Bernoulli **29**(1), 2023, 24–51
https://doi.org/10.3150/21-BEJ1448

Statistics of the two star ERGM

SUMIT MUKHERJEE[a] and YUANZHE XU[b]

Columbia University, New York, NY, USA. [a]*sm3949@columbia.edu,* [b]*yuanzhe.xu@columbia.edu*

In this paper, we explore the two star Exponential Random Graph Model, which is a two parameter exponential family on the space of simple labeled graphs. We introduce auxiliary variables to express the two star model as a mixture of the β model on networks. Using this representation, we study asymptotic distribution of the number of edges, and the sampling variance of the degrees. In particular, the limiting distribution for the number of edges has similar phase transition behavior to that of the magnetization in the Curie-Weiss Ising model of Statistical Physics. Using this, we show existence of consistent estimates for both parameters. Finally, we prove that the centered partial sum of degrees converges as a process to a Brownian bridge in all parameter domains, irrespective of the phase transition.

Keywords: ERGM; auxiliary variables; phase transition; consistent estimation; two-star

1. Introduction

Inference on graphs/networks is a topic of considerable recent interest in Statistics and Machine Learning. Both parametric and non-parametric models have been introduced to study graphs. In the parametric setting, perhaps the simplest model is the celebrated Erdős-Rényi model, where all the edges are independent, and there is only one parameter in the model. However, this model is too simplistic to be able to capture real life networks. Note that an Erdős-Rényi model can be expressed as an exponential family, with the number of edges as a sufficient statistic. As a first step towards modeling dependence between edges, it is natural to consider parametric models where there are more than one sufficient statistic. This motivation led to the introduction and study of exponential families on the space of graphs with finitely many sufficient statistics. Typical sufficient statistics of interest include higher order subgraph counts, such as number of stars, number of cycles, number of cliques, and so on. We will refer to exponential families on the space of graphs as Exponential Random Graph Models. For the sake of convenience the abbreviation ERGM will henceforth be used to refer to Exponential Random Graph Models. ERGMs first appeared in Social Sciences (c.f. [2,13,18,27,34,35] and references there-in), and since then have received a lot of attention in Probability (c.f. [14,16,25] and references there-in), Statistics (c.f. [5,30,31] and references there-in) and Statistical Physics ([10,23,24] and references there-in). One of the main attractions behind studying ERGMs is that they can incorporate non-trivial dependence between the edges, as opposed to the Erdős-Rényi model, where the edges are assumed to be independent.

One of the main difficulties of analyzing ERGMs is that the normalizing constant is not available in closed form. Explicit computation of the normalizing constant is computationally prohibitive. One way out is to resort to MCMC, but mixing rates for ERGMs depend crucially on the parameter values as shown in [3, Theorem 5,6], and can take time which is exponential in the number of vertices to mix. In [5] the authors study ERGMs using a large deviation approach. In particular they show that ERGMs are "close" to mixtures of Erdős-Rényi random graphs (c.f. [5, Theorem 6.4] for a formal statement). Since an Erdős-Rényi random graph has a single parameter, this suggests that consistent estimation of both the parameters (as the size of the graph grows) is not possible in the two star ERGM. To investigate this issue, in this paper we study a particular two parameter ERGM, namely the two star model, which is perhaps the simplest ERGM outside an Erdős-Rényi model. The two star model was first studied in [23] using non rigorous methods. This ERGM has exactly two sufficient statistics, the number of edges, and the number of two stars. We demonstrate that in this model consistent estimation of both the parameters is indeed possible, though there is a loss in efficiency for estimating two parameters instead of one (see Corollary 1.9). In the course of our analysis, we derive asymptotic distribution for the number of edges $E(G)$, which shows interesting phase transitions (see Theorem 1.4). In particular, throughout "most" of the 2D parameter space, $E(G)$ has an asymptotic normal distribution. This regime is termed as Θ_1, and is often referred to as the uniqueness regime. On the other hand, along a one dimensional curve in the 2D parameter space, the distribution of $E(G)$ is bimodal, and is a mixture of two normal distributions. This regime is termed as Θ_2, and is often referred to as the non-uniqueness regime. Finally, there is a single parameter configuration, termed Θ_3 and often referred to as the critical point, which is sandwiched between Θ_1 and Θ_2. In this regime $E(G)$ has an asymptotic distribution which is unimodal but not Gaussian. Similar limiting distributions were observed while studying the magnetization in the Curie-Weiss of Statistical Physics ([12]), and more recently in Ising model on "dense" regular graphs ([9]). In some sense, these phase transitions can be viewed as an explanation for the phenomenon of degeneracy observed in ERGMs, observed in [17,20,29,32]. One of the most common features of degeneracy is that repeated samples from the model do not produce similar samples. This is particularly evident in the regime Θ_2, where the number of edges $E(G)$ in two independent samples can be very different with probability $\frac{1}{2}$. Even if the true parameter is in Θ_1 but is close to the boundary, one expects to see similar behavior.

We also derive asymptotic distribution for the sampling variance of the degrees (1.6), which has a limiting Gaussian behavior in all the three regimes $\Theta_1 \cup \Theta_2 \cup \Theta_3$. Thus this statistic does not see the effect of the phase transition to the same extent, as does $E(G)$. Using our techniques, we are also able to analyze general linear statistics of edges (and not just $E(G)$, the total number of edges). To demonstrate the flexibility of our tools, we show that the partial sums of the centered degrees converge to a Brownian bridge (Theorem 1.11). Again this limiting behavior is universal, and does not depend on the three regimes. The understanding from these results is that statistics which depend on contrasts (linear statistics whose coefficients add up to 0) do not feel the effect of the phase transitions/different regimes, whereas non contrast statistics (such as the total number of edges) do. During the course of our proofs, we develop a natural algorithm for simulating from the two star ERGM, which is based on auxiliary variables. As indicated above, simulating from general ERGMs is a hard problem, so the presence of an auxiliary variable algorithm does help us to simulate "efficiently" from the two star model.

A natural question is to what extent does one expect the results of this paper to carry over to more general ERGMs. The main technique used in this paper is to utilize the quadratic nature of the two star model and visualize it as an Ising model on the line graph of the complete graph, and then introduce auxiliary variables to convert the discrete problem to a continuous one. Similar techniques were used to analyze the Curie-Weiss model ([8,19]). In fact, the techniques of this paper can be used to analyze a more general version of the two star model, where each node has a parameter of its own, and the quadratic term is the same as our two star model. For more general ERGMs beyond quadratic interactions, it is unclear whether one can construct auxiliary variables in a manner similar to this paper. Nevertheless, we expect some of the high level features of the two star ERGM to carry over to more general ERGMs. For example, it seems that estimation of multiple parameters consistently should be possible, but with a loss of efficiency. Also, for "most" of the parameter regime barring special configurations, one expects that the number of edges will have an asymptotic normal distribution. Some very preliminary results in this direction can be found in [14,28].

1.1. Formal set up

By a two star, we mean a path of length 2, which has 3 vertices and 2 edges. We begin by introducing the two star ERGM.

Definition 1.1. For a positive integer n, let $\mathcal{G}_n$ denote the space of all simple graphs with vertices labeled $[n] := \{1, 2, \ldots, n\}$. Since a simple graph is uniquely identified by its adjacency matrix, without loss of generality we can take $\mathcal{G}_n$ to also denote the set of all symmetric $n \times n$ matrices, with 0 on the diagonal elements and $\{0,1\}$ on the off-diagonal elements. By slightly abusing the notation, we use G to denote both a graph and its $n \times n$ adjacency matrix $(G_{ij})_{1 \le i,j \le n}$, defined by

$$G_{ij} = \begin{cases} 1 & \text{If an edge is present between vertices } i \text{ and } j \text{ in } G \\ 0 & \text{Otherwise} \end{cases}$$

Set $G_{ii} = 0$ by convention. Let $E(G) := \sum_{i<j} G_{ij}$ denote the number of edges in G, and let

$$T(G) := \sum_{i=1}^{n} \sum_{j<k} G_{ij} G_{ik}$$

denote the number of two stars in G. A simple calculation shows that the number of two stars can be written as

$$T(G) = \sum_{i=1}^{n} \binom{d_i(G)}{2},$$

where $(d_1(G), \ldots, d_n(G))$ is the labeled degree sequence of the graph G, defined by $d_i(G) := \sum_{j=1}^{n} G_{ij}$. Indeed, this is because given any vertex i of degree $d_i(G)$, there are $\binom{d_i(G)}{2}$ two stars with i as their central vertex.

Given parameters $\omega_1 > 0$ and $\omega_2 \in \mathbb{R}$, the two star ERGM is defined by the following probability mass function on $\mathcal{G}_n$:

$$\mathbb{P}_n(G = g) := \frac{1}{Z_n(\omega_1, \omega_2)} \exp\left\{\left(\omega_2 + \frac{\omega_1}{n-1}\right) E(g) + \frac{\omega_1}{n-1} T(g)\right\}, \tag{1}$$

where $Z_n(\omega_1, \omega_2)$ is the normalizing constant.

Note that if $\omega_1 = 0$, the model reduces to an Erdős-Rényi model with parameter $\frac{e^{\omega_2}}{1+e^{\omega_2}}$. The regime $\omega_1 > 0$ corresponds to the so called "Ferromagnetic regime" of Statistical Physics, which encourages more two stars in the sampled graph than an Erdős-Rényi graph with parameter $\frac{e^{\omega_2}}{1+e^{\omega_2}}$. For the analysis of the two star ERGM, we transform the edge variables from $\{0,1\}$ to $\{-1,1\}$, which converts the two star model into an Ising model (c.f. section 1.3). For the sake of mathematical convenience, throughout the paper we work in a slightly different parametrization, given by

$$\theta := \frac{\omega_1}{4} > 0, \quad \beta := \frac{\omega_1 + \omega_2}{2} \in \mathbb{R}. \tag{2}$$

This makes the subsequent analysis and presentation of results cleaner, and so throughout the rest of the paper we will work with the above re-parametrization. As frequently happens for such models, the two star model undergoes a phase transition, and its behavior is qualitatively different in different parts of the parameter regime. The following lemma introduces the different parameter domains arising out of our analysis. The proof of this lemma follows from straightforward calculus, and is deferred to the supplementary material [21] appendix A.

Lemma 1.2. *Setting*

$$q(x) = \theta x^2 - \log\cosh(2\theta x + \beta), \tag{3}$$

the following hold:

(a) *If either $\theta > 0, \beta \neq 0$ or $\theta \in (0, 1/2), \beta = 0$, the function $q(.)$ has a unique global minimizer at t, where t is the unique root of the equation $x = \tanh(2\theta x + \beta)$ which has the same sign as that of β. Further we have $q''(t) = 2\theta[1 - 2\theta(1 - t^2)] > 0$.*
(b) *If $\theta > 1/2, \beta = 0$, the function $q(.)$ has two global minimizers at $\pm t$, where t is the unique positive root of the equation $x = \tanh(2\theta x)$. Further, we have $q''(\pm t) = 2\theta[1 - 2\theta(1 - t^2)] > 0$.*
(c) *If $\theta = 1/2, \beta = 0$, the function $q(.)$ has a unique global minimizer at $t = 0$, and $q''(0) = 0$.*

Definition 1.3. Let t be as defined in Lemma 1.2, and note that t depends on (θ, β), which we suppress for ease of notation. Also let

$$\Theta_{11} := \{\theta \in (0, 1/2), \beta = 0\}, \quad \Theta_{12} := \{\theta > 0, \beta \neq 0\}$$
$$\Theta_2 := \{\theta > 1/2, \beta = 0\}, \quad \Theta_3 := \{\theta = 1/2, \beta = 0\},$$

and set $\Theta_1 := \Theta_{11} \cup \Theta_{12}$. We will refer to the three regimes $\Theta_1, \Theta_2, \Theta_3$ as uniqueness regime, non uniqueness regime, and critical regime respectively, the reason for this nomenclature follows from Lemma 1.2.

1.2. Main results

Our first main result now gives the asymptotic distribution for the number of edges in all the three domains $\{\Theta_1, \Theta_2, \Theta_3\}$.

