Statistical Models for Degree Distributions of Networks

Kayvan Sadeghi
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213
kayvans@andrew.cmu.edu

Alessandro Rinaldo
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213
arinaldo@cmu.edu

Abstract

We define and study the statistical model in exponential family form whose sufficient statistics are the degree distributions of undirected labelled simple graphs. In addition to formalizing this model, we provide some results for the parameter estimation and the asymptotic behaviour of the proposed model.

1 Introduction and background

Introduction. *Exponential random graph models* (ERGMs) form a flexible and powerful family of statistical models for network data, used in variety of fields and especially in the social scenes; see [16] and other papers within the same special issue. These models are of exponential family form with proposed sufficient statistics that can range from the number of edges of the network [8] to the number of k-stars, or other graphical features of networks; see for example [11, 9].

While ERGMs can be specified using any network statistics, it has been long known that node degrees and statistics thereof have great expressive power in representing and modeling networks, perhaps more than most network statistics. See, e.g., [14] and [10]. Some of the recent literature on ERGMs have focussed on the properties of the *beta model*, for which the degree sequence of a network is the sufficient statistic and postulates independent edges; see [2], [6], and [15]. In this paper we consider instead the less studied class of ERGMs whose sufficient statistics are derived from the *degree distribution* of the nodes, for which the assumption of dyadic independence no longer holds.

Contributions. This paper contains the following contributions. We formally define ERGMs based on the 1K model, and we derive conditions for the existence of the MLE of the model parameter and, therefore, for their estimability. We consider the general case of $m \geq 1$ i.i.d observations, which in particular includes the more interesting and common case of $m = 1$ observed network. We are concerned with the asymptotic behavior of the above model and compare it specifically to the behavior of the dense Erdös-Rényi model. We show that the this model is in fact radically different from the Erdös-Rényi model, an appealing feature that many ERGMs do not always possess, as demonstrated by [5].

2 ERGM for degree distribution

The exponential family form. Denote the set of simple, undirected, labeled graphs with n nodes by $\mathcal{G}_n$. Our goal is to model the probability of observing a network g with n nodes in exponential family form with sufficient statistics $n^{(1)}(g) = (n_0(g), n_1(g), \ldots, n_{n-1}(g))$, where $n_k(g)$ is the number of nodes with degree k in g. Notice that the scaled version of sufficient statistics, $n^{(1)}(g)/n$, is generally called the *degree distribution* of g. We call this model the $1K$ *model*.

Denote also the probability of a node having degree k by p_k. Considering the ordered vector of possible degrees $(0, \ldots, n-1)$ in g, the probability of observing g can be written as

$$P(g) = \varphi(p) \prod_{k=0}^{n-1} p_k^{n_k(g)}, \tag{1}$$

where φ is a *normalizing constant* and $\sum_{k=0}^{n-1} p_k = 1$.

To write (1) in a minimal form, we assume that all p_k are positive, and use the parametrization $\tilde{p}_k = p_k/p_{n-1}$, which by using the fact that $\tilde{p}_{n-1} = 1$, implies $p_{n-1} = 1/(1 + \sum_{k=0}^{n-2} \tilde{p}_k)$. Therefore, (1) can be rewritten as

$$P(g) = \varphi(p) p_{n-1}^{\sum_{k=0}^{n-1} n_k(g)} \prod_{k=0}^{n-1} \tilde{p}_k^{n_k(g)} = \frac{\phi(\tilde{p})}{(1 + \sum_{k=0}^{n-2} \tilde{p}_k)^n} \prod_{k=0}^{n-2} \tilde{p}_k^{n_k(g)}. \tag{2}$$

This in turn can be parametrized in exponential family form as

$$P(g) = \exp\left\{ \sum_{k=0}^{n-2} n_k(g)\alpha_k - \psi(\alpha) \right\}, \tag{3}$$

where $\alpha_k = \log \tilde{p}_k$ and $\psi(\alpha) = n \log(1 + \sum_{k=0}^{n-2} \exp(\alpha_k)) - \log \phi(\tilde{p})$.

We reduced the dimension of the sufficient statistics by arbitrarily removing the element n_{n-1} , to obtain $n_-^{(1)}(g)$ as sufficient statistics. We see that this model assigns the same probability to graphs with the same degree distribution, which is the only information the model collects from the data.

