J. Appl. Prob. **56,** 590–601 (2019)
*doi:*10.1017/jpr.2019.34
© *Applied Probability Trust* 2019

ON THE DENSE PREFERENTIAL ATTACHMENT GRAPH MODELS AND THEIR GRAPHON INDUCED COUNTERPART

ÁGNES BACKHAUSZ,* *Eötvös Loránd University and MTA Alfréd Rényi Institute of Mathematics*
DÁVID KUNSZENTI-KOVÁCS,** *MTA Alfréd Rényi Institute of Mathematics*

Abstract

Letting $\mathcal{M}$ denote the space of finite measures on $\mathbb{N}$, and $\mu_\lambda \in \mathcal{M}$ denote the Poisson distribution with parameter λ, the function $W\colon [0,1]^2 \to \mathcal{M}$ given by $W(x,y) = \mu_{c\log x \log y}$ is called the PAG graphon with density c. It is known that this is the limit, in the multigraph homomorphism sense, of the dense preferential attachment graph (PAG) model with edge density c. This graphon can then in turn be used to generate the so-called W-random graphs in a natural way, and similar constructions also work in the slightly more general context of the so-called PAG_κ models. The aim of this paper is to compare these dense PAG_κ models with the W-random graph models obtained from the corresponding graphons. Motivated by the multigraph limit theory, we investigate the expected jumble-norm distance of the two models in terms of the number of vertices n. We present a coupling for which the expectation can be bounded from above by $\mathrm{O}(\log^{3/2} n \cdot n^{-1/2})$, and provide a universal lower bound that is coupling-independent, but without the logarithmic term.

Keywords: Dense graph limit; Pólya urn process; cut norm; jumble norm

2010 Mathematics Subject Classification: Primary 05C80

1. Introduction

Preferential attachment graphs (PAGs) form a group of random growing graph models that have been studied for a long time ([2], [9], and [12]). The main motivation is modelling randomly evolving large real-world networks, such as online and offline social networks, the internet, or biological networks (e.g. protein–protein interactions). The basic PAG models have been extended by various features, for example duplication steps, weighted edges, vertices with random fitness. The study of this wide family of models provided information about several phenomena in real-world networks (asymptotic degree distribution, clustering, relation of local and global properties, epidemic spread). The limiting behaviour of PAG models has also been investigated from various points of view, depending somewhat on the edge density along the graph sequences. For instance, in [3], Berger, Borgs, Chayes, and Saberi consider a sparse version of the process, with a linear number of edges compared to the number of vertices, and prove convergence in the sense of Benjamini and Schramm to a Pólya point graph. A variation with added randomness is considered by Elwes in [10] and [11], where the preferential attachment model is amended in such a way that the number of edges added at each stage itself is a random variable, but in expectation still preserves linear growth. The limit

Received 27 January 2017; revision received 27 February 2019.
* Postal address: Eötvös Loránd University, Pázmány Péter Sétány 1/c, 1117 Budapest, Hungary.
** Postal address: MTA Alfréd Rényi Institute of Mathematics, PO Box 127, 1364 Budapest, Hungary.

https://doi.org/10.1017/jpr.2019.34 Published online by Cambridge University Press

here is the infinite Rado graph, or a multigraph variant of the same, depending on whether multiple edges are allowed during the process.

At the dense end of the spectrum, when the edge density along the sequence is essentially constant c (i.e. the number of edges is approximately $cn^2/2$), the PAG model corresponds to a Pólya urn model where each urn starts with one ball in it. If, instead of speaking of the number of balls in each urn, we speak of the 'weight' of an urn, it becomes natural to consider an initial state with each weight being equal to a parameter $\kappa \in (0, \infty)$ (restricting to equal weights preserves symmetry and exchangeability), giving rise to the so-called PAG_κ model (see [17]).

Borgs, Chayes, Lovász, Sós, and Vesztergombi [6] considered the model PAG_1 under the convergence notion of injective graph densities. They showed that with probability 1 the graph sequence converges to the graphon $W\colon [0, 1]^2 \to \mathbb{R}$ given by $W(x, y) = c \ln x \ln y$. Later, Ráth and Szakács considered in [17] convergence of a more general family of processes with respect to induced multigraph densities, showing that the limit object of the PAG_κ model is a graphon that now takes Poisson distributions as values instead. Hence the corresponding graphs (generated randomly from this limit object) contain multiple edges with no uniform upper bound on the multiplicity, and the original notions for limits of simple graphs cannot be used any longer. The paper [14] by Kunszenti-Kovács, Lovász, and Szegedy provides a framework for handling homomorphism densities in the context of multigraphs, and makes use of the so-called *jumble norm* to measure distance between graphons.

All of the papers [6], [14], and [17] also deal with W-random graph sequences induced by the limit objects W, and show that with probability 1, the resulting graph sequence converges to W in the respective densities sense. These W-random graph models are thus very similar to the classical graph sequences that gave rise to the limit W, but also exhibit some significant differences.

The specific models we work with are the following. First, we fix two constants $c, \kappa > 0$.

Definition 1. We choose independent random variables ξ_i with gamma(κ, κ) distribution for every $1 \le i \le n$. For $i < j$, let Y_{ij} be a Poisson random variable with parameter $c\xi_i\xi_j$. For every i, let Y_{ii} be a Poisson random variable with parameter $c\xi_i^2/2$. Assume that all Y_{ij} are conditionally independent with respect to the ξ_i. We put Y_{ij} edges between vertices i and j for every $1 \le i \le j \le n$. We obtain the random multigraph $\mathbb{G}_W^\circ(n)$ in this way.

If, compared to $\mathbb{G}_W^\circ(n)$, we do not add the loops (i.e. we set $Y_{ii} \equiv 0$ for all i), we obtain the random multigraph $\mathbb{G}_W(n)$.