Theorem 1.4. *Suppose G is a random graph from the two star model in (1).*

(a) *If $(\theta, \beta) \in \Theta_1$, we have*

$$n\Big(\frac{2E(G)}{n^2} - p\Big) \xrightarrow{D} N\Big(-\mu, \sigma^2\Big), \tag{4}$$

where $p = \frac{1+t}{2}$, $\mu := \frac{\theta t(1-t^2)}{[1-\theta(1-t^2)][1-2\theta(1-t^2)]}$, and $\sigma^2 := \frac{1-t^2}{2-4\theta(1-t^2)}$.
(b) *If $(\theta, \beta) \in \Theta_2$, then we have*

$$\mathbb{P}_n\Big(E(G) > \frac{n^2}{4}\Big) = \mathbb{P}_n\Big(E(G) < \frac{n^2}{4}\Big) \to \frac{1}{2}.$$

Further,

$$\begin{aligned} &\left(n\left(\frac{2E(G)}{n^2} - p\right)\Big|E(G) > \frac{n^2}{4}\right) \xrightarrow{D} N(-\mu,\sigma^2),\\ &\left(n\left(\frac{2E(G)}{n^2} - (1-p)\right)\Big|E(G) < \frac{n^2}{4}\right) \xrightarrow{D} N(\mu,\sigma^2). \end{aligned} \tag{5}$$

in which p,μ,σ^2 have the same formulas as above.

(c) *If $(\theta,\beta) \in \Theta_3$, then we have*

$$2\sqrt{n}\Big(\frac{2E(G)}{n^2} - \frac{1}{2}\Big) \xrightarrow{D} \zeta, \tag{6}$$

where ζ is a random variable on $\mathbb{R}$ with density proportional to $e^{-\zeta^2/2-\zeta^4/24}$.

Remark 1.5. Theorem 1.4 demonstrates that out of the parameter regime $\{\theta > 0, \beta \in \mathbb{R}\}$, throughout Θ_1 (which is almost the whole parameter space, in sense of Lebesgue measure), $E(G)$ has a Gaussian distribution. Only in a one dimensional set Θ_2, $E(G)$ has a bimodal distribution and is asymptotically a mixture of two Gaussian distributions. Finally, at single point Θ_3, $E(G)$ has a non Gaussian limiting distribution. Prior to our work, limiting distribution for $E(G)$ was not understood for any ERGM. See however the recent works of [14,28], which makes some progress in this direction in high temperature regime (i.e. θ small). Also related to our work is the asymptotics of the magnetization/sum of spins in Ising models on dense regular graphs. As explained below in section 1.3, the two star ERGM can be thought of as an Ising model on a d_N regular graph with $N := \binom{n}{2}$ vertices and degree $d_N = 2(n-2)$ (so that $d_N \propto \sqrt{N}$). Further, the number of edges is a linear function of the magnetization. Very recently in [9] it was shown that the magnetization/sum of spins in a Ising model on a regular graphs with degree $d_N \gg \sqrt{N}$ is universal, and is the same as obtained for the Curie-Weiss model ($d_N = N-1$) in [12]. As our results demonstrate, universality breaks at the threshold $d_N \propto \sqrt{N}$, as the distribution above does not match that of the Curie-Weiss model in the domains $\Theta_{12} \cup \Theta_2 \cup \Theta_3$. Only in the domain Θ_{11} the limiting distribution of the magnetization in the two star model matches that of the Curie-Weiss model. The techniques employed in this draft are very different from the techniques of both [14] and [9].

Note that the parameters (t,μ,σ^2) in Theorem 1.4 are implicit function of (θ,β). To help better understand how these parameters depend on (θ,β), in figures 1, 2 and 3 we plot (t,μ,σ^2) versus θ/β respectively, keeping the other one fixed at various values, to capture the effect of the different domains.

We now briefly explain the main features of these three figures 1, 2, and 3. In all the three figures, if $\beta = \pm 1$ is fixed (sub-plots I and II), then we stay inside the uniqueness regime for all values of θ. Consequently, in all the three figures we have a smooth function everywhere. Similar behavior is observed in all the three figures if $\theta = .25$ is fixed (sub-plot IV), as again we are always in the uniqueness regime for all values of β. If θ is fixed at .75 while studying β versus t (figure 1 sub-plot V), then as β approaches 0 from the two sides, the parameter t approaches two different roots of the equation $x = \tanh(2\theta x)$. Consequently the function $\beta \mapsto t$ is discontinuous at 0. A similar argument applies to the corresponding plot of μ as well (figure 2 sub-plot V), which shows a similar discontinuity. The corresponding plot of $\beta \mapsto \sigma^2$ (figure 3 sub-plot V) does not show any discontinuity, as σ^2 is a function of t^2, and $\beta \mapsto t^2$ is a continuous map (as opposed to $\beta \mapsto t$ which is discontinuous). However $\beta \mapsto \sigma^2$ is not differentiable, as $\beta \mapsto t^2$ is not differentiable. Finally, if we fix $\beta = 0$ or $\theta = \frac{1}{2}$ (sub-plots III

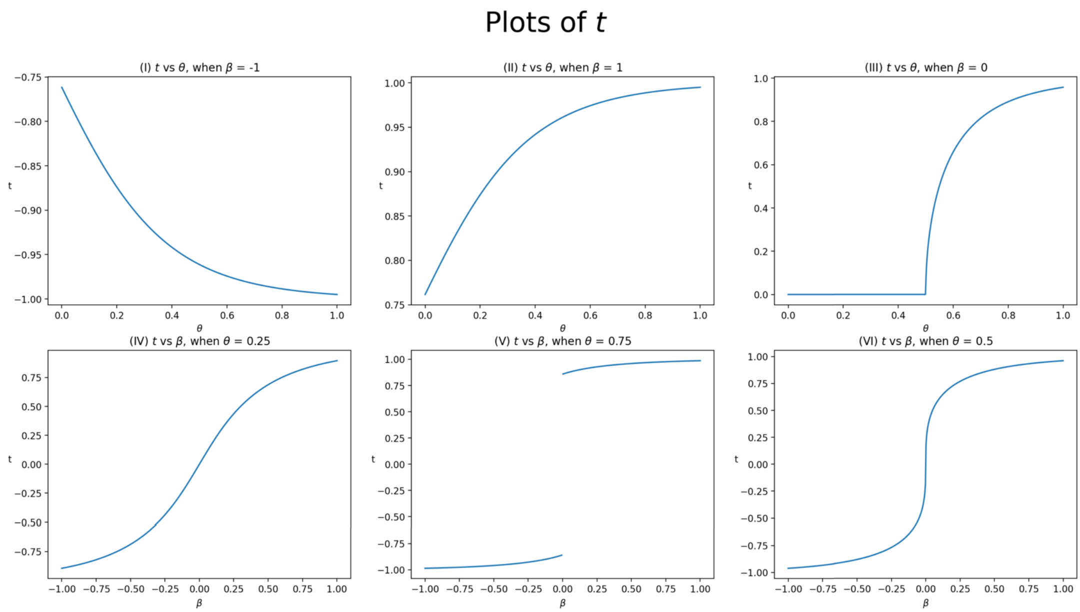

Figure 1. The above figure shows a plot of t versus θ, when β is fixed at $-1, +1, 0$ in sub-plots I, II, III respectively, and a plot of t versus β when θ is fixed at .25, .5, .75 in sub-plots IV, V, VI respectively. Plots I, II and IV are in the uniqueness regime throughout. Plot V corresponds to the non-uniqueness regime as $\beta \to 0$. Plots III and IV demonstrate the effect of the critical point $(\theta, \beta) = (.5, 0)$.

and VI in each figure), we pass through the critical point $\left(\frac{1}{2}, 0\right)$. In figure 1 sub-plots III and VI, the function t is continuous but not differentiable at the critical point. In figure 2 sub-plots III and VI, there is a discontinuity at the critical point, as μ asymptotes to $+\infty$ in one/both directions in sub-plots III/VI respectively. Finally, in figure 3 sub-plots III and VI, we see that the function σ^2 is continuous everywhere, but blows up as the parameters approach the critical point. This illustrates the fact that the asymptotic variance at criticality is of a larger magnitude.

Our second result studies the fluctuations of the empirical variance of the degrees.

Theorem 1.6. *Suppose G is a random graph from the two star model in (1). For all $\theta > 0, \beta \in \mathbb{R}$ we have*

$$\sqrt{n}\left[\frac{4}{n^2}\sum_{i=1}^{n}(d_i(G) - \bar{d}(G))^2 - \tau\right] \xrightarrow{D} N(0, 2\tau^2) \tag{7}$$

where $\bar{d}(G) := \frac{\sum_{i=1}^n d_i(G)}{n}$ and $\tau := \frac{1-t^2}{1-\theta(1-t^2)}$.

Remark 1.7. Note that unlike the total number of edges $E(G)$, the asymptotic distribution of $\sum_{i=1}^n (d_i(G) - \bar{d}(G))^2$ is always Gaussian, and does not change with different regimes $\Theta_1, \Theta_2, \Theta_3$. The only effect of the phase transition is through the parameter τ, which is continuous but not differentiable at $\theta = 1/2$, when $\beta = 0$ is kept fixed. Consequently, if $\beta = 0$, the statistic $\frac{4}{n^2}\sum_{i=1}^n (d_i(G) - \bar{d}(G))^2$ converge in probability to τ, which is not a smooth function of θ. This phenomenon was first observed in [23, Fig 2].

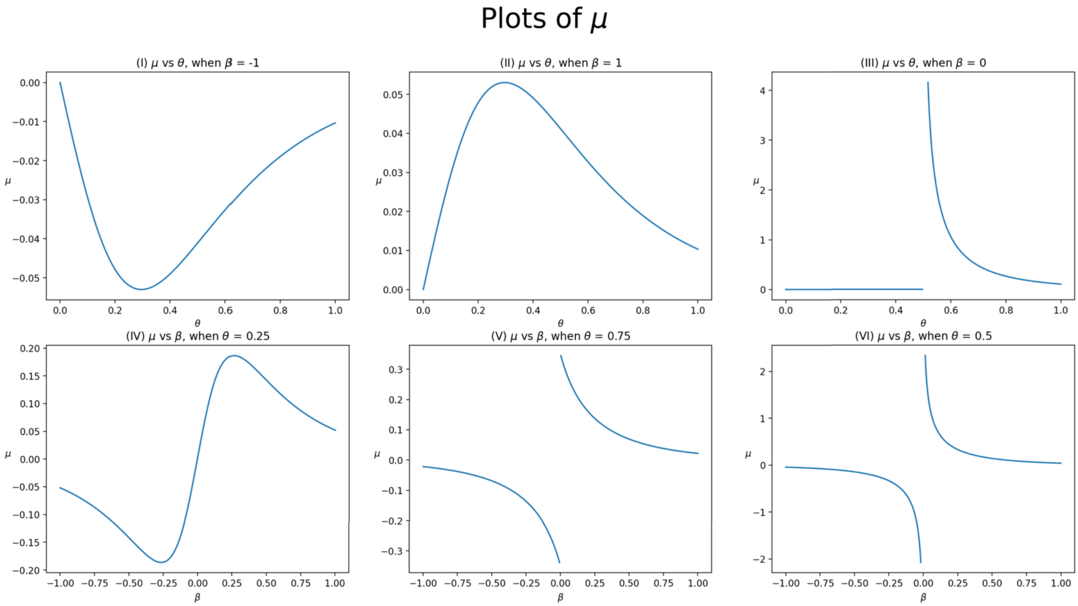

Figure 2. The above figure shows a plot of μ versus θ, when β is fixed at $-1,+1,0$ in sub-plots I, II, III respectively, and a plot of μ versus β when θ is fixed at .25, .5, .75 in sub-plots IV, V, VI respectively. Plots I, II and IV are in the uniqueness regime throughout. Plot V corresponds to the non-uniqueness regime as $\beta \to 0$. Plots III and IV demonstrate the effect of the critical point $(\theta,\beta) = (.5,0)$.

As an application of the two theorems above, we provide consistent estimators of the parameters (θ,β). Stating the estimators require the following definition.

Definition 1.8. Let

$$\hat{t} := \frac{2E(Y)}{n^2}, \quad \hat{\tau} := \frac{1}{n^2}\sum_{i=1}^{n}\left(k_i(Y) - \bar{k}(Y)\right)^2.$$

Corollary 1.9. *Suppose G is a random graph from the two star model in (1), with $(\theta,\beta) \in \Theta := \{(\theta,\beta) : \theta > 0, \beta \in \mathbb{R}\}$.*

(a) *Suppose(θ,β) are both unknown. Let*

$$\hat{\theta} := \frac{1}{1-t^2} - \frac{1}{\hat{\tau}}, \quad \hat{\beta} := \operatorname{arctanh}(\hat{t}) - 2\hat{\theta}\hat{t}.$$

Then $(\hat{\theta},\hat{\beta})$ is jointly $\sqrt{n}$-consistent estimator for (θ,β), i.e. $\sqrt{n}(\hat{\theta}-\theta,\hat{\beta}-\beta) = O_P(1)$.

