Limit laws for self-loops and multiple edges in the configuration model

Omer Angel[a], Remco van der Hofstad[b] and Cecilia Holmgren[c]

[a]*Department of Mathematics, University of British Columbia, Vancouver, BC, V6T 1Z2, Canada. E-mail: angel@math.ubc.ca*
[b]*Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
E-mail: r.w.v.d.hofstad@tue.nl*
[c]*Department of Mathematics, Uppsala University, SE-75310, Uppsala, Sweden. E-mail: cecilia.holmgren@math.uu.se*

Abstract. We consider self-loops and multiple edges in the configuration model as the size of the graph tends to infinity. The interest in these random variables is due to the fact that the configuration model, conditioned on being simple, is a uniform random graph with prescribed degrees. Simplicity corresponds to the absence of self-loops and multiple edges.

We show that the number of self-loops and multiple edges converges in distribution to two independent Poisson random variables when the second moment of the empirical degree distribution converges. We also provide estimations on the total variation distance between the numbers of self-loops and multiple edges and their limits, as well as between the sum of these values and the Poisson random variable to which this sum converges to. This revisits previous works of Bollobás, of Janson, of Wormald and others. The error estimates also imply sharp asymptotics for the number of simple graphs with prescribed degrees.

The error estimates follow from an application of the Stein–Chen method for Poisson convergence, which is a novel method for this problem. The asymptotic independence of self-loops and multiple edges follows from a Poisson version of the Cramér–Wold device using thinning, which is of independent interest.

When the degree distribution has infinite second moment, our general results break down. We can, however, prove a central limit theorem for the number of self-loops, and for the multiple edges between vertices of degrees much smaller than the square root of the size of the graph. Our results and proofs easily extend to directed and bipartite configuration models.

Résumé. Nous considérons les boucles et les arêtes multiples dans le modèle de configuration lorsque la taille du graphe tend vers l'infini. L'intérêt de ces variables aléatoires est dû au fait que le modèle de configuration, conditionné à la simplicité, est un graphe aléatoire uniforme avec des degrés prescrits. La simplicité correspond à l'absence des boucles et des arêtes multiples.

Nous montrons que le nombre des boucles et des arêtes multiples convergent en loi vers deux variables aléatoires indépendantes qui suivent des lois de Poisson lorsque le moment d'ordre 2 de la loi empirique des degrés converge. Nous fournissons aussi des estimations des distances de variation totale entre les nombres des boucles et des arêtes multiples et leurs limites, ainsi qu'entre la somme de ces nombres et la variable aléatoire, qui suit une loi de Poisson, vers laquelle converge cette somme. Cela revisite les œuvres précédentes de Bollobás comme de Janson, de Wormald, et d'autres. Les estimations d'erreur impliquent également une asymptotique précise pour le nombre de graphes simples avec des degrés prescrits.

Les estimations d'erreur découlent d'une application de la méthode de Stein–Chen pour la convergence vers une loi de Poisson, qui est une nouvelle méthode pour ce problème. L'indépendance asymptotique des boucles et des arêtes multiples suit à partir d'une version Poisson du dispositif Cramér–Wold utilisant l'amincissement, qui est intéressant en lui-même.

Lorsque la loi des degrés a un moment d'ordre 2 infini, nos résultats généraux échouent. Nous pouvons, cependant, prouver un théorème de la limite centrale pour le nombre des boucles, et pour les arêtes multiples entre sommets avec degrés beaucoup plus petits que la racine carrée de la taille du graphe. Nos résultats et preuves peuvent facilement s'étendre aux modèles de configuration orientés et bipartis.

1. Introduction and motivation

1.1. *Models and results*

We consider the configuration model $CM_n(d)$, with degrees $d = (d_i)_{i \in [n]}$. The configuration model (CM) is a random graph with vertex set $[n] := \{1, 2, \ldots, n\}$ and with prescribed degrees. Let $d = (d_1, d_2, \ldots, d_n)$ be a given *degree sequence*, i.e., a sequence of n positive integers. The total degree, denoted ℓ_n is

$$\ell_n = \sum_{i \in [n]} d_i, \tag{1.1}$$

and is assumed to be even. The CM on n vertices with degree sequence d is constructed as follows: start with n vertices and d_i half-edges adjacent to each vertex $i \in [n]$. The ℓ_n half-edges are matched in pairs in a uniformly random manner to form the edges of the graph.

Algorithmically, the CM may be sampled as follows. Randomly choose a pair of half-edges, match the chosen pair together to form an edge and remove the two half-edges. Continue until all half-edges are paired. We denote the resulting graph on $[n]$ by $CM_n(d)$, with corresponding edge set $\mathcal{E}_n$. Although self-loops may occur due to the pairing of half-edges that are incident to the same vertex, in many cases the number of self-loops is much smaller than the total degree as $n \to \infty$ (see e.g. [7,13]). The same applies to multiple edges. We say that $CM_n(d)$ is *simple* when it has no self-loops nor multiple edges.

In this paper, we investigate limit laws for the number of self-loops and multiple edges. Specifically, we study the random vector (S_n, M_n), which is defined as

$$S_n = \sum_{i \in [n]} X_{ii}, \qquad M_n = \sum_{1 \le i < j \le n} \binom{X_{ij}}{2}. \tag{1.2}$$

Here, for $i, j \in [n]$, X_{ij} denotes the number of edges between vertices i and j. For clarity, note that we have

$$d_i = X_{ii} + \sum_{j \in [n]} X_{ij} = 2X_{ii} + \sum_{j \neq i} X_{ij}. \tag{1.3}$$

We note that M_n is not precisely equal to the number of multiple edges. This number instead may be written as

$$\widetilde{M}_n = \sum_{1 \le i < j \le n} (X_{ij} - 1)_+, \tag{1.4}$$

where $x_+ = \max\{0, x\}$. However, $M_n = 0$ precisely when there are no self-loops, i.e., when $\widetilde{M}_n = 0$. Moreover, if $X_{ij} = 2$ then the pair i, j contributes 1 to M_n, so $M_n = \widetilde{M}_n$ in the absence of triple edges between vertices.

Furthermore, we let λ_n^S and λ_n^M be the means of the random variables S_n and M_n, i.e.,

$$\lambda_n^S = \mathbb{E}[S_n], \qquad \lambda_n^M = \mathbb{E}[M_n]. \tag{1.5}$$

We can compute that

$$\lambda_n^S = \frac{\sum_{i \in [n]} d_i(d_i - 1)}{2(\ell_n - 1)}, \qquad \lambda_n^M = \frac{\sum_{1 \le i < j \le n} d_j(d_j - 1)d_i(d_i - 1)}{2(\ell_n - 1)(\ell_n - 3)}. \tag{1.6}$$

The calculation of λ_n^S follows since the probability of a connection between any two half-edges is $1/(\ell_n - 1)$ and there are $\binom{d_i}{2}$ choices for the two half-edges that will form a self-loop incident to the vertex i. The calculation of λ_n^M follows since the probability for any two half-edges incident to the vertex i to connect (in order) to any two half-edges incident to the vertex j is $1/[(\ell_n - 1)(\ell_n - 3)]$. Further, there are $\binom{d_i}{2}$ choices for the two-half edges incident to i and $\binom{d_j}{2}$ choices for the two-half edges incident to j. Finally, there are two possible pairings of the two chosen half-edges incident to i to the two chosen half-edges incident to j.

Throughout the paper, we write $f(n) = o(g(n))$ as $n \to \infty$ when $g(n) > 0$ and $\lim_{n\to\infty} |f(n)|/g(n) = 0$. We write $f(n) = O(g(n))$ as $n \to \infty$ when $g(n) > 0$ and $\limsup_{n\to\infty} |f(n)|/g(n) < \infty$. Finally, we write $f(n) = \Theta(g(n))$ as $n \to \infty$ when $f(n) = O(g(n))$ and $g(n) = O(f(n))$.

In many cases and using this notation, we will approximate

$$\lambda_n^S = (\nu_n/2)\big(1 + O(1/n)\big), \qquad \lambda_n^M = \big(\nu_n^2/4\big)\big(1 + O(1/n)\big) - \chi_n, \tag{1.7}$$

where

$$\nu_n = \frac{\sum_{i\in[n]} d_i(d_i - 1)}{\ell_n}, \qquad \chi_n = \frac{\sum_{i\in[n]}[d_i(d_i-1)]^2}{4(\ell_n - 1)(\ell_n - 3)}. \tag{1.8}$$

For future purposes, we also define

$$\mu_n^{(r)} = \frac{\sum_{i\in[n]}(d_i)_r}{\ell_n}, \tag{1.9}$$

where, for an integer m, we let $(m)_r = m(m - 1)\cdots(m - r + 1)$ denote the rth factorial moment. (In particular, $\nu_n = \mu_n^{(2)}$.) We write $\mathcal{L}(X)$ for the distribution of X, and we write $(a \vee b)$ to denote the maximum of a and b. Our main result is the following Poisson approximation for the number of self-loops and multiple edges in the configuration model under certain conditions on the degree sequence:

Theorem 1.1 (Poisson approximation of self-loops and cycles). *For* $\mathrm{CM}_n(\boldsymbol{d})$, *there exists a universal constant* $C > 0$ *such that*

$$\big\|\mathcal{L}(S_n) - \mathsf{Po}\big(\lambda_n^S\big)\big\|_{\mathrm{TV}} \le \frac{C}{(\nu_n/2 \vee 1)} \frac{\nu_n^2}{\ell_n}, \tag{1.10}$$

$$\big\|\mathcal{L}(M_n) - \mathsf{Po}\big(\lambda_n^M\big)\big\|_{\mathrm{TV}} \le \frac{C}{(\lambda_n^M \vee 1)} \frac{(\mu_n^{(3)})^2 + \nu_n^4}{\ell_n} \tag{1.11}$$

and

$$\big\|\mathcal{L}(S_n + M_n) - \mathsf{Po}\big(\lambda_n^S + \lambda_n^M\big)\big\|_{\mathrm{TV}} \le \frac{C}{((\lambda_n^S + \lambda_n^M) \vee 1)} \frac{(\mu_n^{(3)})^2 + \nu_n^4}{\ell_n}. \tag{1.12}$$

In particular,

$$\mathbb{P}\big(\mathrm{CM}_n(\boldsymbol{d}) \text{ simple}\big) = \mathbb{P}(S_n + M_n = 0) = \mathrm{e}^{-\lambda_n^S - \lambda_n^M} + r, \quad \text{where } |r_n| \le \frac{C}{((\lambda_n^S + \lambda_n^M) \vee 1)} \frac{(\mu_n^{(3)})^2 + \nu_n^4}{\ell_n}. \tag{1.13}$$

Let us discuss some of the history of this problem. The configuration model was introduced by Bollobás in [6] to count the number of regular graphs, and provides a very nice example of the *probabilistic method* (see also [2]). Subsequently, the configuration model has been used successfully to analyze many properties of random regular graphs. The number of simple graphs can be rather directly obtained from the probability of simplicity of $\mathrm{CM}_n(\boldsymbol{d})$ (see e.g., [22, Proposition 7.6]). The introduction of the configuration model was inspired by, and generalized the results in, the work of Bender and Canfield [4]. See also Wormald [25] and McKay and Wormald [16] for previous work. Before giving further historical comments about Theorem 1.1, we discuss its implications on the number of simple graphs with a prescribed degree sequence:

Corollary 1.2 (Number of simple graphs with prescribed degrees). *The number* $N_n(\boldsymbol{d})$ *of simple graphs with degrees* $\boldsymbol{d} = (d_i)_{i\in[n]}$ *satisfies*

$$N_n(\boldsymbol{d}) = \big(\mathrm{e}^{-\lambda_n^S - \lambda_n^M} + r_n\big) \frac{(\ell_n - 1)!!}{\prod_{i\in[n]} d_i!} \quad \text{where } |r_n| \le \frac{C}{((\lambda_n^S + \lambda_n^M) \vee 1)} \frac{(\mu_n^{(3)})^2 + \nu_n^4}{\ell_n}. \tag{1.14}$$

In particular, the number $N_n(r)$ of r-regular graphs with n vertices satisfies, when rn is even,

$$N_n(r) = \left(e^{-(r-1)/2-(r-1)^2/4} + O(1/n)\right)\frac{(rn-1)!!}{(r!)^n}. \tag{1.15}$$

The proof of Corollary 1.2 follows directly from [22, Proposition 7.6], which implies that

$$\mathbb{P}\big(\mathrm{CM}_n(\boldsymbol{d}) \text{ simple}\big) = N_n(\boldsymbol{d})\frac{\prod_{i\in[n]} d_i!}{(\ell_n - 1)!!}. \tag{1.16}$$

Let us continue the discussion of the history of the configuration model and Theorem 1.1. The configuration model, as well as uniform random graphs with a prescribed degree sequence, were studied in greater generality by Molloy and Reed in [17] and [18], where they focus on the existence of a giant component. The Poisson approximation for the number of self-loops and multiple edges was first employed by Bollobás [7] in the case of random regular graphs. Janson [13] uses a Poisson approximation relying on the method of moments for the number of vertices having self-loops and the pairs of vertices having multiple edges between them. He investigates the case where the second moment of the degrees remains uniformly bounded, but not necessarily being uniformly integrable. In [14], Janson revisits the problem for S_n and M_n in (1.2) and uses the method of moments as well on the boundary case where the maximal degree is of order $\sqrt{n}$. Similar results were proved previously in earlier versions of [22]. Janson's extension in [14] is inspired by the wish to deal with multiple edges and self-loops for SIR epidemics on the configuration model in joint work with Luczak and Windridge [15].