Remark. Note that using gamma variables instead of the uniform [0, 1]-valued ones is compensated by the loss of F^{-1} in the parameter.

These are the random models we wish to compare to the following version of the PAG model.

Definition 2. Given vertices $\{1, 2, \dots, n\}$, we assign an urn to each vertex, initially with one single ball in each of them. Then we run a linear Pólya urn process for $2\lfloor cn^2/2 \rfloor$ steps. That is, at each step we choose an urn with probabilities proportional to κ plus the actual number of balls, and put an extra ball into it. More precisely, for $t = 1, 2, \dots, 2\lfloor cn^2/2 \rfloor$, at step t, an urn containing d balls is chosen with probability

$$\frac{d + \kappa}{n + t - 1 + n\kappa}.$$

Then we put a new ball into this urn (each random choice is conditionally independent of the previous steps, given the actual distribution of the balls). Finally, for $k = 1, 2, \dots, \lfloor cn^2/2 \rfloor$,

https://doi.org/10.1017/jpr.2019.34 Published online by Cambridge University Press

we add an edge between the vertices where the balls at step $t = 2k - 1$ and at step $t = 2k$ have been placed. We obtain a random multigraph $\mathbb{G}_{\mathrm{PAG}}(n)$ in this way; multiple edges and loops may occur.

To compare these models, we shall use the following distance notion between multigraphs. It may be defined more generally for graphons (which essentially are weighted graphs with vertex set [0, 1]), but that will not be needed here, and we refer to [14] for more details.

Definition 3. Let G and H be two (multi-)graphs on the same vertex set $[n] := \{1, \dots, n\}$ for some positive integer n. Then we define their *jumble-norm distance* as

$$d_{\boxtimes}(G, H) = \frac{1}{n} \cdot \max_{S,T \subseteq [n]} \frac{1}{\sqrt{|S| \cdot |T|}} \left| \sum_{i \in S, j \in T} U_{ij} - V_{ij} \right|,$$

where U_{ij} and V_{ij} denote the multiplicity of edge ij in G and H, respectively.

Our goal in this paper is to show that with probability 1 they are close (but not too close) in the jumble distance. This result can be interpreted as an upper bound on the speed of convergence of the PAG_κ model, in the sense that it is compared to a finite version of the limit object, and not to the graphon itself. Specifically, we have the following main result.

Theorem 1. *Given κ and c, there exist $K > 0$ and a coupling of the preferential attachment model and the W-random graph on n vertices such that for every $n \geq 1$ we have*

$$\mathbb{E}(d_{\boxtimes}(\mathbb{G}_{\mathrm{PAG}}(n), \mathbb{G}_W(n))) \leq K \cdot \log^{3/2} n \cdot n^{-1/2}.$$

The idea of the proof of the main result is to define a coupling based on a Poisson process, which connects the W-random graph and the PAG_κ model (see Section 3). The proofs can be found in Section 4. In the discussion part (Section 5), we point out some features of the W-random version that can make it more useful in certain applications.

2. Motivation

It was proved in [14] that with probability 1, the random graph $\mathbb{G}^\circ_W(n)$ converges to the original function W with respect to multigraph homomorphism densities. As mentioned, this is also the limit object of the sequence of the preferential attachment random graphs $\mathbb{G}_{\mathrm{PAG}}(n)$ with parameter κ on n vertices with $\lfloor cn^2 \rfloor$ edges. Given that letting n go to infinity, the two random sequences $\mathbb{G}_{\mathrm{PAG}}(n)$ and $\mathbb{G}^\circ_W(n)$ tend to the same limit, it is natural to ask how close these two sequences are as a function of n.

In this section we shall go into more detail about the specific choices made concerning the objects and properties studied. For limits of simple graphs, or even graphs with a uniform bound on the edge multiplicity, the standard distance notion used is that of the cut norm. The cut-norm distance $d_\square$ used in many other papers (see e.g. [6] for details) differs from the jumble-norm distance $d_{\boxtimes}$ in the factor $1/\sqrt{|S| \cdot |T|}$ that is omitted in the former. As such, our current distance notion magnifies the differences that occur on small sets, and we clearly have $d_{\boxtimes} \geq d_\square$. The jumble-norm distance turns out to be a more useful tool when introducing a metric for generalized graphons, in part due to the fact that it can be considered as an L^2-version of the cut-norm distance (since $\sqrt{|S| \cdot |T|}$ corresponds to the L^2-norm of the characteristic function of the set $S \times T$), leading to simpler expressions for bounds (cf. [5] and [14]).

Next, we turn our attention to Definition 1, and our choice of distribution for the variables ξ_i. Whilst the first parameter κ does very much influence the model, it is not clear why the second parameter has to be identical. Indeed it is not necessary, but we shall explain why this particular choice was made, and why no loss of generality incurs.

https://doi.org/10.1017/jpr.2019.34 Published online by Cambridge University Press

Let $\mathcal{M}$ denote the space of finite measures on $\mathbb{N}$, and let $W\colon [0,1]^2 \to \mathcal{M}$ be the function given by

$$W(x,y) = \mu_{cF^{-1}(x)F^{-1}(y)},$$

where

$$F(z) = \int_0^z \frac{1}{\Gamma(\kappa)} \kappa^\kappa t^{\kappa-1}\, \mathrm{e}^{-\kappa t}\, \mathrm{d}t$$

is the CDF of gamma(κ,κ) and μ_λ denotes the Poisson distribution with parameter λ.

