)(B) \xrightarrow{d} \xi'(B),$$

where  $B$  is any finite union of rectangles. For simplicity, let us take  $B = [0, t]^2$ ; the general case follows similarly. Let  $U_1, \dots, U_n \sim \text{Uniform}([0, \sqrt{2e(G_n)}])$  be i.i.d. random variables. Further, define  $U'_i = (\frac{\ell_n}{2e(G_n)})^{1/2}U_i$ ,  $1 \leq i \leq n$ , so that,  $U'_1, \dots, U'_n$  are i.i.d. samples from  $\text{Uniform}([0, \sqrt{\ell_n}])$ . Thus, we have

$$\begin{aligned} \text{Lbl}(G_n)([0, t]^2) &= \sum_{\{i, j\} \in E_n} \mathbb{1}\{U_i \leq t, U_j \leq t\}, \\ \text{Lbl}_{\sqrt{\ell_n}}(G_n)([0, t]^2) &= \sum_{\{i, j\} \in E_n} \mathbb{1}\{U'_i \leq t, U'_j \leq t\} \\ &= \sum_{\{i, j\} \in E_n} \mathbb{1}\left\{U_i \leq \sqrt{\frac{2e(G_n)}{\ell_n}}t, U_j \leq \sqrt{\frac{2e(G_n)}{\ell_n}}t\right\} \end{aligned}$$

Let  $t' = c^{-1/2}t$ . It now follows that

$$\mathbb{E}[|\text{Lbl}(G_n)([0, t]^2) - \text{Lbl}_{\sqrt{\ell_n}}(G_n)([0, t']^2)|] \rightarrow 0.$$

The proof of (D.1) now follows using the the assumption that  $\text{Lbl}_{\sqrt{\ell_n}}(G_n) \xrightarrow{d} \xi_{\mathcal{W}}$ .

Next, we need to show that  $\xi' \stackrel{d}{=} \xi_{\mathcal{W}^c}$ . Recall Definition 5 with all related notation. Thus the multigraphon part in  $\xi'$  is given by

$$\sum_{i \neq j} \zeta_{ij} \delta_{(\sqrt{c}\theta_i, \sqrt{c}\theta_j)} + \sum_i \zeta_{ii} \delta_{(\sqrt{c}\theta_i, \sqrt{c}\theta_i)}.$$

On the other hand, if  $\{(\theta_i, v_i)\}_{i \geq 1}$  is a unit rate Poisson point process in  $\mathbb{R}_+^2$ , then in distribution,  $(\sqrt{c}\theta_i, v_i)_{i \geq 1}$  is equal to  $(\theta_i, \sqrt{c}v_i)_{i \geq 1}$ . This gives the required rescaling of the multigraphon part. The rescaling of the star and isolated parts can be dealt similarly using rescaling properties of Poisson point processes, and thus is omitted here.  $\square$

**Acknowledgments.** We thank Samantha Petti for several suggestions on improving an earlier version of the draft, and in particular suggesting a simplification in the coupling in Section 2.1.

**Funding.** S. Dhara was supported by an internship at Microsoft Research Lab—New England, and by the Netherlands Organisation for Scientific Research (NWO) through Gravitation Networks, Grant 024.002.003.

## SUPPLEMENTARY MATERIAL

**A correction to Kallenberg’s theorem for jointly exchangeable random measures** (DOI: [10.1214/21-AOP1508SUPP](https://doi.org/10.1214/21-AOP1508SUPP); .pdf). In this supplement, we establish a characterization of locally finite random measures, correcting an incomplete characterization from Kallenberg [31]. As a consequence, we establish that every random adjacency measure is the adjacency measure corresponding to some (possibly random) multigraphex.