In contrast to the works above based on the moment method, we use a Poisson approximation with couplings based on Stein's method, which also allows us to give error estimates. This method was recently used by Holmgren and Janson [11,12] to investigate the number of fringe trees in certain random trees. A major advantage is that Stein's method makes the approximation quantitative by giving explicit bounds on the error terms. Contrary to Janson [13, 14], our results do not allow for degrees that are of the order of $\sqrt{n}$.

1.2. *Regularity and moment assumptions on vertex degrees*

We next investigate special cases of Theorem 1.1, under stronger assumptions (on regularity and moments) on the degree distribution.

Let us now describe our regularity assumptions on the degree sequence $\boldsymbol{d}$ as $n \to \infty$. We denote the degree of a uniformly chosen vertex V in $[n]$ by $D_n = d_V$. The random variable D_n has distribution function F_n given by

$$F_n(x) = \frac{1}{n}\sum_{j\in[n]} \mathbb{1}_{\{d_j \le x\}}, \tag{1.17}$$

where $\mathbb{1}_A$ denotes the indicator of the event A. Our regularity condition is as follows:

Condition 1.3 (Regularity conditions for vertex degrees). The random variables D_n converge in distribution to some random variable D, and $\mathbb{E}[D_n] \to \mathbb{E}[D] < \infty$.

Remark 1.4 (Uniform integrability). Condition 1.3 implies that the sequence of random variables $(D_n)_{n\ge 1}$ is uniformly integrable. Conversely, if the sequence $(D_n)_{n\ge 1}$ is uniformly integrable, then every subsequence has a further subsequence that converges in distribution, and Condition 1.3 can be used along that subsequence.

Define

$$\nu = \frac{\mathbb{E}[D(D-1)]}{\mathbb{E}[D]}. \tag{1.18}$$

Under suitable assumptions on the second moment of D_n, we can deduce more precise information about S_n and M_n, and in particular consider their joint distribution. Our main results in the finite-variance case are the following three theorems:

Theorem 1.5 (Poisson approximation of self-loops and cycles). *For* $\mathrm{CM}_n(\boldsymbol{d})$, *where $\boldsymbol{d}$ satisfies the Degree Regularity Condition* 1.3 *and* $\lim_{n\to\infty} \mathbb{E}[D_n^2] = \mathbb{E}[D^2] < \infty$, *it holds that*

$$\big\| \mathcal{L}(S_n, M_n) - \mathsf{Po}(\nu/2) \otimes \mathsf{Po}\big(\nu^2/4\big) \big\|_{\mathrm{TV}} \to 0. \tag{1.19}$$

To prove Theorem 1.5, we introduce a Cramér–Wold device for Poisson random variables, that guarantees the independence of the limiting random variables (see Section 2.1), and that is of independent interest.

Our methods also yield some speed of convergence results:

Theorem 1.6 (Speed of convergence for self-loops and cycles under finite third moment). *For* $\mathrm{CM}_n(\boldsymbol{d})$, *where $\boldsymbol{d}$ satisfies the Degree Regularity Condition* 1.3 *and* $\lim_{n\to\infty} \mathbb{E}[D_n^3] = \mathbb{E}[D^3] < \infty$, *it holds that*

$$\mathbb{P}(S_n = 0) = \mathrm{e}^{-\lambda_n^S} + O(1/n), \qquad \mathbb{P}(M_n = 0) = \mathrm{e}^{-\lambda_n^M} + O(1/n), \tag{1.20}$$

and

$$\mathbb{P}(S_n = M_n = 0) = \mathrm{e}^{-\lambda_n^S - \lambda_n^M} + O(1/n). \tag{1.21}$$

In particular,

$$\mathbb{P}\big(\mathrm{CM}_n(\boldsymbol{d}) \ simple\big) = \mathrm{e}^{-\lambda_n^S - \lambda_n^M} + O(1/n). \tag{1.22}$$

Furthermore, when also $\lim_{n\to\infty} \mathbb{E}[D_n^4] = \mathbb{E}[D^4] < \infty$, $\lambda_n^S = \nu_n/2$ *and* $\lambda_n^M = \nu_n^2/4 + O(1/n)$, *so that* (1.20)–(1.22) *hold with λ_n^S and λ_n^M replaced with $\nu_n/2$ and $\nu_n^2/4$.*

Theorem 1.7 (Speed of convergence for self-loops and cycles with infinite third moment). *For* $\mathrm{CM}_n(\boldsymbol{d})$, *where $\boldsymbol{d}$ satisfies the Degree Regularity Condition* 1.3 *and* $\lim_{n\to\infty} \mathbb{E}[D_n^2] = \mathbb{E}[D^2] < \infty$, *it holds that*

$$\mathbb{P}(S_n = 0) = \mathrm{e}^{-\lambda_n^S} + O(1/n), \qquad \mathbb{P}(M_n = 0) = \mathrm{e}^{-\lambda_n^M} + O\big(d_{\max}^2/n\big), \tag{1.23}$$

and

$$\mathbb{P}(S_n = M_n = 0) = \mathrm{e}^{-\lambda_n^S - \lambda_n^M} + O\big(d_{\max}^2/n\big). \tag{1.24}$$

In particular,

$$\mathbb{P}\big(\mathrm{CM}_n(\boldsymbol{d}) \ simple\big) = \mathrm{e}^{-\lambda_n^S - \lambda_n^M} + O\big(d_{\max}^2/n\big). \tag{1.25}$$

Let us relate the above result to the scale-free behavior as observed in many random graphs. We refer the reader to [22, Chapter 1] for an extensive introduction to real-world networks and their power-law degree sequences. Let F_n denote the empirical distribution function of the degree sequence, so that $F_n(x) \to F(x)$ for every $x \in \mathbb{N}$ when Degree Regularity Condition 1.3 holds. Scale-free behavior is defined as follows:

Definition 1.8 (Scale-free random graphs). A configuration model with a degree sequence satisfying Degree Regularity Condition 1.3 is called *scale free with power-law exponent* τ when its asymptotic degree distribution $F(x) = \mathbb{P}(D \le x)$ satisfies that, as $x \to \infty$,

$$1 - F(x) = x^{-(\tau-1)+o(1)}. \tag{1.26}$$

Often, we assume that the asymptotic degree distribution obeys a pure power law, in which case (1.26) is strengthened to $[1 - F](x) = cx^{-(\tau-1)}(1 + o(1))$ for x large. Note that (1.26) allows for slowly-varying corrections that are not important for our discussion. In a pure power-law setting, one can expect that $[1 - F_n](x) \approx cx^{-(\tau-1)}$ (unless x is too large). Then, the number of vertices of degree at least x equals

$$n[1 - F_n](x) \approx cnx^{-(\tau-1)}. \tag{1.27}$$

This is $\Theta(1)$ precisely when $x = \Theta(n^{1/(\tau-1)})$. Thus, one can expect that $d_{\max} = \Theta(n^{1/(\tau-1)})$, so that the error terms in (1.23)–(1.25) are of order $n^{(3-\tau)/(\tau-1)}$, which is $o(1)$ when $\tau > 3$. In turn, $\tau > 3$ corresponds to $\mathbb{E}[D^2] < \infty$, so that we are in the finite-variance degree setting. However, note that the scale-free property in Definition 1.8 is an *asymptotic* property, and it implies very little about the actual behavior of $d_{\max}$, aside from the obvious statement that $d_{\max} \to \infty$ when (1.26) holds. Thus, in the literature often, and particularly in the infinite-variance case where $\tau \in (2, 3)$, more detailed assumptions are made on the empirical tail distribution of the degrees, such as the statement that (1.27) holds with upper and lower bounds for appropriate ranges of x values. Particularly in the infinite-variance regime where $\tau \in (2, 3)$, the behavior can depend rather sensitively on the precise range when the approximation (1.27) holds. See e.g., [23] where distances in the infinite-variance degree configuration model for $\tau \in (2, 3)$ are found to depend sensitively on the precise degree characteristics. Two canonical examples of degree distributions that satisfy (1.27) for a large range of x values are (a) when the degrees are independent and identically distributed with tail distribution function satisfying $1 - F(x) = cx^{-(\tau-1)}$, or (b) when $d_i = [1-F]^{-1}(i/n)$ where the tail distribution function satisfies $1 - F(x) = cx^{-(\tau-1)}(1 + o(1))$ for x large. In these cases, (1.27) can be seen to hold for all $x = o(n^{1/(\tau-1)})$ and $d_{\max} = \Theta(n^{1/(\tau-1)})$.

In our proof, a concrete bound is given of the error term in the Poisson approximation in Theorem 1.5 in terms of the moments of D_n and n, of the form as in (1.10)–(1.13).

1.3. *Infinite-variance degrees*

In this section, we study the configuration model with infinite-variance degrees. We assume that $v_n \to \infty$, which holds when Definition 1.8 holds with $\tau \in (2, 3)$. When the degrees obey a power-law with exponent $\tau \in (2, 3)$, we will assume that

$$v_n = O\left(n^{2/(\tau-1)-1}\right). \tag{1.28}$$

The main example of this behaviour is a power-law degree distribution with exponent τ (recall the discussion below Definition 1.8), for which often $d_{\max} = \max_{i \in [n]} d_i = \Theta(n^{1/(\tau-1)})$ with $\tau \in (2, 3)$. Note that $v_n \geq d_{\max}^2/n\mathbb{E}[D_n]$. Our main result under this assumption is the following central limit theorem for the number of self-loops:

Theorem 1.9 (CLT for self-loops in $\mathrm{CM}_n(d)$ with infinite-variance degrees). *For $\mathrm{CM}_n(d)$, assume that d satisfies the Degree Regularity Condition 1.3, while $v_n \to \infty$ as in (1.28). Then, for $\tau > 2$,*

$$\frac{S_n - v_n/2}{\sqrt{v_n/2}} \xrightarrow{d} Z. \tag{1.29}$$

Unfortunately, our proof does not apply to the multiple edges, since the number of multiple edges between vertices of degree $d_i \gg \sqrt{n}$ grows too rapidly. In this case, we need to assume that the degrees satisfy that $d_{\max} = o(\sqrt{n})$:

Theorem 1.10 (CLT for multiple edges in $\mathrm{CM}_n(d)$ with infinite-variance degrees). *Let $d_{\max} = o(\sqrt{n})$ and $v_n \to \infty$. Then,*

$$\frac{M_n - \lambda_n^M}{\sqrt{\lambda_n^M}} \xrightarrow{d} Z, \tag{1.30}$$

where Z is a standard normal random variable.

Alternatively, we could also count only the multiple edges between vertices of degree $o(\sqrt{n})$. Indeed, take $m_n = o(\sqrt{n})$ and define

$$M_n^{(1)} = \sum_{1 \leq i < j \leq n} \mathbb{1}_{\{d_i, d_j \leq m_n\}} X_{ij}(X_{ij} - 1)/2. \tag{1.31}$$

Then, Theorem 1.10 also holds for $M_n^{(1)}$ with λ_n^M replaced with $\mathbb{E}[M_n^{(1)}]$ for any $m_n = o(\sqrt{n})$. Note that if there two vertices of degree $\Omega(\sqrt{n})$, then the expected number of edges between these two vertices is already of order 1, and the

contribution just from these vertices will impact the Poisson approximation. It is possible that a more careful analysis taking into account high-degree vertices separately will allow extending our results to graphs with such vertices. (The same holds for self-loops with vertices of degree $\Omega(n)$.) We do not pursue this here.

