The Annals of Applied Probability
2011, Vol. 21, No. 4, 1400–1435
DOI: 10.1214/10-AAP728

RANDOM GRAPHS WITH A GIVEN DEGREE SEQUENCE

BY SOURAV CHATTERJEE[1], PERSI DIACONIS AND ALLAN SLY

New York University, Stanford University and Microsoft Research

Large graphs are sometimes studied through their degree sequences (power law or regular graphs). We study graphs that are uniformly chosen with a given degree sequence. Under mild conditions, it is shown that sequences of such graphs have graph limits in the sense of Lovász and Szegedy with identifiable limits. This allows simple determination of other features such as the number of triangles. The argument proceeds by studying a natural exponential model having the degree sequence as a sufficient statistic. The maximum likelihood estimate (MLE) of the parameters is shown to be unique and consistent with high probability. Thus n parameters can be consistently estimated based on a sample of size one. A fast, provably convergent, algorithm for the MLE is derived. These ingredients combine to prove the graph limit theorem. Along the way, a continuous version of the Erdős–Gallai characterization of degree sequences is derived.

1. Introduction.

1.1. *Graphs with a given degree sequence.* Let G be an undirected simple graph on n vertices and let $d_1, \dots, d_n$ be the degrees of the vertices of G. The vector $\mathbf{d} := (d_1, \dots, d_n)$ is usually called the *degree sequence* of G. Correspondingly, the *degree distribution* of G is the probability distribution function F supported on [0, 1], defined as

$$F(x) := \frac{|\{i : d_i \le nx\}|}{n}.$$

In other words, if a vertex is chosen uniformly at random, then the degree of that vertex, divided by n, is a random variable with probability distribution function F.

In recent years, the degree distributions of real world networks have received wide attention. The surveys [41, 42] contain many references as does the detailed account in [12]. The enthusiasm of some authors for "scale free" or "power law graphs" has also generated much controversy [33, 53] which serves as additional motivation for the present paper.

The interest in degree distributions stems from the fact that the degree sequences of real world networks sometimes appear to have power law behavior

Received May 2010; revised August 2010.

[1]Supported in part by NSF Grant DMS-07-07054 and a Sloan Research Fellowship.

MSC2010 subject classifications. 05A16, 05C07, 05C30, 52B55, 60F05, 62F10, 62F12.

Key words and phrases. Random graph, degree sequence, Erdős–Gallai criterion, threshold graphs, graph limit.

that is very different than those occurring in classical models of random graphs, like the Erdős–Rényi model [21]. Researchers have tried various ways of circumventing this problem. An obvious solution is to build random graph models that are forced to give us the degree distribution that we want and then deduce other features by simulation or mathematics. A natural way to do this is to choose a graph uniformly at random from the set of all graphs with a given degree sequence. One frequent appearance of this model is for random regular graphs [54]. As explained in [12], Section 13, the model also arises in testing if the exponential family with degree sequence as sufficient statistic fits a given data set. See [51] for applications where the number of triangles is wanted. The paper [12] has useful ways of simulating graphs with a given degree sequence and an extensive survey of the (mostly nonrigorous) literature for this model. Some rigorous results are also available in the "sparse case," for example, those in [39, 40].

At this point, a gap between our motivation and our theory must be pointed out: the present paper deals with *dense* graphs with a given degree sequence (roughly, graphs whose number of edges is comparable to the square of the number of vertices), whereas much of the literature cited above, for example, power law graphs, revolves around sparse graphs. As of now, our theorems are not directly applicable in the sparse setting, although there is certainly hope for future progress.

In a recent series of papers [5–10], Barvinok and Hartigan have looked at problems related to the structure of directed and undirected (dense) graphs with given degree sequence. The Barvinok and Hartigan work, especially [10], is related to the present paper. This is explained at the end of this Introduction after we have stated our main theorems.

One of the objectives of this article is to give a rather precise description of the structure of random (dense) graphs with a given degree sequence via the notion of graph limits introduced recently by Lovász and Szegedy [34] and developed by Borgs et al. [13–15]. See also the related work of Diaconis and Janson [19] and Austin [2] which traces this back to work of Aldous [1] and Hoover [28]. This gives, in particular, a way to write down exact formulas for the expected number of subgraphs of a given type without simulation.

Before stating our result, we need to introduce the notion of graph limits. We quote the definition verbatim from [34] (see also [14, 15, 19]). Let G_n be a sequence of simple graphs whose number of nodes tends to infinity. For every fixed simple graph H, let $|\hom(H, G)|$ denote the number of homomorphisms of H into G [i.e., edge-preserving maps $V(H) \to V(G)$, where $V(H)$ and $V(G)$ are the vertex sets]. This number is normalized to get the homomorphism density

$$t(H, G) := \frac{|\hom(H, G)|}{|V(G)|^{|V(H)|}}. \tag{1}$$

This gives the probability that a random mapping $V(H) \to V(G)$ is a homomorphism.

Suppose that the graphs G_n become more and more similar in the sense that $t(H, G_n)$ tends to a limit $t(H)$ for every H. One way to define a limit of the sequence $\{G_n\}$ is to define an appropriate limit object from which the values $t(H)$ can be read off.

The main result of [34] (following the earlier equivalent work of Aldous [1] and Hoover [28]) is that indeed there is a natural "limit object" in the form of a symmetric measurable function $W:[0,1]^2 \to [0,1]$ [we call W symmetric if $W(x,y) = W(y,x)$]. Conversely, every such function arises as the limit of an appropriate graph sequence. This limit object determines all the limits of subgraph densities: if H is a simple graph with $V(H) = [k] = \{1, \dots, k\}$, then

$$t(H, W) = \int_{[0,1]^k} \prod_{(i,j) \in E(H)} W(x_i, x_j)\, dx_1 \cdots dx_k.$$

Here $E(H)$ denotes the edge set of H.

Intuitively, the interval $[0,1]$ represents a "continuum" of vertices and $W(x,y)$ denotes the probability of putting an edge between x and y. For example, for the Erdős–Rényi graph $G_{n,p}$, if p is fixed and $n \to \infty$, then the limit graph is represented by the function that is identically equal to p on $[0,1]^2$.

Convergence of a sequence of graphs to a limit has many consequences. From the definition, the count of fixed size subgraphs converges to the right- hand side of the expression for $t(H, W)$ given above. More global parameters also converge. For example, the degree distribution converges to the law of $\int_0^1 W(U, y)\, dy$ where U is a random variable distributed uniformly on $[0,1]$. Similarly, the distribution function of the eigenvalues of the adjacency matrix converges. More generally, a graph parameter is a function from the space of graphs into a space $\mathcal{X}$ which is invariant under isomorphisms. If $\mathcal{X}$ is a topological space, we may ask which graph parameters are continuous with respect to the topology induced by graph limits. This is called "property testing" in the computer science theory literature which has identified many continuous graph parameters. See the surveys [3, 14] for pointers to the literature.

We are now ready to state our result about the limit of graphs with given degree sequences. Suppose that for each n, a degree sequence $\mathbf{d}^n = (d_1^n, \dots, d_n^n)$ is given. Without loss of generality, assume that $d_1^n \geq d_2^n \geq \cdots \geq d_n^n$. We say that the sequence $\{\mathbf{d}^n\}$ has a *scaling limit* if there is a nonincreasing function f on $[0,1]$ such that

$$\lim_{n\to\infty} \left(\left| \frac{d_1^n}{n} - f(0) \right| + \left| \frac{d_n^n}{n} - f(1) \right| + \frac{1}{n} \sum_{i=1}^n \left| \frac{d_i^n}{n} - f\left(\frac{i}{n}\right) \right| \right) = 0. \tag{2}$$

It is not difficult to prove by a simple compactness argument that any sequence $\{\mathbf{d}^n\}$ of degree sequences has a subsequence that converges to a scaling limit in the above sense. Note that convergence in the above sense can be stated equivalently in terms of convergence of degree distributions: $d_1^n/n \to f(0)$, $d_n^n/n \to f(1)$ and

$D_n/n \to f(U)$ in distribution, where D_n is a randomly (uniformly) chosen d_i^n and U is uniformly distributed on $[0, 1]$.

The need to control d_1^n and d_n^n arises from the need to eliminate "outlier" vertices that connect to too many or too few nodes which takes the degree sequence too close to the Erdős–Gallai boundary (see below). Since d_i^n is decreasing in i, outliers can be eliminated by simply controlling d_1^n and d_n^n. The need to eliminate outliers, on the other hand, arises from technical aspects of our analysis.

Define $D'[0, 1]$ to be the set of nonincreasing functions on $[0, 1]$ which are left continuous on $(0, 1)$. The reason for imposing left-continuity is the following: When the scaling limit of a degree sequence is discontinuous, it is not uniquely defined but there always exists a unique limit in $D'[0, 1]$. We could have as well chosen right-continuous.

For each n, let G_n be a random graph chosen uniformly from the set of all simple graphs with degree sequence $\mathbf{d}^n$. Let f be the scaling limit of the sequence $\{\mathbf{d}^n\}$ in the sense defined above. Our objective is to compute the limit of the sequence $\{G_n\}$ in terms of the scaling limit of $\mathbf{d}^n$. We endow the set of scaling limits (i.e., $D'[0, 1]$) with the topology induced by a modified L^1 norm $\|\cdot\|_{1'}$ given by

$$\|f\|_{1'} := |f(0)| + |f(1)| + \int_0^1 |f(x)|\,dx.$$

The choice of this norm is necessitated by the need to make it compatible with our previous notion of convergence of degree sequences.

Not all functions can be scaling limits of degree sequences. Let $\mathcal{F}$ be the set of functions in $D'[0, 1]$ that can be obtained as scaling limits of degree sequences in the sense stated above. By a simple diagonal argument, it is easy to see that $\mathcal{F}$ is a closed subset of $D'[0, 1]$ under the topology of the modified L^1 norm. It is shown in Proposition 1.2 that $\mathcal{F}$ has nonempty interior.

THEOREM 1.1. *Let G_n and f be as above. Suppose that f belongs to the topological interior of the set $\mathcal{F}$ defined above. Then there exists a unique function $g : [0, 1] \to \mathbb{R}$ in $D'[0, 1]$ such that the function*

$$W(x, y) := \frac{e^{g(x)+g(y)}}{1 + e^{g(x)+g(y)}}$$

satisfies, for all $x \in [0, 1]$,

$$f(x) = \int_0^1 W(x, y)\,dy.$$

In this situation, the sequence $\{G_n\}$ converges almost surely to the limit graph represented by the function W.

Theorem 1.1 can be useful only if we can provide a simple way of checking whether f belongs to the interior of $\mathcal{F}$. (Being the limit of a sequence of degree sequences, it is clear that $f \in \mathcal{F}$. The nontrivial question is whether f is in the interior.) The following result gives an easily verifiable equivalent condition.

PROPOSITION 1.2. *A function* $f:[0,1]\to[0,1]$ *in* $D'[0,1]$ *belongs to the interior of* $\mathcal{F}$ *if and only if*:

(i) *there are two constants* $c_1>0$ *and* $c_2<1$ *such that* $c_1\le f(x)\le c_2$ *for all* $x\in[0,1]$ *and*

(ii) *for each* $x\in(0,1]$,

$$\int_x^1 \min\{f(y),x\}\,dy + x^2 - \int_0^x f(y)\,dy > 0.$$

REMARK 1. Condition (ii) in the above result is a continuum version of the well-known Erdős–Gallai criterion [22]: Suppose $d_1\ge d_2\ge\cdots\ge d_n$ are nonnegative integers. The Erdős–Gallai criterion says that $d_1,\dots,d_n$ can be the degree sequence of a simple graph on n vertices if and only if $\sum_{i=1}^n d_i$ is even and for each $1\le k\le n$,

$$\sum_{i=1}^k d_i \le k(k-1) + \sum_{i=k+1}^n \min\{d_i,k\}.$$

(See [35] for extensive discussions and eight equivalent conditions.)

REMARK 2. When the scaling limit f is continuous, convergence in the modified L^1 norm is the same as supnorm convergence. In particular, for continuous scaling limits Theorem 1.1 and Proposition 1.2 both hold if we replace $D'[0,1]$ with $C[0,1]$ and redefine $\mathcal{F}$ analogously under the supnorm topology.

REMARK 3. As an example, consider the limit of the Erdős–Rényi graph $G(n,p)$ as $n\to\infty$. Here $f(x)=p$ for all x. Condition (ii) becomes $(1-x)\min\{p,x\}+x^2-px>0$ for all x. Considering the two cases $x\ge p$ and $x<p$ it is easy to see that this holds, so Erdős–Rényi graphs are in the interior of $\mathcal{F}$ for any fixed p, $0<p<1$.

REMARK 4. In a recent article [37] (following up on the older work [38]), McKay has computed subgraph counts in random graphs with a given degree sequence. However, McKay's results hold only if either the graph is sparse or the graph is dense but all degrees are within $n^{1/2+\varepsilon}$ of the average degree. Thus, it may be possible to recover Theorem 1.1 from McKay's results when the limit shape is a constant function but not in other cases.

The next natural question is whether one can feasibly compute the function g in Theorem 1.1 for a given f. It turns out that this is a central issue in the whole analysis. In fact, to prove Theorem 1.1 we analyze a related statistical model; computation of the maximum likelihood estimate in that model leads to an algorithm for computing g which, in turn, yields a proof of Theorem 1.1. The statistical model is discussed next.

1.2. *Statistics with degree sequences.* Informally, if the degree sequence captures the information in a graph, different graphs with the same degree sequence are judged equally likely. This can be formalized by saying that the degree sequence is a sufficient statistic for a probability distribution on graphs. The Koopman–Pitman–Darmois theorem forces this distribution to be of exponential form. This approach to model building is explained and developed in [32]. Diaconis and Freedman [17] give a version of the Koopman–Pitman–Darmois theorem for discrete exponential families. The approach is also standard fare in statistical mechanics where the uniform distribution on graphs with fixed degree sequence is called "micro-canonical" and the exponential distribution is called "canonical" (see [43]). It turns out that the exponential model has a simple description in terms of independent Bernoulli random variables.