(b) *Suppose θ is known and β is unknown. Let $\tilde{\beta} :=$ arctanh$(\hat{t}) - 2\theta\hat{t}$. Then $\tilde{\beta}$ is an n consistent estimator for β, i.e. $n(\tilde{\beta}-\beta) = O_P(1)$.*

Remark 1.10. The above corollary shows that there is a loss of efficiency when we are trying to estimate both parameters, as opposed to estimating just one parameter. Joint estimation of parameters in general Ising models has been studied in [15], where the authors give a general upper bound on the rate of consistency of pseudo-likelihood (see [15, Theorem 1.2]). Using their result for a d_N regular

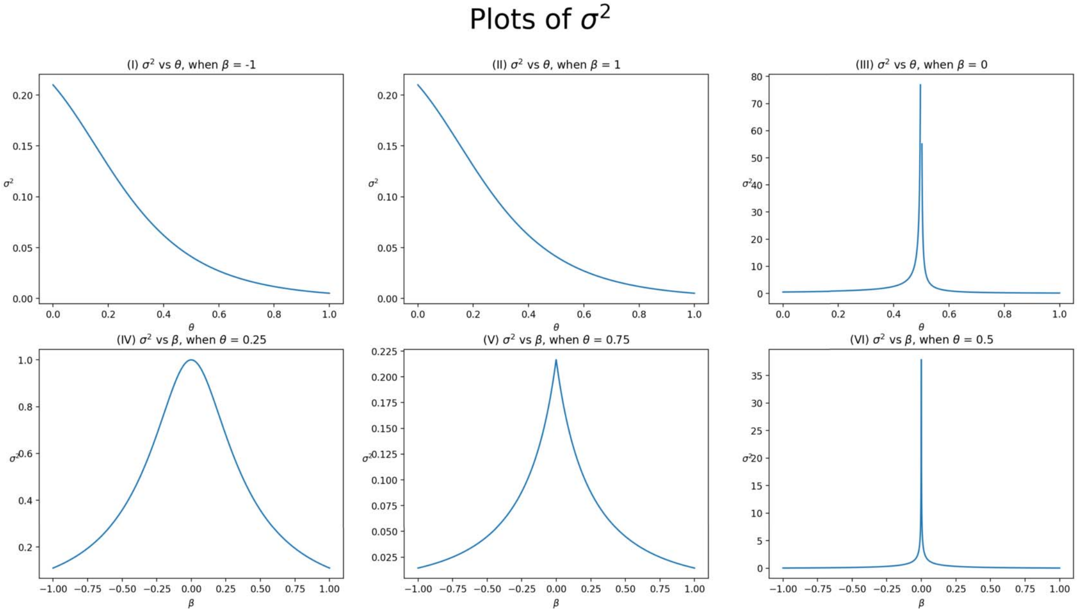

Figure 3. The above figure shows a plot of σ^2 versus θ, when β is fixed at $-1, +1, 0$ in sub-plots I, II, III respectively, and a plot of σ^2 versus β when θ is fixed at .25, .5, .75 in sub-plots IV, V, VI respectively. Plots I, II and IV are in the uniqueness regime throughout. Plot V corresponds to the non-uniqueness regime as $\beta \to 0$. Plots III and IV demonstrate the effect of the critical point $(\theta,\beta) = (.5, 0)$.

graph on N vertices, one concludes that (an upper bound to) the rate of estimation error the pseudo-likelihood estimator is $\frac{d_N}{\sqrt{N}}$. Thus one can consistently estimate (θ,β) on an Ising model on a sequence of d_N regular graph, if $d_N \ll \sqrt{N}$. However, in this case we have $d_N \propto \sqrt{N}$, and so consistency of the bivariate pseudo-likelihood estimator does not follow from [15]. It is unclear whether the bivariate pseudo-likelihood estimator is consistent in this case. On the other hand, the above corollary gives explicit consistent estimator for both parameters, with rates of consistency.

Our final result shows that the partial sums of the (centered) degree distribution converges as a process in $C[0,1]$ to a Brownian bridge under proper scaling, in all the three parameter domains.

Theorem 1.11. *Suppose G is a random graph from the two star model in* (1). *Let* $W_n(.) \in C[0,1]$ *be the linear interpolation of the points* $\{(\frac{i}{n}, \frac{S_i(\mathbf{d})}{n-1}), i \in [n]\}$, *where* $S_i(\mathbf{d}) := \sum_{j=1}^{i} (d_j(G) - \bar{d}(G))$. *Then*

$$W_n(.) \xrightarrow{D} \sqrt{\tau}\{W(.)\},$$

where $W(.) \in C[0,1]$ *is a Brownian bridge.*

This demonstrates that irrespective of the phase transitions, there is significant Gaussian behavior in the model, which is captured in terms of contrasts. Similar Gaussian fluctuations were obtained in [22] for the Curie-Weiss model at criticality. Note that this behavior is universal across all regimes, similar to Theorem 1.6. This is because, as will be clear from the proofs, the distribution of contrasts (linear functions of edges whose coefficients add up to 0) are universal across regimes, which govern the finite dimensional distributions of Theorem 1.11.

1.3. Auxiliary variables

The main technique for proving the results of this paper is a representation of the two star model as a mixture of β models by introducing auxiliary variables, introduced below. We note that introducing auxiliary variables have been proved to be successful in rigorously analyzing the Curie-Weiss model ([8,19]), and have also been used in [23] to study (non-rigorously) the two star model. Before introducing the auxiliary variable, we first transform the edge variables to $\{-1,1\}$ instead of $\{0,1\}$, and show that the transformed variables is a sample from an Ising model on an appropriate graph.

Transform the edge variables from $\{0,1\}$ to $\{-1,1\}$ by setting $Y_{ij} := 2G_{ij} - 1$ for $i \neq j$, and set $Y_{ii} := 0$ as convention. Via this transformation, the Hamiltonian for the matrix $Y := (Y_{ij})_{1\le i,j\le n}$ (up to additive constants) is given by

$$\frac{\omega_1}{4(n-1)}T(Y) + \frac{\omega_1+\omega_2}{2}E(Y) = \frac{\theta}{n-1}T(Y) + \beta E(Y),$$

in which $\theta = \frac{1}{4}\omega_1$, and $\beta = \frac{1}{2}(\omega_1+\omega_2) \in \mathbb{R}$ as in Lemma 1.2, and

$$T(Y) := \sum_{i=1}^{n}\sum_{j<k} Y_{ij}Y_{ik}, \quad E(Y) = \sum_{i<j} Y_{ij}.$$

Thus the model $\mathbb{P}_n$ defined in (1) is an Ising model in the transformed variable Y on the graph $\widetilde{G}_n$ which is the line graph of the complete graph K_n. More precisely, $\widetilde{G}_n$ has $\mathcal{E} := \{(i,j)|1 \le i < j \le n\}$ as its vertex set, and two distinct vertices $e = (i,j)$ and $f = (k,l)$ are connected iff $\{i,j\}\bigcap\{k,l\} \neq \varnothing$, i.e. $i = k$ or $i = l$ or $j = k$ or $j = l$. Thus $\widetilde{G}_n$ is a regular graph on $\binom{n}{2}$ vertices, with degree $2(n-2)$. Setting

$$k_i(Y) := \sum_{j=1}^{n} Y_{ij} = 2d_i(G) - (n-1), \tag{8}$$

the p.m.f. of Y can be written as

$$\mathbb{P}_n(Y=y) = \frac{1}{\widetilde{Z}_n(\theta,\beta)}\exp\left\{\frac{\theta}{2(n-1)}\sum_{i=1}^{n}k_i(y)^2 + \frac{\beta}{2}\sum_{i=1}^{n}k_i(y)\right\} \tag{9}$$

Let $\phi = (\phi_1,\dots,\phi_n)$ be a random vector in $\mathbb{R}^n$ defined by

$$\phi_i = \frac{k_i(Y)}{n-1} + \frac{W_i}{\sqrt{(n-1)\theta}}, \tag{10}$$

where $(W_1,\dots,W_n) \overset{i.i.d.}{\sim} N(0,1)$ are independent of the Y. The following proposition computes the distribution of $(Y|\phi)$, and the marginal density of ϕ. The proof of this Proposition is deferred to the supplementary material [21] appendix A.

Proposition 1.12. *Suppose Y is an observation from the p.m.f. in* (1).

(a) *Given ϕ, the random variables $\{Y_{ij}\}_{1\le i<j\le n}$ are mutually independent, with*

$$\mathbb{P}_n(Y_{ij}=1|\phi)=\frac{e^{\theta(\phi_i+\phi_j)+\beta}}{e^{\theta(\phi_i+\phi_j)+\beta}+e^{-\theta(\phi_i+\phi_j)-\beta}} \tag{11}$$

(b) *The marginal density of ϕ has a density on $\mathbb{R}^n$ which is proportional to $f_n(\phi)$, where $-\log f_n(\phi):=\sum_{i<j}p(\phi_i,\phi_j)$ with*

$$p(x,y)=\frac{\theta}{2}\Big(x^2+y^2\Big)-\log\cosh\Big[\theta(x+y)+\beta\Big]=\frac{\theta}{4}(x-y)^2+q\Big(\frac{x+y}{2}\Big), \tag{12}$$

where $q(x)=\theta x^2-\log\cosh(2\theta x+\beta)$ as in Lemma 1.2.

Remark 1.13. MCMC using auxiliary random variables is a common technique in simulations ([1,11, 33]). Using Proposition 1.12, it follows that the conditional distribution of the graph G given the vector ϕ is the β-model, which has received considerable attention in Statistics [4,6,7,26] and references therein). Thus the two star model (1) can be expressed as a mixture of β-models with random weights. Since both the conditional distributions $(Y|\phi)$ and $(\phi|Y)$ are easy to simulate, one can use a Gibbs sampler to simulate from the two star model, by iteratively simulating from the conditional distributions till the Markov Chain converges.

1.4. Simulation results

In this section we validate Theorem 1.4 and Theorem 1.11 using numerical simulations. For simulating from the two star ERGM we use the Gibbs sampling algorithm of Proposition 1.12. For verifying Theorem 1.4, we work with $n=500$ vertices on the two star ERGM with parameters (θ,β) equal to $(1/4,0)$, $(1/2,0)$, and $(3/4,0)$, which belong to the uniqueness regime, the critical point, and the non-uniqueness regime respectively. For each of these three parameter configurations, we simulate 5000 independent samples from the two star ERGM, by running the Gibbs sampling algorithm with a burn in period of 1000 for each sample. For each sample, we observe the centered and scaled sum of degrees

$$\frac{1}{n}\sum_{i=1}^n k_i(Y)=\frac{2}{n}\left[\sum_{i=1}^n d_i(G)-\frac{n(n-1)}{2}\right]=\frac{4}{n}\left[E(G)-\frac{n(n-1)}{4}\right].$$

The QQ plot of these values for the three regimes are given in figure 4.

As is seen in figure 4, in the uniqueness regime (first picture), the limiting distribution is clearly Gaussian, as there is a strong agreement with normal quantiles. At the critical point, the limiting distribution is no longer Gaussian, as is shown by deviation from the normal quantiles. In the non uniqueness regime, the data is strongly bimodal, and hence cannot be globally Gaussian. This is exactly the behavior predicted by Theorem 1.4. Theorem 1.4 suggests that if we zoom into each of the two modes, we will again see Gaussian fluctuations. To confirm this, we do a QQ plot for the positive and negative values separately. This is given below in figure 5.

For verifying Theorem 1.11, we obtained one sample from the two star ERGM on $n=1000$ vertices at criticality ($(\theta,\beta)=(1/2,0)$), after running the chain for 1000 iterations. Having obtained the graph G, we computed the partial sums

$$\frac{1}{n-1}\sum_{j=1}^i\Big(k_j(Y)-\bar{k}(Y)\Big)=\frac{2}{n-1}\sum_{j=1}^i\Big(d_j(G)-\bar{d}(G)\Big),$$

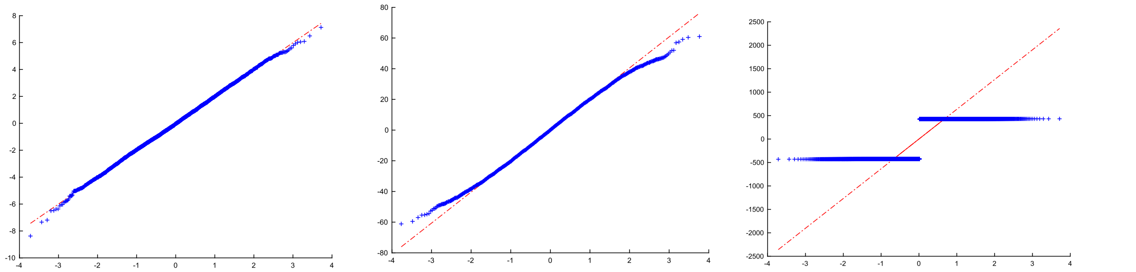

Figure 4. The QQ plot for the centered and scaled sum of degrees is given for 5000 independent samples from the two star ERGM on $n = 500$ vertices, for the three parameter configurations $(\theta, \beta) = (1/4, 0)$, $(1/2, 0)$, and $(3/4, 0)$ respectively. In the uniqueness regime (first picture), the limiting distribution is Gaussian. At the critical point, the limiting distribution is no longer Gaussian. In the non uniqueness regime, the data is strongly bimodal.

and plotted the partial sums versus i for $1 \le i \le n$ in figure 6.

