Combinatorics, Probability and Computing (2025), **34**, pp. 115–130
doi:10.1017/S0963548324000300

ARTICLE

Embedding theorems for random graphs with specified degrees

Pu Gao and Yuval Ohapkin

University of Waterloo, Waterloo, Canada
Corresponding author: Pu Gao; Email: pu.gao@uwaterloo.ca

(Received 2 March 2023; revised 20 April 2024; accepted 5 August 2024; first published online 16 October 2024)

Abstract

Given an $n \times n$ symmetric matrix $W \in [0,1]^{[n]\times[n]}$, let $\mathcal{G}(n, W)$ be the random graph obtained by independently including each edge $jk \in \binom{[n]}{2}$ with probability $W_{jk} = W_{kj}$. Given a degree sequence $\mathbf{d} = (d_1, \ldots, d_n)$, let $\mathcal{G}(n, \mathbf{d})$ denote a uniformly random graph with degree sequence $\mathbf{d}$. We couple $\mathcal{G}(n, W)$ and $\mathcal{G}(n, \mathbf{d})$ together so that asymptotically almost surely $\mathcal{G}(n, W)$ is a subgraph of $\mathcal{G}(n, \mathbf{d})$, where W is some function of $\mathbf{d}$. Let $\Delta(\mathbf{d})$ denote the maximum degree in $\mathbf{d}$. Our coupling result is optimal when $\Delta(\mathbf{d})^2 \ll \|\mathbf{d}\|_1$, that is, W_{ij} is asymptotic to $\mathbb{P}(ij \in \mathcal{G}(n, \mathbf{d}))$ for every $i, j \in [n]$. We also have coupling results for $\mathbf{d}$ that are not constrained by the condition $\Delta(\mathbf{d})^2 \ll \|\mathbf{d}\|_1$. For such $\mathbf{d}$ our coupling result is still close to optimal, in the sense that W_{ij} is asymptotic to $\mathbb{P}(ij \in \mathcal{G}(n, \mathbf{d}))$ for most pairs $ij \in \binom{[n]}{2}$.

Keywords: Kim–Vu's sandwich conjecture; non-regular degree sequence; coupling; embedding; generalized Erdos–Renyi graphs; Chung–Lu model

2020 MSC Codes: Primary: 05C80; Secondary: 05A16

1. Introduction

Given a realisable degree sequence $\mathbf{d} = (d_1, \ldots, d_n)$ (a degree sequence $\mathbf{d}$ is realisable if there exists a simple graph with degree sequence $\mathbf{d}$), let $\mathcal{G}(n, \mathbf{d})$ denote a random graph chosen uniformly from the set of graphs on $[n]$ where vertex i has degree d_i. Random graphs with a specified degree sequence are a popular class of random graphs used in many fields of research such as social network modelling and analysis [29, 33], epidemic analysis [5], and network sciences [9, 34, 35]. While these random graphs have been extensively used to model and analyse real-world networks, such as social networks and the internet, they present several analytical challenges. The most prominent difficulties in analysing $\mathcal{G}(n, \mathbf{d})$ are evaluating the edge probabilities and dealing with edge dependencies. Compared with the classical Erdős–Rényi random graph $\mathcal{G}(n, p)$ where every edge $jk \in \binom{[n]}{2}$ appears independently with probability p, there is no known closed formula for $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}))$ for general $\mathbf{d}$. Although asymptotic formulas exist for $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}))$ for some classes of $\mathbf{d}$, the correlation between the edges poses additional challenges when estimating the probabilities of events in $\mathcal{G}(n, \mathbf{d})$.

We denote $\mathcal{G}(n, \mathbf{d})$ by $\mathcal{G}(n, d)$ when $\mathbf{d} = (d, d, \ldots, d)$, that is, $\mathbf{d}$ is a d-regular degree sequence. The random regular graph $\mathcal{G}(n, d)$ is the most well-understood model among all $\mathcal{G}(n, \mathbf{d})$. For instance, the asymptotic enumeration of d-regular graphs on n vertices has been completely solved following a sequence of benchmark research [2–4, 6, 28, 30–32]. Properties and graph parameters of random regular graphs have been extensively studied, including Hamiltonicity, the chromatic and list chromatic numbers, independence number, and the distribution of subgraphs in

$\mathcal{G}(n, d)$ [13, 13, 22, 27]. We refer the interested readers to a survey of Wormald [36] for many other properties of random regular graphs. However, in general $\mathbf{d}$ can vary from near-regular sequences to heavy-tailed sequences such as power-law sequences. Both enumerating graphs of given degree sequences and analysing properties such as connectivity or Hamiltonicity of $\mathcal{G}(n, \mathbf{d})$ turn out to be challenging, and there are many open problems in this field. In 2004 Kim and Vu proposed the sandwich conjecture, which informally says that $\mathcal{G}(n, d)$ can be well approximated by $\mathcal{G}(n, p = d/n)$ through sandwiching $\mathcal{G}(n, d)$ between two correlated copies of $\mathcal{G}(n, p)$, one with p slightly smaller than d/n and the other with p slightly greater than d/n. To formally state the conjecture, we define a coupling of a finite set of random variables (or graphs) $Z_1, \ldots, Z_k$ as a k-tuple of random variables $(\hat{Z}_1, \ldots, \hat{Z}_k)$ defined in the same probability space such that the marginal distribution of $\hat{Z}_i$ is the same as the distribution of Z_i for every $1 \le i \le k$. The sandwich conjecture is about a 3-tuple coupling of random graphs, proposed by Kim and Vu, given as below. The standard Landau notation is used, and a formal definition of the notation is given before Section 1.1.

Conjecture 1.1 ([25]). *For $d \gg \log n$, there exist probabilities $p_1, p_2 = (1 + o(1))d/n$ and a coupling (G_L, G, G_U) such that marginally, $G_L \sim \mathcal{G}(n, p_1)$, $G \sim \mathcal{G}(n, d)$, $G_U \sim \mathcal{G}(n, p_2)$ and jointly, $\mathbb{P}(G_L \subseteq G \subseteq G_U) = 1 - o(1)$.*

The sandwich conjecture, if proved to be true, is a powerful tool for analysing $\mathcal{G}(n, d)$, and reveals beautiful distributional relations between the two different random graph models. The assumption $d \gg \log n$ is necessary, as for $p = O(\log n/n)$, the maximum and minimum degrees of $\mathcal{G}(n, p)$ differ by some constant factor away from 1, making it impossible for the sandwich conjecture to hold. For simplicity, we refer to a coupling $(\mathcal{G}_n^1, \mathcal{G}_n^2)$ of two random graphs $\mathcal{G}_n^1$ and $\mathcal{G}_n^2$ as *embedding $\mathcal{G}_n^1$ into $\mathcal{G}_n^2$*, if in the coupling $\mathcal{G}_n^1$ is a subgraph of $\mathcal{G}_n^2$ with probability going to 1 as $n \to \infty$. It turns out that embedding $\mathcal{G}(n, d)$ into $\mathcal{G}(n, p = (1 + o(1))d/n)$ is much more difficult than embedding $\mathcal{G}(n, p = (1 - o(1))d/n)$ into $\mathcal{G}(n, d)$. In their paper [25], Kim and Vu established an embedding of $\mathcal{G}(n, p = (1 - o(1))d/n)$ into $\mathcal{G}(n, d)$ for d up to approximately $n^{1/3}$ (and $d \gg \log n$). Subsequent work by Dudek, Frieze, Ruciński, and Šileikis [14] improved this result to $d = o(n)$, and extended it to random uniform hypergraphs as well. The first 2-side sandwich theorem was proved by Isaev, McKay, and the first author [17, 18] in the case that d is linear in n. Klimosová, Reiher, Ruciński, and Šileikis [26] proved the sandwich conjecture for $d \gg (n \log n)^{3/4}$. Finally Isaev, McKay, and the first author [19] confirmed the sandwich conjecture for all $d \ge \log^4 n$. It is worth noting that a more general sandwich theorem was presented in [17–19] for random graphs $\mathcal{G}(n, \mathbf{d})$ where $\mathbf{d}$ is a near-regular degree sequence, where all degrees are asymptotically equal.

While it is not possible to approximate $\mathcal{G}(n, \mathbf{d})$ by $\mathcal{G}(n, p)$ in a useful way for more general forms of $\mathbf{d}$ (since the typical degree distribution of $\mathcal{G}(n, p)$ would be very different from $\mathbf{d}$), a natural alternative is to consider a generalised Erdös–Rényi graph $\mathcal{G}(n, W)$, where W is a symmetric $n \times n$ matrix. In this model, each edge $jk \in \binom{[n]}{2}$ appears in $\mathcal{G}(n, W)$ independently with probability W_{jk}. By choosing different forms of W, we can recover well-studied models in the literature, such as $\mathcal{G}(n, p)$, the Chung–Lu model [9–12], the $\boldsymbol{\beta}$-model [8, 24], and the stochastic block model [23].