Given a vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n$, let $\mathbb{P}_{\boldsymbol{\beta}}$ be the law of the undirected random graph on n vertices defined as follows: for each $1 \le i \ne j \le n$, put an edge between the vertices i and j with probability

$$p_{ij} := \frac{e^{\beta_i+\beta_j}}{1+e^{\beta_i+\beta_j}}$$

independently of all other edges. Thus, if G is a graph with degree sequence $d_1, \ldots, d_n$, the probability of observing G under $\mathbb{P}_{\boldsymbol{\beta}}$ is

$$\frac{e^{\sum_i \beta_i d_i}}{\prod_{i<j}(1+e^{\beta_i+\beta_j})}.$$

Henceforth, this model of random graphs is called the "$\boldsymbol{\beta}$-model." This model was considered by Holland and Lienhardt [27] in the directed case and by Park and Newman [43] and Blitzstein and Diaconis [12] in the undirected case. It is a close cousin to the Bradley–Terry model for rankings [which itself goes back (at least) to Zermelo]. See [29] for extensive references. The $\boldsymbol{\beta}$-model is also a simple version of a host of exponential models actively in use for analyzing network data. We will not try to survey this vast literature but recommend the extensive treatments in [30, 41, 48]. The website for the International Network for Social Network Analysis contains further information.

Suppose a random graph G is generated from the $\boldsymbol{\beta}$-model where $\boldsymbol{\beta} \in \mathbb{R}^n$ is unknown. Is it possible to estimate $\boldsymbol{\beta}$ from the observed G? It is not difficult to show that the maximum likelihood estimate (MLE) $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ must satisfy the system of equations

$$d_i = \sum_{j \ne i} \frac{e^{\hat{\beta}_i+\hat{\beta}_j}}{1+e^{\hat{\beta}_i+\hat{\beta}_j}}, \qquad i = 1, \ldots, n, \tag{3}$$

where $d_1, \ldots, d_n$ are the degrees of the vertices in the observed graph G. Questions may arise about the existence, uniqueness and accuracy of the MLE. Since the dimension of the parameter space grows with n, it is not clear if this is a "good" estimate of $\boldsymbol{\beta}$ in the traditional sense of consistency in statistical estimation theory.

The following theorem shows that under certain mild assumptions on $\boldsymbol{\beta}$, there is a high chance that the MLE exists, is unique and estimates $\boldsymbol{\beta}$ with uniform accuracy in all coordinates.

THEOREM 1.3. *Let G be drawn from the probability measure $\mathbb{P}_{\boldsymbol{\beta}}$ and let $d_1, \ldots, d_n$ be the degree sequence of G. Let $L := \max_{1 \le i \le n} |\beta_i|$. Then there is a constant $C(L)$ depending only on L such that with probability at least $1 - C(L)n^{-2}$, there exists a unique solution $\hat{\boldsymbol{\beta}}$ of the maximum likelihood equations* (3), *that satisfies*

$$\max_{1 \le i \le n} |\hat{\beta}_i - \beta_i| \le C(L)\sqrt{\frac{\log n}{n}}.$$

It may seem surprising that all n parameters can be accurately estimated from a single realization of the graph. However, one needs to observe that there are, in fact, $n(n-1)/2$ independent random variables lurking in the background (namely, the indicators whether edges are present or not). There is a well-known heuristic that in a p-parameter model with m observations, "the usual asymptotics" work provided that p^2/m tends to zero as m tends to infinity. See [44–47] for details (and counter examples). In our model $p = n$ and $m = n(n-1)/2$, so p^2/m does not tend to zero but stays bounded. The heuristic, although not directly applicable, hints at a reason why one can expect estimability of parameters.

In work closer to the present paper, Simons and Yao [50] studied the Bradley–Terry model for comparing n contestants. Here a random orientation of the complete graph on n vertices is chosen based on "player a beats player b with probability $\theta(a)/[\theta(a) + \theta(b)]$." They show that MLE is consistent here as well. Hunter [29] shows that the MM algorithm also behaves well in this problem.

The next theorem characterizes all possible expected degree sequences of the $\boldsymbol{\beta}$-model as $\boldsymbol{\beta}$ ranges over $\mathbb{R}^n$. The nice feature is that no degree sequence is left out.

THEOREM 1.4. *Let $\mathcal{R}$ denote the set of all expected degree sequences of random graphs following the law $\mathbb{P}_{\boldsymbol{\beta}}$ as $\boldsymbol{\beta}$ ranges over $\mathbb{R}^n$. Let $\mathcal{D}$ denote the set of all possible degree sequences of undirected graphs on n vertices. Then*

$$\mathrm{conv}(\mathcal{D}) = \overline{\mathcal{R}},$$

where $\mathrm{conv}(\mathcal{D})$ *denotes the convex hull of $\mathcal{D}$ and $\overline{\mathcal{R}}$ is the topological closure of $\mathcal{R}$.*

Incidentally, the convex hull of $\mathcal{D}$ is a well-studied polytope. For example, its extreme points are the threshold graphs. (A graph is a threshold graph if there is a real number S and for each vertex v a real vertex weight $w(v)$ such that, for any two vertices v, u, (u, v), there is an edge if and only if $w(u) + w(v) \geq S$. See [35] for much more on this.)

A self-contained proof of Theorem 1.4 is given in Section 3. However, it is possible to derive it from classical results about the mean space of exponential families (see, e.g., [16] or [4]; in particular, see [52], Theorem 3.3).

Finally, let us describe a fast algorithm for computing the MLE if it exists. Recall that the L^∞ norm of a vector $\mathbf{x} = (x_1, \ldots, x_n)$ is defined as

$$|\mathbf{x}|_\infty := \max_{1 \leq i \leq n} |x_i|.$$

For $1 \leq i \neq j \leq n$ and $\mathbf{x} \in \mathbb{R}^n$, let

$$r_{ij}(\mathbf{x}) := \frac{1}{e^{-x_j} + e^{x_i}}. \tag{4}$$

Given a realization of the random graph G with degree sequence $d_1, \ldots, d_n$, define for each i the function

$$\varphi_i(\mathbf{x}) := \log d_i - \log \sum_{j \neq i} r_{ij}(\mathbf{x}). \tag{5}$$

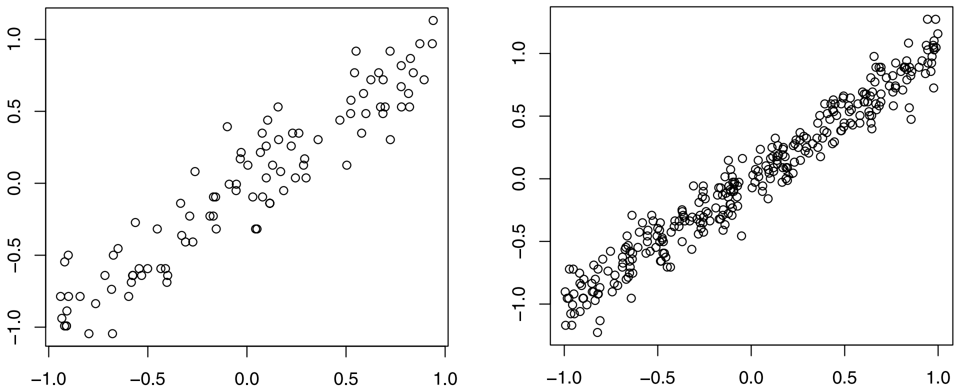
Let $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ be the function whose ith component is φ_i. An easy rearrangement of terms shows that the fixed points of φ are precisely the solutions of (3). The following theorem exploits this to give an algorithm for computing the MLE in the $\boldsymbol{\beta}$-model.

THEOREM 1.5. *Suppose the ML equations* (3) *have a solution* $\hat{\boldsymbol{\beta}}$. *Then* $\hat{\boldsymbol{\beta}}$ *is a fixed point of the function* φ. *Starting from any* $\mathbf{x}_0 \in \mathbb{R}^n$, *define* $\mathbf{x}_{k+1} = \varphi(\mathbf{x}_k)$ *for* $k = 0, 1, 2, \ldots$. *Then* $\mathbf{x}_k$ *converges to* $\hat{\boldsymbol{\beta}}$ *geometrically fast in the* L^∞ *norm where the rate depends only on* $(|\hat{\boldsymbol{\beta}}|_\infty, |\mathbf{x}_0|_\infty)$. *In particular,* $\hat{\boldsymbol{\beta}}$ *must be the unique solution of* (3). *Moreover,*

$$|\mathbf{x}_0 - \hat{\boldsymbol{\beta}}|_\infty \leq C|\mathbf{x}_0 - \mathbf{x}_1|_\infty,$$

where C *is a continuous function of the pair* $(|\hat{\boldsymbol{\beta}}|_\infty, |\mathbf{x}_0|_\infty)$. *Conversely, if the ML equations* (3) *do not have a solution, then the sequence* $\{\mathbf{x}_k\}$ *must have a divergent subsequence.*

There are many other algorithms available for calculating the MLE. For example, Holland and Leinhardt [27] use an iterative scaling algorithm and discuss the method of scoring and weighted least squares. Hunter [29] develops the MM algorithm for a similar task. Markov chain Monte Carlo algorithms and the Robbins–Monro stochastic approximation approach are also used for computing the MLE in exponential random graph models. See [31], Section 6.5.2, for examples and literature. The iterative algorithm we use is a hybrid of standard algorithms which



(a) $n = 100,\ \beta_i \sim \text{Unif}[-1, 1]$. (b) $n = 300,\ \beta_i \sim \text{Unif}[-1, 1]$.

FIG. 1. *Simulation results: plot of $\hat{\beta}_i$ vs. β_i.*

works well in practice and allows the strong conclusions of Theorem 1.5. We hope that variants can be developed for related high dimensional problems.

Let us now look at the results of some simulations. The left-hand panel in Figure 1 shows the plot of $\hat{\beta}_i$ versus β_i for a graph with 100 vertices, where $\beta_1, \dots, \beta_n$ were chosen independently at uniform from the interval $[-1, 1]$. The right-hand panel is the same, except that n has been increased to 300. The increased accuracy for larger n is clearly visible.

We have also compared our results with the simulation results from the importance sampling algorithm of Blitzstein and Diaconis [12] for a variety of other examples. The results of Figure 1 are typical. This convinces us that the procedures developed in this paper are useful for practical problems.

Comparison to the Barvinok and Hartigan work. As mentioned before, the present work is closely related to a recent series of papers by Barvinok and Hartigan [5–10]. The work was initiated by Barvinok who looked at directed and bipartite graphs in [6]. In their most recent article [10] (uploaded to arXiv when our paper was near completion), they study uniform random (undirected) graphs on n vertices with a given degree sequence $\mathbf{d} = (d_1, \dots, d_n)$ and work with an exponential model as in Section 1.2 with β_i chosen so that the expected degree at i under the $\boldsymbol{\beta}$-model is d_i. Let $G_{\mathbf{d}}$ be a uniformly chosen random graph with the degree sequence $\mathbf{d}$ and $G_{\boldsymbol{\beta}}$ be a random graph chosen from the $\boldsymbol{\beta}$-model. One of their main results shows that (under hypothesis) these two graphs are close together in the following sense: Fix a set of edges S in the complete graph on n vertices. Let $X_{\mathbf{d}}$ be the number of edges of $G_{\mathbf{d}}$ in S. Let $X_{\boldsymbol{\beta}}$ be the number of edges of $G_{\boldsymbol{\beta}}$ in S. They prove that $X_{\mathbf{d}}/n^2$ and $X_{\boldsymbol{\beta}}/n^2$ are each concentrated about their means (using results from the earlier work [5]) and that these means are approximately equal. Their theorem is proved under a condition on the degree sequences that they call "delta tame."

While the two sets of results (i.e., ours and those of Barvinok and Hartigan) were proved independently and the methods of proof are quite different in certain parts (but similar in others), the possible connections are tantalizing. We believe that their mode of convergence ($G_{\mathbf{d}}$ and $G_{\boldsymbol{\beta}}$ contain about the same number of edges in a given set) is equivalent to the graph limit convergence used here. Perhaps this can be established using the "cut-metric" of Frieze and Kannan, as expounded in [14]. We further conjecture, based on Lemma 4.1 in this paper, that their delta tame condition is equivalent to our condition that the limiting degree sequence f is in the interior of $\mathcal{F}$. If this is so, then Proposition 1.2 (or more accurately, Lemma 4.1) gives a necessary and sufficient condition for a degree sequence to be delta tame, showing that essentially all degree sequences except the ones close to the Erdős–Gallai boundary are delta tame.

In summary, the Barvinok and Hartigan work [10] contains elegant estimates of the number of graphs with a given degree sequence and extensions to bipartite graphs under a condition called delta tameness; we work in the emerging language of graph limits and prove a limit theorem under a continuum version of the easily verifiable Erdős–Gallai criterion. Our work contains an efficient algorithm for computing the maximum likelihood estimates of $\boldsymbol{\beta}$ for a given degree sequence with proofs of convergence of the algorithm and consistency of the estimates.

The rest of the paper is organized as follows. In Section 2 we prove Theorem 1.5. This is followed by the proof of Theorem 1.4 in Section 3. Both of these theorems are required for the proof of Theorem 1.3, which is given in Section 4. Proposition 1.2 is proved in Section 5. Finally, the proof of Theorem 1.1, which uses all the other theorems, is given in Section 6.