## REFERENCES

- [1] ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.* **11** 581–598. [MR0637937 https://doi.org/10.1016/0047-259X\(81\)90099-3](https://doi.org/10.1016/0047-259X(81)90099-3)
- [2] AUSTIN, T. (2008). On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.* **5** 80–145. [MR2426176 https://doi.org/10.1214/08-PS124](https://doi.org/10.1214/08-PS124)
- [3] BARABÁSI, A. L. (2016). *Network Science*, 1st ed. Cambridge Univ. Press, Cambridge.
- [4] BARBOUR, A. D. (1988). Stein’s method and Poisson process convergence. *J. Appl. Probab.* **25** 175–184.
- [5] BOLLOBÁS, B. (1980). A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.* **1** 311–316. [MR0595929 https://doi.org/10.1016/S0195-6698\(80\)80030-8](https://doi.org/10.1016/S0195-6698(80)80030-8)
- [6] BOLLOBÁS, B., JANSON, S. and RIORDAN, O. (2007). The phase transition in inhomogeneous random graphs. *Random Structures Algorithms* **31** 3–122. [MR2337396 https://doi.org/10.1002/rsa.20168](https://doi.org/10.1002/rsa.20168)
- [7] BOLLOBÁS, B. and RIORDAN, O. (2009). Metrics for sparse graphs. In *Surveys in Combinatorics 2009. London Mathematical Society Lecture Note Series* **365** 211–287. Cambridge Univ. Press, Cambridge. [MR2588543](https://doi.org/10.1017/S0004972609005434)
- [8] BORGS, C., CHAYES, J., LOVÁSZ, L., SÓS, V. and VESZTERGOMBI, K. (2011). Limits of randomly grown graph sequences. *European J. Combin.* **32** 985–999. [MR2825531 https://doi.org/10.1016/j.ejc.2011.03.015](https://doi.org/10.1016/j.ejc.2011.03.015)
- [9] BORGS, C., CHAYES, J., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2006). Counting graph homomorphisms. In *Topics in Discrete Mathematics: Dedicated to Jarik Nešetřil on the Occasion of His 60th Birthday* 315–371. Springer, Berlin.
- [10] BORGS, C., CHAYES, J. T., COHN, H. and HOLDEN, N. (2017). Sparse exchangeable graphs and their limits via graphon processes. *J. Mach. Learn. Res.* **18** Paper No. 210, 71. [MR3827098](https://doi.org/10.1162/jmlr-2017-18-021)
- [11] BORGS, C., CHAYES, J. T., COHN, H. and LOVÁSZ, L. M. (2019). Identifiability for graphexes and the weak kernel metric. In *Building Bridges II—Mathematics of László Lovász* (A. Sali, I. Bárány and G. Katona eds.). *Bolyai Society Mathematical Studies* **28**. Springer, Berlin.
- [12] BORGS, C., CHAYES, J. T., COHN, H. and VEITCH, V. (2019). Sampling perspectives on sparse exchangeable graphs. *Ann. Probab.* **47** 2754–2800. [MR4021236 https://doi.org/10.1214/18-AOP1320](https://doi.org/10.1214/18-AOP1320)
- [13] BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2018). An  $L^p$  theory of sparse graph convergence II: LD convergence, quotients and right convergence. *Ann. Probab.* **46** 337–396. [MR3758733 https://doi.org/10.1214/17-AOP1187](https://doi.org/10.1214/17-AOP1187)
- [14] BORGS, C., CHAYES, J. T., COHN, H. and ZHAO, Y. (2019). An  $L^p$  theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Trans. Amer. Math. Soc.* **372** 3019–3062. [MR3988601 https://doi.org/10.1090/tran/7543](https://doi.org/10.1090/tran/7543)
- [15] BORGS, C., CHAYES, J. T., DHARA, S. and SEN, S. (2021). Supplement to “Limits of sparse configuration models and Beyond: Graphexes and MultiGraphexes.” <https://doi.org/10.1214/21-AOP1508SUPP>
- [16] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math.* **219** 1801–1851. [MR2455626 https://doi.org/10.1016/j.aim.2008.07.008](https://doi.org/10.1016/j.aim.2008.07.008)
- [17] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2012). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. of Math. (2)* **176** 151–219. [MR2925382 https://doi.org/10.4007/annals.2012.176.1.2](https://doi.org/10.4007/annals.2012.176.1.2)
- [18] BRITTON, T., DEIFEN, M. and MARTIN-LÖF, A. (2006). Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.* **124** 1377–1397. [MR2266448 https://doi.org/10.1007/s10955-006-9168-x](https://doi.org/10.1007/s10955-006-9168-x)
- [19] CARON, F. and FOX, E. B. (2017). Sparse graphs using exchangeable random measures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 1295–1366. [MR3731666 https://doi.org/10.1111/rssb.12233](https://doi.org/10.1111/rssb.12233)
- [20] CHATTERJEE, S., DIACONIS, P. and SLY, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. [MR2857452 https://doi.org/10.1214/10-AAP728](https://doi.org/10.1214/10-AAP728)