The objective of this paper is to establish an embedding of $\mathcal{G}(n, W)$ into $\mathcal{G}(n, \mathbf{d})$, where $\mathbf{d}$ is a degree sequence that may deviate significantly from being regular, for a suitable choice of W. What constitutes a good W? Intuitively one would like the entry W_{jk} to be approximately $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}))$, the true probability that jk is an edge in $\mathcal{G}(n, \mathbf{d})$. To formalise this notion, we define $W^* = W^*(\mathbf{d})$ as the $n \times n$ matrix where $W^*_{jk} = \mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}))$ for every $i, j \in [n]$.

Question 1.2. *Given $\mathbf{d}$, does there exists $W = (1 - o(1))W^*(\mathbf{d})$ such that $\mathcal{G}(n, W)$ can be embedded into $\mathcal{G}(n, \mathbf{d})$?*

Note that the asymptotic value of $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}))$ is not known for all realisable degree sequences $\mathbf{d}$. The coupling scheme we use in this paper heavily relies on the expression of the

https://doi.org/10.1017/S0963548324000300 Published online by Cambridge University Press

asymptotic formula for $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}))$. Therefore, we are restricted to degree sequences for which an asymptotic evaluation of $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}))$ is available in the literature. Without loss of generality we may assume that $d_1 \geq \cdots \geq d_n \geq 1$. Let $\Delta(\mathbf{d}) = d_1$ and $\delta(\mathbf{d}) = d_n$. Another important parameter of $\mathbf{d}$ is $J(\mathbf{d}) = \sum_{i=1}^{d_1} d_i$. The following proposition, a direct corollary of [20, Theorem 1], estimates the edge probabilities in $\mathcal{G}(n, \mathbf{d})$ when $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$.

Proposition 1.3. (Corollary of [20, Theorem 1]) *Suppose that* $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$. *Then*

$$\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d})) = \left(1 + O\left(\frac{J(\mathbf{d})}{\|\mathbf{d}\|_1}\right)\right) \frac{d_j d_k}{\|\mathbf{d}\|_1 + d_j d_k}.$$

Remark 1.4. The condition $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$ requires $\mathbf{d}$ to be a sparse degree sequence. It also restrict the tail of $\mathbf{d}$ so that there cannot be too many degrees that are very large. See Proposition 3.2 in the Appendix for properties of $\mathbf{d}$ that satisfies this technical condition. The edge probability in Proposition 1.3 is in general not correct without the condition $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$. See an example like this in Remark 1.9.

Remark 1.5. Although Proposition 1.3 does not give asymptotic estimate of $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}))$ for every degree sequence, there are many interesting families of degree sequences satisfying $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$, including examples of

- all d-regular, and near-d-regular degree sequences (namely, all degrees are asymptotic to d) where $d = o(n)$;
- all perturbed sequences from a near-d-regular degree sequence by arbitrarily decreasing at most $(1 - c)n$ entries, where $c > 0$ is fixed, and $d = o(n)$;
- heavy-tailed degree sequences such as power-law sequences. We refer interested readers to [22, Section 2] for a formal definition of power-law sequences and other examples of heavy-tailed degree sequences.

Assume that $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$. It is easy to show that for most pairs jk, $d_j d_k = o(\|\mathbf{d}\|_1)$ (e.g. see Proposition 3.2 in the Appendix), and thus by Proposition 1.3 their edge probabilities are $o(1)$. On the other hand, if there exist pairs jk such that $d_j d_k = \Omega(\|\mathbf{d}\|_1)$, then Proposition 1.3 implies that the edge probabilities in $\mathcal{G}(n, \mathbf{d})$ will not be uniformly $o(1)$. In particular, for every such pair jk the edge probability of jk is bounded from below by some constant $c > 1$. In fact, there exist degree sequences $\mathbf{d}$ satisfying $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$ whose edge probabilities takes values ranging from $o(1)$ to constant $0 < c < 1$ and to $1 - o(1)$. The power-law sequences mentioned in Remark 1.5 are good examples like this. As we will see later, it is the presence of such diverse edge probabilities that poses a challenge to embedding $\mathcal{G}(n, W)$ into $\mathcal{G}(n, \mathbf{d})$.

In light of Proposition 1.3, we define matrix $P(\mathbf{d})$, which is asymptotic to $W^*(\mathbf{d})$ under the condition $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$.

Definition 1.6. *Given* $\mathbf{d} = (d_1, \ldots, d_n)$, *let* $P(\mathbf{d})$ *be the symmetric* $n \times n$ *matrix defined by* $P_{ij} = P_{ji} = \frac{d_i d_j}{\|\mathbf{d}\|_1 + d_i d_j}$, *for every* $1 \leq i < j \leq n$, *and* $P_{ii} = 0$ *for every* $1 \leq i \leq n$.

One of our main results is the following theorem, which gives a positive answer to Question 1.2 for degree sequences satisfying $\Delta(\mathbf{d})^2 = o(\|\mathbf{d}\|_1)$, a condition that is stronger than $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$.

Theorem 1.7. *Assume that* $\boldsymbol{d} = \boldsymbol{d}(n)$ *is a degree sequence satisfying* $\Delta(\boldsymbol{d})^2 = o(\|\boldsymbol{d}\|_1)$ *and* $\delta(\boldsymbol{d}) \gg \log n$. *Then, there exist* $\varepsilon = o(1)$ *and a coupling* (G_L, G), *where* $G_L \sim \mathcal{G}(n, W)$, $W = (1 - \varepsilon)P(\boldsymbol{d})$, *and* $G \sim \mathcal{G}(n, \boldsymbol{d})$, *such that* $\mathbb{P}(G_L \subseteq G) = 1 - o(1)$.

Under the stronger condition that $\Delta(\mathbf{d})^2 = o(\|\mathbf{d}\|_1)$, $d_j d_k = o(\|\mathbf{d}\|_1)$ for all $jk \in \binom{[n]}{2}$, and thus $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d})) = o(1)$ for all $jk \in \binom{[n]}{2}$ by Proposition 1.3. This property is a crucial element in our

https://doi.org/10.1017/S0963548324000300 Published online by Cambridge University Press

coupling scheme which allows us to achieve an optimal embedding. Our next theorem embeds $\mathcal{G}(n, W)$ into $\mathcal{G}(n, \mathbf{d})$ for $\mathbf{d}$ satisfying $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$. In this case, $W_{jk} = (1 - o(1))P(\mathbf{d})_{jk}$ for all jk where $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d})) = o(1)$. However, W_{jk} is not asymptotic to $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}))$ for jk where $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}))$ is bounded away from 0. Given a function $f:\mathbb{R} \to \mathbb{R}$, let $f(P)$ denote the matrix $(f(P_{i,j}))_{i,j\in[n]}$.

Theorem 1.8. *Assume that* $\mathbf{d}$ *is a degree sequence satisfying* $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$ *and* $\delta(\mathbf{d}) \gg \log n$. *Then, there exist* $\varepsilon = o(1)$ *and a coupling* (G_L, G), *where* $G_L \sim \mathcal{G}(n, W)$, $W = (1 - \varepsilon)f(P(\mathbf{d}))$ *where* $f(x) = 1 - e^{-x}$, *and* $G \sim \mathcal{G}(n, \mathbf{d})$, *such that* $\mathbb{P}(G_L \subseteq G) = 1 - o(1)$.

We see that Theorem 1.7 immediately follows from Theorem 1.8.

Proof of Theorem 1.7. Note that $\Delta(\mathbf{d})^2 = o(\|\mathbf{d}\|_1)$ implies $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$ since $J(\mathbf{d}) \le \Delta(\mathbf{d})^2$. Moreover, $\Delta(\mathbf{d})^2 = o(\|\mathbf{d}\|_1)$ implies that $P(\mathbf{d})_{jk} = o(1)$ for all jk by Definition 1.6. Now Theorem 1.7 follows as a straightforward corollary of Theorem 1.8 by noticing that $1 - e^{-x} = (1 + O(x))x$. □

Remark 1.9. Note that $P(\mathbf{d}) = (1 + o(1))W^*(\mathbf{d})$ is not true in general. For instance, consider the d-regular degree sequence where $d = \Theta(n)$. By symmetry we know that $W^*_{jk} = d/(n-1) \sim d/n$ for every jk. On the other hand, by Definition 1.6, $P_{jk} = d^2/(dn + d^2) = d/(n + d)$ which is not asymptotic to d/n as $d = \Theta(n)$. Thus, in this example, P_{jk} is significantly smaller than W^*_{jk}. There are two open problems below that can be very interesting for future research. The first question is a weaker version of Question 1.2, replacing $W^*(\mathbf{d})$ by $P(\mathbf{d})$.

(a) Does Theorem 1.8 hold with some $W = (1 + o(1))P(\mathbf{d})$?