2. Proof of Theorem 1.5.

For a matrix $A = (a_{ij})_{1\le i,j\le n}$, the L^∞ operator norm is defined as

$$|A|_\infty := \max_{|x|_\infty \le 1} |A\mathbf{x}|_\infty.$$

It is a simple exercise to verify that

$$|A|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|.$$

Given $\delta > 0$, let us say the matrix A belongs to the class $\mathcal{L}_n(\delta)$ if $|A|_\infty \le 1$ and for each $1 \le i \ne j \le n$,

$$a_{ii} \ge \delta \quad \text{and} \quad a_{ij} \le -\frac{\delta}{n-1}.$$

Lemma 2.1 is our key tool.

LEMMA 2.1. *Let $\mathcal{L}_n(\delta)$ be defined as above. If $A, B \in \mathcal{L}_n(\delta)$, then*

$$|AB|_\infty \le 1 - \frac{2(n-2)\delta^2}{n-1}.$$

PROOF. Fix $1 \le i \ne k \le n$. By the definition of $\mathcal{L}_n(\delta)$,

$$\sum_{j \notin \{i,k\}} a_{ij} b_{jk} \ge \frac{(n-2)\delta^2}{(n-1)^2} \quad \text{and} \quad a_{ii} b_{ik} + a_{ik} b_{kk} \le -\frac{2\delta^2}{n-1}.$$

Now, if x, y are two positive real numbers, then $|x - y| = |x| + |y| - 2\min\{x, y\}$. Taking $x = \sum_{j \notin \{i,k\}} a_{ij} b_{jk}$ and $y = -(a_{ii} b_{ik} + a_{ik} b_{kk})$, we get

$$\begin{aligned} \left|\sum_{j=1}^n a_{ij} b_{jk}\right| &\le \sum_{j=1}^n |a_{ij} b_{jk}| - 2\min\left\{\sum_{j \notin \{i,k\}} a_{ij} b_{jk}, -(a_{ii} b_{ik} + a_{ik} b_{kk})\right\} \\ &\le \sum_{j=1}^n |a_{ij} b_{jk}| - \frac{2(n-2)\delta^2}{(n-1)^2}. \end{aligned}$$

Combining this with the hypothesis that $|A|_\infty \le 1$ and $|B|_\infty \le 1$, we get

$$\begin{aligned} |AB|_\infty &= \max_{1 \le i \le n} \sum_{k=1}^n \left|\sum_{j=1}^n a_{ij} b_{jk}\right| \\ &\le \max_{1 \le i \le n} \sum_{j=1}^n \sum_{k=1}^n |a_{ij} b_{jk}| - \frac{2(n-2)\delta^2}{n-1} \\ &\le 1 - \frac{2(n-2)\delta^2}{n-1}. \end{aligned}$$

The proof is complete. □

Now recall the functions r_{ij} defined in (4). Let

$$q_{ij}(\mathbf{x}) := \frac{r_{ij}(\mathbf{x})}{\sum_{k \ne i} r_{ik}(\mathbf{x})}.$$

Note that for each i and $\mathbf{x}$, $\sum_{j \ne i} q_{ij}(\mathbf{x}) = 1$. Again, for each i

$$\frac{\partial \varphi_i}{\partial x_i} = -\frac{\sum_{j \ne i} \partial r_{ij}/\partial x_i}{\sum_{j \ne i} r_{ij}} = \sum_{j \ne i} \frac{e^{x_i}}{e^{-x_j} + e^{x_i}} q_{ij}$$

and similarly for each distinct i and j,

$$\frac{\partial \varphi_i}{\partial x_j} = -\frac{e^{-x_j}}{e^{-x_j} + e^{x_i}} q_{ij}.$$

Now, if $|\mathbf{x}|_\infty \le K$, then clearly

$$\tfrac{1}{2} e^{-K} \le r_{ij}(\mathbf{x}) \le \tfrac{1}{2} e^K \qquad \text{for all } 1 \le i \ne j \le n.$$

Thus,

$$\frac{e^{-2K}}{n-1} \le q_{ij}(\mathbf{x}) = \frac{r_{ij}(\mathbf{x})}{\sum_{k\neq i} r_{ik}(\mathbf{x})} \le \frac{e^{2K}}{n-1}.$$

It follows that for every $1 \le i \neq j \le n$

$$-\frac{e^{2K}}{n-1} \le \frac{\partial \varphi_i}{\partial x_j} \le -\frac{e^{-4K}}{2(n-1)} \tag{6}$$

and also, for every $1 \le i \le n$,

$$\frac{1}{2}e^{-4K} \le \frac{\partial \varphi_i}{\partial x_i} \le e^{2K}. \tag{7}$$

Now take any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and let K be the maximum of the L^∞ norms of $\mathbf{x}$, $\mathbf{y}$, $\varphi(\mathbf{x})$ and $\varphi(\mathbf{y})$. Let $J(\mathbf{x}, \mathbf{y})$ be the matrix whose (i, j)th element is

$$J_{ij}(\mathbf{x}, \mathbf{y}) = \int_0^1 \frac{\partial \varphi_i}{\partial x_j}\big(t\mathbf{x} + (1-t)\mathbf{y}\big)\, dt.$$

It is a simple calculus exercise to verify that

$$\varphi(\mathbf{x}) - \varphi(\mathbf{y}) = J(\mathbf{x}, \mathbf{y})(\mathbf{x} - \mathbf{y}). \tag{8}$$

For each $i \neq j$, $\partial\varphi_i/\partial x_j$ is negative everywhere and for each i, $\partial\varphi_i/\partial x_i$ is positive everywhere. Moreover, for each i,

$$\sum_{j=1}^n \left|\frac{\partial \varphi_i}{\partial x_j}\right| = \frac{\partial \varphi_i}{\partial x_i} - \sum_{j\neq i} \frac{\partial \varphi_i}{\partial x_j} \equiv 1.$$

It follows that for i and any $\mathbf{x}$, $\mathbf{y}$, $\sum_{j=1}^n |J_{ij}(\mathbf{x}, \mathbf{y})| = 1$. In particular, $|J(\mathbf{x}, \mathbf{y})|_\infty = 1$. From (6) and (7) and the fact that $|J(\mathbf{x}, \mathbf{y})|_\infty = 1$, we see that $J(\mathbf{x}, \mathbf{y}) \in \mathcal{L}_n(\delta)$ for $\delta = \frac{1}{2}e^{-4K}$. Similarly,

$$\begin{aligned}\varphi(\varphi(\mathbf{x})) - \varphi(\varphi(\mathbf{y})) &= J(\varphi(\mathbf{x}), \varphi(\mathbf{y}))\big(\varphi(\mathbf{x}) - \varphi(\mathbf{y})\big)\\ &= J(\varphi(\mathbf{x}), \varphi(\mathbf{y}))J(\mathbf{x}, \mathbf{y})(\mathbf{x} - \mathbf{y})\end{aligned}$$

and $J(\varphi(\mathbf{x}), \varphi(\mathbf{y})) \in \mathcal{L}_n(\delta)$ also. Applying Lemma 2.1, we get

$$|J(\varphi(\mathbf{x}), \varphi(\mathbf{y}))J(\mathbf{x}, \mathbf{y})|_\infty \le 1 - \frac{2(n-2)\delta^2}{n-1}.$$

Thus,

$$|\varphi(\varphi(\mathbf{x})) - \varphi(\varphi(\mathbf{y}))|_\infty \le \left(1 - \frac{2(n-2)\delta^2}{n-1}\right)|\mathbf{x} - \mathbf{y}|_\infty. \tag{9}$$

The quantity inside the brackets will henceforth be denoted by $\theta(\mathbf{x}, \mathbf{y})$. Note that $0 \le \theta(\mathbf{x}, \mathbf{y}) < 1$ and θ is uniformly bounded away from 1 on subsets of $\mathbb{R}^n \times \mathbb{R}^n$. Moreover, since $|J(\mathbf{x}, \mathbf{y})|_\infty = 1$, we also have the trivial but useful bound

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{y})|_\infty \le |\mathbf{x} - \mathbf{y}|_\infty.$$

Now suppose φ has a fixed point $\hat{\boldsymbol{\beta}}$. If we start with arbitrary $\mathbf{x}_0$ and define $\mathbf{x}_{k+1} = \varphi(\mathbf{x}_k)$ for each $k \geq 0$, then for each k, we have

$$|\mathbf{x}_{k+1} - \hat{\boldsymbol{\beta}}|_\infty = |\varphi(\mathbf{x}_k) - \varphi(\hat{\boldsymbol{\beta}})|_\infty \leq |\mathbf{x}_k - \hat{\boldsymbol{\beta}}|_\infty.$$

In particular, the sequence $\{\mathbf{x}_k\}_{k\geq 0}$ remains bounded. Therefore, by (9), there is a single $\theta \in [0, 1)$, depending only on $|\hat{\boldsymbol{\beta}}|_\infty$ and $|\mathbf{x}_0|_\infty$ in a continuous manner, such that for all $k \geq 0$, we have

$$|\mathbf{x}_{k+3} - \mathbf{x}_{k+2}|_\infty \leq \theta |\mathbf{x}_{k+1} - \mathbf{x}_k|_\infty \tag{10}$$

and

$$|\mathbf{x}_{k+2} - \hat{\boldsymbol{\beta}}|_\infty \leq \theta |\mathbf{x}_k - \hat{\boldsymbol{\beta}}|_\infty.$$

The second inequality shows that $\mathbf{x}_k$ converges to $\hat{\boldsymbol{\beta}}$ geometrically fast and the first inequality gives

$$\begin{aligned}
|\mathbf{x}_0 - \hat{\boldsymbol{\beta}}|_\infty &\leq \sum_{k=0}^{\infty} |\mathbf{x}_k - \mathbf{x}_{k+1}|_\infty \\
&\leq \frac{1}{1-\theta}(|\mathbf{x}_0 - \mathbf{x}_1|_\infty + |\mathbf{x}_1 - \mathbf{x}_2|_\infty) \\
&\leq \frac{2}{1-\theta}|\mathbf{x}_0 - \mathbf{x}_1|_\infty.
\end{aligned}$$

Finally, note that if $\hat{\boldsymbol{\beta}}$ does not exist, then the sequence $\{\mathbf{x}_k\}$ must have a divergent subsequence. Otherwise, (9) would imply that (10) must hold for all k for some $\theta \in [0, 1)$. This, in turn, would imply that $\mathbf{x}_k$ must converge to a limit as $k \to \infty$, which would then be a fixed point of φ and, therefore, a solution of the ML equations. The proof is complete.

Before moving to the next section we will prove a technical lemma which will be of use in the proof of Theorem 1.1 based on the above calculations.

LEMMA 2.2. *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *such that* $\max\{|\mathbf{x}|_\infty, |\mathbf{y}|_\infty\} \leq K$. *Then*

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{y})|_1 \leq 2e^{2K}|\mathbf{x} - \mathbf{y}|_1,$$

where $|\cdot|_1$ *is the usual* L^1 *norm on* $\mathbb{R}^n$.

PROOF. By equation (8),

$$\begin{aligned}
|\varphi(\mathbf{x}) - \varphi(\mathbf{y})|_1 &= |J(\mathbf{x}, \mathbf{y})(\mathbf{x} - \mathbf{y})|_1 \\
&= \sum_{i=1}^{n} \left| \sum_{j=1}^{n} (x_j - y_j) \cdot \int_0^1 \frac{\partial \varphi_i}{\partial x_j}(t\mathbf{x} + (1-t)\mathbf{y})\, dt \right|
\end{aligned}$$

$$\leq \sum_{j=1}^{n} |x_j - y_j| \cdot \left(\sum_{i=1}^{n} \sup_{t\in[0,1]} \left| \frac{\partial \varphi_i}{\partial x_j} (t\mathbf{x} + (1-t)\mathbf{y}) \right| \right)$$

$$\leq \sum_{j=1}^{n} 2e^{2K} |x_j - y_j| = 2e^{2K} |\mathbf{x} - \mathbf{y}|_1,$$

where the second inequality follows from equations (6) and (7). □

3. Proof of Theorem 1.4. We need the following simple technical lemma.

LEMMA 3.1. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a twice differentiable function such that $M := \sup_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) < \infty$. Let ∇f and $\nabla^2 f$ denote the gradient vector and the Hessian matrix of f and suppose there is a finite constant C such that the L^2 operator norm of $\nabla^2 f$ is uniformly bounded by C. Then for any $\mathbf{x} \in \mathbb{R}^n$,*

$$|\nabla f(\mathbf{x})|^2 \leq 2C\big(M - f(\mathbf{x})\big),$$

where $|\cdot|$ denotes the Euclidean norm. In particular, there exists a sequence $\{\mathbf{x}_k\}_{k\geq 1}$ such that $\lim_{k\to\infty} \nabla f(\mathbf{x}_k) = 0$.