1.4. *Directed and bipartite configuration models*

In this section, we discuss the directed and bipartite configuration model.

Self-loops and multiple edges in the directed configuration model. For a general description of the directed configuration model we refer to Cooper and Frieze [9] and van der Hofstad [22, Section 7.8]. Fix $\boldsymbol{d}^{(\text{in})} = (d_i^{(\text{in})})_{i \in [n]}$ and $\boldsymbol{d}^{(\text{out})} = (d_i^{(\text{out})})_{i \in [n]}$ to be sequences of in-degrees and out-degrees of the vertices $i \in [n]$, respectively. For a graph with in- and out-degree sequence $\boldsymbol{d} = (\boldsymbol{d}^{(\text{in})}, \boldsymbol{d}^{(\text{out})})$ to exist, we need to assume that

$$\hat{\ell}_n = \sum_{i \in [n]} d_i^{(\text{in})} = \sum_{i \in [n]} d_i^{(\text{out})}. \tag{1.32}$$

The directed configuration model $\text{DCM}_n(\boldsymbol{d}^{(\text{in})}, \boldsymbol{d}^{(\text{out})})$ is obtained by pairing each in-half-edge to a uniformly chosen out-half-edge. Similarly as for the undirected case, we may investigate limit laws for the number of self-loops and multiple edges $(\hat{S}_n, \hat{M}_n)$. Self-loops occur if an in-half-edge is paired to an out-half-edge incident to the same vertex. Multiple edges occur between a pair of vertices, if two in-half-edges that are incident to one of the vertices are paired to two out-half-edges that are incident to the other vertex. Note that we are not considering two edges with opposite directions between two vertices as a pair of multiple-edges (since this phenomenon actually often happens in real-world networks), but only pairs of edges with the same direction. Thus, we define

$$\hat{M}_n = \sum_{1 \leq i < j \leq n} \left[X_{ij}^{(\text{in})}(X_{ij}^{(\text{in})} - 1)/2 + X_{ij}^{(\text{out})}(X_{ij}^{(\text{out})} - 1)/2 \right]$$

$$= \sum_{i \neq j, i, j \in [n]} X_{ij}^{(\text{in})}(X_{ij}^{(\text{in})} - 1)/2 = \sum_{i \neq j, i, j \in [n]} X_{ij}^{(\text{out})}(X_{ij}^{(\text{out})} - 1)/2, \tag{1.33}$$

where $X_{ij}^{(\text{in})}$ are the number of edges between i and j that are directed from j to i and $X_{ij}^{(\text{out})}$ are the number of edges between i and j that are directed from i to j, and the last equality follows by the symmetry $X_{ij}^{(\text{in})} = X_{ji}^{(\text{out})}$.

Let

$$\hat{\lambda}_n^S = \mathbb{E}[\hat{S}_n], \qquad \hat{\lambda}_n^M = \mathbb{E}[\hat{M}_n]. \tag{1.34}$$

By similar calculations as in the undirected case we get

$$\hat{\lambda}_n^S = \frac{\sum_{i \in [n]} d_i^{(\text{in})} d_i^{(\text{out})}}{\hat{\ell}_n}, \qquad \hat{\lambda}_n^M = \frac{\sum_{i \neq j, i, j \in [n]} d_i^{(\text{in})}(d_i^{(\text{in})} - 1) d_j^{(\text{out})}(d_j^{(\text{out})} - 1)}{2\hat{\ell}_n(\hat{\ell}_n - 1)}. \tag{1.35}$$

We also define

$$\mu_n^{(r,\text{in})} = \frac{\sum_{i \in [n]} (d_i^{(\text{in})})_r}{\hat{\ell}_n}, \qquad \mu_n^{(r,\text{out})} = \frac{\sum_{i \in [n]} (d_i^{(\text{out})})_r}{\hat{\ell}_n}. \tag{1.36}$$

Then, our main result for the directed CM is as follows:

Theorem 1.11 (Poisson approximation of self-loops and multiple edges in directed CM). *For* $\text{DCM}_n(\boldsymbol{d}^{(\text{in})}, \boldsymbol{d}^{(\text{out})})$, *there exists a universal constant* $C > 0$ *such that*

$$\left\| \mathcal{L}(\hat{S}_n) - \text{Po}(\hat{\lambda}_n^S) \right\|_{\text{TV}} \leq \frac{C}{(\hat{\lambda}_n^S \vee 1)} \frac{(\hat{\lambda}_n^S)^2}{\hat{\ell}_n}, \tag{1.37}$$

$$\left\| \mathcal{L}(\hat{M}_n) - \mathsf{Po}\big(\hat{\lambda}_n^M\big) \right\|_{\mathrm{TV}} \le \frac{C}{(\hat{\lambda}_n^M \vee 1)} \frac{\mu_n^{(3,\mathrm{in})} \mu_n^{(3,\mathrm{out})} + (\hat{\lambda}_n^M)^2}{\hat{\ell}_n} \tag{1.38}$$

and

$$\left\| \mathcal{L}(\hat{S}_n + \hat{M}_n) - \mathsf{Po}\big(\hat{\lambda}_n^S + \hat{\lambda}_n^M\big) \right\|_{\mathrm{TV}} \le \frac{C}{((\hat{\lambda}_n^S + \hat{\lambda}_n^M) \vee 1)} \frac{\mu_n^{(3,\mathrm{in})} \mu_n^{(3,\mathrm{out})} + (\hat{\lambda}_n^M)^2}{\hat{\ell}_n}. \tag{1.39}$$

In particular,

$$\mathbb{P}\big(\mathrm{DCM}_n\big(\boldsymbol{d}^{(\mathrm{in})}, \boldsymbol{d}^{(\mathrm{out})}\big) \text{ simple}\big) = \mathbb{P}(\hat{S}_n + \hat{M}_n = 0) = \mathrm{e}^{-\hat{\lambda}_n^S - \hat{\lambda}_n^M} + \hat{r}_n,$$

$$\text{where } |\hat{r}_n| \le \frac{C}{((\hat{\lambda}_n^S + \hat{\lambda}_n^M) \vee 1)} \frac{\mu_n^{(3,\mathrm{in})} \mu_n^{(3,\mathrm{out})} + (\hat{\lambda}_n^M)^2}{\hat{\ell}_n}. \tag{1.40}$$

Multiple edges in the bipartite configuration model. We continue with a discussion of multiple edges in the bipartite configuration model. For a general description of the bipartite configuration model we refer e.g. to Blanchet and Stauffer [5] and Janson [14]. Let $n^{(1)}$ denote the number of vertices on the left side of the bipartite graph, and $n^{(r)}$ the number of vertices on the right side of the bipartite graph. Fix $\boldsymbol{d}^{(1)} = (d_i^{(1)})_{i \in [n^{(1)}]}$ and $\boldsymbol{d}^{(r)} = (d_j^{(r)})_{j \in [n^{(r)}]}$ degrees sequences for the two left and right parts, with

$$\bar{\ell}_n = \sum_{i \in [n^{(1)}]} d_i^{(1)} = \sum_{j \in [n^{(r)}]} d_j^{(r)}. \tag{1.41}$$

The bipartite configuration model $\mathrm{BCM}_n(\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(r)})$ is obtained by pairing each half-edge incident to one of the vertices in $n^{(1)}$ to a uniformly chosen half-edge of those incident to the vertices in $n^{(r)}$. Thus, in this model there are obviously no self-loops. However, there could exist multiple edges $\bar{M}_n$. Multiple edges occur between a pair of vertices (i, j) when two half-edges incident to a vertex $i \in [n^{(1)}]$ are paired to two half-edges that are incident to a vertex $j \in [n^{(r)}]$.

To study the number of multiple edges, we define

$$\bar{M}_n = \sum_{i \in [n^{(1)}], j \in [n^{(r)}]} \bar{X}_{ij}(\bar{X}_{ij} - 1)/2, \tag{1.42}$$

where $\bar{X}_{ij}$ denotes the number of edges between i and j. Let $\bar{\lambda}_n^M = \mathbb{E}[\bar{M}_n]$. By similar calculations as for the standard configuration model we get

$$\bar{\lambda}_n^M = \frac{\sum_{i \in [n^{(1)}], j \in [n^{(r)}]} d_i^{(1)}(d_i^{(1)} - 1) d_i^{(r)}(d_i^{(r)} - 1)}{2\bar{\ell}_n(\bar{\ell}_n - 1)}. \tag{1.43}$$

We also define

$$\mu_n^{(k,1)} = \frac{\sum_{i \in [n^{(1)}]} (d_i^{(1)})_k}{\bar{\ell}_n}, \qquad \mu_n^{(k,r)} = \frac{\sum_{j \in [n^{(r)}]} (d_j^{(r)})_k}{\bar{\ell}_n}. \tag{1.44}$$

Then, our main result for the bipartite CM is as follows:

Theorem 1.12 (Poisson approximation of multiple edges in bipartite CM). *For* $\mathrm{BCM}_n(\boldsymbol{d}^{(1)}, \boldsymbol{d}^{(r)})$, *there exists a universal constant* $C > 0$ *such that*

$$\left\| \mathcal{L}(\bar{M}_n) - \mathsf{Po}\big(\bar{\lambda}_n^M\big) \right\|_{\mathrm{TV}} \le \frac{C}{(\bar{\lambda}_n^M \vee 1)} \frac{\mu_n^{(3,1)} \mu_n^{(3,r)} + (\bar{\lambda}_n^M)^2}{\bar{\ell}_n}. \tag{1.45}$$

In particular,

$$\mathbb{P}\big(\text{BCM}_n\big(\boldsymbol{d}^{(\text{l})}, \boldsymbol{d}^{(\text{r})}\big) \text{ simple}\big) = \mathbb{P}(\bar{M}_n = 0) = e^{-\bar{\lambda}_n^M} + \bar{r}_n, \quad \text{where } |\bar{r}_n| \leq \frac{C}{(\bar{\lambda}_n^M \vee 1)} \frac{\mu_n^{(3,\text{l})} \mu_n^{(3,\text{r})} + (\bar{\lambda}_n^M)^2}{\bar{\ell}_n}. \quad (1.46)$$

Remark 1.13. For the directed configuration model $\text{DCM}_n(\boldsymbol{d}^{(\text{in})}, \boldsymbol{d}^{(\text{out})})$ we can also prove results that correspond to Theorems 1.5–1.10 for the undirected case. Similarly, for the bipartite configuration model $\text{BCM}_n(\boldsymbol{d}^{(\text{l})}, \boldsymbol{d}^{(\text{r})})$, we can prove results that correspond to Theorems 1.6, 1.7 and 1.10 (recalling that there are no self-loops in this model). We leave these statements of the other results to the reader.

1.5. *Discussion and open problems*

In this section, we discuss our results and provide open problems.

Instead of investigating S_n and M_n as in (1.2), one could also investigate other random variables that imply simplicity when the variable equals zero. An example would be to study $\widetilde{M}_n$ in (1.4). Another natural example would be

$$S_n^{(\text{l})} = \sum_{i \in [n]} \mathbb{1}_{\{X_{ii} \geq 1\}}, \qquad M_n^{(\text{l})} = \sum_{1 \leq i < j \leq n} \mathbb{1}_{\{X_{ij} \geq 2\}}, \qquad (1.47)$$

as Janson does in [13]. Both alternatives are of interest, since they all quantify different aspects of how many self-loops and multiple edges there are, and might satisfy central limits theorems for infinite-variance degrees for *different* values of the power-law exponent τ. However, application of Stein's method to these random variables is more difficult. For $M_n^{(\text{l})}$, this is primarily due to the fact that the probability of $\{X_{kl} \geq 2\}$ conditionally on $\{X_{ij} \geq 2\}$ is quite involved. This effect is best seen in the conditions on $d_{\max}$. Indeed, our results necessitate the fact that $d_{\max} = o(\sqrt{n})$, since often error terms are of the order $d_{\max}^2/n$ (see e.g., (1.23) in Theorem 1.7). Janson [13] and [14] allows for settings where $d_{\max} = \Theta(\sqrt{n})$. In particular, in [14], Janson proves a Poisson approximation for $S_n + M_n$ as in (1.2) in the regime where $d_{\max} = \Theta(\sqrt{n})$. In particular, Janson proves that the difference between the moments of $S_n + M_n$ and $\widehat{Z}$, which is given in terms of Poisson random variables, is $O(n^{-1/2})$ or $O(d_{\max}/n)$. Here,

$$\widehat{Z} = \sum_{i \in [n]} \widehat{X}_i + \sum_{1 \leq i < j \leq n} \binom{\widehat{X}_{ij}}{2}, \qquad (1.48)$$

where $\widehat{X}_i$ is Poisson with mean $\nu_n/2$, while $\widehat{X}_{ij}$ is Poisson with mean λ_{ij} with

$$\lambda_{ij} = \frac{\sqrt{d_i(d_i - 1)d_j(d_j - 1)}}{\ell_n}. \qquad (1.49)$$

Note that $\widehat{Z}$ is not quite a Poisson random variable. Thus, this result describes deviations from a Poisson random variable in the regime where $d_{\max} = \Theta(\sqrt{n})$, and shows that the Poisson approximation attempted in our paper can not be expected to hold in this regime.