Note that Theorem 1.7 answers the question above only for degree sequences satisfying $\Delta^2 = o(\|\mathbf{d}\|_1)$. However, in the more general setting as Theorem 1.8, there can be pairs jk where $d_jd_k = \Omega(\|\mathbf{d}\|_1)$, and for such pairs, our current embedding requires W where W_{jk} is significantly smaller than P_{jk}. Hence we do not have a complete answer to question (a).

(b) Can Question 1.2 be answered for $\mathbf{d}$ where $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$ is not satisfied? In particular, is there a way to embed $\mathcal{G}(n, W)$ into $\mathcal{G}(n, \mathbf{d})$ for some reasonable W without knowing the asymptotic value of $W^*(\mathbf{d})$ or the conditional edge probabilities as in Proposition 1.3?

Remark 1.10. To obtain a sandwich-type result as for $\mathcal{G}(n, d)$, we would hope to embed $\mathcal{G}(n, \mathbf{d})$ into $\mathcal{G}(n, W)$ for some $W = (1 + o(1))W^*$. If we approach in the same way as for $\mathcal{G}(n, d)$, we would embed $\mathcal{G}(n, (J - I) - (1 + o(1))W^*)$ into $\mathcal{G}(n, (n-1)\mathbf{1} - \mathbf{d})$ where $\mathbf{1}$ is the all-one vector, $J = \mathbf{1}\mathbf{1}^T$ and I is the identity matrix. Both the proofs in [17, 27] for embedding $\mathcal{G}(n, 1 - (1 + o(1))d/n)$ into $\mathcal{G}(n, n - 1 - d)$ use some counting arguments that heavily rely on the fact that the underlying graphs (during the construction of the coupling) have almost equal degrees, and we do not think this argument can extend to graphs that are far away from being regular. In fact, we do not have enough intuition to support a sandwich conjecture for $\mathcal{G}(n, \mathbf{d})$ as for $\mathcal{G}(n, d)$, especially for degree sequences $\mathbf{d}$ where the values of edge probabilities range from $o(1)$ to $1 - o(1)$.

Throughout the paper n is assumed to be sufficiently large. For two sequences of real numbers $(a_n)_{n=0}^{\infty}$ and $(b_n)_{n=0}^{\infty}$, we write $a_n = O(b_n)$ if there is $C > 0$ such that $|a_n| < C|b_n|$ for every n. We say $a_n = \Omega(b_n)$ if $a_n > 0$ and $b_n = O(a_n)$. We say $b_n = o(a_n)$, or $b_n \ll a_n$, or $a_n \gg b_n$ if $a_n > 0$ and $\lim_{n\to\infty} b_n/a_n = 0$. We say a sequence of events A_n indexed by n holds a.a.s. (asymptotically almost surely) if $\mathbb{P}(A_n) = 1 - o(1)$. All asymptotics are with respect to $n \to \infty$.

https://doi.org/10.1017/S0963548324000300 Published online by Cambridge University Press

1.1 Relation between $\mathcal{G}(n, P(\boldsymbol{d}))$ and the Chung–Lu model

In 2002, Chung and Lu introduced the random graph model with given expected degrees [9–12]. Given a sequence of nonnegative real numbers $\mathbf{w} = (w_1, \ldots, w_n)$ satisfying that

$$\max_i w_i^2 \le \|\mathbf{w}\|_1, \tag{1}$$

the random graph $G(\mathbf{w})$ is defined by $\mathcal{G}(n, \hat{W}(\mathbf{w}))$ where $\hat{W}_{jk}(\mathbf{w}) = w_j w_k / \|\mathbf{w}\|_1$. To avoid relying on the assumption [1] one can define $\hat{W}_{jk} = \min\{w_j w_k / \|\mathbf{w}\|_1, 1\}$. However, most of the work about the Chung–Lu model assumed [1]. Notice that the three matrices $P(\mathbf{d})$, $W^*(\mathbf{d})$, and $\hat{W}(\mathbf{d})$ are all asymptotically equal if $\mathbf{d}$ satisfies $\Delta(\mathbf{d})^2 = o(\|\mathbf{d}\|_1)$, as under this assumption, $P(\mathbf{d})_{jk} \sim W^*(\mathbf{d})_{jk} \sim \hat{W}(\mathbf{d})_{jk} = d_j d_k / \|\mathbf{d}\|_1$ for all jk, by definition of these three matrices and Proposition 1.3. However, if there exists jk such that $d_j d_k = \Omega(\|\mathbf{d}\|_1)$ then $P(\mathbf{d})_{jk}$, $W^*(\mathbf{d})_{jk}$, and $\hat{W}(\mathbf{d})_{jk}$ may be all asymptotically distinct.

1.2 Applications

Given two $n \times n$ matrices W_1 and W_2, we say $W_1 \le W_2$ if $(W_1)_{ij} \le (W_2)_{ij}$ for every $i, j \in [n]$. It is well known that $\mathcal{G}(n, W)$ has the nice "nesting property", meaning that $\mathcal{G}(n, W_1)$ can be embedded into $\mathcal{G}(n, W_2)$ provided that $W_1 \le W_2$. However, $\mathcal{G}(n, \mathbf{d})$ does not have the nesting property. Given two degree sequences $\mathbf{d} \preceq \mathbf{g}$, it is in general not true that $\mathcal{G}(n, \mathbf{d})$ can be embedded into $\mathcal{G}(n, \mathbf{g})$. In fact, it is easy to construct degree sequences $\mathbf{d} \preceq \mathbf{g}$ for which there exists jk such that $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d})) = 1 - o(1)$ and $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{g})) = o(1)$ (see an example like this in the Appendix). Consequently and perhaps rather surprisingly, it is difficult to prove some rather "intuitive" results about $\mathcal{G}(n, \mathbf{d})$. For instance, although it is known that a.a.s. $\mathcal{G}(n, 3)$ is connected, to our knowledge it is not known if $\mathcal{G}(n, \mathbf{d})$ is a.a.s. connected for every $\mathbf{d}$ provided that $\delta(\mathbf{d}) \ge 3$ [20]. Most such results are restricted to certain families of degree sequences for which either some enumeration results are known, or some enumeration proof techniques can be applied. Theorem 1.8 gives a powerful tool to obtain such results by first embedding $\mathcal{G}(n, W)$ into $\mathcal{G}(n, \mathbf{d})$ and then applying the nesting property of $\mathcal{G}(n, W)$. We show a few examples below.

Theorem 1.11. *Assume $\boldsymbol{d}$ is a degree sequence satisfying $J(\boldsymbol{d}) = o(\|\boldsymbol{d}\|_1)$ and $\delta(\boldsymbol{d}) \gg \log n$. Then,*

- *(a) if $\delta(\boldsymbol{d})^2 / \|\boldsymbol{d}\|_1 \ge (1 + c) \log n / n$ for some fixed $c > 0$ then a.a.s. $\mathcal{G}(n, \boldsymbol{d})$ is Hamiltonian and k-connected for every fixed k;*
- *(b) if $\delta(\boldsymbol{d})^2 / \|\boldsymbol{d}\|_1 \ge c/n$ for some fixed $c > 1$ then a.a.s. $\mathcal{G}(n, \boldsymbol{d})$ contains H as a minor for every fixed graph H;*
- *(c) if $\delta(\boldsymbol{d})^2 / \|\boldsymbol{d}\|_1 \ge n^{-2/(k+1)+c}$ for some fixed $c > 0$ where $k > 0$ is a fixed integer, then a.a.s. simultaneously for all graphs H on $[n]$ with maximum degree at most k, $\mathcal{G}(n, \boldsymbol{d})$ has a subgraph isomorphic to H.*

Proof. By definition, $P(\mathbf{d})_{jk} \ge (1 - o(1))\delta(\mathbf{d})^2 / \|\mathbf{d}\|_1$ for every jk. Thus, by Theorem 1.8 and the nesting property of $\mathcal{G}(n, W)$, $\mathcal{G}(n, p)$ can be embedded into $\mathcal{G}(n, \mathbf{d})$ for some $p = (1 - o(1))\delta(\mathbf{d})^2 / \|\mathbf{d}\|_1$. Parts (a,b) follow as for every fixed $c > 0$, $\mathcal{G}(n, (1 + c) \log n / n)$ is a.a.s. Hamiltonian and k-connected [7]; and $\mathcal{G}(n, c/n)$ contains every fixed graph minor when $c > 1$ [15]. Part (c) follows from Theorem 7.2 of [16]. $\square$

Remark 1.12. It should be possible to improve or even remove some assumptions such as $\delta(\mathbf{d})^2 / \|\mathbf{d}\|_1 \ge (1 + c) \log n / n$ for some fixed $c > 0$ in part (a), by directly analysing $\mathcal{G}(n, (1 + o(1)) f(P(\mathbf{d})))$. For instance, the sharp threshold of Hamiltonicity was studied for the stochastic block model [1]. We did not attempt it as the main objective of this paper is to prove the embedding theorems.