This is the graphon that represents the limit of the $\mathbb{G}_{\mathrm{PAG}}(n)$ sequence, and is well-defined fixing κ and c. Now the natural way of generating the so-called W-random graph model from it is as follows. We consider the set $[0,1]$ as the vertex set of the infinite graph with 'adjacency function' W, and sample a random spanned subgraph on n vertices by choosing its vertices independently uniformly from $[0,1]$. After this first randomization, we obtain a 'graph' on n vertices where each 'edge' is a Poisson distribution (or, equivalently, each pair of not necessarily different vertices is labelled with the appropriate Poisson distribution). To obtain a true multigraph, we then independently sample an edge multiplicity for each pair of different vertices from the corresponding Poisson distribution. As for pairs of identical vertices, if we allow loops, for technical reasons we first halve the parameter of the Poisson distribution, then double the sampled value (in this way the graphon takes a simpler form; typically changing W on a set of measure zero does not influence densities if the test-graphs are loopless). This will correspond to the random graph $\mathbb{G}^\circ_W(n)$, whereas if loops are disallowed, we obtain the random graph $\mathbb{G}_W(n)$.

Note that changing the scale parameter for the ξ_i will not change anything: as the values of $W(x,y)$ do not change, it would merely render the conversion from W to the model in Definition 1 more complicated, hence our choice of keeping the two parameters equal. As an added bonus, the distributions $\Gamma(\kappa,\kappa)$ all have expected value 1, simplifying the computations (and this is certainly also the reason for the specific choice of Γ distributions used in the formula of W in [17]).

In the last section, we provide a universal, coupling-independent lower bound of $\Theta(n^{-1/2})$ for the same expectation. The exponents match, but the logarithmic term still leaves room for improvement.

3. Coupling

In order to prove Theorem 1, we need to construct a particular coupling for which the distance of $\mathbb{G}_{\mathrm{PAG}}$ and $\mathbb{G}_{\mathrm{W}}$ is smaller than the upper bound. We build a coupling of appropriate realizations of $\mathbb{G}_{\mathrm{PAG}}$ and $\mathbb{G}^\circ_{\mathrm{W}}$, then we find a realization of $\mathbb{G}_{\mathrm{W}}$ in the same probability space.

We will use the following well-known fact (which can also be considered as an application of de Finetti's theorem on exchangeable random variables). Let $(\xi_i)_{i=1}^n$ be independent gamma(κ,κ)-distributed random variables, as in Definition 1. We consider N random variables that are independent and identically distributed conditionally with respect to $(\xi_i)_{i=1}^n$, and each of them takes value i with probability

$$\rho_i := \frac{\xi_i}{\sum_{k=1}^n \xi_k}$$

for $i = 1, 2, \ldots, n$. Then we get the joint distribution of the sequence of the first N draws of the n-colour Pólya urn. Indeed, these N random variables have a Dirichlet joint distribution (see e.g. [8, Theorem XI.4.1.]), and using the moments of Dirichlet distributions, it can be seen that this distribution is the same as the joint distribution of the first N draws (see [4], [7], or [18, Section 2]).

https://doi.org/10.1017/jpr.2019.34 Published online by Cambridge University Press

We construct the coupling based on a sequence of independent random variables $(\xi_i)_{i=1}^n$ with distribution gamma(κ, κ). Conditionally with respect to $(\xi_i)_{i=1}^n$, we assign independent Poisson processes to each pair of vertices. For $1 \le i < j \le n$, the rate of the process is $2\rho_i\rho_j$ for the pair (i, j) (with ρ_i defined above), and for $1 \le i \le n$, the rate is ${\rho_i}^2$ for (i, i). We denote by $N_s^{(ij)}$ the number of events until time s in the process assigned to (i, j). The cumulation of these processes is also a Poisson process. Let τ be the time when the total number of events reaches $\lfloor cn^2/2 \rfloor$. We get a realization of $\mathbb{G}_{\mathrm{PAG}}(n)$ if we put $N_\tau^{(ij)}$ edges between i and j for all $1 \le i \le j \le n$. Indeed, all $\lfloor cn^2/2 \rfloor$ events are distributed among the pairs of vertices independently, with probabilities proportional to the rates $2\rho_i\rho_j$ and ρ_i^2. Thus this is the same as the joint distribution of consecutive pairs of $2\lfloor cn^2/2 \rfloor$ independent elements chosen according to $(\rho_i)_{i=1}^n$, which provides a Pólya urn by the argument above. On the other hand, we get a realization of $\mathbb{G}_W^\circ(n)$ if we put $N^{(ij)}_{c(\sum_{k=1}^n \xi_k)^2/2}$ edges between i and j, because the number of edges between the pairs of vertices are independent Poisson random variables with the appropriate parameters:

$$\frac{1}{2}c\left(\sum_{k=1}^n \xi_k\right)^2 \cdot 2\rho_i\rho_j = c\xi_i\xi_j \quad \text{and} \quad \frac{1}{2}c\left(\sum_{k=1}^n \xi_k\right)^2 \cdot \rho_i^2 = c\xi_i^2/2.$$

Hence this provides a coupling of the two models.

Proposition 1. *In the coupling above, for some $K_{1,2} > 0$ we have*

$$\mathbb{E}(d_{\boxtimes}(\mathbb{G}_{\mathrm{PAG}}(n), \mathbb{G}_W^\circ(n))) \le K_{1,2} \cdot n^{-1/2} \cdot \log^{3/2} n, \qquad n = 1, 2, \ldots.$$

As for $\mathbb{G}_W(n)$, we can use the same probability space as before, generate $G_W^\circ(n)$, and then delete the loops.

Proposition 2. *There exists $K_{2,3} > 0$ such that*

$$\mathbb{E}(d_{\boxtimes}(\mathbb{G}_W^\circ(n), \mathbb{G}_W(n))) \le K_{2,3} \cdot n^{-3/4}, \qquad n = 1, 2, \ldots$$

holds when $\mathbb{G}_W$ is obtained from $\mathbb{G}_W^\circ(n)$ by deleting the loops.