As predicted, the plot looks like a Brownian curve starting and ending at 0. Similar pictures were obtained in all parameter regimes.

The rest of the paper is as follows: Sections 2 proves Theorem 1.4, Theorem 1.6, Corollary 1.9, and Theorem 1.11. The lemmas necessary for proving the main results are proved in section 3 for the uniqueness and non-uniqueness domains (i.e. $(\theta, \beta) \in \Theta_1 \cup \Theta_2$), and in section 4 for the critical domain (i.e. $(\theta, \beta) \in \Theta_3$). The supplementary material [21] collects the proofs of supporting lemmas and propositions.

2. Proof of main results (Theorems 1.4, 1.6, 1.11 and Corollary 1.9)

For proving our main results we need the following lemmas, the proof of which is deferred to sections 3 and 4 for $(\theta, \beta) \in \Theta_1 \cup \Theta_2$ and $(\theta, \beta) \in \Theta_3$ respectively.

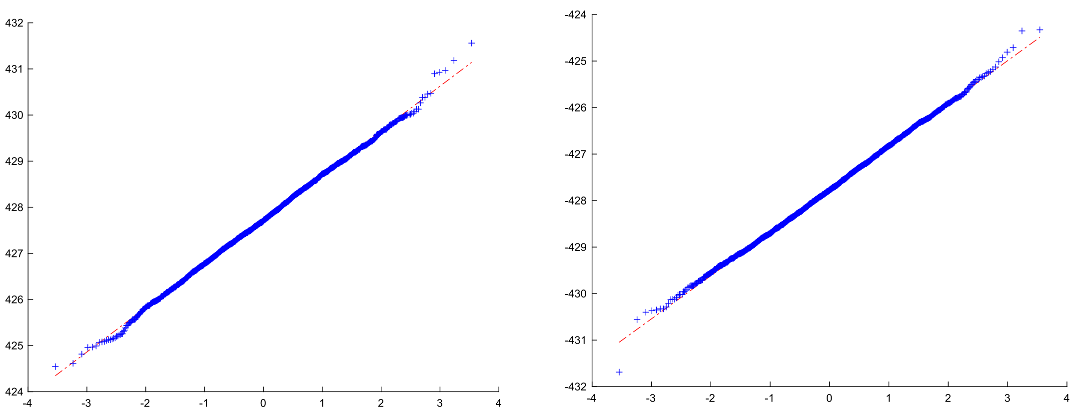

Figure 5. Starting from 5000 independent samples from the two star ERGM on $n = 500$ vertices with parameter $(\theta, \beta) = (3/4, 0)$, the QQ plot for the positive and negative values are given separately. Both the individual plots show agreement with Gaussian quantiles, which shows conditional Gaussian behavior near each of the two models.

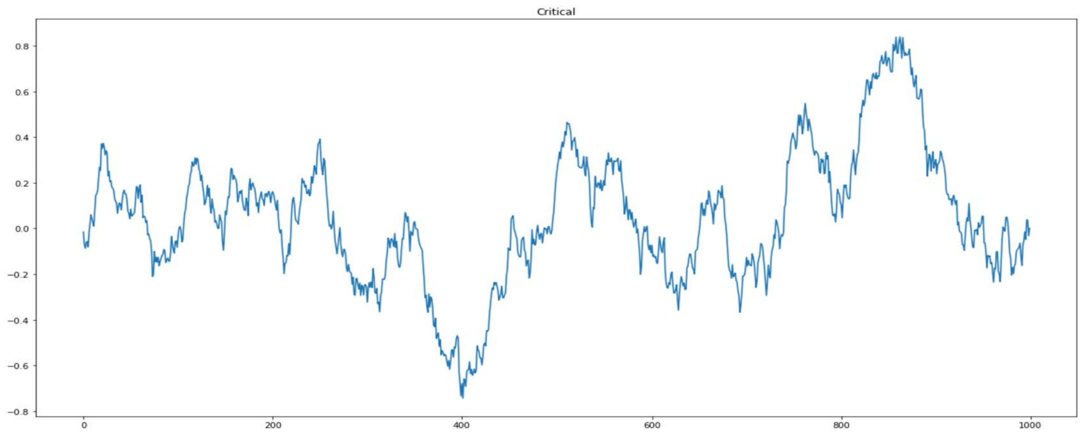

Figure 6. From one sample from the two star ERGM on $n = 1000$ vertices with parameter $(\theta, \beta) = (1/2, 0)$, the centered and partial sums of the degrees upto vertex i is plotted against i, for $1 \le i \le 1000$. The figure roughly resembles a Brownian curve starting and ending at the origin.

Lemma 2.1.

(a) *For* $(\theta, \beta) \in \Theta_1$, *we have*

$$n(\bar{\phi} - t) \xrightarrow{D} N\left(-\frac{2\theta t(1-t^2)}{[1-\theta(1-t^2)][1-2\theta(1-t^2)]}, \frac{1}{\theta - 2\theta^2(1-t^2)}\right).$$

(b) *For* $(\theta, \beta) \in \Theta_2$, *we have*

$$\mathbb{P}_n(\bar{\phi} > 0) = \mathbb{P}_n(\bar{\phi} < 0) = \frac{1}{2},$$

and further

$$\left[n(\bar{\phi} - t)\Big|\bar{\phi} > 0\right] \xrightarrow{D} N\left(-\frac{2\theta t(1-t^2)}{[1-\theta(1-t^2)][1-2\theta(1-t^2)]}, \frac{1}{\theta - 2\theta^2(1-t^2)}\right),$$

$$\left[n(\bar{\phi} + t)\Big|\bar{\phi} < 0\right] \xrightarrow{D} N\left(\frac{2\theta t(1-t^2)}{[1-\theta(1-t^2)][1-2\theta(1-t^2)]}, \frac{1}{\theta - 2\theta^2(1-t^2)}\right).$$

(c) *For* $(\theta, \beta) \in \Theta_3$, *we have* $\sqrt{n}\bar{\phi} \xrightarrow{D} \zeta$, *where* ζ *is a random variable on* $\mathbb{R}$ *with density proportional to* $e^{-\frac{\zeta^2}{2} - \frac{\zeta^4}{24}}$.

Lemma 2.2. *Setting*

$$a_1 := \theta - \theta^2(1-t^2), \tag{13}$$

for all $(\theta, \beta) \in \Theta$ *we have the following conclusions:*

(a) $\sqrt{n}\left[\sum_{i=1}^{n}(\phi_i - \bar{\phi})^2 - a_1^{-1}\right] \xrightarrow{D} N(0, 2a_1^{-2})$.

(b) *For any triangular array of real numbers* $(c_n(i), c_n(2), \ldots, c_n(n))$ *such that* $\sum_{i=1}^{n} c_n(i) = 0$ *and* $n^{-1}\sum_{i=1}^{n} c_n(i)^2 \to 1$, *we have* $\sum_{i=1}^{n} c_n(i)\phi_i \xrightarrow{D} N(0, a_1^{-1})$.

(c) *Setting* $S_i(\phi) := \sum_{j=1}^{i}(\phi_j - \bar{\phi})$, *for every* $\varepsilon > 0$ *we have*

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}_n\Big(\max_{i,j\in[n]:|i-j|\le n\delta}|S_i(\phi) - S_j(\phi)| > \varepsilon\Big) = 0.$$

2.1. Asymptotic notation

Throughout the rest of the paper we will use the following notations. Let $\{r_n\}_{n\ge 1}$ and $\{s_n\}_{n\ge 1}$ be two sequences of positive real numbers. Then we will say

- $r_n = o(s_n)$ if $\lim_{n\to\infty}\frac{r_n}{s_n} = 0$,
- $r_n = O(s_n)$ or $r_n \lesssim s_n$ if $\limsup_{n\to\infty}\frac{r_n}{s_n} < \infty$,
- $r_n = \Omega(s_n)$ if $\liminf_{n\to\infty}\frac{r_n}{s_n} > 0$.

If $\{R_n\}_{n\ge 1}$ and $\{S_n\}_{n\ge 1}$ are sequences of random variables, we will say

- $R_n = o_P(S_n)$ if $\frac{R_n}{S_n} \xrightarrow{P} 0$.
- $R_n = O_P(S_n)$, if $\frac{R_n}{S_n}$ is tight.

2.2. Proof of Theorem 1.4

(a) $(\theta, \beta) \in \Theta_1$.

To begin, using (10) we have

$$n(\bar{\phi} - t) = \frac{n}{n-1}\Big[\bar{k}(Y) - (n-1)t\Big] + \frac{n\bar{W}}{\sqrt{(n-1)\theta}}.$$

Using this along with part (a) of Lemma 2.1 and the observation $\frac{n\bar{W}}{\sqrt{(n-1)\theta}} \xrightarrow{D} N(0, \frac{1}{\theta})$ gives

$$\bar{k}(Y) - (n-1)t \xrightarrow{D} N\left(-\frac{2\theta t(1-t^2)}{[1-\theta(1-t^2)][1-2\theta(1-t^2)]}, \frac{2(1-t^2)}{1-2\theta(1-t^2)}\right)$$

Finally use (8) to note that $\bar{k}(Y) = 2\bar{d}(G) - (n-1)$, and so

$$\bar{d}(G) - (n-1)p \xrightarrow{D} N\left(-\frac{\theta t(1-t^2)}{[1-\theta(1-t^2)][1-2\theta(1-t^2)]}, \frac{1-t^2}{2[1-2\theta(1-t^2)]}\right),$$

which verifies Theorem 1.4 for $(\theta, \beta) \in \Theta_1$.

(b) The conclusion $\mathbb{P}_n(\bar{\phi} > 0) = \mathbb{P}_n(\bar{\phi} < 0) = \frac{1}{2}$ follows from symmetry. On the set $\bar{\phi} > 0$ ($\bar{\phi} < 0$) we have $\bar{\phi} \xrightarrow{P} t$ ($\bar{\phi} \xrightarrow{P} -t$) respectively, by invoking Lemma 2.1 part (b). On the set $\bar{\phi} > 0$, using (10) it follows that

$$\frac{1}{n}\bar{k}(Y) \xrightarrow{P} t \Rightarrow \frac{2E(G)}{n^2} \xrightarrow{P} \frac{1+t}{2} = p > \frac{1}{2}.$$

A similar argument gives that on the set $\bar{\phi} < 0$ we have

$$\frac{2E(G)}{n^2} \xrightarrow{P} \frac{1-t}{2} = 1 - p < \frac{1}{2}.$$

Thus $\mathbb{P}_n(\{\bar{\phi} > 0\}\Delta\{E(G) > \frac{n^2}{4}\}) \to 0$ (here Δ represents symmetric difference between the two sets), and so without loss of generality we can replace the conditioning set $E(G) > \frac{n^2}{4}$ by $\bar{\phi} > 0$. From then, using part (b) of Lemma 2.1 and mimicking the proof of part (a) above gives the desired conclusion. A similar proof works when we condition on the set $E(G) < \frac{n^2}{4}$.