2. The coupling procedure

2.1 The old and the new

Assume that we aim to embed $\mathcal{G}(n,p)$ into $\mathcal{G}(n,d)$ where $p=(1-o(1))d/n$. Regardless of a few minor differences, the coupling procedures employed in [14, 17–19, 25, 26] all use the following approach, which was introduced in the original paper of Kim and Vu [25]: let $x_1, x_2, \ldots, x_m$ be a sequence of random edges, each uniformly and independently chosen from K_n (here m is a carefully chosen integer-valued random variable that is concentrated around $\|\mathbf{d}\|_1/2$). Sequentially add edges in this sequence to G and to G_L, respectively. With a small probability $\epsilon_i = o(1)$, edge x_i is rejected by G; and with a slightly larger (than ϵ_i) but still rather small probability ζ, x_i is rejected by G_L. The parameter ϵ_i is chosen to be proportional to $p_i(x_i)$, the probability that x_i is an edge of $\mathcal{G}(n,\mathbf{d})$ conditional on the event that all the edges that have been added to G are edges of $\mathcal{G}(n,\mathbf{d})$. The key idea of the coupling procedure is that, until m gets very close to $dn/2$, with high probability, $p_i(jk)$ is approximately the same for every edge jk that has not been added to G yet. Since x_i is uniformly chosen, a small rejection probability ϵ_i suffices to ensure that x_i is added to G according to the correct conditional probability. We can prove that with high probability, $\epsilon_i = o(1)$ for every $1 \le i \le m$, where m, as commented earlier, is concentrated around $(1-o(1))dn/2$. Hence we can choose some $\zeta = o(1)$ such that $\epsilon_i \le \zeta$ for every $1 \le i \le m$. Moreover, (a) G_L obtained after the m-th iteration is a uniformly random graph conditional on the number of edges it contains, as every edge in $\binom{[n]}{2}$ has an equal probability to be added to G_L; and (b) G_L is a subgraph of G, as the rejection probability ζ for G_L is slightly larger than that for G in every step.

Now we are considering $\mathbf{d}$ where the degrees are not all asymptotically the same. The most natural way to extend the previous coupling procedure is to generate the sequence of i.i.d. random edges $x_1, x_2, \ldots, x_m$ where each edge is chosen with probability proportional to $W^*(\mathbf{d})$. This coupling procedure works well when $\Delta(\mathbf{d})^2 = o(\|\mathbf{d}\|_1)$. This is because the conditional probability $p_i(jk)$ of jk being an edge of $\mathcal{G}(n,\mathbf{d})$ in step i of the procedure turns out to be proportional to $W^*(\mathbf{d})_{jk}$ uniformly for all jk during the whole coupling process, resulting a small rejection probability ϵ_i. However, rather surprisingly, if there exists jk such that $d_jd_k = \Omega(\|\mathbf{d}\|_1)$ then the ratio $p_i(jk)/W^*_{jk}$ changes in a non-uniform way over jk and over time i of the coupling procedure, resulting larger and larger rejection probability ϵ_i. As ζ has to be chosen uniformly during the process, we can only choose $\zeta = 1-o(1)$, meaning that almost all edges are rejected by G_L. The coupling procedure thus fails.

To overcome the challenges, we develop two novel ideas.

- Instead of using the same rejection probability ζ for every edge to be added to G_L, we use different rejection probabilities for different edges. However, for a given edge jk, the rejection probability ζ_{jk} remains uniform throughout the procedure. This uniformity is needed for the output G_L to have the correct distribution.
- During the coupling procedure, the value of $p_i(jk)/W^*_{jk}$ decreases significantly for edges jk where $d_jd_k = O(\|\mathbf{d}\|_1)$, but changes little for jk where $d_jd_k \gg \|\mathbf{d}\|_1$. Consequently, many rejections (particularly those where $d_jd_k \ll \|\mathbf{d}\|_1$) occur already for the construction of G, and thus the first idea above would not help, as even more rejections would occur for G_L than for G. To reduce the rejection probability for the construction of G, we "intentionally" boost the probability (by an $\omega(1)$ factor) of generating edges jk where $d_jd_k \gg \|\mathbf{d}\|_1$ in the sequence of random edges $x_1, x_2, \ldots$. This probability boosting strategy magically reduces the rejection probability $\epsilon_i(jk)$ for jk such that $d_jd_k \ll \|\mathbf{d}\|_1$ (note that most $jk \in \binom{[n]}{2}$ are this type; see Proposition 3.2 in the Appendix). However, the rejection probabilities $\epsilon_i(jk)$ will be high (close to 1) for jk where $d_jd_k \gg \|\mathbf{d}\|_1$. Nonetheless, the first idea mentioned earlier will be applicable in this case — we reject these jk more often than the others for the construction of G_L. Consequently, the edge probability for such jk in the final construction

of G_L turns out to be close to $1 - e^{-1}$ instead of 1. Note that for pairs jk where $d_jd_k \gg \|\mathbf{d}\|_1$, their edge probability W^*_{jk} in $\mathcal{G}(n, \mathbf{d})$ is asymptotic to 1 by Proposition 1.3. Thus, their edge probabilities in G_L (which is asymptotic to $1 - e^{-1}$) is smaller than W^*_{jk}, implying that our embedding is not "tight". The good news is that only a small proportion of $jk \in \binom{[n]}{2}$ are of this type.

We define the procedure for sequential generation of $\mathcal{G}(n, W)$ in Section 2.2, the procedure for sequential generation of $\mathcal{G}(n, \mathbf{d})$ in Section 2.3. In Section 2.4 we couple the two procedures together to sequentially generate the coupled pair (G_L, G). The parameters used by the coupling procedure will be specified in Section 2.5. Finally the proof of Theorem 1.8 is given in Section 3.

2.2 Sequential generation of $\mathcal{G}(n, W)$

We define a sequential sampling procedure SeqApprox-P$(\mathbf{d}, \lambda, \Lambda)$, which adds a sequence of edges one at a time, and outputs a random graph where every edge in $\binom{[n]}{2}$ appears independently. The input λ is a positive real number, and $\Lambda \in [0, 1]^{[n]\times[n]}$ is a symmetric $n \times n$ matrix. We may think of $\mathbf{d}$ as the target degree sequence. As mentioned before, instead of weighting edge probabilities according to their true probabilities in $\mathcal{G}(n, \mathbf{d})$, we boost the probability of edge jk if d_jd_k is large. To formalise this notion, we define matrix Q as follows, which provides the probability distribution for the edges to be sequentially sampled in the procedure SeqApprox-P$(\mathbf{d}, \lambda, \Lambda)$.

Definition 2.1. *Given $\mathbf{d} = (d_1, \ldots, d_n)$, let $Q(\mathbf{d})$ be the symmetric $n \times n$ matrix defined by $Q_{ij} = Q_{ji} = \left(\sum_{1\le k<\ell\le n} d_kd_\ell\right)^{-1} d_id_j$, for every $1 \le i < j \le n$, and $Q_{ii} = 0$ for every $1 \le i \le n$.*

Procedure SeqApprox-P$(\mathbf{d}, \lambda, \Lambda)$ is given below. We use $\mathbf{Po}(\lambda)$ to denote the Poisson distribution with mean λ.

---

procedure SeqApprox-P$(\mathbf{d}, \lambda, \Lambda)$
  Let $\mathcal{I} \sim \mathbf{Po}(\lambda)$.
  Let G_0 be the empty graph on $[n]$.
  for i in $1, \ldots, \mathcal{I}$ **do**
    Pick an edge $jk \in \binom{[n]}{2}$ with probability proportional to d_jd_k (i.e. with probability $Q(\mathbf{d})_{jk}$).
    $G_i = G_{i-1} \cup \{jk\}$ with probability Λ_{jk}, and $G_i = G_{i-1}$ with probability $1 - \Lambda_{jk}$.
  end for
  Return $G_{\mathcal{I}}$.
end procedure

The parameters λ and Λ will be set in Section 2.5. Roughly speaking, λ is approximately $\|\mathbf{d}\|_1/2$ so that the number of edges in the output of SeqApprox-P$(\mathbf{d}, \lambda, \Lambda)$ is close to the number of edges in $\mathcal{G}(n, \mathbf{d})$, if only a small proportion of them are to be rejected, which we will prove later. The matrix Λ is chosen so that Λ_{jk} is approximately $\|\mathbf{d}\|_1/(d_jd_k + \|\mathbf{d}\|_1)$. Since each edge jk is selected with probability proportional to d_jd_k by the definition of Q, and due to the choice of

https://doi.org/10.1017/S0963548324000300 Published online by Cambridge University Press

Λ, edge jk is selected and accepted with probability proportional to $d_jd_k/(d_jd_k + \|\mathbf{d}\|_1)$, which is approximately W^*_{jk} as desired. The exact values of λ and Λ are only important when we couple the two sampling processes together. At this point, regardless of their values, we prove that the output of SeqApprox-P($\mathbf{d}$, λ, Λ) has distribution $\mathcal{G}(n, W)$ for some W as a function of $\mathbf{d}$, λ and Λ. For two matrices A and B of the same dimension, we denote by $A \odot B$ the Hadamard product of A and B, defined by $(A \odot B)_{ij} = A_{ij}B_{ij}$ for every entry ij.