Calculating the normalizing constant. We know that $\sum_{g \in \mathcal{G}_n} P(g) = 1$, where $\mathcal{G}_n$ is the set of all non-isomorphic graphs with n nodes. We use this to calculate the normalizing constant $\psi(\alpha) = \psi_n(\alpha)$ for a fixed n. In this case the normalizing constant can be calculated directly from the set of all graphs with n nodes. Denote all these graphs by $\mathcal{G}_n = \{g_1, \ldots, g_M\}$, where $M = 2^{\binom{n}{2}}$. It is easy to prove that the normalizing constant for the 1K model can be written as

$$\psi(\alpha) = \log(1 + \cdots + \prod_{k=0}^{n-2} \tilde{p}_k^{n_k(g_i)} + \cdots + \tilde{p}_0^n) = \log(1 + \cdots + e^{\sum_{k=0}^{n-2} n_k(g_i)\alpha_k} + \cdots + e^{n\alpha_0}), \tag{4}$$

where there are M terms corresponding to each $g_i \in \mathcal{G}_n$, and the first and the last terms correspond to the complete and the null graphs.

Notice that in (4) there are repeated terms for different labeling of isomorphic graphs as well as non-isomorphic graphs with the same degree distribution. We illustrate the proposition with an example; see also [12] for asymptotic estimates of the number of graphs in $\mathcal{G}_n$ with a prescribed degree distribution.

Example 1. *There are* 4 *non-isomorphic graphs with* 3 *nodes, where* g_2 *and* g_3 *are repeated three times for different labeling:*

Figure 1: All non-isomorphic graphs with 3 nodes and the number of times they are repeated due to labeling.

Now (4) implies

$$\psi(\alpha) = \log(1 + 3\tilde{p}_1^2 + 3\tilde{p}_1^2\tilde{p}_0 + \tilde{p}_0^3) = \log(1 + 3e^{2\alpha_1} + 3e^{2\alpha_1+\alpha_0} + e^{3\alpha_0}). \quad (5)$$

This can be verified by considering (2), which implies that $P(g_1) = \phi(\tilde{p})/(1 + \tilde{p}_0 + \tilde{p}_1)^3$*,* $P(g_2) = \tilde{p}_1^2 \cdot \phi(\tilde{p})/(1 + \tilde{p}_0 + \tilde{p}_1)^3$*,* $P(g_3) = \tilde{p}_1^2 \cdot \tilde{p}_0 \cdot \phi(\tilde{p})/(1 + \tilde{p}_0 + \tilde{p}_1)^3$*, and* $P(g_4) = \tilde{p}_0^3 \cdot \phi(\tilde{p})/(1 + \tilde{p}_0 + \tilde{p}_1)^3$*.*

Degeneracy. Heuristically, the degree distribution model should be regarded as relatively unaffected by degeneracy issues, for at least two reasons. First, the change statistic [see 17] corresponding to adding an edge between two nodes of degrees k and k' is $\alpha_{k+1} + \alpha_{k'+1} - (\alpha_k + \alpha_{k'})$, suggesting in general a lack of significant correlation with the number of edges in the graph. Secondly, the growing (in n) number of parameters and the fact that the degree distributions are negatively correlated prevent the distribution from concentrating on very few configurations as it pushes the probability mass to spread out. See also [13] for an approach for dealing with degeneracy of this model based on the large deviation principle.

Existence of the maximum likelihood estimator. Suppose that $\{g_1, \ldots, g_m\}$, $m \geq 1$ are m iid observations of the networks with n nodes. The *log-likelihood* function can be written as

$$l(\alpha|g_1, \ldots, g_m) = (\sum_{k=0}^{n-2} \alpha_k \sum_{l=1}^{m} n_k(g_l)) - m\psi(\alpha). \quad (6)$$

We define the *average observed sufficient statistics* for the 1K model to be $\bar{n}^{(1)} = (\sum_{r=1}^{m} n^{(1)}(g_r))/m$ for m iid observations $\{g_1, \ldots, g_m\}$. It is known that, for distributions in exponential families, the MLE exists if and only if the average observed sufficient statistics lie on the interior of the model polytope; see [1] and [4].

We first explore the corresponding *model polytope* $A_{n-1} = convhull(\{n_-^{(1)}(g), g \in \mathcal{G}_n\})$ to derive a necessary and sufficient condition for the existence of the maximum likelihood estimator (MLE) of α, that is, $\{\hat{\alpha} : l(\hat{\alpha}|g_1, \ldots, g_m) = \max_{\alpha \in \mathbb{R}} l(\alpha|g_1, \ldots, g_m)\}$. We then exploit the well-studied theory of exponential family to calculate the MLE when existing.

In order to study A_{n-1}, we need the two following notations and lemmas: Denote the k-regular graphs with n nodes by R_k. In addition, by R_{kl} denote the graphs with n nodes such that $n_k(R_{kl}) = n - 1$ and $n_l(R_{kl}) = 1$.

Lemma 1. *For* $0 < k < n$*, there exists an* R_k *if and only if* kn *is even.*

Lemma 2. *Suppose that* n *and* k *are odd numbers and* $0 \leq k, l < n$*. There exists an* R_{kl} *if and only if* l *is even.*

Proof. Suppose that l is even. We know that there is a k-regular graph H with $n-1$ nodes. For this graph, consider a matching $\{(i_1, j_1), \ldots, (i_{(n-1)/2}, j_{(n-1)/2}\}$. Now H and an isolated node h provides R_{k0}. By removing the edge between i_1 and j_1 and connecting h to both i_1 and j_1, we obtain R_{k2}. By this method, we can inductively generate all R_{kl}.