(c) Again using (10) we have

$$\sqrt{n}\bar{\phi} = \frac{\sqrt{n}\bar{k}(Y)}{n-1} + \frac{\sqrt{n}\bar{W}}{\sqrt{(n-1)\theta}}.$$

Since $\bar{W} \xrightarrow{P} 0$, it follows from part (c) of Lemma 2.1 that $\frac{\bar{k}(Y)}{\sqrt{n}} \xrightarrow{D} \zeta$. The desired result then follows from on noting that

$$\frac{\bar{k}(Y)}{\sqrt{n}} = \sqrt{n}\Big[\frac{2\bar{d}(G) - (n-1)}{n}\Big] = 2\sqrt{n}\Big[\frac{\bar{d}(G)}{n} - \frac{1}{2}\Big] + O\Big(\frac{1}{\sqrt{n}}\Big).$$

2.3. Proof of Theorem 1.6

Using (10) we can write

$$\sum_{i=1}^{n}(\phi_i - \bar{\phi})^2 - a_1^{-1} = A_n + B_n + C_n, \tag{14}$$

where

$$A_n := \Big[\frac{1}{(n-1)\theta}\sum_{i=1}^{n}\big(W_i - \bar{W}\big)^2 - \frac{1}{\theta}\Big], \quad B_n := \frac{2}{\sqrt{(n-1)^3\theta}}\sum_{i=1}^{n}\big(k_i(Y) - \bar{k}(Y)\big)\big(W_i - \bar{W}\big),$$

$$C_n := \Big[\frac{1}{(n-1)^2}\sum_{i=1}^{n}\big(k_i(Y) - \bar{k}(Y)\big)^2 - \tau\Big].$$

Here we have used the fact that

$$a_1^{-1} = \tau + \theta^{-1}, \tag{15}$$

where a_1 is as in (13). We now claim that given the graph G, the random variables $\sqrt{n}A_n$ and $\sqrt{n}B_n$ are asymptotically independent, i.e. for any $s \in \mathbb{R}$ we have

$$\Big|\mathbb{E}(e^{is\sqrt{n}(A_n+B_n)}|G) - \mathbb{E}(e^{is\sqrt{n}A_n}|G)\mathbb{E}(e^{is\sqrt{n}B_n}|G)\Big| \xrightarrow{P} 0. \tag{16}$$

Given (16), we have

$$\mathbb{E}e^{is\sqrt{n}(A_n+B_n+C_n)} = \mathbb{E}[\mathbb{E}(e^{is\sqrt{n}A_n}|G)\mathbb{E}(e^{is\sqrt{n}B_n}|G)e^{is\sqrt{n}C_n}] + o_P(1). \tag{17}$$

Also, note that

$$A_n \xrightarrow{P} 0, \quad \mathsf{Var}(B_n|G) = \frac{4}{(n-1)^3\theta}\sum_{i=1}^n \big(k_i(Y) - \bar{k}(Y)\big)^2 \xrightarrow{P} 0,$$

where the second convergence uses the observation $\sum_{i=1}^n (k_i - \bar{k})^2 = O_P(n^2)$. This, along with (14) gives

$$\frac{1}{n^2}\sum_{i=1}^n \big(k_i(Y) - \bar{k}(Y)\big)^2 \xrightarrow{P} \tau.$$

Consequently, given G the random variable $\sqrt{n}B_n$ has a Normal distribution with mean 0, and variance D_n, where

$$D_n := \frac{4n}{(n-1)^3\theta}\sum_{i=1}^n \big(k_i(Y) - \bar{k}(Y)\big)^2 \xrightarrow{P} \frac{4\tau}{\theta} =: \sigma_2^2. \tag{18}$$

Finally, it is straightforward to check that

$$\sqrt{n}A_n \xrightarrow{D} N(0, \sigma_1^2), \text{ where } \sigma_1^2 := 2\theta^{-2} \tag{19}$$

Combining (17) along with (18) and (19) gives

$$\mathbb{E}e^{is\sqrt{n}(A_n+B_n+C_n)} = e^{-\frac{s^2}{2}(\sigma_1^2+\sigma_2^2)}\mathbb{E}[e^{is\sqrt{n}C_n}] + o_P(1).$$

Since $\sqrt{n}(A_n + B_n + C_n) \xrightarrow{D} N\big(0, 2a_1^{-2}\big)$ by Lemma 2.2 part (a), it follows hat $\sqrt{n}C_n \xrightarrow{D} N(0, \sigma_3^2)$, where

$$2a_1^{-2} = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 = 2\theta^{-2} + 4\tau\theta^{-1} + \sigma_3^2.$$

Using (15) it follows that $\sigma_3^2 = 2\tau^2$, as desired.

To complete the proof, it suffices to verify (16). To this effect, given the graph G construct an orthogonal matrix $\mathbf{O}_n$ whose first row is proportional to the constant vector $\mathbf{1}$, and second row is proportional to the vector $(k_1 - \bar{k}, \ldots, k_n - \bar{k})$. Then with $\mathbf{U} := \mathbf{O}_n\mathbf{W} \sim N(\mathbf{0}, \mathbf{I}_n)$ we have

$$\sum_{i=2}^n U_i^2 = \sum_{i=1}^n \big(W_i - \bar{W}\big)^2, \quad U_2 = \frac{1}{\sqrt{\sum_{i=1}^n \big(k_i - \bar{k}\big)^2}}\sum_{i=1}^n \big(k_i - \bar{k}\big)W_i,$$

and so

$$\begin{aligned}\sqrt{n}(A_n, B_n) =& \left\{\sqrt{n}\Big[\frac{1}{(n-1)\theta}\sum_{i=2}^n U_i^2 - \frac{1}{\theta}\Big], U_2\sqrt{D_n}\right\} \\ =& \left\{\sqrt{n}\Big[\frac{1}{(n-1)\theta}\sum_{i=3}^n U_i^2 - \frac{1}{\theta}\Big], U_2\sqrt{D_n}\right\} + o_P(1),\end{aligned}$$

and so (16) follows.

2.4. Proof of Corollary 1.9

(a) Using Theorem 1.4 and Theorem 1.6, it follows that for all $(\theta,\beta) \in \Theta$ we have $n(\hat{t}^2 - t^2) = O_P(1)$, and $\sqrt{n}(\hat{\tau} - \tau) = O_P(1)$. This gives

$$\begin{aligned} |\hat{\theta} - \theta| \leq & \left| \frac{1}{1-\hat{t}^2} - \frac{1}{1-t^2} \right| + \left| \frac{1}{\tau} - \frac{1}{\hat{\tau}} \right| \\ = & |\hat{t}^2 - t^2| \frac{1}{(1-t^2)(1-\hat{t}^2)} + \frac{\tau - \hat{\tau}}{\tau\hat{\tau}} = O_P(|\hat{t}^2 - t^2|) + O_P(|\hat{\tau} - \tau|) = O_P(n^{-1/2}). \end{aligned}$$

Similarly,

$$\begin{aligned} |\hat{\beta} - \beta| \leq & |\text{arctanh}(\hat{t}) - \text{arctanh}(t)| + 2\hat{\theta}|\hat{t} - t| + 2|t|(|\hat{\theta} - \theta|) \\ = & O_P(|\hat{t} - t|) + O_P(|\hat{\theta} - \theta|) = O_P(n^{-1/2}), \end{aligned}$$

as desired.

(b) With $\tilde{\beta}$ as defined, we have

$$|\tilde{\beta} - \beta| \leq |\text{arctanh}(\hat{t}) - \text{arctanh}(t)| + 2\theta|\hat{t} - t| = O_P(|\hat{t} - t|) = O_P(n^{-1}),$$

as desired.

2.5. Proof of Theorem 1.11

Proof. We first check the convergence of finite dimensional distributions. For the sake of simplicity we check it for 2 dimensional distributions. Fixing $0 < s_1 < s_2 < 1$, it suffices to show that for any $(r_1, r_2) \in \mathbb{R}^2$,

$$r_1 W_n(s_1) + r_2 W_n(s_2) \xrightarrow{D} N(0, \tau\psi), \text{ where } \psi := r_1^2 s_1(1-s_1) + r_2^2 s_2(1-s_2) + 2r_1 r_2 s_1(1-s_2).$$

With

$$b_n(i) := \frac{(r_1+r_2)}{\sqrt{n}} 1_{\{1 \leq i < ns_1\}} + \frac{r_2}{\sqrt{n}} 1_{\{ns_1 < i \leq ns_2\}} \text{ and } c_n(i) := b_n(i) - \bar{b}_n,$$

in which $\bar{b}_n := \frac{1}{n} \sum_{i=1}^{n} b_n(i)$. We have

$$\begin{aligned} r_1 W_n(s_1) + r_2 W_n(s_2) = & \frac{(r_1+r_2)}{n\sqrt{n}} \sum_{1 \leq i \leq ns_1} \left(k_i(Y) - \bar{k}(Y) \right) + \frac{r_2}{n\sqrt{n}} \sum_{ns_1 < i \leq ns_2} \left(k_i(Y) - \bar{k}(Y) \right) + O\left(\frac{1}{\sqrt{n}} \right) \\ = & \frac{1}{n} \sum_{i=1}^{n} b_n(i) \left(k_i(Y) - \bar{k}(Y) \right) + O\left(\frac{1}{\sqrt{n}} \right) = \frac{1}{n} \sum_{i=1}^{n} c_n(i) k_i + O\left(\frac{1}{\sqrt{n}} \right). \end{aligned} \tag{20}$$

Since $\sum_{i=1}^{n} c_n(i) = 0$ and

$$\sum_{i=1}^{n} c_n(i)^2 = \sum_{i=1}^{n} b_n(i)^2 - n\bar{b}_n^2 \to (r_1+r_2)^2 s_1 + r_2^2(s_2 - s_1) - (r_1 s_1 + r_2 s_2)^2 = \psi,$$

by part (c) of Lemma 2.2 we have $\sqrt{n}\sum_{i=1}^{n} c_n(i)\phi_i \xrightarrow{D} N\left(0, \frac{\psi}{a_1}\right)$. This, along with (10), gives $\frac{1}{n}\sum_{i=1}^{n} c_n(i)k_i \xrightarrow{D} N(0, \tau\psi)$. This, along with (20) verifies convergence of finite dimensional distributions.

It thus suffices to show tightness, for which using Arzela-Ascoli Theorem it suffices to verify that for every $\varepsilon > 0$ we have

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}_n\left(\max_{i,j\in[n]:|i-j|\le n\delta}|S_i(\mathbf{d}) - S_j(\mathbf{d})| > (n-1)\varepsilon\right) = 0. \tag{21}$$

To verify (21), first use (10) to note that

$$\begin{aligned}&\frac{1}{n-1}\max_{i,j\in[n]:|i-j|\le n\delta}|S_i(\mathbf{d}) - S_j(\mathbf{d})|\\ \le&\frac{1}{2}\left[\max_{i,j\in[n]:|i-j|\le n\delta}|S_i(\phi) - S_j(\phi)| + \frac{1}{\sqrt{(n-1)\theta}}\max_{i,j\in[n]:|i-j|\le n\delta}|S_i(\mathbf{W}) - S_j(\mathbf{W})|\right],\end{aligned}$$

where $\mathbf{W} = (W_1, \ldots, W_n)$ is a sequence of i.i.d. $N(0,1)$ random variables, and $S(\mathbf{W}) = \sum_{j=1}^{i} W_j$. This in turn gives the following bound to the RHS of (21):

$$\begin{aligned}&\mathbb{P}_n\left(\max_{i,j\in[n]:|i-j|\le n\delta}|S_i(\mathbf{d}) - S_j(\mathbf{d})| > \varepsilon(n-1)\right)\\ \le&\mathbb{P}_n\left(\max_{i,j\in[n]:|i-j|\le n\delta}|S_i(\phi) - S_j(\phi)| > \frac{\varepsilon}{4}\right)\\ +&\mathbb{P}_n\left(\frac{1}{\sqrt{(n-1)\theta}}\max_{i,j\in[n]:|i-j|\le n\delta}|S_i(\mathbf{W}) - S_j(\mathbf{W})| > \frac{\varepsilon\sqrt{(n-1)\theta}}{4}\right).\end{aligned}$$

The first term in the RHS above converges to 0 as $n \to \infty$ followed by $\delta \to 0$ using part (c) of Lemma 2.2, and the second term converges to 0 under the same double limit by tightness of sample paths for partial sums of i.i.d. random variables. Thus we have verified (21), and hence the proof of the theorem is complete. □

3. Proof of Lemma 2.1 and Lemma 2.2 for $(\theta, \beta) \in \Theta_1 \cup \Theta_2$

We first state a general approximation result, which will be used to analyze the marginal distribution of $(\phi_1, \ldots, \phi_n)$ by approximating the un-normalized density $f_n(.)$ of Proposition (1.12) by something more tractable. The approximating measure will change across the three parameter regimes $\Theta_1 \cup \Theta_2 \cup \Theta_3$.

Lemma 3.1. *For an interval $U \subseteq \mathbb{R}$, let $h_n(.), g_n(.) : U^n \mapsto \mathbb{R}$ be non negative and integrable. Define the probability measures $\mathbb{G}_n$ and $\mathbb{H}_n$ on U^n by setting*

$$\frac{d\mathbb{G}_n}{d\lambda^{\otimes n}} := \frac{g_n}{\int_{U^n} g_n d\lambda^{\otimes n}}, \quad \frac{d\mathbb{H}_n}{d\lambda^{\otimes n}} := \frac{h_n}{\int_{U^n} h_n d\lambda^{\otimes n}},$$

where $\lambda^{\otimes n}$ is Lebesgue measure on $\mathbb{R}^n$. Setting $L_n(.) = \log\frac{g_n}{h_n}$, suppose that L_n is $O_P(1)$ under both measures $\mathbb{G}_n, \mathbb{H}_n$.

(a) *Then the sequence of probability measures* $\mathbb{G}_n$ *and* $\mathbb{H}_n$ *are mutually contiguous.*

(b) *If* $(X_n, L_n) \overset{d,\mathbb{G}_n}{\longrightarrow} N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\sigma_{12})$ *then* $X_n \overset{d,\mathbb{H}_n}{\longrightarrow} N(\mu_1+\sigma_{12},\sigma_1^2)$.