Lemma 2.2. *SeqApprox-P($\mathbf{d}$, λ, Λ) returns a random graph G with distribution $\mathcal{G}(n, f(\Lambda \odot Q))$, where $Q = Q(\mathbf{d})$ and $f(x) = 1 - \exp(-\lambda x)$.*

Proof. Let $e_1 = j_1k_1, \ldots, e_N = j_Nk_N$ be an enumeration of the edges in $\binom{[n]}{2}$, where $N = \binom{n}{2}$. For $1 \le i \le N$, let X_i denote the number of times that edge e_i is sampled throughout SeqApprox-P($\mathbf{d}$, λ, Λ). We prove that the components of $\mathbf{X} = (X_1, \ldots, X_N)$ are mutually independent, and consequently, every edge e_i appears independently in G with probability $\mathbb{P}(X_i \ge 1)$. For each edge $e_i \in \binom{[n]}{2}$, the probability that e_i is in G is thus given by

$$\begin{aligned}\mathbb{P}(X_i \ge 1) = 1 - \mathbb{P}(X_i = 0) &= 1 - \sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!}\sum_{j=0}^{m}\binom{m}{j}Q_{e_i}^j(1-Q_{e_i})^{m-j}\left(1-\Lambda_{e_i}\right)^j \\ &= 1 - \sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!}(1 - Q_{e_i} + Q_{e_i}(1-\Lambda_{e_i}))^m \\ &= 1 - \exp\left(-\lambda + \lambda(1 - Q_{e_i} + Q_{e_i}(1-\Lambda_{e_i}))\right) = 1 - \exp(-\lambda Q_{e_i}\Lambda_{e_i}),\end{aligned}$$

and the probability generating function for X_i is given by

$$\begin{aligned}\mathbb{E}z^{X_i} &= \sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!}\sum_{j=0}^{m}\binom{m}{j}Q_{e_i}^j(1-Q_{e_i})^{m-j}\sum_{k=0}^{j}\binom{j}{i}\Lambda_{e_i}^k(1-\Lambda_{e_i})^{j-k}z^k \\ &= \sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!}\sum_{j=0}^{m}\binom{m}{j}Q_{e_i}^j(1-Q_{e_i})^{m-j}(1-\Lambda_{e_i}+\Lambda_{e_i}z)^j \\ &= \sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!}(1 - Q_{e_i} + Q_{e_i}(1-\Lambda_{e_i}+\Lambda_{e_i}z))^m = \exp(-\lambda Q_{e_i}\Lambda_{e_i}(1-z)).\end{aligned}$$

On the other hand, the probability generating function for the random vector $\mathbf{X}$ is given by

$$\begin{aligned}\sum_{j_1,\ldots,j_N}\mathbb{P}(X_1 = j_1, \ldots, X_N = j_N)z_1^{j_1}\cdots z_N^{j_N} &= \sum_{m=0}^{\infty} e^{-\lambda}\frac{\lambda^m}{m!}\left(\sum_{i=1}^{N}\left(Q_{e_i}\Lambda_{e_i}z_i + Q_{e_i}(1-\Lambda_{e_i})\right)\right)^m \\ &= e^{-\lambda}\exp\left(\sum_{i=1}^{N}\left(\lambda Q_{e_i}\Lambda_{e_i}z_i + \lambda Q_{e_i}(1-\Lambda_{e_i})\right)\right) \\ &= \prod_{i=1}^{N}\exp\left(-\lambda\, Q_{e_i}\Lambda_{e_i} + \lambda Q_{e_i}\Lambda_{e_i}z_i\right) = \prod_{i=1}^{N}\mathbb{E}z^{X_i},\end{aligned}$$

where the second last equation above holds because $\sum_{i=1}^{N} Q_{e_i} = 1$ by definition of Q. Hence we have shown that the components of $\mathbf{X}$ are mutually independent. Thus, $G \sim \mathcal{G}(n, W)$ where W is the symmetric $n \times n$ matrix given by $W_{ij} = 1 - \exp(-\lambda Q_{ij}\Lambda_{ij})$ for every $i, j \in [n]$. □

2.3 Sequential generation of $\mathcal{G}(n, \mathbf{d})$

We define procedure SeqSample-D($\mathbf{d}$) which sequentially generates a random graph with distribution $\mathcal{G}(n, \mathbf{d})$. This procedure is essentially the same as previously used in [17–19, 26]. Let $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}) \mid H)$ denotes the probability that jk is an edge of $\mathcal{G}(n, \mathbf{d})$ conditional on H being a subgraph of $\mathcal{G}(n, \mathbf{d})$.

---

procedure SeqSample-D($\mathbf{d}$)
  Let G_0 be the empty graph on $[n]$.
  for i in $1, \ldots, \|\mathbf{d}\|_1/2$ **do**
    Pick an edge $jk \in \binom{[n]}{2} \setminus G_{i-1}$ with probability proportional to $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}) \mid G_{i-1})$.
    $G_i = G_{i-1} \cup \{jk\}$.
  end for
  Return $G_{\|\mathbf{d}\|_1/2}$.
end procedure

Given $\mathbf{d}$, and an integer $0 \leq m \leq \|\mathbf{d}\|_1/2$, let $\mathcal{G}(n, \mathbf{d}, m)$ denote a uniformly random subgraph of $\mathcal{G}(n, \mathbf{d})$ with exactly m edges. The following lemma follows by a simple counting argument and was proved in [18, Lemma 3].

Lemma 2.3. *Let $0 \leq m \leq \|\boldsymbol{d}\|_1/2$ and G_m be the graph obtained after m iterations of SeqSample-D($\boldsymbol{d}$). Then $G_m \sim \mathcal{G}(n, \boldsymbol{d}, m)$.*

Lemma 2.3 (with $m = \|\mathbf{d}\|_1/2$) immediately implies the following corollary.

Corollary 2.4. *Let G be the output of* SeqSample − D($\boldsymbol{d}$). *Then $G \sim \mathcal{G}(n, \boldsymbol{d})$.*

We need the following lemma, which follows by standard concentration results and can be found in [18, Lemma 4]. We use $\mathbf{Bin}(K, p)$ to denote the Binomial distribution with K trials and success probability p. (Note that Lemma 4(c) of [18] only states the upper tail bound of part (c) below; but its proof gives both the upper and lower tail bounds.)

Lemma 2.5. *Let $Y \sim \mathbf{Bin}(K, p)$ for some positive integer K and $p \in [0, 1]$.*

(a) For any $\epsilon \geq 0$, $\mathbb{P}(|Y - pK| > \epsilon pK) \leq 2e^{-\frac{\epsilon^2}{2+\epsilon}pK}$.
(b) If $p = j/K$ for some integer $j \in (0, K)$, then $\mathbb{P}(Y = j) \geq \frac{1}{3}(p(1-p)K)^{-1/2}$.
(c) Let $\mathcal{I} \sim \mathbf{Po}(\mu)$ for some $\mu > 0$. Then, for any $\epsilon \geq 0$, $\mathbb{P}(|\mathcal{I} - \mu| \geq \epsilon\mu) \leq 2e^{-\frac{\epsilon^2}{2+\epsilon}\mu}$.

The following lemma is similar to [18, Lemma 6], whose proof easily extends to more general degree sequences $\mathbf{d}$ discussed in this paper through straightforward adjustments. We include a short proof here. We will use this lemma to gain information on the remaining degree sequence of $\mathcal{G}(n, \mathbf{d})$ when part of it has been constructed. Note that the probability lower bound $1 - \exp\left(-\Omega(\xi^2 p_m d_j) + 2\log n\right)$ below is not necessarily nonnegative. If it is negative then the assertion is trivially true.

https://doi.org/10.1017/S0963548324000300 Published online by Cambridge University Press

Lemma 2.6. *Let $0 \le m < \|\boldsymbol{d}\|_1/2$ and $p_m = (\|\boldsymbol{d}\|_1 - 2m)/\|\boldsymbol{d}\|_1$. Let $d_j^{G_m}$ denote the degree of j in G_m. For any $\xi = \xi_n \in (0, 1)$, $|d_j - d_j^{G_m} - p_m d_j| \le \xi p_m d_j$ for all $j \in [n]$ with probability at least $1 - \exp\left(-\Omega(\xi^2 p_m d_j) + 2\log n\right)$.*