4. Proofs

Proof of Theorem 1. The statement follows from Propositions 1 and 2. □

We shall therefore now turn our attention to proving the bounds connecting each pair of models. Since the jumble-norm distance is not always easy to work with, we shall make use of the following lemma.

Lemma 1. *Let G and H be two (undirected) multigraphs on the vertex set $\{1, 2, \ldots, n\}$. Let U_{ij} be the number of edges between i and j in G, and V_{ij} the same quantity in H. Then the following holds:*

$$d_{\boxtimes}(G, H) = \frac{1}{n} \cdot \max_{S,T} \frac{1}{\sqrt{|S| \cdot |T|}} \left| \sum_{i \in S, j \in T} U_{ij} - V_{ij} \right| \le \frac{1}{n} \cdot \max_{1 \le i \le n} \sum_{j=1}^n |U_{ij} - V_{ij}|.$$

Proof. Let $\sigma_i = \sum_{j=1}^n |U_{ij} - V_{ij}|$. Notice that if $|S| = s$, $|T| = t$, and $s \le t$, then

$$\left| \sum_{i \in S, j \in T} U_{ij} - V_{ij} \right| \le \sum_{i \in S, j \in T} |U_{ij} - V_{ij}| \le \sum_{i \in S} \sigma_i \le s \max_{1 \le i \le n} \sigma_i.$$

https://doi.org/10.1017/jpr.2019.34 Published online by Cambridge University Press

Hence

$$\frac{1}{\sqrt{st}}\left|\sum_{i\in S, j\in T} U_{ij} - V_{ij}\right| \le \frac{s\max_{1\le i\le n}\sigma_i}{\sqrt{st}} = \frac{\sqrt{s}}{\sqrt{t}}\max_{1\le i\le n}\sigma_i \le \max_{1\le i\le n}\sigma_i,$$

as we assumed that $s \le t$. In the reverse case $s \ge t$, we get the same with the bound

$$\max_{1\le j\le n}\sum_{i=1}^{n}|U_{ij} - V_{ij}|.$$

Since $U_{ij} = U_{ji}$ and $V_{ij} = V_{ji}$, this is equal to the previous maximum. This finishes the proof. □

4.1. Models $\mathbb{G}_{\mathbf{PAG}}$ and $\mathbb{G}_W^\circ$

Proof of Proposition 1. Let U_{ij} be the number of edges between i and j in the random graph $\mathbb{G}_{\mathrm{PAG}}$, and let V_{ij} be the number of edges between them in model $\mathbb{G}_W^\circ$. By using the notations introduced for the coupling of the two models, we have $U_{ij} - V_{ij} = N_\tau^{(ij)} - N_\sigma^{(ij)}$, where $\sigma = c(\sum_{k=1}^n \xi_k)^2/2$. If $\tau \ge \sigma$, then all these differences are nonnegative, and all of them are negative if $\tau < \sigma$. Thus

$$\Delta_{n,i} = \sum_{j=1}^{n}|U_{ij} - V_{ij}| = \left|\sum_{j=1}^{n} N_\tau^{(ij)} - N_\sigma^{(ij)}\right|, \qquad 1 \le i \le n.$$

We will use the fact that by cumulating the independent Poisson processes assigned to the pairs of vertices we get a Poisson process with rate $2\sum_{i<j}\rho_i\rho_j + \sum_i \rho_i^2 = 1$. Let N_s be the total number of events until time s, that is, $N_s = \sum_{i\le j} N_s^{(ij)}$, which has Poisson distribution with parameter s. Since there are $\lfloor cn^2/2\rfloor$ events in the cumulated process until τ, there are $|\lfloor cn^2/2\rfloor - N_\sigma|$ events between τ and σ. On the other hand, independently of each other, all these events increase $\Delta_{n,i}$ by 1 with probability $p_i = \rho_i(\rho_i + 2\sum_{j\ne i}\rho_j) \le 2\rho_i$, because we can think of the process as the cumulated Poisson process with the types of the events chosen independently with probabilities proportional to the rates. We conclude that, conditionally with respect with respect to N_σ and $(\xi_i)_{i=1}^n$, the quantity $\Delta_{n,i}$ has binomial distribution with parameters $|\lfloor cn^2/2\rfloor - N_\sigma|$ and p_i.

We will give an upper bound on $\Delta_{n,i}$ on an appropriately chosen event which has large probability. First, since each ξ_i is a random variable with distribution $\mathrm{gamma}(\kappa,\kappa)$, and they are independent, the proportion $\rho_i = \xi_i/(\sum_{k=1}^n \xi_k)$ has distribution $\mathrm{beta}(1, n-1)$. Therefore, for $n \ge 2$ we have

$$\begin{aligned}\mathbb{P}\left(\rho_i > \frac{16}{n}\log n\right) &= \int_{16\log n/n}^{1}(n-1)(1-x)^{n-2}\,\mathrm{d}\,x\\ &= \int_0^{1-16\log n/n}(n-1)x^{n-2}\,\mathrm{d}\,x\\ &= (1-16\log n/n)^{n-1}\\ &\le \exp(-8\log n)\\ &= n^{-8}.\end{aligned}$$

Since $(\xi_k)_{k=1}^n$ are independent and identically distributed random variables satisfying the conditions of Cramér's theorem (or the Chernoff bound), we can choose $a > 0$ such that

$$\mathbb{P}\left(\sum_{k=1}^{n}\xi_k - n > a\cdot n^{1/2}\sqrt{\log n}\right) = \mathrm{O}(n^{-8})$$

https://doi.org/10.1017/jpr.2019.34 Published online by Cambridge University Press

holds. Below, we fix such an a, and define $s_n = \lfloor cn^2/2 \rfloor - 2acn^{3/2}\sqrt{\log n}$ and $t_n = \lfloor cn^2/2 \rfloor + 2acn^{3/2}\sqrt{\log n}$. Then N_{s_n} and N_{t_n} have Poisson distributions with parameters s_n and t_n, respectively. By a standard application of Chernoff's bound for these random variables, we find that there exists $b > 0$ such that $N_{s_n} - s_n < -bn\sqrt{\log n}$ and $N_{t_n} - t_n > bn\sqrt{\log n}$ both have probability $\mathrm{O}(n^{-8})$.