We next discuss configuration models in the power-law setting where $\tau \in (2, 3)$ in some more detail. As we see in Theorems 1.9–1.10, the number of self-loops and multiple edges in this case tend to infinity in probability, so that it is highly unlikely that there are none. This makes that the approach to obtain simple graphs by conditioning the configuration model to be simple is no viable option. However, real-world networks with power-law degrees with $\tau \in (2, 3)$ are often observed, see e.g. the surveys in [1] and [19]. For example, Newman [20] proposes, amongst others, the configuration model with power-law degrees as a model for real-world networks, while Newman, Strogatz and Watts [21] investigate the graph distances of such models. There is ample evidence that practitioners do wish to obtain simple graphs as a null-model for many real-world networks. There are many papers using the configuration model as null-models for real-world networks. In the case of infinite-variance degrees, this gives rise to an enormous problem. One possible solution to resolve this issue is to consider, instead, the *erased configuration model*, as suggested by Britton, Deijfen and Martin-Löf [8]. In this erased model, self-loops are simply erased and multiple-edges merged

to make the graph simple. While this does not produce a graph that has a uniform distribution over the space of all simple graphs, this model is highly practical, and we see that we only remove a small proportion of the edges so that the degree distribution is virtually unaltered (see e.g. [22, Chapter 7] for more details). This explains our interest in configuration models with power-law degrees. Theorem 1.9–1.10 can thus be seen as quantifications of the statement that we 'do not remove many edges'. For example, Van der Hoorn and Litvak [24] investigate the number of removed edges in the erased configuration model in the setting where the degrees are i.i.d. with distribution function F satisfying $[1 - F](x) = cx^{-(\tau-1)}$ for some $\tau \in (2, 3)$ (in fact, they even allow for extra slowly-varying functions, but we refrain from discussing this generalization). The number of removed edges corresponds to

$$R_n = S_n + \sum_{1 \leq i < j \leq n} \left(X_{ij} - \mathbb{1}_{X_{ij} \geq 1} \right), \tag{1.50}$$

which, for $\tau \in (2, 3)$, is significantly smaller than $S_n + M_n$. Interestingly, the upper bound proved by van der Hoorn and Litvak [24] is *different* for $\tau \in (5/2, 3)$ compared to $\tau \in (2, 5/2)$. It would be of interest to investigate whether such a phase transition is an artifact of the proof, or whether it really is there.

Gao and Wormald [10] take a different approach. Indeed, they investigate the number of simple graphs in the power-law case with $\tau \in (2, 3)$. Under assumptions on $\mu_n^{(r)}$, they investigate sharp asymptotics for $\mathbb{P}(\mathrm{CM}_n(\boldsymbol{d})$ simple). Let $M_r = \ell_n \mu_n^{(r)}$, then [10, Theorem 1] assumes that $M_2 = o(M_1^{9/8})$. [10, Theorem 2] assumes that the number of vertices of degree i can be uniformly bounded by a constant times $nk^{-1/\tau}$ for $\tau > 5/2$, which, in particular, implies that $d_{\max} = O(n^{1/\tau})$. The result that most closely relates to ours is [10, Theorem 3], where it is assumed that the number of vertices i with degree $d_i > x$ is bounded by $Cnx^{-(\tau-1)}$, for some $C > 0$ and uniformly in x, where $\tau \in (1 + \sqrt{3}, 3)$ (note that $1 + \sqrt{3} \approx 2.732$). This is close to a scale-free upper bound, recall Definition 1.27. Gao and Wormald call the latter setting a *power-law distribution-bounded* degree sequence. Mind that this setting allows for a maximal degree $d_{\max} = \Theta(n^{1/(\tau-1)})$, which can be well above $\sqrt{n}$.

These results are highly interesting, and show in particular that $\mathbb{P}(S_n = M_n = 0) = \mathrm{e}^{-(\nu_n/2 + \nu_n^2/4)(1+o(1))}$ while at the same time giving an asymptotic expression of $o(1)$ in the exponent in terms of $\mu_n^{(r)}$ with $r = 1, 2$ and 3. These can be used to compute the asymptotic number of simple graphs with the given degree sequence. The proofs rely on *switching methods*, which have been used in the literature to study various settings in which the connection between simple graphs and the configuration model cannot be used, such as in [16], where McKay and Wormald study the number of regular graphs with degrees r_n with $r_n \to \infty$ and $r_n = o(\sqrt{n})$. Also Janson [13] relies on the switching method.

The results by Gao and Wormald cannot be obtained from ours, since the event of simplicity is an extreme-value event when $\tau \in (2, 3)$ with vanishing probability, whereas we study weak limits with increasing means and variances. On the other hand, the assumption that $d_{\max} = O(n^{1/\tau}) = o(\sqrt{n})$ with $\tau \in (2, 3)$ implies that we obtain a CLT for both the number of self-loops as well as the number of multiple edges by Theorem 1.10.

Instead of assuming that $d_{\max} = O(n^{1/\tau})$, we prefer to work with cases where $\sum_{i \geq k} n_i = O(nk^{-1/(\tau-1)})$, as in the power-law distribution-bounded case of Gao and Wormald [10]. This preference is inspired by the fact that the maximum of n i.i.d. random variables with tail distribution function $1 - F(x) = cx^{-(\tau-1)}$ is of order $n^{1/(\tau-1)}$ rather than $n^{1/\tau}$. Recall Definition 1.27 and the definition below it. In turn, a natural choice of deterministic power-law degrees arises when we take the number n_k of vertices of degree k to be equal to $n_k = \lceil nF(k) \rceil - \lceil nF(k-1) \rceil$, where again $1 - F(x) = cx^{-(\tau-1)}$ is a power-law distribution and also $d_{\max} = \Theta(n^{1/(\tau-1)})$.

Another interesting problem is to investigate whether the CLT for the number of multiple edges M_n can be extended to the full range $\tau \in (2, 3)$ without the restriction that $d_{\max} = o(\sqrt{n})$. Our current proof relies on a Poisson approximation, which in particular can only be used when the mean and the variance of the asymptotic normal distribution are comparable. We believe that this is false for some $\tau \in (2, 3)$. It would be interesting to investigate whether instead Stein's method for normal asymptotic distributions can be applied. This open problem is also interesting for other sums of indicators, for example for $\widetilde{M}_n$ in (1.4) and for $M_n^{(1)}$ in (1.47).

Organization. The remainder of this paper is organised as follows. In Section 2, we present the preliminaries used in this paper, which include a novel Poisson Cramér-Wold device as well as bounds on Poisson approximations. In Section 3, we present couplings of dependent indicators that will be crucial in applying the Stein-Chen method. In Section 4, we present the proofs of our main results.

2. Preliminaries

2.1. *Poisson Cramér–Wold device*

In this section, we show that convergence of two random variables to two *independent* Poisson variables follows when we can prove convergence of sums of their thinned versions. This is to Poisson random variables, as the Cramér–Wold device is for Normal variables. We start by explaining this method, which is of independent interest.

Let (X, Y) be two integer-valued random variables. Fix $p, q \in [0, 1]$ and define

$$X_p = \mathrm{Bin}(X, p), \qquad Y_q = \mathrm{Bin}(Y, q) \tag{2.1}$$

to be two binomial random variables, independent conditioned on X, Y. Then, the Poisson Cramér–Wold device is the following theorem:

Theorem 2.1 (Poisson Cramér–Wold device). *Suppose that, for every $p, q \in [0, 1]$, $X_p + Y_q$ has a Poisson distribution with mean $p\mu_X + q\mu_Y$. Then (X, Y) are two independent Poisson random variables with means μ_X and μ_Y, respectively.*

Proof. Let $M_{X,Y}(s, t) = \mathbb{E}[e^{sX+tY}]$ denote the joint moment generating function of a random vector (X, Y) and $M_X(t) = \mathbb{E}[e^{tX}]$ the moment generating function of the random variable X. Recall that the moment generating function of a binomial random variable X with parameters n and p equals $M_X(t) = (pe^t + (1 - p))^n$ and that of a Poisson random variable Y with parameter λ equals $M_Y(t) = e^{\lambda(e^t - 1)}$.

We know that $M_{X_p + Y_q}(t) = \mathbb{E}[e^{t(X_p + Y_q)}] = e^{(p\mu_X + q\mu_Y)(e^t - 1)}$. We wish to show that $M_{X,Y}(t, s) = e^{\mu_X(e^t - 1) + \mu_Y(e^s - 1)}$, and it suffices to prove this for $s, t \geq 0$.

We rewrite the moment generating function of $X_p + Y_q$ as

$$M_{X_p + Y_q}(t) = \mathbb{E}\big[e^{t(X_p + Y_q)}\big] = \mathbb{E}\big[\big(pe^t + (1 - p)\big)^X \big(qe^t + (1 - q)\big)^Y\big]$$
$$= M_{X,Y}\big(\log(pe^t + (1 - p)), \log(qe^t + (1 - q))\big). \tag{2.2}$$

Without loss of generality, we may assume that $t \geq s$. We also assume that $s \geq 0$. We take $p = 1$, so that $\log(pe^t + (1 - p)) = t$, and q such that $\log(qe^t + (1 - q)) = s$. Solving gives $q = (e^s - 1)/(e^t - 1) \in [0, 1]$, since $0 \leq s \leq t$. Then we get that

$$M_{X,Y}(t, s) = e^{\mu_X(e^t - 1) + q\mu_Y(e^t - 1)} = e^{\mu_X(e^t - 1) + \mu_Y(e^s - 1)}, \tag{2.3}$$

as required. We conclude that X and Y are independent Poisson variables with means μ_X and μ_Y, respectively. $\quad\square$

Remark 2.2. The proof above only uses the assumption in the case $p = 1$ or $q = 1$ (according to whether $t \geq s$ or $t \leq s$). By analyticity, $M(s, t)$ for $s \leq t$ determines M and the laws of X, Y completely, so we can even restrict to $p = 1$.

Corollary 2.3 (Poisson Cramér–Wold device for convergence). *Let $(X^{(n)}, Y^{(n)})_{n \geq 1}$ be non-negative integer random variables. Suppose that, for every $p, q \in [0, 1]$, $X_p^{(n)} + Y_q^{(n)}$ converges in distribution to a Poisson distribution with mean $p\mu_X + q\mu_Y$. Then $(X^{(n)}, Y^{(n)})$ converges in distribution to a pair of independent Poisson random variables with means μ_X and μ_Y, respectively.*

This could be proved directly using characteristic functions in a similar way as in Theorem 2.1 (where instead moment generating functions were used). Instead, we deduce this from Theorem 2.1.