Proof. Take $G \sim \mathcal{G}(n, \mathbf{d})$ and let $\mathbf{h} = (h_1, \dots, h_n)$ be the degree sequence of the graph H obtained by independently keeping every edge of G with probability p_m. Then, $h_j \sim \mathbf{Bin}(d_j, p_m)$ for every $j \in [n]$. By Lemma 2.3, conditioned on the event that $|E(H)| = \|\mathbf{d}\|_1/2 - m$, $\mathbf{h}$ has the same distribution as $\mathbf{d} - \mathbf{d}^{G_m}$. Therefore, by Lemma 2.5 (a,b), for every $j \in [n]$,

$$\mathbb{P}(|d_j - d_j^{G_m} - p_m d_j| \ge \xi p_m d_j) \le \frac{\mathbb{P}(|h_j - p_m d_j| \ge \xi p_m d_j)}{\mathbb{P}(|E(H)| = \|\mathbf{d}\|_1/2 - m)} \le \frac{2e^{-\frac{\xi^2}{2+\xi} p_m d_j}}{\frac{1}{3}(p_m(1-p_m)\|\mathbf{d}\|_1/2)^{-1/2}}$$
$$\le \sqrt{\|\mathbf{d}\|_1} \cdot e^{-\Omega(\xi^2 p_m d_j)} = \exp\left(-\Omega(\xi^2 p_m d_j) + \log n\right),$$

as $\|\mathbf{d}\|_1 \le n^2$. The lemma follows by taking the union bound over $j \in [n]$. □

2.4 Couple the two sequential sampling procedures

Finally we couple the two aforementioned procedures and define procedure Coupling($\mathbf{d}$, λ, Λ) which sequentially constructs $\mathcal{G}(n, W)$ and $\mathcal{G}(n, \mathbf{d})$ together. The central idea of Coupling is that marginally, the constructions of $G_L \sim \mathcal{G}(n, W)$ and $G \sim \mathcal{G}(n, \mathbf{d})$ follow precisely the procedures SeqApprox-P and SeqSample-D, respectively. As in SeqApprox-P, edges jk in K_n are sequentially sampled independently with probability proportional to $d_j d_k$. If jk was already added to G then jk is accepted by G_L with probability Λ_{jk} as in SeqApprox-P. If jk was not in G, then jk is accepted to G with some probability η_{jk} so that the probability that jk is selected and accepted by G is proportional to the probability that jk is in $\mathcal{G}(n, \mathbf{d})$, conditional on the current construction of G, as desired in SeqSample-D. On the other hand, jk is accepted by G_L with probability Λ_{jk} as required by SeqApprox-P. If $\eta_{jk} \ge \Lambda_{jk}$ then we can couple the two operations so that jk is added to G_L only when it is added to G. On the other hand, if $\eta_{jk} < \Lambda_{jk}$ then we simply sample $G_L \sim \mathcal{G}(n, W)$ and $G \sim \mathcal{G}(n, \mathbf{d})$ independently. We can prove that the latter case happens rarely. The formal description of Coupling is given below.

---

procedure Coupling($\mathbf{d}$, λ, Λ)
  Let L_0 and G_0 be empty graphs on vertex set $[n]$.
  Let $\mathcal{I} \sim \mathbf{Po}(\lambda)$.
  for i in $1, \dots, \mathcal{I}$ **do**
    Pick an edge jk of K_n with probability proportional to $d_j d_k$ (i.e. with probability Q_{jk}).
    if $jk \in G_{i-1}$ **then**
      $G_i = G_{i-1}$;
      $L_i = L_{i-1} \cup \{jk\}$ with probability Λ_{jk},
      $L_i = L_{i-1}$ with probability $1 - \Lambda_{jk}$.

---

https://doi.org/10.1017/S0963548324000300 Published online by Cambridge University Press

---

        else
            Let $\eta_{jk}^{(i)} = \frac{\rho_i(jk)}{\max_{h\ell \notin G_{i-1}} \rho_i(h\ell)}$, where $\rho_i(h\ell) = (d_h d_\ell)^{-1}\mathbb{P}(h\ell \in \mathcal{G}(n, \mathbf{d}) \mid G_{i-1})$.
            if $\eta_{jk}^{(i)} < \Lambda_{jk}$ **then**
                Return IndSample($\mathbf{d}, \lambda, \Lambda$)
            else
                $G_i = G_{i-1} \cup \{jk\}$ and $L_i = L_{i-1} \cup \{jk\}$ with probability Λ_{jk},
                $G_i = G_{i-1} \cup \{jk\}$ and $L_i = L_{i-1}$ with probability $\eta_{jk}^{(i)} - \Lambda_{jk}$,
                $G_i = G_{i-1}$ and $L_i = L_{i-1}$ with probability $1 - \eta_{jk}^{(i)}$.
            end if
        end if
    end for
    for $i \geq \mathcal{I} + 1$, while G_{i-1} has fewer edges than $\mathcal{G}(n, \mathbf{d})$ **do**
        Pick an edge $jk \notin G_{i-1}$ with probability proportional to $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}) \mid G_{i-1})$;
        $G_i = G_{i-1} \cup \{jk\}$.
    end for
    Return (G_L, G), where $G = G_i$ and $G_L = L_{\mathcal{I}}$.
end procedure
procedure IndSample($\mathbf{d}, \lambda, \Lambda$)
    Independently sample $G \sim \mathcal{G}(n, \mathbf{d})$, and $G_L \sim \mathcal{G}(n, f(\Lambda \odot Q))$ where
        $f(x) = 1 - \exp(-\lambda x)$.
    Return (G_L, G)
end procedure

Lemma 2.7. *Let (G_L, G) be the output of Coupling($\mathbf{d}, \lambda, \Lambda$).*

(a) $G_L \sim \mathcal{G}(n, f(\Lambda \odot Q))$ *where* $Q = Q(\mathbf{d})$ *and* $f(x) = 1 - \exp(-\lambda x)$.
(b) $G \sim \mathcal{G}(n, \mathbf{d})$.
(c) If IndSample($\mathbf{d}, \lambda, \Lambda$) is not called during the execution of Coupling($\mathbf{d}, \lambda, \Lambda$), then $G_L \subseteq G$ *in the output of Coupling($\mathbf{d}, \lambda, \Lambda$).*

Proof. Parts (a,b) are trivially true if IndSample($\mathbf{d}, \zeta$) is called. Now assume that IndSample($\mathbf{d}, \zeta$) is not called. Part (c) follows directly by the coupling procedure, as every edge is added to G_L only when it is, or has been added to G. For (a), note that if IndSample($\mathbf{d}, \zeta$) is not called then the edges are sequentially added to G_L exactly as in SeqApprox-P($\mathbf{d}, \lambda, \Lambda$), and thus part (a) follows by Lemma 2.2. For part (b), notice that in each step i, edge jk is added to G_{i-1} with probability

https://doi.org/10.1017/S0963548324000300 Published online by Cambridge University Press

$Q_{jk}\eta_{jk}^{(i)}$, which is proportional $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}) \mid G_{i-1})$. Thus, the edges are added to G exactly as in the execution of SeqSample-D($\mathbf{d}$), and part (b) follows by Corollary 2.4. □

2.5 Specify λ and Λ for the coupling procedure

By the assumptions $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$ and $\delta(\mathbf{d}) \gg \log n$, there exist $\xi, \zeta, \zeta' = o(1)$ that go to zero sufficiently slowly so that

$$\zeta' \gg \|\mathbf{d}\|_1^{-1/3} \tag{2}$$

$$\zeta'\xi^2 \gg \frac{\log n}{\delta(\mathbf{d})} \tag{3}$$

$$\zeta' \gg \frac{J(\mathbf{d})}{\|\mathbf{d}\|_1} \tag{4}$$

$$\zeta \gg \frac{J(\mathbf{d})}{\zeta'\|\mathbf{d}\|_1} + \xi. \tag{5}$$

To see why they exist, notice that there exists $\zeta' = o(1)$ satisfying [4] since $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$. Moreover, since $\delta(\mathbf{d}) \gg \log n$, we may assume that ζ' goes to 0 sufficiently slowly so that [2] is satisfied and there exists ξ satisfying [3]. Finally, given [4] and $\xi = o(1)$, the right-hand side of [5] is $o(1)$, and hence there exists ζ satisfying [5].

Choose ξ, ζ, ζ' that satisfy all the conditions above. For the coupling procedure, we set

$$\lambda = (1-\zeta')\|\mathbf{d}\|_1/2 \tag{6}$$

$$\Lambda_{jk} = (1-\zeta)\frac{\|\mathbf{d}\|_1}{\|\mathbf{d}\|_1 + d_jd_k} \quad \text{for every } 1 \le j < k \le n. \tag{7}$$

3. Proof of Theorem 1.8

Let $\Delta = \Delta(\mathbf{d})$, $Q = Q(\mathbf{d})$, and $P = P(\mathbf{d})$. By Lemma 2.7, it suffices to show that

$$\mathbb{P}(\text{IndSample}(\mathbf{d}, \zeta) \text{ is called}) = o(1) \tag{8}$$

$$\lambda\Lambda \odot Q = (1 + o(1))P, \tag{9}$$

as $1 - \exp\left(-(1+o(1))P_{ij}\right) = (1+o(1))(1-e^{-P_{ij}})$ for every ij.