The arguments above imply that we have

$$\mathbb{P}(F_{n,i}) \geq 1 - \mathrm{O}(n^{-8}), \tag{1}$$

where we define

$$F_{n,i} = \left\{ \rho_i \leq \frac{16}{n} \log n \right\} \cap \left\{ \left| \sum_{k=1}^{n} \xi_k - n \right| < a \cdot n^{1/2} \sqrt{\log n} \right\}$$
$$\cap \{ N_{s_n} > s_n - bn\sqrt{\log n} \} \cap \{ N_{t_n} < t_n + bn\sqrt{\log n} \}.$$

Now we will bound $\Delta_{n,i}$ on this event $F_{n,i}$ (for fixed i). Notice that (for sufficiently large n) the second condition in this definition implies that

$$\left| \left(\sum_{k=1}^{n} \xi_k \right)^2 - n^2 \right| \leq an^{1/2}\sqrt{\log n} \cdot (2n + an^{1/2}\sqrt{\log n}),$$

and thus $s_n \leq \sigma = c(\sum_{k=1}^{n} \xi_k)^2/2 \leq t_n$. Therefore, for sufficiently large n, on the event $F_{n,i}$ we have

$$N_\sigma - \lfloor cn^2/2 \rfloor \geq N_{s_n} - \lfloor cn^2/2 \rfloor \geq s_n - bn\sqrt{\log n} - \lfloor cn^2/2 \rfloor \geq -3acn^{3/2}\sqrt{\log n}$$

and

$$N_\sigma - \lfloor cn^2/2 \rfloor \leq N_{t_n} - \lfloor cn^2/2 \rfloor \leq t_n + bn\sqrt{\log n} - \lfloor cn^2/2 \rfloor \leq 3acn^{3/2}\sqrt{\log n},$$

and consequently

$$|N_\sigma - \lfloor cn^2/2 \rfloor| \leq 3acn^{3/2}\sqrt{\log n}.$$

Recall that, conditionally with respect to $(\xi_i)_{i=1}^n$ and N_σ, the random variable $\Delta_{n,i}$ has binomial distribution with parameters $|N_\sigma - \lfloor cn^2/2 \rfloor|$ and p_i, which was at most $2\rho_i \leq 32 \log n/n$ on $F_{n,i}$. Let $\mathcal{F}$ be the σ-algebra generated by $(\xi_i)_{i=1}^n$ and the process N. Then, since $F_{n,i}$ is measurable with respect to $\mathcal{F}$, we have

$$\begin{aligned}
&\mathbb{P}(\{\Delta_{n,i} > 192ac \cdot n^{1/2} \log^{3/2} n\} \cap F_{n,i}) \\
&\quad \leq \mathbb{E}\left(\mathbb{E}\left(\frac{\mathrm{e}^{\Delta_{n,i}}}{\exp(192ac \cdot n^{1/2} \log^{3/2} n)} \,\middle|\, \mathcal{F} \right) \cdot \mathbb{I}_{F_{n,i}} \right) \\
&\quad = \mathbb{E}\left(\frac{(1 + (e-1)p_i)^{|N_\sigma - \lfloor cn^2/2 \rfloor|}}{\exp(192ac \cdot n^{1/2} \log^{3/2} n)} \cdot \mathbb{I}_{F_{n,i}} \right) \\
&\quad \leq \exp((e-1)(32 \log n/n) \cdot 3ac \cdot n^{3/2} \log^{1/2} n - 192ac \cdot n^{1/2} \log^{3/2} n) \\
&\quad = \mathrm{O}(n^{-8}).
\end{aligned}$$

Putting this together with equation (1) we conclude that

$$\mathbb{P}\left(\sum_{j=1}^{n} |U_{ij} - V_{ij}| > 192ac \cdot n^{1/2} \log^{3/2} n \right) = \mathrm{O}(n^{-8})$$

holds for every i with $1 \leq i \leq n$.

https://doi.org/10.1017/jpr.2019.34 Published online by Cambridge University Press

Finally, let A_n be the event that $\sum_{j=1}^{n} |U_{ij} - V_{ij}| > 192ac \cdot n^{1/2} \log^{3/2} n$ holds for some $1 \le i \le n$. This last inequality implies $\mathbb{P}(A_n) = \mathrm{O}(n^{-7})$. By using Lemma 1, the Cauchy–Schwarz inequality and the fact that N_σ has Poisson distribution conditionally with respect to σ, we obtain

$$\begin{aligned}\mathbb{E}(d_{\boxtimes}(\mathbb{G}_{\mathrm{PAG}}(n), \mathbb{G}_W^\circ(n)) &\le \frac{1}{n}\mathbb{E}\left(\mathbb{I}_{\overline{A}_n} \cdot \max_{1\le i\le n} \sum_{j=1}^{n} |U_{ij} - V_{ij}|\right) \\ &\quad + \frac{1}{n}\mathbb{P}(A_n)^{1/2} \cdot \left(\mathbb{E}\left(\left(\sum_{i,j=1}^{n} |U_{ij} - V_{ij}|\right)^2\right)\right)^{1/2} \\ &\le \frac{1}{n} \cdot 192ac \cdot n^{1/2} \log^{3/2} n + \mathrm{O}(n^{-4})(\mathbb{E}(|N_\sigma - \lfloor cn^2/2 \rfloor|^2))^{1/2} \\ &\le 192ac \cdot n^{-1/2} \log^{3/2} n + \mathrm{O}(n^{-4})(2\mathbb{E}(\sigma^2 + \sigma + c^2 n^4))^{1/2}.\end{aligned}$$