Proof. Taking $p = 1$ and $q = 0$ we find that $X^{(n)}$ is a tight sequence, and similarly so is $Y^{(n)}$, and thus there are subsequences of $(X^{(n)}, Y^{(n)})$ that converge in distribution. Let (X, Y) be some subsequential limit. By Theorem 2.1 we find that (X, Y) are independent Poisson variables as claimed. Since every subsequential limit has the same law, the convergence is along the entire sequence. $\quad\square$

2.2. *Poisson approximation*

We will make extensive use of Poisson approximations. For this, we rely on [3, Theorem 2.C], which we quote for convenience. We start by introducing some notation. Let

$$W = \sum_{\alpha \in \Lambda} I_\alpha \tag{2.4}$$

be a sum of (possibly dependent) indicator functions indexed by some set Λ. Let $(J_{\beta\alpha})_{\beta \in \Lambda \setminus \{\alpha\}}$ be a collection of indicator variables with the joint distribution of $((I_\beta)_{\beta \in \Lambda} \mid I_\alpha = 1)$, i.e., the conditional distribution of all other indicators *given* that $I_\alpha = 1$. Let $p_\alpha = \mathbb{P}(I_\alpha = 1)$ and

$$\lambda = \sum_{\alpha \in \Lambda} p_\alpha = \mathbb{E}[W]. \tag{2.5}$$

Note that the while the joint distribution of the variables $(J_{\beta\alpha})$ is specified, the coupling with the family (I_α) can be chosen arbitrarily. Then we have the following Poisson approximation:

Theorem 2.4 (Poisson approximations [3]). *With $(J_{\beta\alpha})_{\beta \in \Lambda \setminus \{\alpha\}}$ as above,*

$$\|\mathcal{L}(W), \mathsf{Po}(\lambda)\|_{\mathrm{TV}} \leq (1 \wedge \lambda^{-1})\left(\sum_{\alpha \in \Lambda} p_\alpha^2 + \sum_{\alpha \in \Lambda} \sum_{\beta \neq \alpha} p_\alpha \mathbb{E}\big[|I_\beta - J_{\beta\alpha}|\big]\right). \tag{2.6}$$

As these are indicator variables, we can compute that $\mathbb{E}[|I_\beta - J_{\beta\alpha}|] = \mathbb{P}(I_\beta \neq J_{\beta\alpha})$. Our proof is based on finding an efficient coupling of I_β and $J_{\alpha\beta}$, i.e., one for which $J_{\beta\alpha} = I_\beta$ holds with high probability.

3. Couplings for the number of self-loops and multiple edges in the CM

In this section, we investigate (S_n, M_n) as defined in (1.2). We will rely on Poisson approximations as in Theorem 2.4, for which it is convenient to rewrite (S_n, M_n) as

$$S_n = \sum_{i \in [n]} \sum_{1 \leq s < t \leq d_i} L_{st}, \qquad M_n = \sum_{1 \leq i < j \leq n} \sum_{1 \leq s_1 < s_2 \leq d_i} \sum_{1 \leq t_1 \neq t_2 \leq d_j} L_{s_1 t_1, s_2 t_2}. \tag{3.1}$$

Here L_{st} is the indicator that the half-edges s and t that are incident to the same vertex are paired to form a self-loop, while $L_{s_1 t_1, s_2 t_2}$ is the indicator that $s_1 t_1$ and $s_2 t_2$ are paired together, where s_1, s_2 are incident to the same vertex, as are t_1, t_2. Note that we may assume $s_1 < s_2$, but swapping t_1, t_2 will lead to different configurations, so their order is not given.

We use the Poisson approximation in Theorem 2.4, jointly with the Poisson Cramér–Wold device in Theorem 2.1 and thus deal with

$$W = \sum_{i \in [n]} \sum_{1 \leq s < t \leq d_i} L_{st} K_{st} + \sum_{1 \leq i < j \leq n} \sum_{1 \leq s_1 < s_2 \leq d_i} \sum_{1 \leq t_1 \neq t_2 \leq d_j} L_{s_1 t_1, s_2 t_2} K_{s_1 t_1, s_2 t_2}, \tag{3.2}$$

where (K_{st}) are i.i.d. Bernoulli's with probability p and $(K_{s_1 t_1, s_2 t_2})$ are i.i.d. Bernoulli's with probability q. We need to describe the law of the indicators conditioned on $I_\alpha = 1$. Here α can be st or $s_1 t_1, s_2 t_2$ and $I_\alpha = L_\alpha K_\alpha$. Note that the K_α's are completely independent of everything else, so they do not change the story. For simplicity, we will just deal with $K_\alpha \equiv 1$, though all computations below are valid for any set of K's.

The success probabilities. We start by analyzing the "success" probabilities p_α. For α corresponding to a self-loop,

$$p_\alpha = \frac{1}{\ell_n - 1} \quad \text{where } \ell_n = \sum_{i \in [n]} d_i. \tag{3.3}$$

For α corresponding to a pair of edges,

$$p_\alpha = \frac{1}{(\ell_n - 1)(\ell_n - 3)}. \tag{3.4}$$

This gives us that

$$\lambda = \sum_{i \in [n]} \sum_{1 \le s \le t \le d_i} \frac{1}{\ell_n - 1} + \sum_{1 \le i < j \le n} \sum_{1 \le s_1 \le s_2 \le d_i} \sum_{1 \le t_1 \ne t_2 \le d_j} \frac{1}{(\ell_n - 1)(\ell_n - 3)}$$

$$= \frac{1}{2(\ell_n - 1)} \sum_{i \in [n]} d_i(d_i - 1) + \frac{1}{4(\ell_n - 1)(\ell_n - 3)} \sum_{i \ne j \in [n]} d_i(d_i - 1) d_j(d_j - 1)$$

$$= \left[\nu_n/2 + \nu_n^2/4 \right] (1 + o(1)). \tag{3.5}$$

Further,

$$\sum_\alpha p_\alpha^2 = \left[\frac{\nu_n}{2\ell_n} + \frac{\nu_n^2}{4\ell_n^2} \right] (1 + o(1)) = o(1) \tag{3.6}$$

as long as $d_{\max} = o(n)$. This bounds the easier term $\sum p_\alpha^2$ on the right-hand side of (2.6). We now turn to the more involved contribution in the right-hand side of (2.6), for which the task is to give a convenient and efficient description of the distribution of $(J_{\beta\alpha})_\beta = ((I_\beta)_{\beta \in \Lambda} \mid I_\alpha = 1)$. This means that we need to study the distribution of L_α *conditionally* on $L_\beta = 1$. More precisely, below we define a coupling of the L_α and the conditioned L_α for each β.

Before describing the coupling in the different cases that can arise, we make the following observation. For some pairs α, β it is the case that $L_\alpha = 1$ and $L_\beta = 1$ are incompatible. This happens whenever these events require the same half-edge to be matched in two different ways. In that case, conditioning on $L_\alpha = 1$ makes $L_\beta = 0$. In all other cases, it is easy to see that L_α and L_β are positively correlated. For example, if both α, β are self-loop events, then $\mathbb{E}[L_\alpha L_\beta] = \frac{1}{(\ell_n-1)(\ell_n-3)} > \frac{1}{(\ell_n-1)^2}$. Similar inequalities hold when one or both of the two are multiple edge events. There are also pairs α, β which include a common matched pair, in which case the correlation is very strong. In light of this positive correlation, for any compatible pair α, β, the conditioned L_β stochastically dominates the unconditioned L_β. Our coupling realizes this stochastic domination: If α, β are compatible, then forcing L_α can only increase L_β.

Since α and β can be of two distinct types, corresponding to self-loops and multiple edges, this gives rise to four different cases. We start with the conditional law of $L_{s't'}$ conditionally on $L_{st} = 1$:

(a) *Conditional law of $L_{s't'}$ conditionally on $L_{st} = 1$.*

To create the conditional law of $(J_{\beta\alpha})$, which is the same as the joint law of (L_β) given $L_\alpha = 1$ with $\alpha = st$, we start with $CM_n(\mathbf{d})$, giving us the unconditional law of (L_β). When the half-edges s and t have been paired to one another, we do nothing, because then $L_{st} = 1$ already. When $L_{st} = 0$, we break open the two edges containing s and t respectively, pair s and t, and pair the two other half-edges that are now unpaired to each other. We refer to this as *rewiring to create α*. It is clear that the resulting graph is $CM_n(\mathbf{d})$ conditioned on half-edges s, t being matched, and thus produces the required distribution $(J_{\beta\alpha})$, and also couples it with the unconditioned law (L_β). We now compute $\mathbb{E}[|L_\beta - J_{\beta\alpha}|] = \mathbb{P}(L_\beta \ne J_{\beta\alpha})$ (and in this first case we assume that $\beta = s't'$ corresponds to a self-loop).

We note that $J_{\beta\alpha} = L_\beta$, unless the self-loop β is present and is destroyed, or the self-loop β is absent and is created. We have two different cases depending on whether α and β are incident to two distinct vertices or they are incident to the same vertex. We will now examine the contributions to each of these two cases.

Case (a1): *The self-loops α and β are incident to the distinct vertices i and j*: We start with the case where $\alpha = st$ and $\beta = s't'$ are incident to two distinct vertices i and j. We first note that rewiring can never destroy the self-loop β, since the half-edges s' and t' that are incident to the vertex j can only be affected if before the rewiring they were paired to the half-edges s and t that are incident to the vertex i. From this fact it also follows that β is created exactly if before the rewiring, the half-edges s' and t' in β are paired (in some order) to the half-edges s and t in α. This has
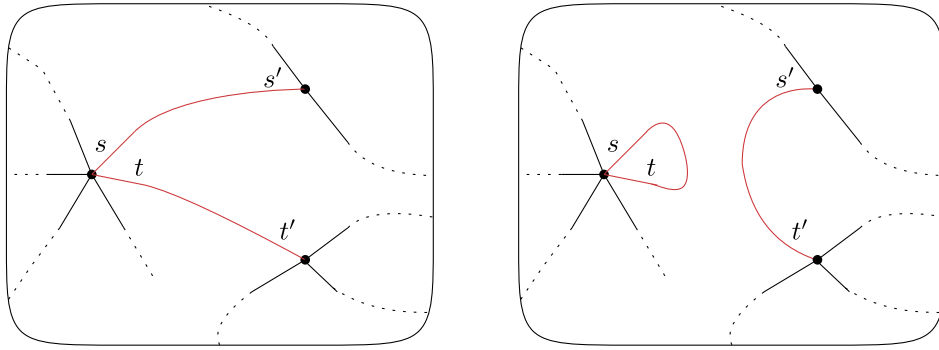
Fig. 1. Rewiring the left configuration so as to create the self-loop with half-edges $\{s, t\}$ results in creation of the edge $\{s', t'\}$. If a second edge $\{s'_2, t'_2\}$ is already present (not shown) then a multiple edge will be formed.

probability $\frac{2}{(\ell_n - 1)(\ell_n - 3)}$. Note that for two distinct vertices i and j, there are $\binom{d_i}{2}\binom{d_j}{2}$ choices for the pair (s, t) incident to i and the pair (s', t') incident to j.

Case (a2): *The self-loops α and β are incident to the same vertex i, and are disjoint*: We now consider the case where $\alpha = st$ and $\beta = s't'$ are incident to the same vertex i, and do not share any half-edge, so that $\{s, t\} \cap \{s', t'\} = \varnothing$. As in case (a1), rewiring cannot destroy the self-loop $s't'$, and creates the self-loop $s't'$ precisely when s, t are paired to $\{s', t'\}$ in some order. This occurs with probability $\frac{2}{(\ell_n - 1)(\ell_n - 3)}$. Note that there are $6\binom{d_i}{4}$ choices for half-edges s, t and half-edges s', t' incident to the vertex i.

Case (a3): *The self-loops α and β are incident to the same vertex i, and overlap*: Consider finally the case of self loops with $\{s, t\} \cap \{s', t'\} \neq \varnothing$. If $s't'$ is a self-loop before rewiring, it is destroyed by the rewiring. This occurs with probability $\frac{1}{(\ell_n - 1)}$. Note that there are $6\binom{d_i}{3}$ choices for the pairs $\{s, t\}$ and $\{s', t'\}$ with an overlap.

Recall the notation $(m)_k = m(m - 1) \cdots (m - k + 1)$. Using that $p_\alpha = 1/(\ell_n - 1)$, the total contribution from cases (a1)–(a3) to the second sum in (2.6) is thus equal to

$$\sum_{i \neq j \in [n]} \frac{(d_i)_2 (d_j)_2}{4} \cdot \frac{2}{(\ell_n - 1)^2 (\ell_n - 3)} + \sum_{i \in [n]} \frac{(d_i)_4}{4} \cdot \frac{2}{(\ell_n - 1)^2 (\ell_n - 3)} + \sum_{i \in [n]} (d_i)_3 \cdot \frac{1}{(\ell_n - 1)^2}$$

$$= O(v_n^2 / \ell_n) + O(\mu_n^{(3)} / \ell_n), \tag{3.7}$$

where the first sum corresponds to the total contribution for the case when α and β are incident to two distinct vertices i and j, and the last two sums correspond to the total contribution for the case when α and β are incident to the same vertex i.