Proof of (9). For every $1 \le j < k \le n$,

$$Q_{jk} = \frac{d_jd_k}{\sum_{1\le h<\ell\le n} d_hd_\ell} = \frac{d_jd_k}{\frac{1}{2}(\|\mathbf{d}\|_1^2 - \sum_{i=1}^n d_i^2)} = \frac{2d_jd_k}{\|\mathbf{d}\|_1^2 - O(\Delta)\|\mathbf{d}\|_1}.$$

Since $\lambda = (1-\zeta')\|\mathbf{d}\|_1/2$ by [6], and by [7]

$$\begin{aligned}\lambda\Lambda_{jk}Q_{jk} &= \frac{(1-\zeta')\|\mathbf{d}\|_1}{2}\frac{(1-\zeta)\|\mathbf{d}\|_1}{\|\mathbf{d}\|_1 + d_jd_k}\frac{2d_jd_k}{\|\mathbf{d}\|_1^2(1+O(\Delta/\|\mathbf{d}\|_1)}\\ &= \left(1 + O\left(\zeta + \zeta' + \frac{\Delta}{\|\mathbf{d}\|_1}\right)\right)\frac{d_jd_k}{\|\mathbf{d}\|_1 + d_jd_k},\end{aligned}$$

https://doi.org/10.1017/S0963548324000300 Published online by Cambridge University Press

and (9) follows by noting that $\zeta, \zeta', \Delta/\|\mathbf{d}\|_1 = o(1)$. □

Proof of (8). By Lemma 2.5(c) (with $\mu = \lambda$ and $\epsilon = \lambda^{-1/3}$) and [6], a.a.s. $\mathcal{I} = (1 + O(\lambda^{-1/3}))\lambda = (1 - \zeta' + O(\|\mathbf{d}\|_1^{-1/3}))\|\mathbf{d}\|_1/2$. It suffices then to show that a.a.s. throughout the execution of Coupling($\mathbf{d}, \lambda, \Lambda$), $\eta^{(i)}_{x_i} \geq \Lambda_{x_i}$ for every $1 \leq i \leq \mathcal{I}$, where x_i is the random edge sampled in the ith iteration. Let m_i be the number of edges in G_i. Since a.a.s. $\mathcal{I} = (1 - \zeta' + O(\|\mathbf{d}\|_1^{-1/3}))\|\mathbf{d}\|_1/2$ and $m_i \leq \mathcal{I}$ for every $1 \leq i \leq \mathcal{I}$, it follows then by [2] that a.a.s.

$$p_{m_i} \geq p_{m_\mathcal{I}} \geq \zeta'/2 \quad \text{for every } 1 \leq i \leq \mathcal{I}, \tag{10}$$

where p_{m_i} is defined by $1 - 2m_i/\|\mathbf{d}\|_1$ as in Lemma 2.6. □

By Lemma 2.6 (with ξ chosen in Section 2.5), [10] and [3], with probability at least

$$1 - \exp\left(-\Omega(\xi^2 p_m d_j) + 2\log n\right) \geq 1 - \exp\left(-\Omega(\xi^2\zeta'\delta(\mathbf{d})) + 2\log n\right) = 1 - \exp\left(-\omega(\log n)\right),$$

we have that $|d_j - d_j^{G_i} - p_{m_i}d_j| \leq \xi p_{m_i} d_j$ for all $j \in [n]$. Now take the union bound over all the $\mathcal{I} \leq \|\mathbf{d}\|_1/2$ steps, we conclude that

$$d_j - d_j^{G_i} = (1 + O(\xi))p_{m_i}d_j \quad \text{for every } j \in [n], \text{and for every } i \leq \mathcal{I}. \tag{11}$$

Next, we estimate $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}) \mid G_i)$. Note that Proposition 1.3 gives $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}) \mid G_i)$ when $G_i = \emptyset$. Here, we apply another corollary of [20], [Theorem 1], given in Theorem 3.1 below, which estimates the edge probabilities in $\mathcal{G}(n, \mathbf{d})$ when conditioned on a set of edges H being present in $\mathcal{G}(n, \mathbf{d})$.

For two degree sequences $\mathbf{d}$ and $\mathbf{g}$, we say $\mathbf{d} \preceq \mathbf{g}$ if $d_i \leq g_i$ for every $i \in [n]$. Given a graph H on $[n]$, let $\mathbf{d}^H$ denote the degree sequence of H.

Theorem 3.1. *Suppose H is a graph on $[n]$ with $\boldsymbol{d}^H \preceq \boldsymbol{d}$, and let $\boldsymbol{t} = \boldsymbol{d} - \boldsymbol{d}^H$. Suppose that $J(\boldsymbol{d}) = o(\|\boldsymbol{t}\|_1)$ and suppose $jk \notin H$. Then*

$$\mathbb{P}(jk \in \mathcal{G}(n, \boldsymbol{d}) \mid H) = \left(1 + O\left(\frac{J(\boldsymbol{d})}{\|\boldsymbol{t}\|_1}\right)\right)\frac{t_j t_k}{\|\boldsymbol{t}\|_1 + t_j t_k},$$

where $\mathbb{P}(jk \in \mathcal{G}(n, \boldsymbol{d}) \mid H)$ denotes the probability that $jk \in \mathcal{G}(n, \boldsymbol{d})$ conditional on the event that $\mathcal{G}(n, \boldsymbol{d})$ contains H as a subgraph.

We first verify that the assumption $J(\mathbf{d}) = o(\|\mathbf{t}\|_1)$ of Theorem 3.1 is satisfied for every $i \leq \mathcal{I}$ with $H = G_i$. By (10) we may assume that $p_{m_i} \geq \zeta'/2$ for every $1 \leq i \leq \mathcal{I}$ (which holds a.a.s.). Let $\mathbf{t}_i = \mathbf{d} - \mathbf{d}^{G_i}$. Then,

$$\|\mathbf{t}_i\|_1 = \|\mathbf{d}\|_1 - 2m_i = \|\mathbf{d}\|_1(1 - 2m_i/\|\mathbf{d}\|_1) = \|\mathbf{d}\|_1 p_{m_i} \geq \|\mathbf{d}\|_1\zeta'/2. \tag{12}$$

By [4], $J(\mathbf{d}) \ll \zeta'\|\mathbf{d}\|_1$ and consequently $J(\mathbf{d})/\|\mathbf{t}_i\|_1 = O(J(\mathbf{d})/\zeta'\|\mathbf{d}\|_1) = o(1)$ for every $1 \leq i \leq \mathcal{I}$. By Theorem 3.1, [11] and [12], a.a.s.

$$\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}) \mid G_i) = \left(1 + O\left(\frac{J(\mathbf{d})}{p_{m_i}\|\mathbf{d}\|_1} + \xi\right)\right)\frac{p_{m_i}d_j d_k}{\|\mathbf{d}\|_1 + p_{m_i}d_j d_k}, \quad \text{for every } 1 \leq i \leq \mathcal{I},$$

where $J(\mathbf{d})/p_{m_i}\|\mathbf{d}\|_1 + \xi = O(J(\mathbf{d})/\zeta'\|\mathbf{d}\|_1 + \xi) = o(1)$ as shown before. Recall that

$$\rho_i(jk) = \frac{\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d}) \mid G_i)}{d_j d_k}.$$

It follows that

$$\rho_i(jk) = \left(1 + O\left(\frac{J(\mathbf{d})}{\zeta'\|\mathbf{d}\|_1} + \xi\right)\right)\frac{p_{m_i}}{\|\mathbf{d}\|_1 + p_{m_i}d_j d_k}.$$

https://doi.org/10.1017/S0963548324000300 Published online by Cambridge University Press

Notice that

$$\frac{p_{m_i}}{\|\mathbf{d}\|_1 + p_{m_i} d_j d_k} \le \frac{p_{m_i}}{\|\mathbf{d}\|_1} \quad \text{for all } jk \notin G_{i-1}.$$

Thus,

$$\max_{h\ell \notin G_{i-1}} \rho_i(h\ell) \le \left(1 + O\left(\frac{J(\mathbf{d})}{\zeta' \|\mathbf{d}\|_1} + \xi\right)\right) p_{m_i} / \|\mathbf{d}\|_1.$$

Hence,

$$\eta_{jk}^{(i)} \ge \left(1 + O\left(\frac{J(\mathbf{d})}{\zeta' \|\mathbf{d}\|_1} + \xi\right)\right) \frac{\|\mathbf{d}\|_1}{\|\mathbf{d}\|_1 + p_{m_i} d_j d_k} \ge (1 - \zeta) \frac{\|\mathbf{d}\|_1}{\|\mathbf{d}\|_1 + d_j d_k},$$

by [5]. Now [8] follows.

Acknowledgement

We thank two anonymous referees for their valuable advice in improving the presentation of the paper.

Funding statement

Research supported by NSERC RGPIN-04173-2019.