This implies the statement if we calculate the moments of $\sigma = c(\sum_{k=1}^{n} \xi_k)^2/2$ by using that the sum has gamma$(n\kappa, \kappa)$ distribution, and hence its rth moment is

$$\frac{\Gamma(n\kappa + r)}{\kappa^r \Gamma(n\kappa)}. \qquad \square$$

4.2. Models $\mathbb{G}_W^\circ$ and $\mathbb{G}_W$

Proof of Propositon 2. We note that $\mathbb{G}_W^\circ(n)$ and $\mathbb{G}_W(n)$ coincide everywhere but the main diagonal, and it is then easy to see that

$$d_{\boxtimes}(\mathbb{G}_W^\circ(n), \mathbb{G}_W(n)) = \frac{1}{n} \cdot \max_{1\le i\le n} Y_{ii}.$$

Recall that Y_{ii} has Poisson distribution with parameter $c\xi_i^2$, where ξ_i has gamma(κ, κ) distribution. Assume first that $\zeta > 0$ is fixed, and $X \sim \mathrm{Pois}(\zeta)$. Then

$$\sum_{k=y+1}^{\infty} \mathbb{P}(X \ge k) \le \sum_{k=y+1}^{\infty} k\mathbb{P}(X = k) = \sum_{k=y+1}^{\infty} \frac{\zeta^k}{(k-1)!} \mathrm{e}^{-\zeta} \le \zeta \sum_{k=z}^{\infty} \frac{\zeta^k}{k!} \mathrm{e}^{-\zeta} = \zeta \mathbb{P}(X \ge y).$$

We will use the factorial moments of the Poisson distribution. We use the notation $x^{(b)} = x(x-1)\cdots(x-b+1)$. For every fixed i and integers $y > b > 0$, we have

$$\begin{aligned}\sum_{k=y+1}^{\infty} \mathbb{P}(Y_{ii} \ge k) &= \mathbb{E}\left(\mathbb{E}\left(\sum_{k=y+1}^{\infty} \mathbb{P}(Y \ge k) \,\middle|\, \xi_i\right)\right) \\ &\le \mathbb{E}(c\xi_i^2 \mathbb{P}(Y^{(b)} \ge y^{(b)} | \xi_i)) \\ &\le \mathbb{E}\left(c\xi_i^2 \frac{\xi_i^{(b)}}{y^{(b)}}\right) \\ &= \frac{K(b)}{y^{(b)}}\end{aligned}$$

for some $K(b) > 0$, because the gamma distribution has finite moments.

https://doi.org/10.1017/jpr.2019.34 Published online by Cambridge University Press

For an arbitrary function $f : \mathbb{N}^+ \to \mathbb{N}^+$ we may apply the above inequality to obtain

$$\mathbb{E}\Big(\max_{1\le i\le n}|Y_{ii}|\Big) = \sum_{k=1}^{\infty}\mathbb{P}\Big(\max_{1\le i\le n}|Y_{ii}|\ge k\Big) \le f(n) + \sum_{k=f(n)+1}^{\infty} n\mathbb{P}(Y_{11}\ge k) \le f(n) + n\frac{K(b)}{f(n)^{(b)}}.$$

Now let $N \in \mathbb{N}^+$ be fixed, set $f(n) := n^{1/5}$ and $b := 4$. For sufficiently large n (such that $n-3 \ge n/2$ and $f(n)^{(4)} \ge f(n)^4/16$), this yields

$$\mathbb{E}\Big(\max_{1\le i\le n}|Y_{ii}|\Big) \le n^{1/5} + 16K(4)n^{1-4/5} \le n^{1/4}(K(4)+1)).$$

Lemma 1 concludes the proof. □

5. Discussion

Our main theorem shows that the classical dense preferential attachment graph model yields random graphs that are close to the random graph model obtained through the PAG_κ-graphon, the limit object in the multigraph homomorphism sense of the random sequence $\mathbb{G}_{\mathrm{PAG}}$. They are not indistinguishable, however (we provide a lower bound on their distance below), and they each have their own advantages for applications.

The random graphs $\mathbb{G}_{\mathrm{PAG}}$ have the advantage that the number of edges is deterministic, but, contrary to the sparse PAG models, one cannot easily generate a growing family of graphs $\mathbb{G}_{\mathrm{PAG}}(n)$. For the graphon induced $\mathbb{G}^\circ_W$, the number of edges is random, though still asymptotically concentrated around the expected value. Also, the way it is generated does not carry the preferential attachment flavour. This may be an advantage from the simulation point of view: the random variables in the model can be generated simultaneously, without the cn^2 steps that have to be performed after each other in the PAG_κ model.

However, it is possible to couple the elements of the sequence $\mathbb{G}^\circ_W(n)$ (or $\mathbb{G}_W(n)$) so that we obtain a growing sequence (and still keep the convergence with probability 1). Indeed, passing from n to $n+1$ only means that we have to generate the random variable ξ_{n+1}, independently of the previous ξ_i, and then generate the appropriate Poisson random variables $Y_{j(n+1)}$ for $1 \le j \le n+1$. This coupling shows that adding an extra vertex and extending $\mathbb{G}^\circ_W(n)$ to $\mathbb{G}^\circ_W(n+1)$ can be performed easily. It seems that this does not hold for the $\mathbb{G}_{\mathrm{PAG}}$ model.