(b) *Conditional law of $L_{s'_1 t'_1, s'_2 t'_2}$ conditionally on $L_{st} = 1$*

Continuing our analysis of the above coupling, we now consider $\beta = \{s'_1 t'_1, s'_2 t'_2\}$ corresponding to a pair of parallel edges. We have different cases depending on whether the half-edges in α and β are incident to three distinct vertices or only two different vertices. We will now examine the contributions to each of these two cases.

Case (b1): *The self-loop α and the multiple-edge β are incident to three vertices*: We start with the case where $\alpha = st$ is incident to a vertex i, and $\beta = \{s'_1 t'_1, s'_2 t'_2\}$ is incident to two other vertices, so that s'_1, s'_2 are incident to vertex j and the pair t'_1, t'_2 are incident to vertex k, with $\{i, j, k\}$ all distinct. Note that (as in case (a1)), rewiring cannot destroy the multiple edge β, since β is only affected if there is some half-edge in β that is paired to some half-edge in α before the rewiring, in that case β could not have been present in $CM_n(\boldsymbol{d})$. However, rewiring can create the multiple edge β. This occurs when before the rewiring either the edge $s'_1 t'_1$ or the edge $s'_2 t'_2$ already existed and the half-edges s and t are paired to the two remaining half-edges from β; see Figure 1 for an illustration.

We thus have four symmetric cases: One of these cases is when s was paired to s'_1 and t to t'_1, while s'_2 was paired to t'_2. Thus, the total probability for these four symmetric cases is $\frac{4}{(\ell_n - 1)(\ell_n - 3)(\ell_n - 5)}$. Note that there are $\binom{d_i}{2}$ choices

for the pair of half-edges (s, t) in the vertex i and then $2\binom{d_j}{2}\binom{d_k}{2}$ to choose the multiple edge $\beta = \{s_1' t_1', s_2' t_2'\}$ between vertices j and k. When we sum over all vertices i, j and k, we could either assume that $j < k$ or divide the total sum by 2, since we can permute j and k. In total, using that $p_\alpha = 1/(\ell_n - 1)$, the contribution from Case (b1) to the second sum in (2.6) is

$$\sum_{i \neq j \neq k} \frac{(d_i)_2 (d_j)_2 (d_k)_2}{8} \cdot \frac{4}{(\ell_n - 1)^2 (\ell_n - 3)(\ell_n - 5)} = \frac{6(d_i)_2 (d_j)_2 (d_k)_2}{2(\ell_n - 1)^2 (\ell_n - 3)(\ell_n - 5)}. \tag{3.8}$$

Here, the 6 in the first term comes from the possible orders of i, j, k.

Case (b2): *The self-loop α and the multiple-edge β are incident to two vertices*: We now consider the case when $\alpha = st$ and $\beta = \{s_1' t_1', s_2' t_2'\}$ are incident to only two distinct vertices. Specifically, we assume that both the half-edges s, t and s_1', s_2' are all incident to the vertex i, while the half-edges t_1', t_2' are incident to a different vertex j. This is split into three sub-cases, depending on whether $\{s, t\}$ and $\{s_1', s_2'\}$ have zero, one, or two elements in common.

If $\{s_1', s_2'\} \cap \{s, t\} = \varnothing$ then we cannot destroy β since no half-edge in β could have been paired to s or t before the rewiring. However, we can create β. This again occurs when before the rewiring either the edge $s_1' t_1'$ or the edge $s_2' t_2'$ already existed and the half-edges s and t are paired to the two remaining half-edges in β. The total probability for these four symmetric cases is the same as in case (b1): $\frac{4}{(\ell_n-1)(\ell_n-3)(\ell_n-5)}$. Note that there are $2\binom{d_i}{2}\binom{d_j}{2}\binom{d_i - 2}{2}$ choices for the multiple edge β incident to the vertices i and j and the self-loop α incident to vertex i.

If $\{s_1', s_2'\} \cap \{s, t\} \neq \varnothing$ then rewiring cannot create β, since one of the half-edges in β must be part of the self-loop α after the rewiring. In case $\{s_1', s_2', s, t\}$ are three distinct half-edges so that $s = s_1'$ or $s = s_2'$, then we destroy β if β existed before the rewiring (while the final half-edge t incident to vertex i was paired to an arbitrary half-edge). These two symmetric cases thus have probability $\frac{2}{(\ell_n-1)(\ell_n-3)}$, and there are $2\binom{d_i}{2}\binom{d_j}{2}(d_i - 2)$ choices for the multiple edge β incident to the vertices i and j and the remaining half-edge t in the self-loop α incident to vertex i.

Finally we consider the case when $s = s_1'$ and $t = s_2'$. Then we destroy β if β existed before the rewiring, which occurs with probability $\frac{1}{(\ell_n-1)(\ell_n-3)}$. Note that there are $2\binom{d_i}{2}\binom{d_j}{2}$ choices for the multiple edge β incident to the vertices i and j and then the self-loop α incident to vertex i is also decided from that choice.

In total, again using $p_\alpha = 1/(\ell_n - 1)$, the contribution from Cases (b1) and (b2) to the second sum in (2.6) is

$$\sum_{i < j < k \in [n]} \frac{6(d_i)_2 (d_j)_2 (d_k)_2}{2(\ell_n - 1)^2 (\ell_n - 3)(\ell_n - 5)} + \sum_{i \neq j \in [n]} \frac{(d_i)_4 (d_j)_2}{(\ell_n - 1)^2 (\ell_n - 3)(\ell_n - 5)} + \sum_{i \neq j \in [n]} \frac{2(d_i)_3 (d_j)_2 + (d_i)_2 (d_j)_2}{2(\ell_n - 1)^2 (\ell_n - 3)}$$

$$= O\big(v_n^3 / \ell_n\big) + O\big(\mu_n^{(4)} v_n / \ell_n^2\big) + O\big(\mu_n^{(3)} v_n / \ell_n\big). \tag{3.9}$$

Note that in this paper we do not consider the joint distribution of S_n and M_n when $v_n \to \infty$, so this term will only be used for $v_n = O(1)$.

(c) *Conditional law of $L_{s't'}$ conditionally on $L_{s_1 t_1, s_2 t_2} = 1$*

To deal with Case (c), we rely on symmetry that is present in our setting. The simple observation in the lemma below is described in [3, p.25], but we prove it for completeness.

Lemma 3.1 (Symmetry). *With the notation in Section* 2.2,

$$p_\alpha \mathbb{E}[I_\beta - J_{\beta\alpha}] = p_\beta \mathbb{E}[I_\alpha - J_{\alpha\beta}] = -\mathrm{Cov}(I_\alpha, I_\beta).$$

Proof. We have that $p_\alpha \mathbb{E}[I_\beta] = p_\alpha p_\beta$, and

$$p_\alpha \mathbb{E}[J_{\beta\alpha}] = \mathbb{P}(I_\alpha = I_\beta = 1) = \mathbb{E}[I_\alpha I_\beta].$$

Thus, the difference is the covariance (multiplied by -1) and is invariant to swapping α and β. □

In our setting, for compatible α and β the difference $I_\beta - J_{\beta\alpha}$ is never positive, while for incompatible α, β it is never negative. Thus,

$$p_\alpha \big| \mathbb{E}[I_\beta - J_{\beta\alpha}] \big| = p_\alpha \mathbb{E}\big[|I_\beta - J_{\beta\alpha}| \big].$$

We conclude that $p_\alpha \mathbb{E}[|I_\beta - J_{\beta\alpha}|]$ is also invariant to swapping α and β. In particular the sum over self-loops α and multiple edges β is the same as the sum over multiple edges α and self-loops β, and thus the contribution from Case (c) is *equal* to the contribution from Case (b).

(d) *Conditional law of $L_{s_1' t_1', s_2' t_2'}$ conditionally on $L_{s_1 t_1, s_2 t_2} = 1$*

We now turn our coupling to the case where $\alpha = \{s_1 t_1, s_2 t_2\}$ is a pair of parallel edges. We rewire $\mathrm{CM}_n(\boldsymbol{d})$ to create the coupled variables $(J_{\beta\alpha})$, with the joint law of (L_β) given $L_\alpha = 1$. Start with $\mathrm{CM}_n(\boldsymbol{d})$, giving us the unconditioned (L_β). If the pairs of half-edges (s_1, t_1) and (s_2, t_2) are already paired, then $L_{s_1 t_1, s_2 t_2} = 1$ already, and there is nothing to be done.

When $L_{s_1 t_1, s_2 t_2} = 0$, we break open all the edges containing s_1, s_2, t_1, t_2. This leaves these and at most four additional half-edges unmatched. We then pair s_1 to t_1 and pair s_2 to t_2. The additional unmatched half-edges (of which there are zero, two, or four) are paired randomly. This produces $(J_{\beta\alpha})$ with the needed distribution, coupled with the original (L_β). We shall now estimate $\mathbb{E}[|L_\beta - J_{\beta\alpha}|]$. We note that $J_{\beta\alpha} = L_\beta$, unless the multiple-edge β is present and is destroyed, or the multiple-edge β is absent and is created. Note that we have several cases depending on how the multiple-edges α and β intersect, and whether they are incident to two, three or four distinct vertices.

Case (d1): *The multiple edges α and β are incident to four distinct vertices*: We start with the case when α and β are incident to four different vertices, α to i, j, and β to k, l. Note that in this case rewiring cannot destroy the multiple-edge β, since if β existed then the half-edges in β could not have been paired to the half-edges in α before the rewiring. However, rewiring can create β. This can happen in two different ways. In the first way, all half-edges in α are paired to all half-edges in β, which just has the probability

$$P_1 = \frac{4!}{(\ell_n - 1)(\ell_n - 3)(\ell_n - 5)(\ell_n - 7)}. \tag{3.10}$$

The second way β can be created is if one of the edges of β was present before rewiring, while the remaining two half-edges in β were paired to two-half-edges in α (the remaining two half-edges in α can be paired arbitrarily).
This has probability

$$P_2 = \frac{2 \cdot 4 \cdot 3}{(\ell_n - 1)(\ell_n - 3)(\ell_n - 5)}. \tag{3.11}$$

Note that there are

$$4 \binom{d_i}{2} \binom{d_j}{2} \binom{d_k}{2} \binom{d_l}{2}$$

ways to choose $\alpha = \{s_1 t_1, s_2 t_2\}$ incident to vertices i and j, and $\beta = \{s_1' t_1', s_2' t_2'\}$ incident to vertices k and l. (Also note that in both of the cases just described, after the rewiring we could possibly have created β, but in neither of the cases it is certain that β has been created.) See Figure 2 for an illustration of these two possibilities for rewiring edges.

When we sum over all vertices i, j, k and l we have to divide the total sum by 4 similarly as we divided by 2 in the previous Case (b) when there were three vertices that were incident to α and β.

Using that $p_\alpha = \frac{1}{(\ell_n - 1)(\ell_n - 3)}$, with a factor of 24 for permuting i, j, k, l, we find that the total contribution to the second sum in (2.6) due to case (d1) is bounded by

$$24 \sum_{i < j < k < l \in [n]} \frac{(d_i)_2 (d_j)_2 (d_k)_2 (d_l)_2}{16} p_\alpha (P_1 + P_2) = O\big(v_n^4 / \ell_n \big). \tag{3.12}$$

Case (d2): *The multiple edges α and β are incident to three distinct vertices*: We continue with the case when α and β are incident to only three different vertices i, j and i, k. We can assume that s_1, s_2 and s_1', s_2' are incident to vertex
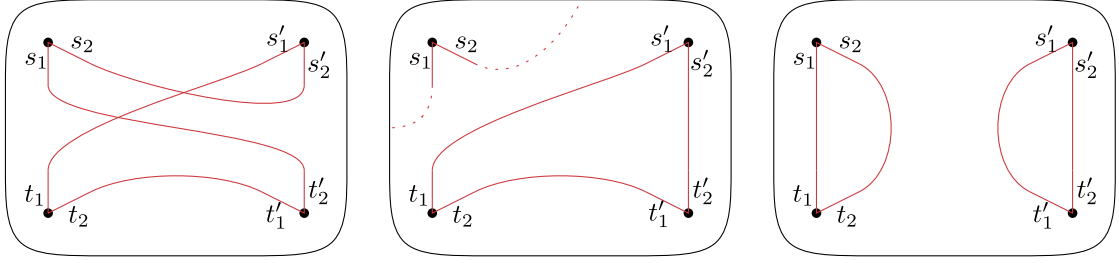
Fig. 2. In the left and centre configurations, rewiring to create the parallel edges $\{s_1 t_1, s_2 t_2\}$ between the two left vertices will create (with positive probability) the parallel edges $\{s_1' t_1', s_2' t_2'\}$ between the right vertices, as shown on the right. Other half-edges on the same vertices are not shown.

i, that t_1, t_2 are incident to j and t_1', t_2' are incident to k. There are sub-cases, according to the number of common half-edges among α and β.