(c) *If* $L_n \overset{d,\mathbb{G}_n}{\to} c$ *where c is a constant, then* $\|\mathbb{G}_n - \mathbb{H}_n\|_{TV} \to 0$.

Our plan is use Lemma 3.1 to approximate the distribution of ϕ by a multivariate Gaussian distribution. The following lemma summarizes some estimates under the approximating Gaussian distribution.

Lemma 3.2. *Let* $\mathbb{G}_{1n}$ *be a multivariate Gaussian distribution on* $\mathbb{R}^n$*, with density proportional to* g_{1n}*, where*

$$-\log g_{1n}(\phi) = \frac{n(n-1)}{2}p(t,t) + \frac{a_1 n}{2}\sum_{i=1}^n(\phi_i - t)^2 - \frac{a_2 n^2}{2}(\bar{\phi}-t)^2,$$

with $a_1 = \theta - \theta^2(1-t^2)$ *as in (13), and* $a_2 := \theta^2(1-t^2)$*. Then the following conclusions hold under* $\mathbb{G}_{1n}$*.*
(a) $\mathbb{E}_{\mathbb{G}_{1n}}|\phi_i - t|^\ell \lesssim_\ell n^{-\ell/2}$.
(b) $\sum_{1\le i<j\le n}\Big(\phi_i+\phi_j-2t\Big)^4 \overset{P}{\to} \frac{6}{a_1^2}$,
(c) Suppose $\mathbf{c} = (c_n(1),\dots,c_n(n))$ *be a vector such that* $\sum_{i=1}^n c_n(i) = 0$*, and* $\frac{1}{n}\sum_{i=1}^n c_n(i)^2 \to 1$*. Then we have*

$$\left[n(\bar{\phi}-t), \sqrt{n}\Big(\sum_{i=1}^n(\phi_i-\bar{\phi})^2 - a_1^{-1}\Big), n\sum_{i=1}^n(\phi_i-t)^3, \sum_{i=1}^n c_n(i)\phi_i\right] \overset{D}{\to} N(\mathbf{0},\Sigma)$$

where

$$\Sigma := \begin{bmatrix} \frac{1}{a_1-a_2} & 0 & \frac{3}{a_1(a_1-a_2)} & 0 \\ 0 & \frac{2}{a_1^2} & 0 & 0 \\ \frac{3}{a_1(a_1-a_2)} & 0 & \frac{15a_1-6a_2}{a_1^3(a_1-a_2)} & 0 \\ 0 & 0 & 0 & \frac{1}{a_1} \end{bmatrix}.$$

(d) For every $\varepsilon > 0$*, setting* $S_i(\phi) = \sum_{j=1}^i(\phi_j - \bar{\phi})$ *as before, we have*

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}_{\mathbb{G}_{1,n}}\big(\max_{i,j\in[n]:|i-j|\le n\delta}|S_i(\phi)-S_j(\phi)| > \varepsilon\big) = 0.$$

The final result we need for proving Lemma 2.1 in the regime $(\theta,\beta) \in \Theta_1 \cup \Theta_2$ is the following:

Lemma 3.3. *Let* $U = \mathbb{R}$ *if* $(\theta,\beta)\in\Theta_1$*, and* $U = (0,\infty)$ *if* $(\theta,\beta)\in\Theta_2$*.*

(a) *Then exists positive constants* $\lambda_1 \ge \lambda_2$ *such that for all* $(x,y)\in U^2$ *we have*

$$\frac{\lambda_2}{2}[(x-t)^2+(y-t)^2] \le p(x,y) - p(t,t) \le \frac{\lambda_1}{2}[(x-t)^2+(y-t)^2],$$

(b) *There exists M large enough such that*

$$\log\mathbb{P}_{n,U}\Big(\sum_{i=1}^n(\phi_i-t)^2 > M\Big) \lesssim -n.$$

where $\mathbb{P}_{n,U}$ *denotes the conditional law of* ϕ *under* $\mathbb{P}_n$ *given* $\phi\in U^n$.

(c) *For any* $l \in \mathbb{N}$ *we have* $\mathbb{E}_{n,U}|\phi_i - t|^l \lesssim_l n^{-\ell/2}$,

(d) $\mathbb{E}_{n,U}\big[\sum_{i=1}^n (\phi_i - t)\big]^2 \lesssim 1$.

The proof of the three Lemmas 3.1, 3.2 and 3.3 are deferred to the supplementary material [21] appendix B.

3.1. Proof of Lemma 2.1 and Lemma 2.2 for $(\theta, \beta) \in \Theta_1$

Let $\mathbb{F}_n$ denote the marginal distribution of ϕ on $\mathbb{R}^n$ under $\mathbb{P}_n$, i.e. $\mathbb{F}_n$ is induced by the unnormalized density $f_n(.)$ defined in Proposition 1.12. We begin by showing the following proposition:

Proposition 3.4. *If* $(\theta,\beta) \in \Theta_1$*, the probability measures* $\mathbb{F}_n$ *and* $\mathbb{G}_{1n}$ *are mutually contiguous, where* $\mathbb{G}_{1n}$ *is the multivariate Gaussian distribution introduced in Lemma 3.2.*

Proof. To this effect, with $q(.)$ as in Lemma 1.2, use a Taylor's series expansion to get

$$q\Big(\frac{x+y}{2}\Big) = q(t) + \frac{q''(t)}{2}\Big(\frac{x+y}{2} - t\Big)^2 + \frac{q'''(t)}{3!}\Big(\frac{x+y}{2} - t\Big)^3 + \frac{q''''(t)}{4!}\Big(\frac{x+y}{2} - t\Big)^4 + R(x,y),$$

where $|R(x,y)| \lesssim |x-t|^5 + |y-t|^5$. Recalling that $p(x,y) = q\Big(\frac{x+y}{2}\Big) + \frac{\theta}{4}(x-y)^2$ then gives

$$\begin{aligned} p(x,y) =& \frac{\theta}{4}(x-y)^2 + q(t) + \frac{q''(t)}{2}\Big(\frac{x+y}{2} - t\Big)^2 + \frac{q'''(t)}{3!}\Big(\frac{x+y}{2} - t\Big)^3 \\ &+ \frac{q''''(t)}{4!}\Big(\frac{x+y}{2} - t\Big)^4 + R(x,y) \\ =& p(t,t) + \frac{1}{2}\Big[a_1(x-t)^2 + a_1(y-t)^2 - 2a_2(x-t)(y-t)\Big] + \frac{a_3}{3!}(x+y-2t)^3 \\ &+ \frac{a_4}{4!}(x+y-2t)^4 + R(x,y), \end{aligned}$$

where $a_1 = \theta - \theta^2(1-t^2), a_2 = \theta^2(1-t^2)$ as in Lemma 3.2, and $a_3 := \frac{q'''(t)}{8}, a_4 := \frac{q''''(t)}{16}$. Adding $p(\phi_i,\phi_j)$ over $i<j$, this gives

$$-\log f_n(\phi) = \sum_{\ell=1}^4 R_{\ell, f_n} + \sum_{i<j} R(\phi_i,\phi_j),$$

where

$$
\begin{aligned}
R_{1,f_n} :=& \sum_{i<j} \frac{a_1}{2}[(\phi_i - t)^2 + (\phi_j - t)^2] = \frac{a_1(n-1)}{2} \sum_{i=1}^n (\phi_i - t)^2 \\
R_{2,f_n} :=& - a_2 \sum_{i<j} (\phi_i - t)(\phi_j - t) = -\frac{a_2 n^2}{2}(\bar{\phi} - t)^2 + \frac{a_2}{2} \sum_{i=1}^n (\phi_i - t)^2 \\
R_{3,f_n} :=& \frac{a_3}{6} \sum_{i<j} (\phi_i + \phi_j - 2t)^3 = \frac{a_3}{12} \left[\sum_{i,j=1}^n (\phi_i + \phi_j - 2t)^3 - 8 \sum_{i=1}^n (\phi_i - t)^3 \right] \\
=& \frac{(n-4)a_3}{6} \sum_{i=1}^n (\phi_i - t)^3 + \frac{3na_3}{6}(\bar{\phi} - t) \sum_{i=1}^n (\phi_i - t)^2, \\
R_{4,f_n} :=& \frac{a_4}{4!} \sum_{1 \le i<j \le n} (\phi_i + \phi_j - 2t)^4.
\end{aligned} \tag{22}
$$

Consequently we have

$$
\begin{aligned}
&\left| - \log f_n(\phi) - \frac{a_1 n}{2} \sum_{i=1}^n (\phi_i - t)^2 - \frac{a_3 n}{6} \sum_{i=1}^n (\phi_i - t)^3 \right| \\
\lesssim & n^2(\bar{\phi} - t)^2 + n(\bar{\phi} - t) \sum_{i=1}^n (\phi_i - t)^2 + \sum_{\ell=2}^5 \sum_{i=1}^n |\phi_i - t|^\ell
\end{aligned} \tag{23}
$$

Fixing $b_4 > a_3^2/3a_1$, define the function $h_{1n}(\phi)$ by

$$
\begin{aligned}
-\log h_{1n}(\phi) :=& \frac{n(n-1)}{2} p(t,t) + \frac{a_1 n}{2} \sum_{i=1}^n (\phi_i - t)^2 + \frac{a_3 n}{3!} \sum_{i=1}^n (\phi_i - t)^3 + \frac{b_4 n}{4!} \sum_{i=1}^n (\phi_i - t)^4 \\
=& \frac{n(n-1)}{2} p(t,t) + n \sum_{i=1}^n \eta(\phi_i - t), \quad \eta(x) := \frac{a_1}{2!} x^2 + \frac{a_3}{3!} x^3 + \frac{b_4}{4!} x^4,
\end{aligned} \tag{24}
$$

and note that $(\phi_1 - t, \ldots, \phi_n - t)$ are i.i.d. under $\mathbb{H}_{1n}$ with density proportional to $e^{-n \sum \eta(.)}$, where $\mathbb{H}_{1n}$ denotes the probability measure induced by h_{1n}. It follows from straightforward calculus that

$$
\begin{aligned}
\left| \int_{\mathbb{R}} x^\ell e^{-n\eta(x)} dx \right| \lesssim_\ell & \frac{1}{n^{\frac{\ell+1}{2}}} \text{ if } \ell \text{ is even,} \\
\lesssim_\ell & \frac{1}{n^{\frac{\ell+3}{2}}} \text{ if } \ell \text{ is odd,}
\end{aligned}
$$

and so

$$
\begin{aligned}
\mathbb{E}_{\mathbb{H}_{1n}} (\phi_i - t)^\ell \lesssim & \frac{1}{n^{\frac{\ell}{2}}} \text{ if } \ell \text{ is even,} \\
\lesssim_\ell & \frac{1}{n^{\frac{\ell}{2}+1}} \text{ if } \ell \text{ is odd.}
\end{aligned} \tag{25}
$$

Also, comparing (23) and (24) we have

$$
\begin{aligned}
&|\log f_n(\phi) - \log h_{1n}(\phi)| \\
&\lesssim n^2(\bar{\phi}-t)^2 + \Big|n(\bar{\phi}-t)\sum_{i=1}^n(\phi_i-t)^2\Big| + \sum_{\ell=2}^5\sum_{i=1}^n|\phi_i-t|^\ell + n\sum_{i=1}^n(\phi_i-t)^4.
\end{aligned} \tag{26}
$$

Using parts (c) and (d) of Lemma 3.3 it follows that $\log f_n(\phi) - \log h_{1n}(\phi)$ is $O_P(1)$ under $\mathbb{F}_n$. To show the same conclusion under $\mathbb{H}_{1n}$, it suffices to note that

$$
\mathbb{E}_{\mathbb{H}_{1n}}\Big[\sum_{i=1}^n(\phi_i-t)\Big]^2 \lesssim 1, \quad \mathbb{E}_{\mathbb{H}_{1n}}|\phi_i-t|^\ell \lesssim n^{-\ell/2}, \tag{27}
$$

both of which follow from (25). It thus follows from Lemma 3.1 that $\mathbb{F}_n$ and $\mathbb{H}_{1n}$ are mutually contiguous. To complete the proof, it suffices to show that $\mathbb{G}_{1n}$ and $\mathbb{H}_{1n}$ are mutually contiguous. Proceeding to verify this, note that

$$
\begin{aligned}
\Big|\log\frac{g_{1n}(\phi)}{h_{1n}(\phi)}\Big| \lesssim & n(\bar{\phi}-t)^2 + \Big|n(\bar{\phi}-t)\sum_{i=1}^n(\phi_i-t)^2\Big| \\
& + n\Big|\sum_{i=1}^n(\phi_i-t)^3\Big| + \sum_{\ell=2}^5\sum_{i=1}^n|\phi_i-t|^\ell + n\sum_{i=1}^n(\phi_i-t)^4.
\end{aligned} \tag{28}
$$