- [21] CHUNG, F. and LU, L. (2006). Concentration inequalities and martingale inequalities: A survey. *Internet Math.* **3** 79–127. MR2283885
- [22] DALEY, D. J. and VERE-JONES, D. (2008). *An Introduction to the Theory of Point Processes. Vol. II: General Theory and Structure*, 2nd ed. *Probability and Its Applications (New York)*. Springer, New York. MR2371524 <https://doi.org/10.1007/978-0-387-49835-5>
- [23] DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl.* (7) **28** 33–61. MR2463439
- [24] VAN DER HOFSTAD, R. (2017). *Random Graphs and Complex Networks, I*. Cambridge Univ. Press, Cambridge.
- [25] VAN DER HOFSTAD, R., HOOGHIEMSTRA, G. and VAN MIEGHEM, P. (2005). Distances in random graphs with finite variance degrees. *Random Structures Algorithms* **27** 76–123. MR2150017 <https://doi.org/10.1002/rsa.20063>
- [26] HOOVER, D. N. (1979). Relations on probability spaces and arrays of random variables. Preprint. Institute for Advanced Study, Princeton.
- [27] JACOD, J. and SHIRYAEV, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **288**. Springer, Berlin. MR1943877 <https://doi.org/10.1007/978-3-662-05265-5>
- [28] JANSON, S. (2017). On convergence for graphexes. Available at [arXiv:1702.06389](https://arxiv.org/abs/1702.06389).
- [29] JANSON, S., ŁUCZAK, T. and RUCINSKI, A. (2000). *Random Graphs. Wiley-Interscience Series in Discrete Mathematics and Optimization*. Wiley Interscience, New York. MR1782847 <https://doi.org/10.1002/9781118032718>
- [30] KALLENBERG, O. (1990). Exchangeable random measures in the plane. *J. Theoret. Probab.* **3** 81–136. MR1031426 <https://doi.org/10.1007/BF01063330>
- [31] KALLENBERG, O. (2005). *Probabilistic Symmetries and Invariance Principles. Probability and Its Applications (New York)*. Springer, New York. MR2161313
- [32] KINGMAN, J. F. C. (1967). Completely random measures. *Pacific J. Math.* **21** 59–78. MR0210185
- [33] LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes. Springer Series in Statistics*. Springer, New York. MR0691492
- [34] LIPTSER, R. S. and SHIRYAYEV, A. N. (1989). *Theory of Martingales. Mathematics and Its Applications (Soviet Series)* **49**. Kluwer Academic, Dordrecht. MR1022664 <https://doi.org/10.1007/978-94-009-2438-3>
- [35] LOVÁSZ, L. (2012). *Large Networks and Graph Limits. American Mathematical Society Colloquium Publications* **60**. Amer. Math. Soc., Providence, RI. MR3012035 <https://doi.org/10.1090/coll/060>
- [36] LOVÁSZ, L. and SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957. MR2274085 <https://doi.org/10.1016/j.jctb.2006.05.002>
- [37] LOVÁSZ, L. and SZEGEDY, B. (2007). Szemerédi’s lemma for the analyst. *Geom. Funct. Anal.* **17** 252–270. MR2306658 <https://doi.org/10.1007/s00039-007-0599-6>
- [38] MOLLOY, M. and REED, B. (1995). A critical-point for random graphs with a given degree sequence. *Random Structures Algorithms* **6** 161–179.
- [39] NEWMAN, M. E. J. (2010). *Networks: An Introduction*. Oxford Univ. Press, Oxford. MR2676073 <https://doi.org/10.1093/acprof:oso/9780199206650.001.0001>
- [40] PITTEL, B. (2010). On a random graph evolving by degrees. *Adv. Math.* **223** 619–671. MR2565544 <https://doi.org/10.1016/j.aim.2009.08.015>
- [41] RÁTH, B. and SZAKÁCS, L. (2012). Multigraph limit of the dense configuration model and the preferential attachment graph. *Acta Math. Hungar.* **136** 196–221. MR2945218 <https://doi.org/10.1007/s10474-012-0217-4>
- [42] VEITCH, V. and ROY, D. M. (2015). The class of random graphs arising from exchangeable random measures. Available at [arXiv:1512.03099](https://arxiv.org/abs/1512.03099).
- [43] VEITCH, V. and ROY, D. M. (2019). Sampling and estimation for (sparse) exchangeable graphs. *Ann. Statist.* **47** 3274–3299. MR4025742 <https://doi.org/10.1214/18-AOS1778>
- [44] WORMALD, N. C. (1999). Models of random regular graphs. In *Surveys in Combinatorics, 1999 (Canterbury)* (J. D. Lamb and D. A. Preece eds.). *London Mathematical Society Lecture Note Series* **267** 239–298. Cambridge Univ. Press, Cambridge. MR1725006