When $\{s_1, s_2\} \cap \{s_1', s_2'\} = \varnothing$, we cannot destroy the multiple-edge β, since we have eight different half-edges in α and β. In this case we can again create β if the half-edges are paired as described in the previous case i.e., with probability P_1 in (3.10) and probability P_2 in (3.11) respectively (there is a possibility that β is created). Note that there are

$$4 \binom{d_i}{2} \binom{d_i - 2}{2} \binom{d_j}{2} \binom{d_k}{2}$$

ways to choose $\alpha = \{s_1 t_1, s_2 t_2\}$ incident to vertices i and j and $\beta = \{s_1' t_1', s_2' t_2'\}$ incident to vertices i and k. If $\{s_1, s_2\} \cap \{s_1', s_2'\} \neq \varnothing$, then we cannot create the multiple edge β incident to the vertices i and k since after the rewiring at least one of the half-edges s_1', s_2' in β is paired to a half-edge in α that is incident to the vertex j. However, when β existed before the rewiring, it is destroyed by the same reason. Hence, we have two possibilities for this to happen i.e., $|\{s_1, s_2\} \cap \{s_1', s_2'\}|$ is equal to 1 or 2. The multiple edge β exists with probability $\frac{1}{(\ell_n - 1)(\ell_n - 3)}$. Note that in the case when α and β contain three distinct half-edges incident to i and two distinct half-edges incident to j and k, respectively, we have

$$8 \binom{d_i}{2} (d_i - 2) \binom{d_j}{2} \binom{d_k}{2}$$

ways to choose α and β, whereas in the case when α and β contain two distinct half-edges incident to i, j and k, respectively, we have

$$4 \binom{d_i}{2} \binom{d_j}{2} \binom{d_k}{2}$$

ways to choose α and β. Using that $p_\alpha = \frac{1}{(\ell_n - 1)(\ell_n - 3)}$, the total contribution to the second sum in (2.6) due to case (d2) is thus bounded by

$$\sum_{i \neq j \neq k \in [n]} \frac{(d_i)_4 (d_j)_2 (d_k)_2}{4} p_\alpha (P_1 + P_2) + \sum_{i \neq j \neq k \in [n]} \frac{2(d_i)_3 (d_j)_2 (d_k)_2 + (d_i)_2 (d_j)_2 (d_k)_2}{(\ell_n - 1)^2 (\ell_n - 3)^2}$$

$$= O\big(\mu_n^{(4)} v_n^2 / \ell_n^2\big) + O\big(\mu_n^{(3)} v_n^2 / \ell_n\big) + O\big(v_n^3 / \ell_n\big). \tag{3.13}$$

Case (d3): *The multiple edges α and β are compatible and incident to two vertices*: We finally consider the case when α and β are incident to two different vertices i and j. In this case rewiring can both create and destroy β when forcing α. In the case when all the eight half-edges are distinct we can again not destroy β. As in case (d1) and (d2) when there were eight distinct half-edges in α and β and as described above there were two different scenarios when β could possibly be created, i.e., the first scenario has probability P_1 in (3.10) and the second scenario has probability

P_2 in (3.11). Note that there are

$$4\binom{d_i}{2}\binom{d_i - 2}{2}\binom{d_j}{2}\binom{d_j - 2}{2} = \frac{(d_i)_4(d_j)_4}{4}$$

ways to choose α and β incident to the vertices i and j. In this case, β is created with probability $1/3$ (as there are three ways of rewiring the loose half-edges).

There is one other case where β can be created by rewiring, namely when α and β share a common edge (e.g. $s_1' = s_1$ and $t_1' = t_1$, while the remaining half-edges are all distinct. In that case, rewiring will create β precisely when (s_2', t_2') are joined prior to rewiring, which has probability $1/(\ell_n - 1)$. The number of pairs α, β in this class is

$$(d_i)_3(d_j)_3.$$

The total contribution from case (d3) is therefore

$$\sum_{i < j \in [n]} \frac{(d_i)_4(d_j)_4}{4} p_\alpha (P_1 + P_2) + (d_i)_3(d_j)_3 p_\alpha \cdot \frac{1}{\ell_n - 1} = O\big((\mu_n^{(4)})^2/\ell_n^3\big) + O\big((\mu_n^{(3)})^2/\ell_n\big). \tag{3.14}$$

Case (d4): *The multiple edges α and β are incompatible and incident to only two vertices*: In all other configurations of α and β involving only two vertices i, j, α and β cannot coexist, and so rewiring destroys β whenever it is present, which occurs with probability $p_\beta = \frac{1}{(\ell_n - 1)(\ell_n - 3)}$.

This can occur in various ways: Using four half-edges from i and two or three from j (or the other way around), or having some overlap in both i and j. Instead of carefully enumerating all the ways this can happen, let us just observe that the number is dominated by $O(d_i^4 d_j^3 + d_i^3 d_j^4)$, since at most three half-edges are chosen at one vertex and at most four at the other. Since i, j may be swapped, the total contribution from this case is at most

$$\sum_{i \neq j \in [n]} C d_i^3 d_j^4 p_\alpha p_\beta = O\big(\mu_n^{(4)} \mu_n^{(3)}/\ell_n^2\big). \tag{3.15}$$

In total, the contribution due to case (d) is thus equal to

$$O\big(\nu_n^4/\ell_n\big) + O\big(\mu_n^{(4)} \nu_n^2/\ell_n^2\big) + O\big(\mu_n^{(3)} \nu_n^2/\ell_n\big) + O\big(\nu_n^3/\ell_n\big) + O\big((\mu_n^{(3)})^2/\ell_n\big) = O(1)\frac{\nu_n^4 + (\mu_n^{(3)})^2}{\ell_n}. \tag{3.16}$$

Here we use that $\nu_n^2 \mu_n^{(3)} \leq \nu_n^4 + (\mu_n^{(3)})^2$ and $\nu_n = O(\ell_n)$. Thus, we note that the largest contributions in case (d) are due to one sum that appears in case (d1) and one sum that appears in case (d3), i.e.,

$$O(1) \sum_{i \neq j \neq k \neq l \in [n]} \frac{(d_i)_2(d_j)_2(d_k)_2(d_l)_2}{(\ell_n - 1)^2(\ell_n - 3)^2(\ell_n - 5)} + O(1) \sum_{i \neq j \in [n]} \frac{(d_i)_3(d_j)_3}{(\ell_n - 1)^2(\ell_n - 3)^2} = O(1)\frac{\nu_n^4 + (\mu_n^{(3)})^2}{\ell_n}. \tag{3.17}$$

4. Proofs of main theorems

4.1. *Proofs for configuration model*

Conclusion to the proof of Theorem 1.1. To conclude the proof of Theorem 1.1, we distinguish between the proofs of (1.10), (1.11) and (1.12), and note that (1.13) is a direct consequence of (1.12). For each of these cases, we need to sum up the corresponding contributions in the above cases (a)–(d).

To prove (1.10), we only need to consider the contribution due to case (a), which is $O(\nu_n^2/\ell_n)$. The contribution due to $\sum_\alpha p_\alpha^2$ equals $O(\nu_n/\ell_n) = O(\nu_n^2/\ell_n)$, while $\lambda_n^S = (\nu_n/2)(1 + O(1/n))$. Thus, Theorem 2.4 gives that

$$\big\|\mathcal{L}(S_n) - \mathsf{Po}\big(\lambda_n^S\big)\big\|_{\mathrm{TV}} \leq \frac{O(1)}{(\nu_n/2 \vee 1)} \frac{\nu_n^2}{\ell_n}, \tag{4.1}$$

which completes the proof of (1.10).

To prove (1.11), we only need to consider the contribution due to case (d), which is $O((\nu_n^4 + (\mu_n^{(3)})^2)/\ell_n)$. The contribution due to $\sum_\alpha p_\alpha^2 = \lambda_n^M/(\ell_n - 1)(\ell_n - 3)$ equals $O(\nu_n^2/\ell_n^2)$. Thus, Theorem 2.4 gives that

$$\big\| \mathcal{L}(M_n) - \mathsf{Po}\big(\lambda_n^M\big) \big\|_{\mathrm{TV}} \leq \frac{O(1)}{(\lambda_n^M \vee 1)} \frac{\nu_n^4 + (\mu_n^{(3)})^2}{\ell_n}, \tag{4.2}$$

which completes the proof of (1.11).

To prove (1.12), we need to consider the contribution due to cases (a)–(d), which is $O((\nu_n^4 + (\mu_n^{(3)})^2)/\ell_n)$. The contribution due to $\sum_\alpha p_\alpha^2 = \lambda_n^S/(\ell_n - 1) + \lambda_n^M/(\ell_n - 1)(\ell_n - 3)$ equals $O(\nu_n/\ell_n)$. Thus, Theorem 2.4 gives that

$$\big\| \mathcal{L}(S_n + M_n) - \mathsf{Po}\big(\lambda_n^S + \lambda_n^M\big) \big\|_{\mathrm{TV}} \leq \frac{O(1)}{((\lambda_n^S + \lambda_n^M) \vee 1)} \frac{\nu_n^4 + (\mu_n^{(3)})^2}{\ell_n}, \tag{4.3}$$

which completes the proof of (1.12), and thus of Theorem 1.1.

Conclusion to the proof of Theorem 1.5. For Theorem 1.5, we use the Poisson approximation for W in (3.2), and rely on Corollary 2.3. Since $\lim_{n\to\infty} \mathbb{E}[D_n^2] = \mathbb{E}[D^2] < \infty$, $d_{\max} = o(\sqrt{n})$, so that $(\mu_n^{(3)})^2/\ell_n \leq d_{\max}^2 \nu_n^2/\ell_n = o(1)$. Thus, $W \xrightarrow{d} W_p$, which has a Poisson distribution with parameter $p\nu/2 + q\nu^2/4$, so that the assumptions in Corollary 2.3 are satisfied. We conclude that $(S_n, M_n) \xrightarrow{d} (S, M)$, where S and M are independent Poisson variables with parameters $\nu/2$ and $\nu^2/4$ respectively. This implies that $\|\mathcal{L}(S_n, M_n) - \mathsf{Po}(\nu/2) \otimes \mathsf{Po}(\nu^2/4)\|_{\mathrm{TV}} \to 0$, since for integer-valued random vectors, the two notions of convergence are equivalent.

Conclusion to the proof of Theorems 1.6–1.7. For Theorem 1.6, we note that $\nu_n \to \nu$ and $\mu_n^{(3)} \to \mu^{(3)} \equiv \mathbb{E}[(D)_3]/\mathbb{E}[D]$ under the assumptions of Theorem 1.6. For Theorem 1.7, we note that $\nu_n \to \nu$ under the assumptions of Theorem 1.7, while $\mu_n^{(3)} \leq d_{\max} \nu_n$.

Conclusion to the proof of Theorems 1.9–1.10. Theorem 1.9 follows from the fact that $\nu_n \to \infty$, so that the bound in (1.10) is $O(\nu_n/\ell_n)$. Since $\mathbb{E}[D_n]$ is bounded it follows from (1.28) that $d_{\max} \leq Cn^{1/(\tau-1)} = o(n)$. Since $\nu_n \leq d_{\max}$ and $d_{\max} = o(n)$ when $\mathbb{E}[D_n] \to \mathbb{E}[D]$, we obtain that $\|\mathcal{L}(S_n) - \mathsf{Po}(\lambda_n^S)\|_{\mathrm{TV}} = o(1)$. Since $\nu_n \to \infty$, by the CLT, $(\mathsf{Po}(\lambda_n^S) - \lambda_n^S)/\sqrt{\lambda_n^S} \xrightarrow{d} Z$. Since $\lambda_n^S = (\nu_n/2)(1 + O(1/n))$, this completes the proof.

The proof of Theorem 1.10 is similar, now using $\lambda_n^M = \Theta(\nu_n^2) \to \infty$, so that

$$\frac{O(1)}{((\lambda_n^S + \lambda_n^M) \vee 1)} \frac{\nu_n^4 + (\mu_n^{(3)})^2}{\ell_n} \leq \frac{\nu_n^2 + (\mu_n^{(3)}/\nu_n)^2}{\ell_n} \leq d_{\max}^2/\ell_n = o(1), \tag{4.4}$$

since we assume that $d_{\max} = o(\sqrt{n})$.