We need to show that the RHS of (28) is $O_P(1)$ under both $\mathbb{H}_{1n}$ and $\mathbb{G}_{1n}$. Again the desired conclusion for $\mathbb{H}_{1n}$ follows (27), and using (25) to note that

$$
n\mathbb{E}_{\mathbb{H}_{1n}}\Big[\sum_{i=1}^n(\phi_i-t)^3\Big]^2 \lesssim 1. \tag{29}
$$

To complete the proof, it suffices to verify (27) and (29) under $\mathbb{G}_{1n}$. But this follows from parts (a) and (c) of Lemma 3.2. This shows that $\mathbb{F}_n$ and $\mathbb{G}_{1n}$ are mutually continuous, and so we have verified the proposition. □

Proof of Lemma 2.1 for $(\theta,\beta)\in\Theta_1$. Use (22) to note that

$$
-\log\frac{f_n}{g_{1n}} = \frac{a_2-a_1}{2}\sum_{i=1}^n(\phi_i-t)^2 + R_{3,f_n} + R_{4,f_n} + \sum_{1\le i<j\le n}R(\phi_i,\phi_j).
$$

Invoking parts (a) and (b) of Lemma 3.2, under $\mathbb{G}_{1n}$ we have

$$
\Big(\sum_{i=1}^n(\phi_i-t)^2, R_{4,f_n}, \sum_{1\le i<j\le n}R(\phi_i,\phi_j)\Big) \xrightarrow{P} \Big(\frac{1}{a_1}, \frac{a_4}{4a_1^2}, 0\Big). \tag{30}
$$

Also, using (22), a direct expansion gives

$$
R_{3,f_n} = \frac{(n-4)a_3}{6}\sum_{i=1}^n(\phi_i-t)^3 + \frac{3a_3}{6}n(\bar{\phi}-t)\sum_{i=1}^n(\phi_i-t)^2
$$

$$= \frac{na_3}{6}\sum_{i=1}^n (\phi_i - t)^3 + \frac{a_3}{2a_1} n(\bar{\phi} - t) + o_p(1), \tag{31}$$

where the last equality again uses (30). Combining (30) and (31) along with (22) gives that under $\mathbb{G}_{1n}$,

$$-\log \frac{f_n}{g_{1n}} = \frac{a_2 - a_1}{2a_1} + \frac{a_4}{4a_1^2} + \frac{na_3}{6}\sum_{i=1}^n (\phi_i - t)^3 + \frac{a_3}{2a_1} n(\bar{\phi} - t) + o_p(1). \tag{32}$$

Using part (c) of Lemma 3.2, it follows that $\log \frac{f_n}{g_{1n}}$ is $O_p(1)$ under $\mathbb{G}_{1n}$. Since $\mathbb{F}_n$ and $\mathbb{G}_{1n}$ are mutually contiguous by Proposition 3.4, it follows that $\log \frac{f_n}{g_{1n}}$ is $O_p(1)$ under $\mathbb{F}_n$ as well. Thus to find out the limiting distribution of $n(\bar{\phi} - t)$ under $\mathbb{F}_n$, invoking Lemma 3.1 part (b) and (32) it suffices to find out the joint limiting distribution of $\left[n(\bar{\phi} - t), n\sum_{i=1}(\phi_i - t)^3\right]$ under $\mathbb{G}_{1n}$.

To this effect, using part (c) of Lemma 3.2, under $\mathbb{G}_{1n}$ we have

$$\left[n(\bar{\phi} - t), n\sum_{i=1}(\phi_i - t)^3\right] \xrightarrow{D} N\left(\mathbf{0}, \begin{bmatrix} \frac{1}{a_1 - a_2} & \frac{3}{a_1(a_1 - a_2)} \\ \frac{3}{a_1(a_1 - a_2)} & \frac{15a_1 - 6a_2}{a_1^3(a_1 - a_2)} \end{bmatrix}\right).$$

By the mutual contiguity of $\mathbb{F}_n$ and $\mathbb{G}_{1n}$ (Proposition 3.4) and part (b) of Lemma 3.1, it follows that under $\mathbb{F}_n$ we have $n(\bar{\phi} - t) \xrightarrow{D} N(-\mu, \frac{1}{a_1 - a_2})$, where

$$\mu = \frac{a_3}{2a_1} \times \frac{1}{a_1 - a_2} + \frac{a_3}{6} \times \frac{3}{a_1(a_1 - a_2)} = \frac{a_3}{a_1(a_1 - a_2)} = \frac{2\theta t(1 - t^2)}{[1 - \theta(1 - t^2)][1 - 2\theta(1 - t^2)]}$$

as desired. □

Proof of Lemma 2.2 for $(\theta, \beta) \in \Theta_1$.

(a) Using part (c) of Lemma 3.2 along with (32) it follows that the random variables $\sqrt{n}[\sum_{i=1}^n (\phi_i - t)^2 - \frac{1}{a_1}]$ and $\frac{\log f_n}{\log g_{1n}}$ are asymptotically mutually independent and Gaussian under $\mathbb{G}_{1n}$. The desired result then follows from part (c) of Lemma 3.2 along with part (b) of Lemma 3.1.

(b) The proof of part (b) follows on similar lines as the proof of part (a), and is not repeated here.

(c) By Proposition 3.4 the two distributions $\mathbb{F}_n$ and $\mathbb{G}_{1n}$ are mutually contiguous, and so it suffices to verify the result under $\mathbb{G}_{1n}$. But this is precisely part (d) of Lemma 3.2, and so the proof is complete. □

3.2. Proof of Lemma 2.1 and Lemma 2.2 for $(\theta, \beta) \in \Theta_2$

We begin by stating the following proposition, the proof of which is deferred to the supplementary material [21] appendix C.

Proposition 3.5. *For $(\theta, \beta) \in \Theta_2$, we have*

$$\left|\frac{1}{2} - \mathbb{P}_n(\phi_i \geq 0, 1 \leq i \leq n)\right| \leq e^{-\Omega(n)}.$$

Proof of Lemma 2.1 and Lemma 2.2 for $(\theta,\beta)\in\Theta_2$. By symmetry, we have $\mathbb{P}_n(\bar\phi>0)=\frac{1}{2}$, which along with Proposition 3.5 gives that conditioned on $\bar\phi>0$ we have

$$\mathbb{P}_n(\phi_i\le 0 \text{ for some } i, 1\le i\le n|\bar\phi>0)\le e^{-\Omega(n)}.$$

Thus at an exponentially vanishing cost we can replace the event $\bar\phi>0$ by the event $\{\phi_i>0, 1\le i\le n\}$. Consequently, invoking Lemma 3.3 with $U=(0,\infty)$ and proceeding exactly as in the uniqueness domain we get the following conclusions:

$$\begin{aligned}
&[n(\bar\phi-t)|\bar\phi>0]\xrightarrow{D} N\left(-\frac{2\theta t(1-t^2)}{[1-\theta(1-t^2)][1-2\theta(1-t^2)]},\frac{1}{\theta-2\theta^2(1-t^2)}\right),\\
&\left(\sqrt{n}\Big[\sum_{i=1}^n(\phi_i-\bar\phi)^2-a_1^{-1}\Big]\Big|\bar\phi>0\right)\xrightarrow{D} N(0,2a_1^{-2}),\\
&\left(\sqrt{n}\sum_{i=1}^n c_n(i)(\phi_i-\bar\phi)|\bar\phi>0\right)\xrightarrow{D} N(0,a_1^{-1}),\\
&\limsup_{\delta\to0}\limsup_{n\to\infty}\mathbb{P}_n(\max_{i,j\in[n]:|i-j|\le n\delta}|S_i(\phi)-S_j(\phi)|>\varepsilon|\bar\phi>0)=0.
\end{aligned}$$

Here the last line above holds for any $\varepsilon>0$. Since ϕ and $-\phi$ have the same distribution, we get using symmetry that

$$\begin{aligned}
&[n(\bar\phi+t)|\bar\phi<0]\xrightarrow{D} N\left(\frac{2\theta t(1-t^2)}{[1-\theta(1-t^2)][1-2\theta(1-t^2)]},\frac{1}{\theta-2\theta^2(1-t^2)}\right),\\
&\left(\sqrt{n}\Big[\sum_{i=1}^n(\phi_i-\bar\phi)^2-a_1^{-1}\Big]\Big|\bar\phi<0\right)\xrightarrow{D} N(0,2a_1^{-2}),\\
&\left(\sqrt{n}\sum_{i=1}^n c_n(i)(\phi_i-\bar\phi)|\bar\phi<0\right)\xrightarrow{D} N(0,a_1^{-1}),\\
&\limsup_{\delta\to0}\limsup_{n\to\infty}\mathbb{P}_n(\max_{i,j\in[n]:|i-j|\le n\delta}|S_i(\phi)-S_j(\phi)|>\varepsilon|\bar\phi<0)=0.
\end{aligned}$$

This readily proves Lemma 2.1. Lemma 2.2 follows on noting that the conditional distribution in the second, third and fourth lines in the above display is the same for $\bar\phi>0$ and $\bar\phi<0$. □

4. Proof of Lemmas 2.1 and 2.2 for $(\theta,\beta)\in\Theta_3$

We first state two lemmas which we will use to prove Lemma 2.1 and Lemma 2.2 for $(\theta,\beta)\in\Theta_3$. The first lemma is the analogue of Lemma 3.3 parts (c) and (d), and the second lemma is the analogue of Lemma 3.2. The proof of the two lemmas are deferred to the supplementary material [21] appendix D.

Lemma 4.1. *Suppose* $(\theta,\beta)\in\Theta_3$.

(a) *For any positive integer ℓ we have*

$$\mathbb{E}|\phi_i-\bar\phi|^l\lesssim_\ell\frac{1}{n^{l/2}}.$$

(b) $\limsup_{n\to\infty} n^2\mathbb{E}\bar{\phi}^4 < \infty$.

Lemma 4.2. *Suppose*

$$-\log g_{3n}(\phi) := \frac{(n-1)\theta}{4}\sum_{i=1}^n(\phi_i-\bar{\phi})^2 - \frac{1}{2}n\bar{\phi}^2 - \frac{1}{24}n^2\bar{\phi}^4,$$

and let $\mathbb{G}_{3n}$ *denote the corresponding probability measure on* $\mathbb{R}^n$. *Then the following conclusions under* $\mathbb{G}_{3n}$*:*

(a)

$$n\mathbb{E}_{\mathbb{G}_{3n}}\bar{\phi}^2 \lesssim 1, \quad n\mathbb{E}_{\mathbb{G}_{3n}}(\phi_i-\bar{\phi})^2 \lesssim 1. \tag{33}$$

(b)

$$\sum_{i=1}^n(\phi_i-\bar{\phi})^2 \xrightarrow{P} 4, \tag{34}$$

$$n^{-1/2}\sum_{1\le i<j\le n}(\phi_i+\phi_j-2\bar{\phi})^3 \xrightarrow{P} 0, \tag{35}$$

$$\sum_{1\le i<j\le n}(\phi_i+\phi_j-2\bar{\phi})^4 \xrightarrow{P} 96. \tag{36}$$

(c) $\sqrt{n}\bar{\phi} \xrightarrow{D} \zeta$, *where* ζ *is a continuous random variable on* $\mathbb{R}$ *with density proportional to* $e^{-\frac{\zeta^2}{2}-\frac{\zeta^4}{24}}$ *with respect to Lebesgue measure.*

(d)

$$\sqrt{n}\Big[\sum_{i=1}^n(\phi_i-\bar{\phi})^2 - \frac{1}{a_1}\Big] \xrightarrow{D} N\Big(0, \frac{2}{a_1^2}\Big).$$

(e) *For any triangular array* $(c_n(1),\dots,c_n(n))$ *with* $\sum_{i=1}^n c_n(i)=0, \frac{1}{n}\sum_{i=1}^n c_n(i)^2 \to 1$ *we have*

$$\sum_{i=1}^n c_n(i)\phi_i \xrightarrow{D} N\Big(0,\frac{1}{a_1}\Big).$$

(f) *For every* $\varepsilon > 0$ *we have*

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}\mathbb{P}_n(\max_{i,j\in[n]:|i-j|\le n\delta}|S_i(\phi)-S_j(\phi)|>\varepsilon)=0.$$

Proceeding to verify Lemma 2.1 and Lemma 2.2, we begin by showing the following proposition, which is the analogue of Proposition 3.4 for $(\theta,\beta)\in\Theta_3$.