The following proposition provides a lower bound for the jumble distance of the $\mathbb{G}_{\mathrm{PAG}}(n)$ and $\mathbb{G}^\circ_W(n)$ models, which holds for every coupling, and differs from the upper bound of Theorem 1, i.e. $\mathrm{O}(n^{-1/2}\log^{3/2} n)$, only in logarithmic terms.

Proposition 3. *There exists $L_1, L_2 > 0$ such that for every coupling of the random graph models $\mathbb{G}_{\mathrm{PAG}}(n)$ and $\mathbb{G}^\circ_W(n)$ we have*

$$\mathbb{E}(d_{\boxtimes}(\mathbb{G}_{\mathrm{PAG}}(n), \mathbb{G}^\circ_W(n))) \ge L_1 n^{-1/2},$$

and for every coupling of the random graph models $\mathbb{G}_{\mathrm{PAG}}$ and $\mathbb{G}_W(n)$ we have

$$\mathbb{E}(d_{\boxtimes}(\mathbb{G}_{\mathrm{PAG}}(n), \mathbb{G}_W(n))) \ge L_2 n^{-1/2}.$$

Proof. If we take $S = T = \{1, 2, \dots, n\}$ in Definition 3, then we obtain a lower bound for the jumble-norm distance of $\mathbb{G}_{\mathrm{PAG}}$ and $\mathbb{G}_{\mathrm{W}}$ by understanding the difference of the number of edges. The main point is that the distribution of this quantity does not depend on the coupling. In $\mathbb{G}_{\mathrm{PAG}}(n)$, the number of edges is deterministic and it is equal to $\lfloor cn^2/2 \rfloor$, which we denote by

https://doi.org/10.1017/jpr.2019.34 Published online by Cambridge University Press

m_n. We denote by $\mathcal{E}^\circ$ the number of edges in the $\mathbb{G}^\circ_{\mathrm{W}}(n)$ graph model. It has Poisson distribution with parameter

$$c \sum_{1\le i<j\le n} \xi_i\xi_j + \frac{c}{2}\sum_{i=1}^n \xi_i^2 = \frac{c}{2}\left(\sum_{i=1}^n \xi_i\right)^2$$

conditionally with respect to $\mathcal{G}$ (recall Definition 1). Therefore, we have

$$\begin{aligned}
\mathbb{E}(|\mathcal{E}^\circ - m_n|) &= \mathbb{E}(|\mathcal{E}^\circ - m_n|)|\mathcal{G}) \\
&\ge \mathbb{E}(|\mathbb{E}(\mathcal{E}^\circ - m_n|\mathcal{G})|) \\
&= \mathbb{E}\left(\left|\frac{c}{2}\left(\sum_{i=1}^n \xi_i\right)^2 - m_n\right|\right) \\
&= \frac{c}{2}\cdot\mathbb{E}\left(\left|\sum_{i=1}^n \xi_i - \sqrt{\frac{2m_n}{c}}\right|\cdot\left(\sum_{i=1}^n \xi_i + \sqrt{\frac{2m_n}{c}}\right)\right) \\
&\ge \sqrt{\frac{cm_n}{2}}\,\mathbb{E}\left(\left|\sum_{i=1}^n \xi_i - \sqrt{\frac{2m_n}{c}}\right|\right) \\
&\ge \sqrt{\frac{cm_n}{2}}\cdot\left(\mathbb{E}\left(\left|\sum_{i=1}^n \xi_i - n\right|\right) - 1\right),
\end{aligned}$$

when n is sufficiently large, in particular, when $n-1 \le \sqrt{2m_n/c} = \sqrt{2\lfloor cn^2/2\rfloor/c} \le n$ holds. Since $\sum_{i=1}^n \xi_i$ has gamma distribution with parameters $n\kappa$ and κ, using the formula for the mean absolute deviation of the gamma distribution, we find that

$$\mathbb{E}(|\mathcal{E}^\circ - m_n|) \ge \frac{cn}{2}\cdot\sqrt{\frac{2}{\pi}}\cdot\sqrt{n/\kappa} \ge L_1 n^{3/2} \tag{1}$$

holds with appropriate $L_1 > 0$. Therefore, we have

$$\mathbb{E}(d_{\boxtimes}(\mathbb{G}_{\mathrm{PAG}}(n), \mathbb{G}^\circ_W(n))) \ge \frac{1}{n}\cdot\mathbb{E}\left(\left|\sum_{1\le i,j\le n} U_{ij} - V_{ij}\right|\right) = \frac{1}{n^2}\cdot\mathbb{E}(|\mathcal{E}^\circ - m_n|) = \Theta(n^{-1/2}).$$

On the other hand, since we have

$$\mathbb{E}(|\mathcal{E}^\circ - \mathcal{E}|) = \mathbb{E}\left(\frac{c}{2}\sum_{i=1}^n \xi_i^2\right) = \frac{c(1/\kappa+1)}{2}\cdot n,$$

a similar statement follows for the graph model $\mathbb{G}_W(n)$. □

Remark 1. Although it is not the focus of the present paper, we shall say a few words on the influence of the constants c and κ on the bounds. Since the aim was not to produce the sharpest possible expressions for the bounds, the explanations will be somewhat heuristic in nature, and cannot necessarily directly be read off the formulas.

It follows from equation (1) that the larger c is, the larger L_1 in Proposition 3 is, because cm_n is an increasing function of c, while the distribution of ξ does not depend on c. On the other hand, the proof of the upper bound also produces an expression which is an increasing function of c, if we keep every other parameter fixed (for example, the constant a depends only

https://doi.org/10.1017/jpr.2019.34 Published online by Cambridge University Press

on κ). This suggests that, although the upper and lower bounds are not the same, the jumble distance in the optimal coupling increases as c is larger. Equation (1) also implies that, at least for the W-random graph model with loops, the lower bound is smaller for larger κ.