PROOF. Fix a point $\mathbf{x} \in \mathbb{R}^n$ and let $\mathbf{y} = \nabla f(\mathbf{x})$. Suppose C is a uniform bound on the L^2 operator norm of $\nabla^2 f$. Then for any $t \geq 0$,

$$|\nabla f(\mathbf{x} + t\mathbf{y}) - \nabla f(\mathbf{x})| \leq Ct|\mathbf{y}|. \tag{11}$$

Now let $g(t) = f(\mathbf{x} + t\mathbf{y})$. Then for all t,

$$g(t) - g(0) \leq M - f(\mathbf{x}).$$

Again, note that

$$\begin{aligned} g'(t) &= \langle \mathbf{y}, \nabla f(\mathbf{x} + t\mathbf{y}) \rangle \\ &= \langle \mathbf{y}, \nabla f(\mathbf{x} + t\mathbf{y}) - \nabla f(\mathbf{x}) \rangle + \langle \mathbf{y}, \nabla f(\mathbf{x}) \rangle \\ &\geq -Ct|\mathbf{y}|^2 + |\mathbf{y}|^2. \end{aligned}$$

[The last step follows by (11) and Cauchy–Schwarz.] Thus, for any $t \geq 0$,

$$M - f(\mathbf{x}) \geq g(t) - g(0) = \int_0^t g'(s)\, ds \geq |\mathbf{y}|^2 \int_0^t (1 - Cs)\, ds.$$

Taking $t = 1/C$ gives the desired result. □

PROOF OF THEOREM 1.4. Let $g = (g_1, \ldots, g_n) : \mathbb{R}^n \to \mathbb{R}^n$ be the function defined as

$$g_i(\mathbf{x}) = \sum_{j\neq i} \frac{e^{x_i + x_j}}{1 + e^{x_i + x_j}}, \qquad i = 1, \ldots, n.$$

Then $\mathcal{R}$ is the range of g. This is because the expected degree of vertex i of a random graph following the law $\mathbb{P}_{\mathbf{x}}$ is $g_i(\mathbf{x})$. In particular, the vector $g(\mathbf{x})$ is a weighted average of degree sequences and hence,

$$\mathrm{conv}(\mathcal{D}) \supseteq \overline{\mathcal{R}}.$$

Now, for every $\mathbf{y} \in \mathbb{R}^n$, let $f_{\mathbf{y}} : \mathbb{R}^n \to \mathbb{R}$ be the function

$$f_{\mathbf{y}}(\mathbf{x}) = \sum_{i=1}^{n} x_i y_i - \log \sum_{1 \le i < j \le n} (1 + e^{x_i + x_j}).$$

Note that under $\mathbb{P}_{\mathbf{x}}$, the probability of obtaining a given graph with degree sequence $d = (d_1, \ldots, d_n)$ is exactly

$$\frac{e^{\sum_i x_i d_i}}{\prod_{i<j}(1 + e^{x_i + x_j})}.$$

Thus, the above quantity must be bounded by 1 and hence, taking logs, we get $f_d(\mathbf{x}) \le 0$. Since $f_{\mathbf{y}}(\mathbf{x})$ depends linearly on $\mathbf{y}$, this implies that

$$f_{\mathbf{y}}(\mathbf{x}) \le 0 \qquad \text{for all } \mathbf{y} \in \mathrm{conv}(\mathcal{D}), \mathbf{x} \in \mathbb{R}^n.$$

Now fix $\mathbf{y} \in \mathrm{conv}(\mathcal{D})$. Then $f_{\mathbf{y}}(\mathbf{x}) \le 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Moreover, it is easy to show that $\nabla^2 f$ is uniformly bounded. Hence, it follows from Lemma 11 that there exists a sequence $\{\mathbf{x}_k\}_{k \ge 1}$ such that $\lim_{k\to\infty} \nabla f_{\mathbf{y}}(\mathbf{x}_k) = 0$. But

$$\nabla f_{\mathbf{y}}(\mathbf{x}) = \mathbf{y} - g(\mathbf{x}).$$

Thus, $\mathbf{y} = \lim_{k\to\infty} g(\mathbf{x}_k)$. This shows that

$$\mathrm{conv}(\mathcal{D}) \subseteq \overline{\mathcal{R}}$$

and hence, completes the proof of the claim that $\mathrm{conv}(\mathcal{D}) = \overline{\mathcal{R}}$. □

4. Proof of Theorem 1.3 (Existence and consistency of the MLE). This section uses the notation of Section 1 without explicit reference. The proof consists of two lemmas. The first lemma gives a condition for the "tightness" of the MLE. This result is closely related to the Erdős–Gallai characterization of degree sequences. The second lemma shows that conditions needed for the first lemma are satisfied with high probability. An addenda at the end of the section contains some results about existence of the MLE and the closely related topic of conjugate Bayesian analysis.

In this section we will repeatedly encounter statements like "c is a positive constant depending only on $a, b, \ldots$." Such a statement should be interpreted as "c can be expressed as a function of $a, b, \ldots$ that is bounded away from 0 and ∞ on compact subsets of the domain of $(a, b, \ldots)$." Sometimes, c will be expressed as $C(a, b, \ldots)$.

LEMMA 4.1. *Let* $(d_1, \ldots, d_n)$ *be a point in the set* $\overline{\mathcal{R}}$ *of Theorem* 1.4. *Suppose there exist* $c_1, c_2 \in (0,1)$ *such that* $c_2(n-1) \le d_i \le c_1(n-1)$ *for all* i. *Suppose* c_3 *is a positive constant such that*

$$\frac{1}{n^2} \inf_{B\subseteq\{1,\ldots,n\}, |B|\ge c_2^2 n} \left\{ \sum_{j\notin B} \min\{d_j, |B|\} + |B|(|B|-1) - \sum_{i\in B} d_i \right\} \ge c_3.$$

Then a solution $\hat{\boldsymbol{\beta}}$ *of* (3) *exists and satisfies* $|\hat{\boldsymbol{\beta}}|_\infty \le c_4$, *where* c_4 *is a constant that depends only on* c_1, c_2, c_3.

PROOF. In this proof, $C(c_1, c_2, c_3)$ denotes positive constants that depend only on c_1, c_2, c_3, in the sense defined above. The argument repeatedly uses the monotonicity of $e^{x+y}/(1+e^{x+y})$ in x for each y.

Assume first that $\hat{\boldsymbol{\beta}}$ exists in the sense that there exists $\hat{\boldsymbol{\beta}} \in \mathbb{R}^n$ such that (3) is satisfied. Let c_1, c_2, c_3 be as in the statement of the lemma. It is proved below that $|\hat{\boldsymbol{\beta}}|_\infty$ is bounded above by $C(c_1, c_2, c_3)$.

Let $d_{\max} := \max_i d_i$ and $d_{\min} := \min_i d_i$. Similarly, let $\hat{\beta}_{\max} := \max_i \hat{\beta}_i$ and $\hat{\beta}_{\min} := \min_i \hat{\beta}_i$. The first step is to prove that $\hat{\beta}_{\max} \le C(c_1, c_2, c_3)$. If $\hat{\beta}_{\max} \le 0$, there is nothing to prove. So assume that $\hat{\beta}_{\max} > 0$. Let

$$m := \left|\{i : \hat{\beta}_i > -\tfrac{1}{2}\hat{\beta}_{\max}\}\right|.$$

Clearly, by the assumption that $\hat{\beta}_{\max} > 0$, it is guaranteed that $m \ge 1$. Let i^* be an index that maximizes $\hat{\beta}_i$. Then by (3), we see that

$$d_{\max} \ge d_{i^*} > (m-1)\frac{e^{(1/2)\hat{\beta}_{\max}}}{1+e^{(1/2)\hat{\beta}_{\max}}}.$$

This implies

$$n - m > n - 1 - d_{\max}\bigl(1 + e^{-(1/2)\hat{\beta}_{\max}}\bigr) \ge n - 1 - c_1(n-1)\bigl(1 + e^{-(1/2)\hat{\beta}_{\max}}\bigr).$$

In particular, this shows that if $\hat{\beta}_{\max} > C(c_1)$ then $m < n$ and hence, there exists i such that $\hat{\beta}_i \le -\frac{1}{2}\hat{\beta}_{\max}$. Suppose this is true and fix any such i. (In particular note that $\hat{\beta}_i < 0$.) Let

$$m_i := \left|\{j : j \ne i, \hat{\beta}_j < -\tfrac{1}{2}\hat{\beta}_i\}\right|.$$

Then by (3),

$$d_{\min} \le d_i < m_i \frac{e^{(1/2)\hat{\beta}_i}}{1+e^{(1/2)\hat{\beta}_i}} + n - 1 - m_i,$$

which gives

$$m_i < (n - 1 - d_{\min})\bigl(1 + e^{(1/2)\hat{\beta}_i}\bigr) \le (n-1)(1-c_2)\bigl(1 + e^{-(1/4)\hat{\beta}_{\max}}\bigr).$$

Note that there are at least $n-m_i$ indices j such that $\hat{\beta}_j \ge -\frac{1}{2}\hat{\beta}_i \ge \frac{1}{4}\hat{\beta}_{\max}$. The last display implies that if $\hat{\beta}_{\max} > C(c_1, c_2)$, then there exists i such that $n - m_i \ge bn$, where

$$b := c_2^2.$$

Consequently, if $\hat{\beta}_{\max} > C(c_1, c_2)$, there is a set $A \subseteq \{1, \dots, n\}$ of size at least bn such that $\hat{\beta}_j \ge \frac{1}{4}\hat{\beta}_{\max}$ for all $j \in A$, where $b = c_2^2$. Henceforth, assume that $\hat{\beta}_{\max}$ is so large that such a set exists. Let

$$h := \sqrt{\hat{\beta}_{\max}}.$$

For each integer r between 0 and $\frac{1}{16}h - 1$, let

$$D_r := \left\{ i : -\frac{\hat{\beta}_{\max}}{8} + rh \le \hat{\beta}_i < -\frac{\hat{\beta}_{\max}}{8} + (r+1)h \right\}.$$

Since $D_0, D_1, \dots$ are disjoint, there exists r such that

$$|D_r| \le \frac{n}{(1/16)h - 1},$$

provided $h > 16$. By assumption, $\hat{\beta}_{\max} > C(c_1, c_2)$. Since we are free to choose $C(c_1, c_2)$ as large as we like, it can be assumed without loss of generality that $h > 16$.

Fix such an r between 0 and $\frac{1}{16}h - 1$. Let

$$B := \left\{ i : \hat{\beta}_i \ge \frac{\hat{\beta}_{\max}}{8} - \left(r + \frac{1}{2}\right)h \right\}.$$

Clearly, the set B contains the previously defined set A and hence,

$$|B| \ge bn. \tag{12}$$

Now, for each $i \ne j$, define

$$\hat{p}_{ij} := \frac{e^{\hat{\beta}_i + \hat{\beta}_j}}{1 + e^{\hat{\beta}_i + \hat{\beta}_j}}.$$

For each i, let

$$d_i^B := \sum_{j \in B \setminus \{i\}} \hat{p}_{ij}.$$

Since $\hat{\beta}_i \ge \frac{\hat{\beta}_{\max}}{16}$ for each $i \in B$, it follows that

$$\begin{aligned} |B|(|B|-1) - \sum_{i \in B} d_i^B &= |B|(|B|-1) - \sum_{i,j \in B, i \ne j} \hat{p}_{ij} \\ &= \sum_{i,j \in B, i \ne j} (1 - \hat{p}_{ij}) \qquad (13) \\ &\le \frac{|B|(|B|-1)}{1 + e^{(1/8)\hat{\beta}_{\max}}}. \end{aligned}$$

The above inequality is the first step of a two-step argument. For the second step, take any $j \notin B$. Consider three cases. First, suppose $\hat{\beta}_j \geq -\frac{\hat{\beta}_{\max}}{8} + (r+1)h$. Then for each $i \in B$, $\hat{\beta}_i + \hat{\beta}_j \geq \frac{h}{2}$ and, therefore,

$$\min\{d_j, |B|\} - d_j^B \leq |B| - \sum_{i \in B} \hat{p}_{ij} = \sum_{i \in B}(1 - \hat{p}_{ij}) \leq \frac{|B|}{1 + e^{h/2}}.$$

Next, suppose $\hat{\beta}_j \leq -\frac{\hat{\beta}_{\max}}{8} + rh$. Then for any $i \notin B$, $\hat{\beta}_i + \hat{\beta}_j \leq -\frac{h}{2}$. Thus,

$$\min\{d_j, |B|\} - d_j^B \leq d_j - d_j^B = \sum_{i \notin B, i \neq j} \hat{p}_{ij} \leq ne^{-h/2}.$$

Finally, the third case covers all $j \notin B$ that do not fall in either of the previous two cases. This is a subset of the set of all j comprising the set D_r. Combining the three cases gives

$$\sum_{j \notin B}(\min\{d_j, |B|\} - d_j^B) \leq \frac{n^2}{1 + e^{h/2}} + n^2 e^{-h/2} + \frac{16n^2}{h - 16}. \tag{14}$$

But

$$\sum_{j \notin B} d_j^B = \sum_{i \in B, j \notin B} \hat{p}_{ij} = \sum_{i \in B}(d_i - d_i^B).$$

Thus, adding (13) and (14),

$$\begin{aligned} &\sum_{j \notin B} \min\{d_j, |B|\} + |B|(|B| - 1) - \sum_{i \in B} d_i \\ &\quad \leq \frac{n^2}{1 + e^{(1/8)\hat{\beta}_{\max}}} + \frac{n^2}{1 + e^{h/2}} + n^2 e^{-h/2} + \frac{16n^2}{h - 16}. \end{aligned} \tag{15}$$

The left-hand side of the above inequality is bounded below by $c_3 n^2$, by the definition of c_3 in the statement of the theorem. The coefficient of n^2 on the right-hand side tends to zero as $\hat{\beta}_{\max} \to \infty$. This shows that $\hat{\beta}_{\max} \leq C(c_1, c_2, c_3)$, where the bound is finite since $c_3 > 0$. Next, note that for any i,

$$d_i \leq \frac{ne^{\hat{\beta}_i + \hat{\beta}_{\max}}}{1 + e^{\hat{\beta}_i + \hat{\beta}_{\max}}}$$

and, therefore, if i^{**} is a vertex that minimizes $\hat{\beta}_i$, then

$$d_{\min} \leq d_{i^{**}} \leq \frac{ne^{\hat{\beta}_{\min} + \hat{\beta}_{\max}}}{1 + e^{\hat{\beta}_{\min} + \hat{\beta}_{\max}}}.$$

Combined with the upper bound on $\hat{\beta}_{\max}$ and the lower bound on $d_{\min}$, this shows that $\hat{\beta}_{\min} \geq -C(c_1, c_2, c_3)$.