Proposition 4.3. *With* $\mathbb{G}_{3n}$ *as defined in Lemma 4.2 above, if* $(\theta,\beta)\in\Theta_3$ *then we have* $\|\mathbb{F}_n-\mathbb{G}_{3n}\|_{TV}\to 0$.

Proof. Expanding $q(x) = \frac{x^2}{2} - \log\cosh(x)$ by a Taylor's series around 0 we get

$$q(x) = \frac{2x^4}{4!} + R(x), \text{ where } |R(x)| \lesssim |x|^6.$$

Since $\theta = \frac{1}{2}$, using (12) and summing over $1 \le i < j \le n$ we get

$$-\log f_n(\phi) = \frac{n}{8}\sum_{i=1}^n(\phi_i - \bar{\phi})^2 + \sum_{i<j}\frac{(\phi_i + \phi_j)^4}{2^3 4!} + \sum_{i<j} R(\phi_i + \phi_j). \tag{37}$$

With $N = \frac{n(n-1)}{2}$ as before, expanding the second term in (37) we get

$$\begin{aligned}\sum_{i<j}(\phi_i + \phi_j)^4 =& 16N\bar{\phi}^4 + 24\sum_{i<j}(\phi_i + \phi_j - 2\bar{\phi})^2\bar{\phi}^2 \\ &+ 8\sum_{i<j}(\phi_i + \phi_j - 2\bar{\phi})^3\bar{\phi} + \sum_{i<j}(\phi_i + \phi_j - 2\bar{\phi})^4,\end{aligned}$$

which along with (37) and the identity $\sum_{i<j}(\phi_i + \phi_j - 2\bar{\phi})^2 = (n-2)\sum_{i=1}^n(\phi_i - \bar{\phi})^2$ gives

$$\begin{aligned}-\log\frac{f_n}{g_{3n}} =& -\frac{n}{24}\bar{\phi}^4 + \frac{1}{8}\bar{\phi}^2\Big[(n-2)\sum_{i=1}^n(\phi_i - \bar{\phi})^2 - 4n\Big] \\ &+ \frac{1}{2^3 4!}\sum_{1\le i<j\le n}\Big[8\bar{\phi}(\phi_i + \phi_j - 2\bar{\phi})^3 + (\phi_i + \phi_j - 2\bar{\phi})^4 + R(\phi_i + \phi_j)\Big]\end{aligned} \tag{38}$$

To bound each term on the RHS of (38) separately, use Lemma 4.1 to get that under $\mathbb{F}_n$,

$$n\bar{\phi}^4 \xrightarrow{P} 0, \quad \sum_{i<j} R(\phi_i + \phi_j) \lesssim n\sum_{i=1}^n(\phi_i - \bar{\phi})^6 + n^2\bar{\phi}^6 \xrightarrow{P} 0, \tag{39}$$

$$\Big|(n-2)\bar{\phi}^2\sum_{i=1}^n(\phi_i - \bar{\phi})^2 - 4n\bar{\phi}^2\Big| \lesssim n\bar{\phi}^2\Big[1 + \sum_{i=1}^n(\phi_i - \bar{\phi})^2\Big] = O_P(1),$$

$$\Big|\sum_{1\le i<j\le n}(\phi_i + \phi_j - 2\bar{\phi})^3(2\bar{\phi})\Big| \lesssim \Big|\sqrt{n}\bar{\phi}\Big|\sqrt{n}\sum_{i=1}^n|\phi_i - \bar{\phi}|^3 = O_P(1)$$

$$\sum_{1\le i<j\le n}(\phi_i + \phi_j - 2\bar{\phi})^4 \lesssim n\sum_{i=1}^n(\phi_i - \bar{\phi})^4 = O_P(1).$$

It thus follows that $\log\frac{f_n}{g_{3n}}$ is $O_P(1)$ under $\mathbb{F}_n$. To show the same conclusion under $\mathbb{G}_{3n}$, it suffices to show that the estimates of Lemma 4.1 hold under $\mathbb{G}_{3n}$ as well, which follows from part (a) of Lemma 4.2. Thus, using Lemma 3.1 we have that $\mathbb{F}_n$ and $\mathbb{G}_{3n}$ are mutually contiguous.

Finally to show that $\mathbb{F}_n$ and $\mathbb{G}_{3n}$ are close in total variation, invoking part (c) of Lemma 3.1 it suffices to show that $\log(f_n/g_{3n})$ converges in probability to a constant under $\mathbb{G}_{3n}$. Invoking (38) and (39), it suffices to show the following conclusions under $\mathbb{G}_{3n}$:

$$\sum_{i=1}^n(\phi_i - \bar{\phi})^2 \xrightarrow{P} 4, \tag{40}$$

$$n^{-1/2} \sum_{1 \le i < j \le n} (\phi_i + \phi_j - 2\bar{\phi})^3 \xrightarrow{P} 0, \tag{41}$$

$$\sum_{1 \le i < j \le n} (\phi_i + \phi_j - 2\bar{\phi})^4 \xrightarrow{P} 96. \tag{42}$$

But this follows from part (b) of Lemma 4.2. Thus we have verified Proposition 4.3. □

Proof of Lemma 2.1 and Lemma 2.2 for $(\theta, \beta) \in \Theta_3$**.** By Proposition 4.3 the two probability measures $\mathbb{F}_n$ and $\mathbb{G}_{3n}$ are close in total variation, and so it suffices to work with $\mathbb{G}_{3n}$. But under $\mathbb{G}_{3n}$ the desired conclusions are immediate from parts (c), (d), (e) and (f) of Lemma 4.2. □

Acknowledgements

We thank the associate editor and two anonymous referees, whose comments and suggestions have greatly improved the quality of presentation in the paper.

Funding

The first author was partially supported by NSF Grants DMS-1712037 and DMS-2113414.

Supplementary Material

Supplement to "Statistics of the two star ERGM" (DOI: 10.3150/21-BEJ1448SUPP; .pdf). The supplementary material contains the proof of helpful lemmas and propositions, which have been used to prove the main results of the paper. The proofs of Lemma 1.2 and Proposition 1.12 are in Appendix A. The proofs of the three Lemmas 3.1, 3.2 and 3.3 are in appendix B. The proof of proposition 3.5 is in appendix C. The proofs of Lemmas 4.1 and 4.2 are deferred to appendix D.

References

[1] Andersen, H.C. and Diaconis, P. (2007). Hit and run as a unifying device. *J. Soc. Fr. Stat. & Rev. Stat. Appl.* **148** 5–28. MR2502361

[2] Anderson, C.J., Wasserman, S. and Crouch, B. (1999). A p* primer: Logit models for social networks. *Soc. Netw.* **21** 37–66.

[3] Bhamidi, S., Bresler, G. and Sly, A. (2008). Mixing time of exponential random graphs. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science* 803–812. IEEE.

[4] Blitzstein, J. and Diaconis, P. (2010). A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Math.* **6** 489–522. MR2809836 https://doi.org/10.1080/15427951.2010.557277

[5] Chatterjee, S. and Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Ann. Statist.* **41** 2428–2461. MR3127871 https://doi.org/10.1214/13-AOS1155

[6] Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *Ann. Appl. Probab.* **21** 1400–1435. MR2857452 https://doi.org/10.1214/10-AAP728

[7] Chatterjee, S. and Mukherjee, S. (2019). Estimation in tournaments and graphs under monotonicity constraints. *IEEE Trans. Inf. Theory* **65** 3525–3539. MR3959003 https://doi.org/10.1109/TIT.2019.2893911

[8] Comets, F. and Gidas, B. (1991). Asymptotics of maximum likelihood estimators for the Curie-Weiss model. *Ann. Statist.* **19** 557–578. MR1105836 https://doi.org/10.1214/aos/1176348111
[9] Deb, N. and Mukherjee, S. (2020). Fluctuations in Mean-Field Ising models. ArXiv Preprint. Available at arXiv:2005.00710.
[10] DeMuse, R., Larcomb, D. and Yin, M. (2018). Phase transitions in edge-weighted exponential random graphs: Near-degeneracy and universality. *J. Stat. Phys.* **171** 127–144. MR3773854 https://doi.org/10.1007/s10955-018-1991-3
[11] Edwards, R.G. and Sokal, A.D. (1988). Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Phys. Rev. D* **38** 2009–2012. MR0965465 https://doi.org/10.1103/PhysRevD.38.2009
[12] Ellis, R.S. and Newman, C.M. (1978). The statistics of Curie-Weiss models. *J. Stat. Phys.* **19** 149–161. MR0503332 https://doi.org/10.1007/BF01012508
[13] Frank, O. and Strauss, D. (1986). Markov graphs. *J. Amer. Statist. Assoc.* **81** 832–842. MR0860518
[14] Ganguly, S. and Nam, K. (2019). Sub-critical Exponential random graphs: Concentration of measure and some applications. ArXiv Preprint. Available at arXiv:1909.11080.
[15] Ghosal, P. and Mukherjee, S. (2020). Joint estimation of parameters in Ising model. *Ann. Statist.* **48** 785–810. MR4102676 https://doi.org/10.1214/19-AOS1822
[16] Götze, F., Sambale, H. and Sinulis, A. (2021). Concentration inequalities for polynomials in α-subexponential random variables. *Electron. J. Probab.* **26** Paper No. 48. MR4247973 https://doi.org/10.1214/21-ejp606
[17] Handcock, M.S., Robins, G., Snijders, T., Moody, J. and Besag, J. (2003). Assessing degeneracy in statistical models of social networks. Technical Report, Working paper.
[18] Holland, P.W. and Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *J. Amer. Statist. Assoc.* **76** 33–65. MR0608176
[19] Mukherjee, R., Mukherjee, S. and Yuan, M. (2018). Global testing against sparse alternatives under Ising models. *Ann. Statist.* **46** 2062–2093. MR3845011 https://doi.org/10.1214/17-AOS1612
[20] Mukherjee, S. (2020). Degeneracy in sparse ERGMs with functions of degrees as sufficient statistics. *Bernoulli* **26** 1016–1043. MR4058359 https://doi.org/10.3150/19-BEJ1135
[21] Mukherjee, S. Xu, Y. (2023). Supplement to "Statistics of the two star ERGM." https://doi.org/10.3150/21-BEJ1448SUPP
[22] Papangelou, F. (1989). On the Gaussian fluctuations of the critical Curie-Weiss model in statistical mechanics. *Probab. Theory Related Fields* **83** 265–278. MR1012501 https://doi.org/10.1007/BF00333150
[23] Park, J. and Newman, M.E.J. (2004). Solution of the two-star model of a network. *Phys. Rev. E (3)* **70** 066146. MR2133810 https://doi.org/10.1103/PhysRevE.70.066146
[24] Park, J. and Newman, M.E.J. (2004). Statistical mechanics of networks. *Phys. Rev. E (3)* **70** 066117. MR2133807 https://doi.org/10.1103/PhysRevE.70.066117
[25] Radin, C. and Yin, M. (2013). Phase transitions in exponential random graphs. *Ann. Appl. Probab.* **23** 2458–2471. MR3127941 https://doi.org/10.1214/12-AAP907
[26] Rinaldo, A., Petrović, S. and Fienberg, S.E. (2013). Maximum likelihood estimation in the β-model. *Ann. Statist.* **41** 1085–1110. MR3113804 https://doi.org/10.1214/12-AOS1078
[27] Robins, G., Pattison, P., Kalish, Y. and Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Soc. Netw.* **29** 173–191.
[28] Sambale, H. and Sinulis, A. (2020). Logarithmic Sobolev inequalities for finite spin systems and applications. *Bernoulli* **26** 1863–1890. MR4091094 https://doi.org/10.3150/19-BEJ1172
[29] Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *J. Amer. Statist. Assoc.* **106** 1361–1370. MR2896841 https://doi.org/10.1198/jasa.2011.tm10747
[30] Schweinberger, M. and Stewart, J. (2020). Concentration and consistency results for canonical and curved exponential-family models of random graphs. *Ann. Statist.* **48** 374–396. MR4065166 https://doi.org/10.1214/19-AOS1810
[31] Shalizi, C.R. and Rinaldo, A. (2013). Consistency under sampling of exponential random graph models. *Ann. Statist.* **41** 508–535. MR3099112 https://doi.org/10.1214/12-AOS1044
[32] Snijders, T.A., Pattison, P.E., Robins, G.L. and Handcock, M.S. (2006). New specifications for exponential random graph models. *Sociol. Method.* **36** 99–153.

[33] Swendsen, R.H. and Wang, J.-S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.* **58** 86.

[34] Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications* **8**. Cambridge: Cambridge University Press.

[35] Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks. I. An introduction to Markov graphs and *p*. *Psychometrika* **61** 401–425. MR1424909 https://doi.org/10.1007/BF02294547

Received February 2021 and revised November 2021