References

[1] Anastos, M., Frieze, A. and Gao, P. (2021) Hamiltonicity of random graphs in the stochastic block model. *SIAM J. Discrete Math.* **35** 1854–1880.
[2] Barvinok, A. (2010) On the number of matrices and a random matrix with prescribed row and column sums and 0-1 entries. *Adv. Math.* **224** 316–339.
[3] Barvinok, A. and Hartigan, J. A. (2013) The number of graphs and a random graph with a given degree sequence. *Random Struct. Algor.* **42** 301–348.
[4] Bender, E. A. and Canfield, E. R. (1978) The asymptotic number of labeled graphs with given degree sequences. *J. Combin. Theory, Ser. A.* **24** 296–307.
[5] Bohman, T. and Picollelli, M. (2012) Sir epidemics on random graphs with a fixed degree sequence. *Random Struct. Algor.* **41** 179–214.
[6] Bollobás, B. (1980) A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.* **1** 311–316.
[7] Bollobás, B. and Bollobás, B. (1998) *Random graphs*. Springer.
[8] Chatterjee, S., Diaconis, P. and Sly, A. (2011) Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435.
[9] Chung, F. and Lu, L. (2002) The average distances in random graphs with given expected degrees. *Proc. Natl. Acad. Sci.* **99** 15879–15882.
[10] Chung, F. and Lu, L. (2002) Connected components in random graphs with given expected degree sequences. *Ann. Comb.* **6** 125–145.
[11] Chung, F. and Lu, L. (2006) The volume of the giant component of a random graph with given expected degrees. *Siam. J. Discrete Math.* **20** 395–411.
[12] Chung, F., Lu, L. and Vu, V. (2004) The spectra of random graphs with given expected degrees. *Internet Math.* **1** 257–275.
[13] Cooper, C., Frieze, A. and Reed, B. (2002) Random regular graphs of non-constant degree: connectivity and hamiltonicity. *Combin. Probab. Comput.* **11** 249–261.
[14] Dudek, A., Frieze, A., Ruciński, A. and Šileikis, M. (2017) Embedding the Erdős–Rényi hypergraph into the random regular hypergraph and Hamiltonicity. *J. Combin. Theory, Ser. B.* **122** 719–740.
[15] Fountoulakis, N., Kühn, D. and Osthus, D. (2008) The order of the largest complete minor in a random graph. *Random Struct. Algor.* **33** 127–141.
[16] Frankston, K., Kahn, J., Narayanan, B. and Park, J. (2021) Thresholds versus fractional expectation-thresholds. *Ann. Math.* **194** 475–495.
[17] Gao, P., Isaev, M. and McKay, B. D. (2020) Sandwiching random regular graphs between binomial random graphs. In Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM. pp. 690–701.

[18] Gao, P., Isaev, M. and McKay, B. D. (2022) Sandwiching dense random regular graphs between binomial random graphs. *Probab. Theory Rel.* **184** 115–158.
[19] Gao, P., Isaev, M. and McKay, B. D. (2023), Kim-vu's sandwich conjecture is true for all $d \geq \log^4 n$, arXiv preprint arXiv:2011.09449.
[20] Gao, P. and Ohapkin, Y. (2023) Subgraph probability of random graphs with specified degrees and applications to chromatic number and connectivity. *Random Struct. Algor.* **62** 911–934.
[21] Gao, P. and Wormald, N. (2016) Enumeration of graphs with a heavy-tailed degree sequence. *Adv. Math.* **287** 412–450.
[22] Gao, Z. and Wormald, N. C. (2008) Distribution of subgraphs of random regular graphs. *Random Struct. Algor.* **32** 38–48.
[23] Holland, P. W., Laskey, K. B. and Leinhardt, S. (1983) Stochastic blockmodels: first steps. *Soc. Networks.* **5** 109–137.
[24] Isaev, M. and McKay, B. D. (2018) Complex martingales and asymptotic enumeration. *Random Struct. Algor.* **52** 617–661.
[25] Kim, J. H. and Vu, V. H. (2004) Sandwiching random graphs: universality between random graph models. *Adv. Math.* **188** 444–469.
[26] Klimošová, T., Reiher, C., Ruciński, A. and Šileikis, M. (2023) Sandwiching biregular random graphs. *Combin. Probab. Comput.* **32** 1–44.
[27] Krivelevich, M., Sudakov, B., Vu, V. H. and Wormald, N. C. (2001) Random regular graphs of high degree. *Random Struct. Algor.* **18** 346–363.
[28] Liebenau, A. and Wormald, N. (2017) Asymptotic enumeration of graphs by degree sequence, and the degree sequence of a random graph, arXiv preprint arXiv: 1702.08373.
[29] Lusher, D., Koskinen, J. and Robins, G. (2013) *Exponential random graph models for social networks: theory, methods, and applications*. Cambridge University Press.
[30] McKay, B. D. (1985) Asymptotics for symmetric 0-1 matrices with prescribed row sums. *Ars. Combin.* **19** 15–25.
[31] McKay, BD. and Wormald, N. C. (1990) Asymptotic enumeration by degree sequence of graphs of high degree. *Eur. J. Combin.* **11** 565–580.
[32] McKay, B. D. and Wormald, N. C. (1991) Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$. *Combinatorica.* **11** 369–382.
[33] Morris, J. F., O'Neal, J. W. and Deckro, R. F. (2014) A random graph generation algorithm for the analysis of social networks. *J. Def. Model. Simul.* **11** 265–276.
[34] Newman, M. E. J., Strogatz, S. H. and Watts, D. J. (2001) Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E.* **64** 026118.
[35] Van Der Hofstad, R. (2024) *Random graphs and complex networks*. Cambridge University Press.
[36] Wormald, N. C., et al. (1999) Models of random regular graphs, *London mathematical society lecture note series*, 239–298.

Appendix

We prove a few properties of degree sequences $\mathbf{d}$ where $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$.

Proposition 3.2. *Let* $\mathbf{d} = \mathbf{d}_n$ *be a sequence of degree sequences such that* $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$. *Let* $\Delta = \max\{d_i, i \in [n]\}$. *Then,*

(a) $\Delta = o(n)$.

(b) *For every* $\epsilon > 0$, $|\{ij{:}d_i d_j \geq \epsilon \|\mathbf{d}\|_1\}| \leq \epsilon n^2$ *for all sufficiently large* n.

Proof. *For (a), suppose on the contrary that* $\Delta \geq \delta n$ *for some absolute constant* $\delta > 0$. *Then,*

$$\|\mathbf{d}\|_1 = \sum_{i=1}^{n} d_i \leq \delta^{-1} \sum_{i=1}^{\delta n} d_i \leq \delta^{-1} J(\mathbf{d}),$$

contradicting with $J(\mathbf{d}) = o(\|\mathbf{d}\|_1)$.

For (b), note that $J(\mathbf{d}) \geq \Delta d_\Delta$ *and thus* $\Delta d_\Delta = o(\|\mathbf{d}\|_1)$. *It follows that for every* $i, j \in [n]$ *such that* $j \geq \Delta$, $d_i d_j \leq \Delta d_\Delta = o(\|\mathbf{d}\|_1)$. *Thus, part (b) follows by* $\Delta = o(n)$ *from part (a).* □

Example 3.3. *We construct an example of degree sequences* $\mathbf{d} \preceq \mathbf{g}$ *for which there exists* jk *such that* $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{d})) = 1 - o(1)$ *and* $\mathbb{P}(jk \in \mathcal{G}(n, \mathbf{g})) = o(1)$. *Let* $\mathbf{t}$ *be a degree sequence where* $t_1 = t_2 = n^{2/3}$, *and* $t_3 = \cdots = t_n = 1$ *(without loss of generality we assume that* $n^{2/3}$ *is an integer and* $\|\mathbf{t}\|_1$ *is even). Let* $\mathbf{g} = (n-1)\mathbf{1} - \mathbf{t}$, *and let* $\mathbf{d}$ *be* $(n - 2n^{2/3})$*-regular. Obviously,* $\mathbf{d} \preceq \mathbf{g}$. *By symmetry,*

https://doi.org/10.1017/S0963548324000300 Published online by Cambridge University Press

the probability that vertices 1 and 2 are adjacent in $\mathcal{G}(n, \mathbf{d})$ *is equal to* $(n - 2n^{2/3})/(n-1) = 1 - o(1)$. *By Proposition 1.3, the probability that these two vertices are adjacent in* $\mathcal{G}(n, \mathbf{t})$ *is* $1 - o(1)$, *and consequently, their adjacency probability in* $\mathcal{G}(n, \mathbf{g})$ *is* $o(1)$, *as* $\mathcal{G}(n, \mathbf{g})$ *has the same distribution as the complement of* $\mathcal{G}(n, \mathbf{t})$.

Cite this article: Gao P and Ohapkin Y (2025). Embedding theorems for random graphs with specified degrees. *Combinatorics, Probability and Computing* **34**, 115–130. https://doi.org/10.1017/S0963548324000300