To complete the proof of the lemma, it must be proved that $\hat{\boldsymbol{\beta}}$ exists. Since $(d_1, \ldots, d_n) \in \overline{\mathcal{R}}$, by Theorem 1.4 there is a sequence of points $\{\mathbf{x}_k\}_{k\geq 0}$ in $\mathbb{R}^n$ that converge to $(d_1, \ldots, d_n)$ for which solutions to (3) exist. Let $\{\hat{\boldsymbol{\beta}}_k\}_{k\geq 0}$ denote a sequence of solutions. The steps above prove that $|\hat{\boldsymbol{\beta}}_k|_\infty \leq C$ for all large enough k where C is some constant depending only on c_1, c_2, c_3. Therefore, the sequence $\{\hat{\boldsymbol{\beta}}_k\}_{k\geq 0}$ must have a limit point. This limit point is clearly a solution to (3) for the original sequence $d_1, \ldots, d_n$. □

The next lemma shows that the degree sequence in a typical realization of our random graph satisfies the conditions of Lemma 4.1.

LEMMA 4.2. *Let G be drawn from the probability measure $\mathbb{P}_{\boldsymbol{\beta}}$ and let $d_1, \ldots, d_n$ be the degree sequence of G. Let $L := \max_{1\leq i\leq n} |\beta_i|$ and let $c \in (0, 1)$ be any constant. Then there are constants $C > 0$ and $c_1, c_2 \in (0, 1)$ depending only on L and a constant $c_3 \in (0, 1)$ depending only on L and c such that if $n > C$, then with probability at least $1 - 2n^{-2}$, $c_2(n-1) \leq d_i \leq c_1(n-1)$ for all i and*

$$\frac{1}{n^2} \inf_{B\subseteq\{1,\ldots,n\}, |B|\geq cn} \left\{ \sum_{j\notin B} \min\{d_j, |B|\} + |B|(|B|-1) - \sum_{i\in B} d_i \right\}$$
$$\geq c_3 - \sqrt{\frac{6\log n}{n}}.$$

PROOF. Let

$$\bar{d}_i := \sum_{j\neq i} \frac{e^{\beta_i+\beta_j}}{1+e^{\beta_i+\beta_j}}.$$

Note that for each i, d_i is a sum of independent indicator random variables and $\mathbb{E}(d_i) = n\bar{d}_i$. Therefore, by Hoeffding's inequality [26],

$$\mathbb{P}(|d_i - \bar{d}_i| > x) \leq 2e^{-x^2/2n}.$$

Thus, if we let E be the event

$$\left\{ \max_i |d_i - \bar{d}_i| > \sqrt{6n\log n} \right\},$$

then by a union bound,

$$\mathbb{P}(E) \leq \frac{2}{n^2}.$$

Now, clearly, there are constants $c_1' < 1$ and $c_2' > 0$ depending only on L such that $c_2'(n-1) \leq \bar{d}_i \leq c_1'(n-1)$ for all i. Therefore, under E^c, if n is sufficiently large (depending on L), we get constants c_1, c_2 depending only on L such that $c_2(n-1) \leq d_i \leq c_1(n-1)$ for all i.

Next, define

$$g(d_1, \dots, d_n, B) := \sum_{j \notin B} \min\{d_j, |B|\} + |B|(|B| - 1) - \sum_{i \in B} d_i.$$

Note that

$$|g(d_1, \dots, d_n, B) - g(\bar{d}_1, \dots, \bar{d}_n, B)| \le \sum_{i=1}^{n} |d_i - \bar{d}_i| \le n \max_i |d_i - \bar{d}_i|.$$

Moreover, following the notation introduced in the proof of Lemma 4.1, we have

$$\begin{aligned} & g(\bar{d}_1, \dots, \bar{d}_n, B) \\ & \quad = \sum_{j \notin B} (\min\{\bar{d}_j, |B|\} - \bar{d}_j^B) + |B|(|B| - 1) - \sum_{i \in B} \bar{d}_i^B \\ & \quad \ge |B|(|B| - 1) - \sum_{i \in B} \bar{d}_i^B \\ & \quad = \sum_{i, j \in B, i \ne j} (1 - p_{ij}) \ge c_4 |B|(|B| - 1), \end{aligned}$$

where $c_4 \in (0, 1)$ is a constant depending only on L. Thus, under E^c, $n > C$ and $|B| \ge cn$ we have

$$g(d_1, \dots, d_n, B) \ge c_3 n^2 - n^{3/2} \sqrt{6 \log n},$$

where $c_3 \in (0, 1)$ is a constant depending only on L and c. The proof is complete. □

PROOF OF THEOREM 1.3. Let E be the event defined in the proof of Lemma 4.2. Let C, c_1, c_2 be as in Lemma 4.2. By lemmas, if E^c happens and $n > C$, then a solution $\hat{\boldsymbol{\beta}}$ of (3) exists and satisfies $|\hat{\boldsymbol{\beta}}|_\infty \le C(L)$, where $C(L)$ generically denotes a constant that depends only on L. This proves the existence of the MLE. The uniqueness follows from Theorem 1.5.

The proof of the error bound uses Theorem 1.5. Let $\mathbf{x}_0 = \boldsymbol{\beta}$ and define $\{\mathbf{x}_k\}_{k \ge 1}$ as in Theorem 1.5. A simple computation shows that the ith component of $\mathbf{x}_0 - \mathbf{x}_1$ is simply $\log(\bar{d}_i / d_i)$. Under E^c and $n > C$, this is bounded by $C(L)\sqrt{n^{-1} \log n}$. The error bound now follows directly from Theorem 1.5.

Finally, to remove the condition $n > C$, we simply increase $C(L)$ in Theorem 1.3 so that $1 - C(L)n^{-2} < 0$ for $n \le C$. The proof of Theorem 1.3 is complete. □

Addenda. (A) *Practical remarks on the MLE.* Theorem 1.3 shows that with high probability, under the $\mathbb{P}_{\boldsymbol{\beta}}$ measure, for large n the MLE exists and is unique. In applications, a graph is given and Theorem 1.3 may be used to test the $\mathbb{P}_{\boldsymbol{\beta}}$ model.

The MLE may fail to exist because the maximum is taken on at $\beta_i = \pm\infty$ for one or more values of i. For example, with $n = 2$ vertices, an observed graph will either have zero edges or one edge. In the first case, the likelihood is $1/(1 + e^{\beta_1+\beta_2})$, maximized at $\beta_1 = \beta_2 = -\infty$. In the second case the likelihood is $e^{\beta_1+\beta_2}/(1 + e^{\beta_1+\beta_2})$, maximized at $\beta_1 = \beta_2 = \infty$. Here, the MLE fails to exist with probability one.

Similar considerations hold when the observed graph has any isolated vertices and for a star graph. We conjecture: Let G be a graph on n vertices. The MLE for the $\boldsymbol{\beta}$-model exists if and only if the degree sequence lies in the interior of the convex polytope $\operatorname{conv}(\mathcal{D})$ defined in Theorem 1.4.

In cases where the MLE does not exist, it is customary to add a small amount to each degree (see the discussion in [11]). This is often done in a convenient and principled way by using a Bayesian argument.

(B) *Conjugate prior analysis for the $\boldsymbol{\beta}$-model.* Background on conjugate priors for exponential families is in [20] and [24, 25]. The $\boldsymbol{\beta}$-model

$$\mathbb{P}_{\boldsymbol{\beta}}(G) = Z(\boldsymbol{\beta})^{-1} e^{\sum_{i=1}^{n} d_i(G)\beta_i}, \qquad \boldsymbol{\beta} \in \mathbb{R}^n, Z(\boldsymbol{\beta}) = \prod_{1 \le i < j \le n} (1 + e^{\beta_i+\beta_j})$$

has sufficient statistic $\mathbf{d} = (d_1, \dots, d_n)$. Here $\mathbf{d}$ takes values in $\mathcal{D}$, the set of degree sequences for graphs on n vertices. Thus, $\mathbb{P}_{\boldsymbol{\beta}}$ induces a natural exponential family on $\mathcal{D}$ with a base measure μ that does not depend on $\boldsymbol{\beta}$. Following notation in [20], write

$$\mathbb{P}_{\boldsymbol{\beta}}(\mathbf{d}) = \mu(\mathbf{d}) e^{\boldsymbol{\beta}\cdot\mathbf{d} - m(\boldsymbol{\beta})} \qquad \text{with } m(\boldsymbol{\beta}) = \log Z(\boldsymbol{\beta}) = \sum_{1 \le i < j \le n} \log(1 + e^{\beta_i+\beta_j}).$$

Following [20], for $\mathbf{d}_0$ in the interior of $\operatorname{conv}(\mathcal{D})$ and $n_0 > 0$, define the conjugate prior of $\mathbb{R}^n$ by

$$\pi_{n_0,\mathbf{d}_0}(\boldsymbol{\beta}) = Z(n_0, \mathbf{d}_0)^{-1} e^{n_0 \mathbf{d}_0\cdot\boldsymbol{\beta} - n_0 m(\boldsymbol{\beta})}.$$

Here $Z(n_0, \mathbf{d}_0)$ is the normalizing constant, shown to be positive and finite in [20]. By the theory in [20], $\nabla m(\boldsymbol{\beta}) = \mathbb{E}_{\boldsymbol{\beta}}(\mathbf{d})$ and

$$\mathbb{E}_{\pi_{n_0,\mathbf{d}_0}}(\nabla m(\boldsymbol{\beta})) = \mathbb{E}_{\pi_{n_0,\mathbf{d}_0}}(\mathbb{E}_{\boldsymbol{\beta}}(\mathbf{d})) = \mathbf{d}_0.$$

This identity characterizes the prior $\pi_{n_0,\mathbf{d}_0}$. The posterior, given an observed degree sequence $\mathbf{d}(G)$, is

$$\pi_{n_0+1,(\mathbf{d}(G)+n_0\mathbf{d}_0)/(n_0+1)}.$$

Clearly, the mode of the posterior can be found by using the iteration of Theorem 1.5. The proof of Theorem 1.5 shows that the mode exists uniquely for any observed $\mathbf{d}(G)$. The posterior mean must be found using standard Markov chain Monte Carlo techniques.

A natural way to obtain feasible prior mean parameters [i.e., values of $\mathbf{d}_0$ that lie within the interior of $\operatorname{conv}(\mathcal{D})$] is to consider a model of random graphs that puts positive mass on every possible graph on n vertices and take its expected degree sequence. For example, the Erdős–Rényi graph $G(n, p)$, for $0 < p < 1$, is one such model. Its expected degree sequence is $(c, c, \dots, c)$ where $c = (n-1)p$. Thus, $(c, c, \dots, c)$ is a feasible mean parameter for every $c \in (0, n-1)$. Similarly, the expected degree sequence in any of the standard models of power law graphs is a feasible value of $\mathbf{d}_0$ that has power law behavior.

5. Proof of Proposition 1.2 (characterization of the interior).

PROOF OF PROPOSITION 1.2. Let us begin by restating the Erdős–Gallai criterion from Section 1. Suppose $d_1 \geq d_2 \geq \cdots \geq d_n$ are nonnegative integers. The Erdős–Gallai criterion says that $d_1, \dots, d_n$ can be the degree sequence of a simple graph on n vertices if and only if $\sum_{i=1}^n d_i$ is even and for each $1 \leq k \leq n$,

$$\sum_{i=1}^{k} d_i \leq k(k-1) + \sum_{i=k+1}^{n} \min\{d_i, k\}.$$

Now take any function $f \in D'[0,1]$ and let

$$G_f(x) := \int_x^1 \min\{f(y), x\}\, dy + x^2 - \int_0^x f(y)\, dy.$$

Clearly, $G_f(x)$ is continuous as a function of x. If $f \in \mathcal{F}$, the E–G criterion clearly shows that G_f must be a nonnegative function. We claim that this implies that if f belongs to the interior of $\mathcal{F}$, then $G_f(x)$ must be strictly positive for every $x \in (0,1]$. Otherwise, there exists $x \in (0,1]$ such that $G_f(x) = 0$. If we show that there exists a sequence $f_n \to f$ in the modified L^1 topology such that $G_{f_n}(x) < 0$ for each n, then we get a contradiction which proves the claim. This is quite easily done by producing f_n that is strictly bigger than f in $[0, x)$ and equal to f elsewhere, all the while maintaining left-continuity.

Similarly, it is clear that any $f \in \mathcal{F}$ must take values in $[0,1]$. If f attains 0 or 1, then we can produce a sequence $f_n \to f$ whose ranges are not contained in $[0,1]$ and, therefore, f cannot belong to the interior of $\mathcal{F}$.

Thus, we have proved that if f belongs to the interior of $\mathcal{F}$, then f must satisfy the two conditions of Proposition 1.2. Let us now prove the converse. Suppose $f \in D'[0,1]$ such that $0 < c_1 < f(x) < c_2 < 1$ for all $x \in [0,1]$ and $G_f(x) > 0$ for all $x \in (0,1]$. We have to show that any function that is sufficiently close to f in the modified L^1 norm must belong to $\mathcal{F}$.

To do that let us first prove that $f \in \mathcal{F}$. Take any n. Let $d_i^n = \lfloor n f(i/n) \rfloor$, $i = 2, \dots, n$, and $d_1^n = \lfloor n f(0) \rfloor$. Since f is nonincreasing, we have $d_1^n \geq d_2^n \geq \cdots \geq d_n^n$. Increase some of the d_i^n's by 1, if necessary, so that $\sum d_i^n$ is even (and monotonicity is maintained). With this construction, it is clear that

$$\left|\frac{d_1^n}{n} - f(0)\right| + \left|\frac{d_n^n}{n} - f(1)\right| + \frac{1}{n}\sum_{i=1}^{n}\left|\frac{d_i^n}{n} - f\left(\frac{i}{n}\right)\right| \leq \frac{4}{n}.$$

Thus, if $\mathbf{d}^n$ denotes the vector $(d_1^n, \dots, d_n^n)$, then $\mathbf{d}^n$ converges to the scaling limit f. We need to show that for all large enough n, $\mathbf{d}^n$ is a valid degree sequence.