4.2. *Proofs for directed and bipartite configuration models*

Conclusion to the proof of Theorem 1.11. The proof is very similar to the proof of Theorem 1.1. We again distinguish between the proofs of (1.37), (1.38) and (1.39), and note that (1.40) is a direct consequence of (1.39). For each of these cases, we again need to sum up the corresponding contributions (of the couplings) in the above cases (a)–(d), but now for the directed configuration model $\mathrm{DCM}_n(\boldsymbol{d}^{(\mathrm{in})}, \boldsymbol{d}^{(\mathrm{out})})$ (instead of $\mathrm{CM}_n(\boldsymbol{d})$). Below, we abbreviate $\hat{\nu}_n = \hat{\lambda}_n^S$, $\hat{\xi}_n = \hat{\lambda}_n^M$.

To prove (1.37), we only need to consider the contribution due to case (a), which now is $O(\hat{\nu}_n^2/\hat{\ell}_n)$. Again the main contribution is when the self-loops α and β are incident to two distinct vertices i and j, and β is created. Note that $p_\alpha = 1/\hat{\ell}_n$. The first sum in (3.7) (which was the main contribution in the undirected model) now corresponds to

$$\sum_{i \neq j \in [n]} \frac{d_i^{(\mathrm{in})} d_i^{(\mathrm{out})} d_j^{(\mathrm{in})} d_j^{(\mathrm{out})}}{\hat{\ell}_n^2(\hat{\ell}_n - 1)} = O\big(\hat{\nu}_n^2/\hat{\ell}_n\big).$$

The contribution due to $\sum_\alpha p_\alpha^2$ equals $O(\hat{v}_n/\hat{\ell}_n) = O(\hat{v}_n^2/\hat{\ell}_n)$, while $\hat{\lambda}_n^S = \hat{v}_n(1 + O(1/n))$. Thus, Theorem 2.4 gives that

$$\left\| \mathcal{L}(\hat{S}_n) - \mathsf{Po}\big(\hat{\lambda}_n^S\big) \right\|_{\mathrm{TV}} \leq \frac{C}{(\hat{\lambda}_n^S \vee 1)} \frac{\hat{v}_n^2}{\hat{\ell}_n}, \tag{4.5}$$

which completes the proof of (1.37).

To prove (1.38), we only need to consider the contribution due to case (d), which now is $O(\frac{\mu_n^{(3,\mathrm{in})}\mu_n^{(3,\mathrm{out})}+\hat{\xi}_n^2}{\hat{\ell}_n})$. Note that $p_\alpha = \frac{1}{\hat{\ell}_n(\hat{\ell}_n-1)}$. The contribution due to $\sum_\alpha p_\alpha^2 = \hat{\lambda}_n^M/\hat{\ell}_n(\hat{\ell}_n - 1)$ equals $O(\hat{\xi}_n/\hat{\ell}_n^2)$. There are two main contributions corresponding to the two main contributions in the undirected model. The first main contribution corresponds to the case when α and β are incident to four different vertices, and β is created. Then the corresponding sum in (3.12) is now

$$\sum_{i \neq j \neq k \neq l \in [n]} O\big(\frac{(d_i^{(\mathrm{in})})_2(d_j^{(\mathrm{out})})_2(d_k^{(\mathrm{in})})_2(d_l^{(\mathrm{out})})_2}{\hat{\ell}_n^2(\hat{\ell}_n - 1)^2(\hat{\ell}_n - 2)} = O\Big(\frac{\hat{\xi}_n^2}{\hat{\ell}_n}\Big).$$

The second main contribution corresponds to the case when α and β are incident to two vertices i and j, and β is created. Then the corresponding main contribution of the sum in (3.14) is now

$$\sum_{i \neq j \in [n]} O\bigg(\frac{(d_i^{(\mathrm{in})})_3(d_j^{(\mathrm{out})})_3}{\hat{\ell}_n^2(\hat{\ell}_n - 1)}\bigg) = O\bigg(\frac{\mu_n^{(3,\mathrm{in})}\mu_n^{(3,\mathrm{out})}}{\hat{\ell}_n}\bigg).$$

Thus, Theorem 2.4 gives that

$$\left\| \mathcal{L}(\hat{M}_n) - \mathsf{Po}\big(\hat{\lambda}_n^M\big) \right\|_{\mathrm{TV}} \leq \frac{C}{(\hat{\lambda}_n^M \vee 1)} \frac{\mu_n^{(3,\mathrm{in})}\mu_n^{(3,\mathrm{out})} + \hat{\xi}_n^2}{\hat{\ell}_n}, \tag{4.6}$$

which completes the proof of (1.38).

To prove (1.39), we need to consider the contribution due to cases (a)–(d), which now is $O(\frac{\mu_n^{(3,\mathrm{in})}\mu_n^{(3,\mathrm{out})}+\hat{\xi}_n^2}{\hat{\ell}_n})$. The contribution due to $\sum_\alpha p_\alpha^2 = \hat{\lambda}_n^S/\hat{\ell}_n + \hat{\lambda}_n^M/\hat{\ell}_n(\hat{\ell}_n - 1)$ equals $O(\hat{v}_n/\hat{\ell}_n)+O(\hat{\xi}_n/\hat{\ell}_n^2)$. Thus, Theorem 2.4 gives that

$$\left\| \mathcal{L}(\hat{S}_n + \hat{M}_n) - \mathsf{Po}\big(\hat{\lambda}_n^S + \hat{\lambda}_n^M\big) \right\|_{\mathrm{TV}} \leq \frac{C}{((\hat{\lambda}_n^S + \hat{\lambda}_n^M) \vee 1)} \frac{\mu_n^{(3,\mathrm{in})}\mu_n^{(3,\mathrm{out})} + \hat{\xi}_n^2}{\hat{\ell}_n}, \tag{4.7}$$

which completes the proof of (1.39), and thus of Theorem 1.11.

Conclusion to the proof of Theorem 1.12. The proof is again very similar to the proof of Theorem 1.1 and that of Theorem 1.11. However, there are no self-loops in the bipartite configuration model. Thus, we only need to consider case (d) above (regarding the couplings for the multiple edges), but now for the bipartite configuration model $\mathrm{BCM}_n(\boldsymbol{d}^{(\mathrm{l})}, \boldsymbol{d}^{(\mathrm{r})})$. Again there are two main contributions corresponding to the main contributions for the undirected configuration model $\mathrm{CM}_n(\boldsymbol{d})$.

Note that $p_\alpha = \frac{1}{\bar{\ell}_n(\bar{\ell}_n-1)}$. The corresponding sum in (3.12) is now

$$\sum_{i,k \in [n^{(\mathrm{l})}], j,l \in [n^{(\mathrm{r})}]} O\big(\frac{(d_i^{(\mathrm{l})})_2(d_j^{(\mathrm{r})})_2(d_k^{(\mathrm{l})})_2(d_l^{(\mathrm{r})})_2}{\bar{\ell}_n^2(\bar{\ell}_n - 1)^2(\bar{\ell}_n - 2)} = O\bigg(\frac{(\bar{\lambda}_n^M)^2}{\bar{\ell}_n}\bigg).$$

The corresponding main contribution of the sum in (3.14) is now

$$\sum_{i \in [n^{(\mathrm{l})}], j \in [n^{(\mathrm{r})}]} O\bigg(\frac{(d_i^{(\mathrm{l})})_3(d_j^{(\mathrm{r})})_3}{\bar{\ell}_n^2(\bar{\ell}_n - 1)}\bigg) = O\bigg(\frac{\mu_n^{(3,\mathrm{l})}\mu_n^{(3,\mathrm{r})}}{\bar{\ell}_n}\bigg).$$

Thus, Theorem 2.4 gives that

$$\big\| \mathcal{L}(\bar{M}_n) - \mathsf{Po}\big(\bar{\lambda}_n^M\big) \big\|_{\mathrm{TV}} \leq \frac{C}{(\bar{\lambda}_n^M \vee 1)} \frac{\mu_n^{(3,\mathrm{l})} \mu_n^{(3,\mathrm{r})} + (\bar{\lambda}_n^M)^2}{\bar{\ell}_n}, \tag{4.8}$$

which completes the proof of (1.45), and thus of Theorem 1.12.

Acknowledgements

References

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Modern Phys.* **74** (1) (2002) 47–97. MR1895096

[2] N. Alon and J. Spencer. *The Probabilistic Method*, 2nd edition. *Wiley-Interscience Series in Discrete Mathematics and Optimization*. John Wiley & Sons, New York, 2000. MR1885388

[3] A. Barbour, L. Holst and S. Janson. *Poisson Approximation. Oxford Studies in Probability.* **2**. The Clarendon Press Oxford University Press, New York, 1992. MR1163825

[4] E. A. Bender and E. R. Canfield. The asymptotic number of labelled graphs with given degree sequences. *J. Combin. Theory Ser. A* **24** (1978) 296–307. MR0505796

[5] J. Blanchet and A. Stauffer. Characterizing optimal sampling of binary contingency tables via the configuration model. *Random Structures Algorithms* **42** (2) (2013) 159–184. MR3019396

[6] B. Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.* **1** (4) (1980) 311–316. MR0595929

[7] B. Bollobás. *Random Graphs*, 2nd edition. *Cambridge Studies in Advanced Mathematics* **73**. Cambridge University Press, Cambridge, 2001. MR1864966

[8] T. Britton, M. Deijfen and A. Martin-Löf. Generating simple random graphs with prescribed degree distribution. *J. Stat. Phys.* **124** (6) (2006) 1377–1397. MR2266448

[9] C. Cooper and A. Frieze. The size of the largest strongly connected component of a random digraph with a given degree sequence. *Combin. Probab. Comput.* **13** (3) (2004) 319–337. MR2056402

[10] P. Gao and N. Wormald. Enumeration of graphs with a heavy-tailed degree sequence. *Adv. Math.* **287** (2016) 412–450. MR3422681

[11] C. Holmgren and S. Janson. Using Stein's method to show Poisson and normal limit laws for fringe subtrees. *Discrete Math. Theor. Comput. Sci.* **BA** (2014) 169–180. MR3525973

[12] C. Holmgren and S. Janson. Limit laws for functions of fringe trees for binary search trees and random recursive trees. *Electron. J. Probab.* **20** (4) (2015) 1–51. MR3311217

[13] S. Janson. The probability that a random multigraph is simple. *Combin. Probab. Comput.* **18** (1–2) (2009) 205–225. MR2497380

[14] S. Janson. The probability that a random multigraph is simple. II. *J. Appl. Probab.* **51A** (2014) 123–137. MR3317354

[15] S. Janson, M. Luczak and P. Windridge. Law of large numbers for the sir epidemic on a random graph with given degrees. *Random Structures Algorithms* **45** (4) (2014) 724–761. MR3275704

[16] B. D. McKay and N. C. Wormald. Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica* **11** (4) (1991) 369–382. MR1137769

[17] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures Algorithms* **6** (2–3) (1995) 161–179. MR1370952

[18] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.* **7** (3) (1998) 295–305. MR1664335

[19] M. E. J. Newman. The structure and function of complex networks. *SIAM Rev.* **45** (2) (2003) 167–256 (electronic). MR2010377

[20] M. E. J. Newman. Random graphs as models of networks. In *Handbook of Graphs and Networks* 35–68. Wiley-VCH, Weinheim, 2003. MR2016118

[21] M. E. J. Newman, S. Strogatz and D. Watts. Random graphs with arbitrary degree distribution and their application. *Phys. Rev. E* **64** (2000) 026118.

[22] R. van der Hofstad. *Random Graphs and Complex Networks*. *Cambridge Series in Statistical and Probabilistic Mathematics* **1**. Cambridge University Press, Cambridge, 2017. MR3617364

[23] R. van der Hofstad and J. Komjáthy. When is a scale-free graph ultra-small?. *Journ. Stat. Phys.* **169** (2017) 223–264. MR3704860

[24] P. van der Hoorn and N. Litvak. Upper bounds for number of removed edges in the erased configuration model. In *Algorithms and Models for the Web Graph* 54–65. *Lecture Notes in Comput. Sci.* **9479**. Springer, Cham, 2015. MR3500668

[25] N. C. Wormald. The asymptotic distribution of short cycles in random regular graphs. *J. Combin. Theory Ser. B* **31** (2) (1981) 168–182. MR0630980