The 'monotonicity' in c and κ, though of opposing nature, stems from the same phenomenon. In particular, the effect of c might seem surprising at first, since it governs the edge density, and thus one might expect that more edges lead to a better concentration. This is true as n increases due to the degree distribution of the urn model stabilizing and the law of large numbers becoming applicable, but not as c changes, as that shifts this distribution. If either c is increased or κ is decreased, both the graphon W and the degree distribution resulting from the urn model become less 'flat', and the variance of the X_i^* has more of an impact on the resulting W-random graph.

Remark 2. The proof of Proposition 3 shows that even the difference in the number of edges can lead to a positive lower bound on the distance of the two models in the jumble norm. This raises the question of what happens if we modify the models such that the number of edges becomes exactly the same. For example, we consider the random graph $\mathbb{G}_W^\circ(n)$ conditioned on the event that it has $\lfloor cn^2/2 \rfloor$ edges. However, the coupling of the two models used in the proof of Theorem 1 answers this. The condition means that the sum of the Poisson random variables assigned to the pairs of vertices is equal to $\lfloor cn^2/2 \rfloor$. As we have seen, this is the same as if we choose the types of $\lfloor cn^2/2 \rfloor$ edges independently with distribution (ρ_i), and according to Section 3, we get back the preferential attachment model in this way. Hence the two models have the same distribution under the condition that both have exactly $\lfloor cn^2/2 \rfloor$ edges.

Remark 3. In this paper we considered the jumble distance between the two random models for the dense PAG_κ graph, as that is the more natural distance notion for multigraphs generated by unbounded graphons (in this particular case, this corresponds to the unboundedness of the parameters of the Poisson distributions). However, as each finite multigraph generated is bounded *per se*, one may wonder if it is possible to say anything about the cut distance between $\mathbb{G}_{\mathrm{PAG}}$ and $\mathbb{G}_W^\circ$, for example.

We recall that the cut distance of two graphs on the same set of n vertices is defined as

$$d_\square(G,H) = \frac{1}{n^2} \cdot \max_{S,T} \left| \sum_{i \in S, j \in T} U_{ij} - V_{ij} \right|.$$

It is easily seen that $d_\square \le d_{\boxtimes}$, hence the upper bounds given for the jumble distance apply *a fortiori* to the cut distance as well. On the other hand, the methods used in this paper do not yield stronger bounds for the cut-norm distance.

Acknowledgements

The authors are very grateful for the remarks and improvements proposed by the anonymous referee, which ultimately lead to stronger upper and lower bounds.

The first author was supported by the Hungarian National Research, Development and Innovation Office, NKFIH grant K108615 and by the MTA Rényi Institute Lendület 'Limits of Structures' Research Group. The second author has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement 617747, the MTA Rényi Institute Lendület 'Limits of Structures' Research Group, and was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

https://doi.org/10.1017/jpr.2019.34 Published online by Cambridge University Press

References

[1] VON BAHR, B. (1965). On the convergence of moments in the central limit theorem. *Ann. Math. Statist.* **36**, 808–818.

[2] BARABÁSI, A.-L. AND ALBERT, R. (1999). Emergence of scaling in random networks. *Science* **286** (5439), 509–512.

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

[4] BLACKWELL, D. AND MACQUEEN, J. B. (1973). Ferguson distributions via Polya urn schemes. *Ann. Statist.* **2**, 353–355. Available at https://projecteuclid.org/euclid.aos/1176342372.

[5] BORGS, C., CHAYES, J., COHN, H. AND LOVÁSZ, L. M. (2018). Identifiability for graphexes and the weak kernel metric. Available at arXiv:1804.03277.

[6] BORGS, C., CHAYES, J., LOVÁSZ, L., SÓS, V. AND VESZTERGOMBI, K. (2011). Limits of randomly grown graph sequences. *Eur. J. Combin.* **32** (7), 985–999.

[7] DEGROOT, M.H. (1970). *Optimal Statistical Decisions*. McGraw-Hill, New York.

[8] DEVROYE, L. (1986). *Non-Uniform Random Variate Generation*. Springer, New York.

[9] DURRETT, R. (2007). *Random Graph Dynamics* (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge University Press, Cambridge.

[10] ELWES, R. (2015). Preferential attachment processes approaching the Rado multigraph. Available at arXiv:1502.05618.

[11] ELWES, R. (2016). A linear preferential attachment process approaching the Rado graph. Available at arXiv:1603.08806.

[12] FRIEZE, A. AND KAROŃSKI, M. (2015). *Introduction to Random Graphs*. Cambridge University Press, Cambridge.

[13] HALL, P. (1983). On the rate of convergence of moments in the central limit theorem for lattice distributions. *Trans. Amer. Math. Soc.* **278** (1), 169–181.

[14] KUNSZENTI-KOVÁCS, D., LOVÁSZ, L. AND SZEGEDY, B. (2014). Multigraph limits, unbounded kernels, and Banach space decorated graphs. Available at arXiv:1406.7846.

[15] LOVÁSZ, L. (2012). *Large Networks and Graph Limits*. American Mathematical Society, Providence, RI.

[16] PEMANTLE, R. (1988). Random processes with reinforcement. Doctoral thesis, Massachusetts Institute of Technology.

[17] RÁTH, B. AND SZAKÁCS, L. (2012). Multigraph limit of the dense configuration model and the preferential attachment graph. *Acta Math. Hung.* **136**, 196–221.

[18] WONG, T.-T. (1998). Generalized Dirichlet distribution in Bayesian analysis. *Appl. Math. Comput.* **97**, 165–181.

https://doi.org/10.1017/jpr.2019.34 Published online by Cambridge University Press