Since f is bounded and nonincreasing,

$$\lim_{n\to\infty} \int_0^1 |f(x) - f(\lceil nx \rceil/n)|\, dx = 0$$

and so uniformly in $1 \le k \le n$,

$$\left| \frac{\sum_{i=k+1}^n \min\{d_i^n, k\} + k(k-1) - \sum_{i=1}^k d_i^n}{n^2} - G_f(k/n) \right| \le \varepsilon(n),$$

where $\varepsilon(n) \to 0$ as $n \to \infty$. Thus, there exists a sequence of integers $\{k_0(n)\}$, where $k_0(n)/n \to 0$ as $n \to \infty$, such that whenever $k \ge k_0(n)$, we have

$$\sum_{i=k+1}^n \min\{d_i^n, k\} + k(k-1) - \sum_{i=1}^k d_i^n > 0.$$

Again, there exists $c_1' < 1$ and $c_2' > 0$ such that if n is sufficiently large, we have $c_2' \le d_i^n/n \le c_1'$ for all i. Suppose n is so large that $k_0(n)/n < c_2'$ and $(1 - c_1')n - k_0(n) > 0$. Then, if $k \le k_0(n)$, we have

$$\begin{aligned}
&\sum_{i=k+1}^n \min\{d_i^n, k\} + k(k-1) - \sum_{i=1}^k d_i^n \\
&\qquad \ge \sum_{i=k+1}^n \min\{c_2' n, k\} + k(k-1) - \sum_{i=1}^k n c_1' \\
&\qquad = (n-k)k + k(k-1) - c_1' nk \\
&\qquad = \big((1 - c_1')n - k\big)k + k(k-1) > 0.
\end{aligned}$$

Thus, for n so large, we have that for all $1 \le k \le n$,

$$\sum_{i=k+1}^n \min\{d_i^n, k\} + k(k-1) - \sum_{i=1}^k d_i^n > 0.$$

By the Erdős–Gallai criterion, this shows that $(d_1^n, \dots, d_n^n)$ is a valid degree sequence.

Thus, we have shown that any f that satisfies the two conditions of Proposition 1.2 must belong to $\mathcal{F}$. Now we only have to show that if f satisfies the two criteria, then any h sufficiently close to f in the modified L^1 norm must also satisfy them.

Note that G_f is a continuous function that is positive in (0, 1]. Moreover, for all $0 \le x \le 1$,

$$|G_f(x) - G_{f'}(x)| \le \|f - f'\|_{1'}$$

so if $f_n \to f$ in the modified L^1 norm, then $G_{f_n} \to G_f$ in the supnorm. Thus, for any $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $\|h - f\|_{1'} < \delta$, we have $G_h(x) > 0$ for all $x \in [\varepsilon, 1]$. We also have that $c_1 - \delta \le h(x) \le c_2 + \delta$ for all $0 \le x \le 1$. Choosing $\delta, \varepsilon > 0$ small as necessary, we can ensure that $c_1 - \delta > \varepsilon$ and $1 - \varepsilon - \delta - c_2 > 0$. Fix such ε, δ and h. Then, for $x \in (0, \varepsilon)$, we have

$$\begin{aligned} G_h(x) &\ge \int_x^1 \min\{c_1 - \delta, x\}\, dy + x^2 - \int_0^x (c_2 + \delta)\, dy \\ &= (1 - x)x + x^2 - (c_2 + \delta)x \\ &= (1 - \varepsilon - \delta - c_2)x + x^2 > 0. \end{aligned}$$

But we also have $G_h(x) > 0$ for $x \in [\varepsilon, 1]$ by the choice of δ. Thus, we have proved that there exists $\delta > 0$ such that whenever $\|h - f\|_{1'} < \delta$, we have $G_h(x) > 0$ for all $x \in (0, 1]$. Choosing δ sufficiently small, we can ensure that the range of h does not contain 0 or 1. The proof of Proposition 1.2 is complete. □

Proposition 1.2 can be extended into a complete version of the Erdős–Gallai criterion for graph limits. Suppose that $W(x, y)$ is a symmetric function from $[0, 1]^2$ into $[0, 1]$. In [18], Section 4, it is shown that the correct analog of the degree distribution for the graph limit W is the distribution of the random variable

$$X = \int_0^1 W(U, y)\, dy, \tag{16}$$

where U is a random variable distributed uniformly in $[0, 1]$. If a sequence of graphs converges to W then the distribution of the random variable d_i/n (where i is chosen uniformly from n vertices and d_i is the degree of i) converges to X in distribution. The following result characterizes limiting degree variates.

PROPOSITION 5.1. *Let X be a random variable with values in* $[0, 1]$. *Let* $D(x) = \sup\{y : P(X > y) \ge x\}$. *Then X has the representation* (16) *if and only if for all* $x \in (0, 1]$

$$\int_0^x D(y)\, dy \le x^2 + \int_x^1 \min\{D(y), x\}\, dy.$$

The proof is essentially as given above, approximating W by a sequence of finite graphs and using the Erdős–Gallai criterion. We omit further details.

6. Proof of Theorem 1.1 (convergence to graph limit).

6.1. *Preliminary lemmas.* We need a couple of probabilistic results before we can embark on the proof of Theorem 1.1. The first one is a simple application of the method of bounded differences for concentration inequalities.

LEMMA 6.1. *Let H be a finite simple graph of size $\leq n$. Let G be a random graph on n vertices with independent edges. Let $t(H,G)$ be the homomorphism density of H in G, defined in* (1). *Then for any $\varepsilon > 0$,*

$$\mathbb{P}\big(|t(H,G) - \mathbb{E}t(H,G)| > \varepsilon\big) \leq 2e^{-C\varepsilon^2 n^2},$$

where C is a constant that depends only on H.

PROOF. The proof is a simple consequence of the bounded difference inequality [36]. Note that the quantity $t(H,G)$ is a function of the edges of G, considered as independent Bernoulli random variables. When a particular edge is added or removed (i.e., the corresponding Bernoulli variable is set equal to 1 or 0), $\hom(H,G)$ is altered by at most $Cn^{|V(H)|-2}$, where C is a constant that depends only on H. This is because when we fix an edge, we are fixing its two endpoints, which leaves us the freedom of choosing the remaining $|V(H)|-2$ vertices arbitrarily when constructing a homomorphism.

Thus, alteration of the status of an edge changes $t(H,G)$ by at most Cn^{-2}. The bounded difference inequality completes the proof. □

The second preliminary result that we need is a kind of local limit theorem that we need to pass from the $\boldsymbol{\beta}$-model to graphs with given degree sequence.

Let $\mathbf{d} = (d_1, \ldots, d_n)$ be a valid degree sequence on a graph of size n. Let $G = (V, E)$ be a random graph on n vertices labeled $1, \ldots, n$ so that edges i, j are connected with probability p_{ij} satisfying $d_i = \sum_{j \neq i} p_{ij}$ and so that $\delta \leq p_{ij} \leq 1 - \delta$ for some fixed $0 < \delta < \frac{1}{2}$. Let w_{ij} denote the indicator that (i, j) is an edge in G. We obtain a lower bound on the probability that G has degree sequence $\mathbf{d}$.

LEMMA 6.2. *For any $\varepsilon > 0$ and large enough n, the random graph G has degree sequence $\mathbf{d}$ with probability at least $\frac{1}{2}\exp(-\log(\delta) n^{(3/2)+\varepsilon})$.*

We first prove the following claim about the existence of 0–1 contingency tables. An $m \times n$ 0–1 contingency table with integer row and column sums $r_1, \ldots, r_m$ and $c_1, \ldots, c_n$ is an $m \times n$ matrix whose entries are 0 or 1 and whose ith row and jth column sum to r_i and c_j, respectively. Denote the conjugate sequences as $r_i^* = \#\{r_j : r_j \geq i\}$ and $c_i^* = \#\{c_j : c_j \geq i\}$. Let $(r_{[i]})$, $(c_{[i]})$ denote the order statistics of (r_i) and (c_i), that is, permutations of the sequences such that $r_{[1]} \geq r_{[2]} \geq \cdots \geq r_{[m]}$ and $c_{[1]} \geq c_{[2]} \geq \cdots \geq c_{[m]}$.

A condition of Gale and Ryser [23, 49] says that there exists a 0–1 contingency table for row and column sums $r_1, \ldots, r_m$ and $c_1, \ldots, c_n$ if and only if $\sum_{i=1}^m r_i = \sum_{i=1}^n c_i$ and

$$\sum_{i=1}^k r_{[i]} \leq \sum_{i=1}^k c_i^*, \qquad 1 \leq k \leq m, \tag{17}$$

$$\sum_{i=1}^{k} c_{[i]} \le \sum_{i=1}^{k} r_i^*, \qquad 1 \le k \le n. \tag{18}$$

CLAIM 6.3. *Let $0 < \delta < \frac{1}{2}$ and let (p_{ij}) be an $m \times n$ matrix such that $\delta \le p_{ij} \le 1 - \delta$. Suppose that (r_i) and (c_i) are integer sequences satisfying the following*:

- $\sum_{i=1}^{m} r_i = \sum_{i=1}^{n} c_i$;
- $|r_i - \sum_{j=1}^{n} p_{ij}| \le \frac{1}{4}\delta^2 n$ *for* $1 \le i \le m$;
- $|c_j - \sum_{i=1}^{m} p_{ij}| \le \frac{1}{4}\delta^2 m$ *for* $1 \le j \le n$.

Then there exists a 0–1 *contingency table with row and column sums* (r_i) *and* (c_i).

PROOF. We establish that the Gale–Ryser conditions hold. Without loss of generality we may assume that $r_1 \ge r_2 \ge \cdots \ge r_m$. Then condition (17) is equivalent to

$$\sum_{i=1}^{k} r_i \le \sum_{i=1}^{k} c_i^* = \sum_{i=1}^{k}\sum_{j=1}^{n} \mathbb{1}_{\{c_j \ge i\}} = \sum_{j=1}^{n} \min\{k, c_j\}. \tag{19}$$

Now

$$\sum_{i=1}^{k} r_i \le \sum_{i=1}^{k}\sum_{j=1}^{n} p_{ij} + \frac{1}{4}\delta^2 kn$$

and hence,

$$\begin{aligned}
\sum_{j=1}^{n} \min\{k, c_j\} &\ge \sum_{j=1}^{n} \min\left\{k, \sum_{i=1}^{m} p_{ij} - \frac{1}{4}\delta^2 m\right\} \\
&\ge \sum_{i=1}^{k}\sum_{j=1}^{n} p_{ij} + \sum_{j=1}^{n} \min\left\{k - \sum_{i=1}^{k} p_{ij}, \sum_{i=k+1}^{m} p_{ij} - \frac{1}{4}\delta^2 m\right\} \\
&\ge \sum_{i=1}^{k}\sum_{j=1}^{n} p_{ij} + \sum_{j=1}^{n} \min\left\{\delta k, (m-k)\delta - \frac{1}{4}\delta^2 m\right\} \\
&\ge \sum_{i=1}^{k} r_i + n\left(\min\left\{\delta k, (m-k)\delta - \frac{1}{4}\delta^2 m\right\} - \frac{1}{4}\delta^2 k\right),
\end{aligned}$$

where we used the fact that $\delta \le p_{ij} \le 1 - \delta$. Now $\delta k \ge \frac{1}{4}\delta^2 k$ and when $1 \le k \le$

$m(1-\delta+\frac{1}{4}\delta^2)$,

$$\begin{aligned}(m-k)\delta-\tfrac{1}{4}\delta^2 m &= m\delta\bigl(1-\tfrac{1}{4}\delta\bigr)-k\delta\\ &\geq m\delta\bigl(1-\delta+\tfrac{1}{4}\delta^2\bigr)\bigl(1+\tfrac{1}{4}\delta\bigr)-k\delta\\ &\geq k\delta\bigl(1+\tfrac{1}{4}\delta\bigr)-k\delta=\tfrac{1}{4}\delta^2 k\end{aligned}$$

and hence, $n(\min\{\delta k, (m-k)\delta-\frac{1}{4}\delta^2 m\}-\frac{1}{4}\delta^2 k)\geq 0$. To establish equation (19) it then suffices to consider $m(1-\delta+\frac{1}{4}\delta^2)\leq k\leq m$. In this case,

$$c_j\leq\sum_{i=1}^m p_{ij}+\frac{1}{4}\delta^2 m\leq(1-\delta)m+\frac{1}{4}\delta^2 m\leq k$$

and so

$$\sum_{j=1}^n\min\{k,c_j\}=\sum_{j=1}^n c_j=\sum_{i=1}^m r_i\geq\sum_{i=1}^k r_{[i]}$$

establishing (17). Condition (18) follows similarly and hence, there exists a 0–1 contingency table with the prescribed row and column sums. □

We are now ready to prove Lemma 6.2.

PROOF. We split the n vertices into subsets $A=1,2,\dots,n-n^a$ and $B=n-n^a+1,\dots,n$ where $a=\frac{1}{2}+\varepsilon$. For $1\leq i<j\leq|A|$, choose w_{ij} according to p_{ij}. Let $\mathcal{G}$ denote the event that the following conditions hold:

- For all $i\in A$

$$\left|\sum_{j\in A\setminus\{i\}}w_{ij}-\sum_{j\in A\setminus\{i\}}p_{ij}\right|<n^{(1+\varepsilon)/2};\tag{20}$$

- That the total number of edges in the subgraph induced by A satisfies

$$\sum_{i\in A}\Bigl(d_i-\sum_{j\in A\setminus\{i\}}w_{ij}\Bigr)<\sum_{i\in B}d_i<\sum_{i\in A}\Bigl(d_i-\sum_{j\in A\setminus\{i\}}w_{ij}\Bigr)+|B|(|B|-1).\tag{21}$$

Both conditions hold with high probability by simple applications of Hoeffding's inequality [26]. For example, the first follows from Hoeffding's inequality as

$$P\biggl(\biggl|\sum_{j\in A\setminus\{i\}}(w_{ij}-\mathbb{E}w_{ij})\biggr|\geq n^{(1+\varepsilon)/2}\biggr)\leq 2e^{-(1/2)n^{\varepsilon}}$$

and taking a union bound over $i\in A$.

We will show that given $\mathcal{G}$ there is always a way to add edges between vertices in $B \times V$ so that the graph has degree sequence $\mathbf{d}$. First we choose any assignment of the edges $(w_{ij})_{i,j\in B}$ in $B \times B$ so that the total number of edges equals

$$\frac{1}{2}\biggl(\sum_{i\in B} d_i - \sum_{i\in A}\biggl(d_i - \sum_{j\in A\setminus\{i\}} w_{ij}\biggr)\biggr)$$

which is an integer because the sum of the degrees is even and is between 0 and $\frac{1}{2}|B|(|B|-1)$ by equation (21).

It remains to assign edges between A and B so that the graph has degree sequence $\mathbf{d}$. This is exactly equivalent to the question of finding a 0–1 contingency table with dimensions $|A| \times |B|$, row sums $r_i = d_i - \sum_{j\in A\setminus\{i\}} w_{ij}$ for $i \in A$ and column sums $c_i = d_i - \sum_{j\in B\setminus\{i\}} w_{ij}$ for $i \in B$.

Condition (20) guarantees that $r_i = [1 + o(1)]\sum_{j\in B} p_{ij}$ and since $|B| = o[|A|)]$, we have that $c_i = [1+o(1)]\sum_{j\in A} p_{ij}$ uniformly in n. Hence, by Claim 6.3 a 0–1 contingency table with row and column sums (r_i) and (c_j) exists.

Hence, whenever the edges $(w_{ij})_{i,j\in A}$ satisfy $\mathcal{G}$ there exists at least one way to assign the other edges so that the graph has degree sequence $\mathbf{d}$. Since any configuration $(w_{ij})_{i\in V, j\in B}$ has probability at least $\delta^{|V||B|}$ and is independent of $\mathcal{G}$, the probability that G has the degree sequence $\mathbf{d}$ is at least $P(\mathcal{G})\exp(-\log(\delta)n^{1+a})$ and the result follows since $\mathcal{G}$ holds with high probability. □

An alternative approach in the above lower bound could be through the enumeration of the number of graphs of a particular degree sequence as carried out in [10]. In fact, this approach would give a better lower bound than the one we obtain. This was brought to our attention recently by Alexander Barvinok.

6.2. *Proof of Theorem* 1.1. Let $\mathbf{d}^n$, G_n and f be as in the statement of the theorem. By Proposition 1.2 we know that f has the following two properties:

A. There are two constants $c_1 > 0$ and $c_2 < 1$ such that $c_1 \le f(x) \le c_2$ for all $x \in [0, 1]$.
B. For each $0 < b \le 1$,

$$\inf_{x\ge b}\biggl\{\int_x^1 \min\{f(y), x\}\,dy + x^2 - \int_0^x f(y)\,dy\biggr\} > 0.$$

(The infimum is positive because the term within the brackets is a positive continuous function of x.) Now fix n and for each $B \subseteq \{1, \dots, n\}$, consider the quantity

$$\mathcal{E}(B) := \sum_{j\notin B} \min\{d_j^n, |B|\} + |B|(|B|-1) - \sum_{i\in B} d_i^n.$$

Under the assumption that $d_1^n \ge d_2^n \ge \cdots \ge d_n^n$, we claim that for each $1 \le k \le n$, $\mathcal{E}(B)$ is minimized over all subsets B of size k when $B = \{1, \dots, k\}$. To prove this, take any B of size k. Suppose there is $a \in B$ and $b \notin B$ such that $b < a$. Let $B' = (B \setminus \{a\}) \cup \{b\}$. Then clearly, since $d_b^n \geq d_a^n$, we have

$$\sum_{j \notin B} \min\{d_j^n, k\} \geq \sum_{j \notin B'} \min\{d_j^n, k\}$$

and

$$\sum_{i \in B} d_i^n \leq \sum_{i \in B'} d_i^n.$$

Thus, $\mathcal{E}(B) \geq \mathcal{E}(B')$, which proves the claim. Now by the definition of convergence of degree sequences and the fact that f is bounded and nonincreasing,

$$\begin{aligned}
&\left|\sum_{i=1}^{k} \frac{1}{n} \cdot \frac{d_i^n}{n} - \int_0^{k/n} f(y)\,dy\right| \\
&\qquad \leq \sum_{i=1}^{n} \left|\frac{1}{n} \cdot \frac{d_i^n}{n} - \int_{(i-1)/n}^{i/n} f(y)\,dy\right| \qquad (22)\\
&\qquad \leq \frac{1}{n} \sum_{i=1}^{n} \left|\frac{d_i^n}{n} - f\left(\frac{i}{n}\right)\right| + \int_0^1 |f(x) - f(\lceil nx \rceil/n)|\,dx \to 0
\end{aligned}$$

for $1 \leq k \leq n$. Similarly

$$\sum_{j=k+1}^{n} \frac{1}{n} \min\left\{\frac{d_j^n}{n}, \frac{k}{n}\right\} - \int_{k/n}^{1} \min\{f(y), k/n\}\,dy \to 0$$

uniformly in $1 \leq k \leq n$ as $n \to \infty$. Hence, we have that for any $b \in (0, 1)$,

$$\begin{aligned}
&\frac{1}{n^2} \min_{B \subseteq \{1,\dots,n\}, |B| \geq bn} \mathcal{E}(B) \\
&\quad = \min_{k \geq bn} \left\{ \sum_{j=k+1}^{n} \frac{1}{n} \min\left\{\frac{d_j^n}{n}, \frac{k}{n}\right\} + \frac{k(k-1)}{n^2} - \sum_{i=1}^{k} \frac{1}{n} \cdot \frac{d_i^n}{n} \right\} \\
&\quad \to \inf_{x \geq b} \left\{ \int_x^1 \min\{f(y), x\}\,dy + x^2 - \int_0^x f(y)\,dy \right\} \qquad \text{as } n \to \infty.
\end{aligned}$$

Thus, we can apply properties A and B of the function f, the definition of scaling limit of degree sequences and Lemma 4.1 to conclude that for all large n, a solution $\boldsymbol{\beta}^n = (\beta_1^n, \dots, \beta_n^n)$ to (3) for $\mathbf{d}^n$ exists and $|\boldsymbol{\beta}^n|_\infty$ is uniformly bounded.

For each n, define a function $g_n : [0, 1] \to \mathbb{R}$ as

$$g_n(x) := \beta_i^n \qquad \text{if } \frac{i-1}{n} < x \leq \frac{i}{n}$$

and let $g_n(0) := \beta_1^n$. Now fix two positive integers m, n and let

$$N := mn.$$

Define a vector $\mathbf{x}_0 = (x_{0,1}, \dots, x_{0,N}) \in \mathbb{R}^N$ as follows:

$$x_{0,i} = \beta_k^n \qquad \text{if } m(k-1)+1 \le i \le mk.$$

In other words,

$$\mathbf{x}_0 = (\beta_1^n, \beta_1^n, \dots, \beta_1^n, \beta_2^n, \beta_2^n, \dots, \beta_2^n, \dots, \beta_n^n, \beta_n^n, \dots, \beta_n^n),$$

where each β_k^n is repeated m times. For $\ell \ge 1$ define $\mathbf{x}_\ell = \varphi(\mathbf{x}_{\ell-1})$ as in Theorem 1.5 (with N in place of n). Equivalently,

$$x_{\ell,i} - x_{\ell-1,i} = \log d_i^N - \log y_{\ell-1,i} = \log \frac{d_i^N/N}{y_{\ell-1,i}/N}, \tag{23}$$

where

$$y_{\ell,i} := \sum_{j \ne i} \frac{e^{x_{\ell,i}+x_{\ell,j}}}{1+e^{x_{\ell,i}+x_{\ell,j}}}.$$

Note that by definition of $y_{0,i}$ and $x_{0,i}$, if $m(k-1)+1 \le i \le mk$,

$$y_{0,i} - m d_k^n = (m-1)\frac{e^{2\beta_k^n}}{1+e^{2\beta_k^n}} \le m.$$

Consequently, if $m(k-1)+1 \le i \le mk$,

$$|y_{0,i}/N - d_k^n/n| \le 1/n. \tag{24}$$

Hence, by equation (2) [similarly to (22)] it follows that

$$\frac{1}{N}\sum_{i=1}^{N} |y_{0,i}/N - d_i^N/N| \le \varepsilon_1(n)$$

uniformly in N where $\varepsilon_1(n) \to 0$ as $n \to \infty$. From (2), (23), (24) (and implicitly using the continuity of log, property A of the function f and the uniform boundedness of $|\boldsymbol{\beta}^n|_\infty$), we see that

$$|\mathbf{x}_0 - \mathbf{x}_1|_1 \le N\varepsilon_2(n)$$

uniformly in m where $\varepsilon_2(n) \to 0$ as $n \to \infty$. Since $|\boldsymbol{\beta}^n|_\infty$ is uniformly bounded in n by Theorem 1.5, it follows that for large enough n, m,

$$|\mathbf{x}_\ell - \boldsymbol{\beta}^N|_\infty \le K\theta^\ell \tag{25}$$

for some K and $0 < \theta < 1$ independent of n and m. Hence, for some K', also independent of n, m,

$$\sup |\mathbf{x}_\ell|_\infty \le K'.$$

Consequently, by Lemma 2.2 we have that

$$|\mathbf{x}_0 - \mathbf{x}_\ell|_1 \le \left(\sum_{i=1}^{\ell} (2e^{2K'})^i\right) |\mathbf{x}_0 - \mathbf{x}_1|_1. \tag{26}$$

Combining equations (25) and (26) and using the fact that $|\mathbf{x}|_1 \le N|\mathbf{x}|_\infty$ we have that

$$|\mathbf{x}_0 - \boldsymbol{\beta}^N|_1 \le |\mathbf{x}_0 - \mathbf{x}_\ell|_1 + |\mathbf{x}_\ell - \boldsymbol{\beta}^N|_1 \le \left(\sum_{i=1}^{\ell}(2e^{2K'})^i\right) N\varepsilon_2(n) + K\theta^\ell N.$$

Now taking $\ell = \ell(n)$ to infinity slowly enough so that

$$\left(\sum_{i=1}^{\ell}(2e^{2K'})^i\right)\varepsilon_2(n) \to 0$$

it follows that

$$|\mathbf{x}_0 - \boldsymbol{\beta}^N|_1 \le N\varepsilon_3(n)$$

uniformly in m where $\varepsilon_1(n) \to 0$ as $n \to \infty$. But

$$|\mathbf{x}_0 - \boldsymbol{\beta}^N|_1 = N\|g_n - g_N\|_1,$$

where $\|\cdot\|_1$ is the usual L^1 norm on functions on [0, 1]. Thus,

$$\|g_n - g_m\|_1 \le \|g_n - g_N\|_\infty + \|g_m - g_N\|_\infty \le \varepsilon_3(n) + \varepsilon_3(m).$$

This shows that the sequence $\{g_n\}$ is Cauchy under the L^1 norm and thus there exists a uniformly bounded function g^* such that $\|g_n - g^*\|_1 \to 0$. Now, for each n define a function f_n as

$$f_n(x) := \int_0^1 \frac{e^{g_n(x)+g_n(y)}}{1+e^{g_n(x)+g_n(y)}}\,dy.$$

Now by the uniform boundedness of the $|g_n|_\infty$,

$$\int_0^1 \left| f_n(x) - \int_0^1 \frac{e^{g^*(x)+g^*(y)}}{1+e^{g^*(x)+g^*(y)}}\,dy \right| dx \to 0$$

as $n \to \infty$. But from the relation between $\boldsymbol{\beta}^n$ and $\mathbf{d}^n$, it is easy to see that for $x \in (0, 1]$ that $f_n(x) = d^n_{\lceil nx\rceil}/n + O(1/n)$ and hence,

$$\lim_n \|f - f_n\|_1 \to 0.$$

It follows that

$$f(x) = \int_0^1 W^*(x, y)\,dy \qquad \text{a.e.,} \tag{27}$$

where

$$W^*(x, y) = \frac{e^{g^*(x)+g^*(y)}}{1+e^{g^*(x)+g^*(y)}}.$$

We now adjust g^* on a set of measure 0 so that equation (27) holds for all x. Set $\psi : \mathbb{R} \to (0, 1)$ as

$$\psi(z) = \int_0^1 \frac{e^{z+g^*(y)}}{1 + e^{z+g^*(y)}}\, dy.$$

By construction and since g^* is uniformly bounded, it follows that $\psi(z)$ is continuous, strictly increasing and bijective. By equation (27) we have that

$$f(x) = \psi(g^*(x)) \qquad \text{a.e.}$$

and hence, if we set

$$g(x) = \psi^{-1}(f(x)),$$

then $g(x) = g^*(x)$ almost everywhere. Then for all $x \in [0, 1]$,

$$f(x) = \int_0^1 W(x, y)\, dy,$$

where

$$W(x, y) = \frac{e^{g(x)+g(y)}}{1 + e^{g(x)+g(y)}}.$$

Moreover, by the properties of ψ and f, we have that $g \in D'[0, 1]$ and its points of discontinuity are the same as f.

Let us now prove that g is the only function in $D'[0, 1]$ with the above relationship with f. Suppose h is another such function. Fix any n. Define a vector $\mathbf{x}_0 = (x_{0,1}, \dots, x_{0,n}) \in \mathbb{R}^n$ as

$$x_{0,i} := h(i/n), \qquad i = 1, \dots, n.$$

For each $1 \le i \le n$, define

$$y_i := \sum_{j \neq i} \frac{e^{x_{0,i}+x_{0,j}}}{1 + e^{x_{0,i}+x_{0,j}}}.$$

Then since $h \in D'[0, 1]$,

$$\sup_i |y_i/n - f(i/n)| = \sup_i \left| y_i/n - \int_0^1 \frac{e^{h(i/n)+h(y)}}{1 + e^{h(i/n)+h(y)}}\, dy \right| \le \varepsilon_4(n), \tag{28}$$

where $\varepsilon_4(n) \to 0$ as $n \to \infty$. Define $\mathbf{x}_1$ in terms of $\mathbf{x}_0$ and $\mathbf{d}^n$ as in Theorem 1.5. Then for each i,

$$x_{1,i} - x_{0,i} = \log d_i^n - \log y_i = \log \frac{d_i^n/n}{y_i/n}.$$

From (2), (28) and the above identity (and implicitly using the property A of f), we see that

$$|\mathbf{x}_1 - \mathbf{x}_0|_\infty \le \varepsilon_5(n),$$

where $\varepsilon_5(n) \to 0$ as $n \to \infty$. Thus, by Theorem 1.5 we get

$$|\mathbf{x}_0 - \boldsymbol{\beta}^n|_\infty \le \varepsilon_6(n),$$

where $\varepsilon_6(n) \to 0$ as $n \to \infty$. This implies that $\|h - g_n\|_1 \to 0$ and hence, that $h = g$ a.e. Since we assumed both h and g are in $D'[0, 1]$ this implies that $g = h$ on $(0, 1]$. To show that $g(0) = h(0)$, observe that since $g = h$ on $(0, 1]$,

$$f(0) = \int_0^1 \frac{e^{h(0)+h(y)}}{1 + e^{h(0)+h(y)}}\, dy = \int_0^1 \frac{e^{h(0)+g(y)}}{1 + e^{h(0)+g(y)}}\, dy = \psi(h(0))$$

and therefore, by the injectivity of ψ, $g(0) = h(0)$.

Now fix a finite simple graph H. Let $\boldsymbol{\beta}^n$ be as above. Let G_n' denote a random graph from the $\boldsymbol{\beta}^n$-model. Let $\mathbf{d}_n'$ be the degree sequence of G'. Then it is easy to see that conditional on the event $\{\mathbf{d}_n' = \mathbf{d}_n\}$ the law of G_n' is the same as that of G_n.

By Lemma 6.1, given any $\varepsilon > 0$, we have that

$$\mathbb{P}\big(|t(H, G_n') - \mathbb{E}t(H, G_n')| > \varepsilon\big) \le e^{-C_1 n^2},$$

where C_1 is a constant that depends only on H and ε. By Lemma 6.2, we know that

$$\mathbb{P}(\mathbf{d}_n' = \mathbf{d}_n) \ge e^{-C_2 n^{7/4}},$$

where C_2 is another constant that depends only on $|\boldsymbol{\beta}|_\infty$. Thus,

$$\begin{aligned}
\mathbb{P}\big(|t(H, G_n) - \mathbb{E}t(H, G_n')| > \varepsilon\big) &= \mathbb{P}\big(|t(H, G_n') - \mathbb{E}t(H, G_n')| > \varepsilon | \mathbf{d}_n' = \mathbf{d}_n\big) \\
&\le \frac{\mathbb{P}(|t(H, G_n') - \mathbb{E}t(H, G_n')| > \varepsilon)}{\mathbb{P}(\mathbf{d}_n' = \mathbf{d}_n)} \\
&\le e^{-C_3 n^2},
\end{aligned}$$

where C_3 is a constant depending on H, ε and $|\boldsymbol{\beta}|_\infty$. Since $g_n \to g$, it is easy to prove that G_n' converges to W almost surely. From the above inequality, it follows that G_n' and G_n must have the same limit almost surely. The proof of the theorem is complete.

Acknowledgments. The authors are indebted to Joe Blitzstein for many helpful tips and pointers to the literature and Martin Wainwright for the references to [4, 16, 52]. We particularly thank Alexander Barvinok for calling our attention to [10] and suggesting possible connections to our work and Svante Janson for a very careful reading of the manuscript and pointing out numerous small errors. Last, we thank the Associate Editor for a number of useful comments.

REFERENCES

[1] ALDOUS, D. J. (1981). Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal*. **11** 581–598. MR0637937

[2] AUSTIN, T. (2008). On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probab. Surv.* **5** 80–145. MR2426176

[3] AUSTIN, T. and TAO, T. (2010). Testability and repair of hereditary hypergraph properties. *Random Structures Algorithms* **36** 373–463. MR2666763

[4] BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory*. Wiley, Chichester. MR0489333

[5] BARVINOK, A. (2010). What does a random contingency table look like? *Combin. Probab. Comput*. **19** 517–539. MR2647491

[6] 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. MR2600999

[7] BARVINOK, A. and HARTIGAN, J. A. (2009). An asymptotic formula for the number of nonnegative integer matrices with prescibed row and column sums. Preprint. Available at http://arxiv.org/abs/0910.2477.

[8] BARVINOK, A. and HARTIGAN, J. A. (2009). Maximum entropy Edgeworth estimates of volumes of polytopes. Preprint. Available at http://arxiv.org/abs/0910.2497.

[9] BARVINOK, A. and HARTIGAN, J. A. (2010). Maximum entropy Gaussian approximations for the number of integer points and volumes of polytopes. *Adv. in Appl. Math*. **45** 252–289. MR2646125

[10] BARVINOK, A. and HARTIGAN, J. A. (2010). The number of graphs and a random graph with a given degree sequence. Preprint. Available at http://arxiv.org/abs/1003.0356.

[11] BISHOP, Y. M. M., FIENBERG, S. E. and HOLLAND, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, MA.

[12] BLITZSTEIN, J. and DIACONIS, P. (2009). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. Preprint. Available at http://www.people.fas.harvard.edu/~blitz/BlitzsteinDiaconisGraphAlgorithm.pdf.

[13] BORGS, C., CHAYES, J., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2006). Counting graph homomorphisms. In *Topics in Discrete Mathematics*. *Algorithms Combin*. **26** 315–371. Springer, Berlin. MR2249277

[14] BORGS, C., CHAYES, J. T., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2008). Convergent sequences of dense graphs. I. Subgraph frequencies, metric properties and testing. *Adv. Math*. **219** 1801–1851. MR2455626

[15] BORGS, C., CHAYES, J., LOVÁSZ, L., SÓS, V. T. and VESZTERGOMBI, K. (2007). Convergent sequences of dense graphs II. Multiway cuts and statistical physics. Preprint. Available at http://research.microsoft.com/en-us/um/people/borgs/papers/conright.pdf.

[16] BROWN, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **9**. IMS, Hayward, CA. MR0882001

[17] DIACONIS, P. and FREEDMAN, D. (1984). Partial exchangeability and sufficiency. In *Statistics: Applications and New Directions (Calcutta, 1981)* 205–236. Indian Statist. Inst., Calcutta. MR0786142

[18] DIACONIS, P., HOLMES, S. and JANSON, S. (2008). Threshold graph limits and random threshold graphs. *Internet Math*. **5** 267–320 (2009). MR2573956

[19] DIACONIS, P. and JANSON, S. (2008). Graph limits and exchangeable random graphs. *Rend. Mat. Appl. (7)* **28** 33–61. MR2463439

[20] DIACONIS, P. and YLVISAKER, D. (1979). Conjugate priors for exponential families. *Ann. Statist*. **7** 269–281. MR0520238

[21] ERDŐS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.* **5** 17–61. MR0125031
[22] ERDŐS, P. and GALLAI, T. (1960). Graphen mit punkten vorgeschriebenen grades. *Mat. Lapok* **11** 264–274.
[23] GALE, D. (1957). A theorem on flows in networks. *Pacific J. Math.* **7** 1073–1082. MR0091855
[24] GUTIÉRREZ-PEÑA, E. and SMITH, A. F. M. (1995). Conjugate parameterizations for natural exponential families. *J. Amer. Statist. Assoc.* **90** 1347–1356. MR1379477
[25] GUTIÉRREZ-PEÑA, E. and SMITH, A. F. M. (1996). Erratum: "Conjugate parameterizations for natural exponential families." *J. Amer. Statist. Assoc.* **91** 1757. MR1439120
[26] HOEFFDING, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30. MR0144363
[27] HOLLAND, P. W. and LEINHARDT, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176
[28] HOOVER, D. N. (1982). Row-column exchangeability and a generalized model for probability. In *Exchangeability in Probability and Statistics (Rome, 1981)* 281–291. North-Holland, Amsterdam. MR0675982
[29] HUNTER, D. R. (2004). MM algorithms for generalized Bradley–Terry models. *Ann. Statist.* **32** 384–406. MR2051012
[30] JACKSON, M. O. (2008). *Social and Economic Networks*. Princeton Univ. Press, Princeton, NJ. MR2435744
[31] KOLACZYK, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer, New York.
[32] LAURITZEN, S. L. (1988). *Extremal Families and Systems of Sufficient Statistics*. *Lecture Notes in Statistics* **49**. Springer, New York. MR0971253
[33] LI, L., ALDERSON, D., DOYLE, J. C. and WILLINGER, W. (2005). Towards a theory of scale-free graphs: Definition, properties, and implications. *Internet Math.* **2** 431–523. MR2241756
[34] LOVÁSZ, L. and SZEGEDY, B. (2006). Limits of dense graph sequences. *J. Combin. Theory Ser. B* **96** 933–957. MR2274085
[35] MAHADEV, N. V. R. and PELED, U. N. (1995). *Threshold Graphs and Related Topics*. *Annals of Discrete Mathematics* **56**. North-Holland, Amsterdam. MR1417258
[36] MCDIARMID, C. (1989). On the method of bounded differences. In *Surveys in Combinatorics, 1989 (Norwich, 1989)* (J. Siemons, ed.). *London Mathematical Society Lecture Note Series* **141** 148–188. Cambridge Univ. Press, Cambridge. MR1036755
[37] MCKAY, B. D. (2010). Subgraphs of dense random graphs with specified degrees. Preprint. Available at http://arxiv.org/abs/1002.3018.
[38] MCKAY, B. D. and WORMALD, N. C. (1990). Asymptotic enumeration by degree sequence of graphs of high degree. *European J. Combin.* **11** 565–580. MR1078713
[39] MOLLOY, M. and REED, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures Algorithms* **6** 161–179. MR1370952
[40] MOLLOY, M. and REED, B. (1998). The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.* **7** 295–305. MR1664335
[41] NEWMAN, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* **45** 167–256 (electronic). MR2010377
[42] NEWMAN, M. E. J., BARABASI, A.-L. and WATTS, D. J. (EDS.) (2006). *The Structure and Dynamics of Networks*. *Princeton Studies in Complexity*. Princeton Univ. Press, Princeton, NJ. MR2352222
[43] PARK, J. and NEWMAN, M. E. J. (2004). Statistical mechanics of networks. *Phys. Rev. E (3)* **70** 066117, 13. MR2133807
[44] PORTNOY, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *Ann. Statist.* **12** 1298–1309. MR0760690

[45] PORTNOY, S. (1985). Asymptotic behavior of M estimators of p regression parameters when p^2/n is large. II. Normal approximation. *Ann. Statist.* **13** 1403–1417. MR0811499

[46] PORTNOY, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16** 356–366. MR0924876

[47] PORTNOY, S. (1991). Correction: "Asymptotic behavior of M estimators of p regression parameters when p^2/n is large. II. Normal approximation." *Ann. Statist.* **19** 2282. MR1135181

[48] ROBINS, G., SNIJDERS, T., WANG, P., HANDCOCK, M. and PATTISON, P. (2007). Recent developments in exponential random graph (p^*) models for social networks. *Social Networks* **29** 192–215.

[49] RYSER, H. J. (1957). Combinatorial properties of matrices of zeros and ones. *Canad. J. Math.* **9** 371–377. MR0087622

[50] SIMONS, G. and YAO, Y.-C. (1999). Asymptotics when the number of parameters tends to infinity in the Bradley–Terry model for paired comparisons. *Ann. Statist.* **27** 1041–1060. MR1724040

[51] TSOURAKAKIS, C. (2008). Fast counting of triangles in large real networks: Algorithms and laws. In *Proc. of ICDM 2008* 608–617. IEEE Computer Society, Los Alamitos, CA.

[52] WAINWRIGHT, M. J. and JORDAN, M. I. (2008). Graphical models, exponential families and variational inference. *Foundations and Trends in Machine Learning* **1** 1–305.

[53] WILLINGER, W., ALDERSON, D. and DOYLE, J. C. (2009). Mathematics and the Internet: A source of enormous confusion and great potential. *Notices Amer. Math. Soc.* **56** 586–599. MR2509062

[54] WORMALD, N. C. (1999). Models of random regular graphs. In *Surveys in Combinatorics, 1999 (Canterbury). London Mathematical Society Lecture Note Series* **267** 239–298. Cambridge Univ. Press, Cambridge. MR1725006

S. CHATTERJEE
COURANT INSTITUTE OF MATHEMATICAL SCIENCES
NEW YORK UNIVERSITY
251 MERCER STREET
NEW YORK, NEW YORK 10012
USA
E-MAIL: sourav@cims.nyu.edu

P. DIACONIS
DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305
USA
E-MAIL: diaconis@math.stanford.edu

A. SLY
THEORY GROUP
MICROSOFT RESEARCH, ONE MICROSOFT WAY
REDMOND, WASHINGTON 98052
USA
E-MAIL: allansly@microsoft.com