If l is odd, the sum of degrees of nodes would be an odd number, which is impossible. □

Let $B_n = convhull(\{n^{(1)}(g), g \in \mathcal{G}_n\})$. In addition, for $0 \leq k < n$, let $e_k = (0,\ldots,0,n,0,\ldots,0)$, where n is the $(k+1)$st element of the vector of length n, and for $0 \leq k,l < n$ and $k \neq l$, let $e_{kl} = (0,\ldots,0,1,0,\ldots,0,n-1,0,\ldots,0)$, where $n-1$ is the $(k+1)$st element of the vector and 1 is the $(l+1)$st element of the vector of length n. We observer that $n^{(1)}(R_k) = e_k$ and $n^{(1)}(R_{kl}) = e_{kl}$. The following lemma characterizes B_n; see also

Lemma 3. *The model polytope B_n is*

(if n is even) $n\triangle^{n-1}$*, that is, the* $(n-1)$*-simplex scaled by* n*; and*

(if n is odd) *the convex hall of the set of extreme points* $\{e_l : 0 \leq k < n, l \text{ even}\} \cup \{e_{k,l} : 0 \leq k < n, 0 \leq l < n, k \text{ odd, } l \text{ even}\}$.

Proof. $\sum_{k=0}^{n-1} n_k(g) = n$ implies that $A_{n-1} \subseteq n\triangle^{n-1}$. We now need to deal with the two cases separately:

1) *n even:* By lemma 1, we know that for each $0 \leq k < n$, there exists an R_k. The corresponding vectors e_k are all the vertices of A_{n-1}; therefore, $A_{n-1} = n\triangle^{n-1}$.

2) *n odd:* For $0 \leq l < n$ and l even, by lemma 1, there exist R_l, and the corresponding e_l are the extreme points. For $0 \leq k,l < n$, odd k and even l, by Lemma 2, there exist R_{kl}, and it is easy to see that the corresponding e_{kl} cannot be generated by the convex combination of other $n^{(1)}(g)$. Therefore, these are the extreme points too. Now only the vectors with integer entries fewer then n that lie on $n\triangle^{n-1}$ but not on the convex combination of extreme points contain an element n as an entry, which, again by Lemma 1, it is not possible. ☐

Let $\mathbf{0}_i$ be the vector of size i consisting only of zero elements. We can now characterize A_{n-1}:

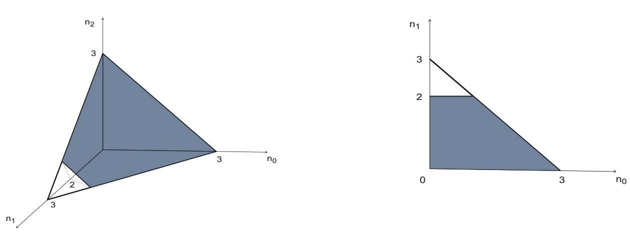
Proposition 1. *The model polytope A_{n-1} is*

(if n is even) *the convex hall of the set of extreme points* $\{e_k : 0 \leq k < n-1\} \cup \{\mathbf{0}_{n-2}\}$*; and*

(if n is odd) *the convex hall of the set of extreme points* $\{e_l : 0 \leq l < n-1, l \text{ even}\} \cup \{e_{k,l} : 0 \leq k < n-1, 0 \leq l < n-1, k \text{ odd, } l \text{ even}\} \cup \{\mathbf{0}_{n-2}\}$.

Proof. By projecting the polytope B_n, given in Lemma 3, onto the $(n-1)$-dimensional Euclidean space with coordinates $(n_0,\ldots n_{n-2})$, we obtain the result. ☐

In Figure 2, the polytopes B_3 and A_3, for graphs with three nodes, are depicted.



(a) (b)

Figure 2: (a) The polytope B_3. (b) The polytope A_3.

Therefore, we have the following theorem:

Theorem 1. *For the* $1K$ *model and* m *iid observations* $\{g_1, \dots, g_m\}$*, the MLE exists if and only if*

(if n is even) $\bar{n}_k^{(1)} \neq 0$ *for all* k*,* $0 \leq k \leq n-2$*, and* $\sum_{k=0}^{n-2} \bar{n}_k^{(1)} < n$*; and*

(if n is odd) *(i)* $\bar{n}_r^{(1)} \neq 0$ *for all* r*,* $0 \leq r \leq n-2$*, and* $\sum_{r=0}^{n-2} \bar{n}_r^{(1)} < n$*; and (ii) for every odd* k*,* $\bar{n}_k^{(1)} < n-1$*.*

Proof. The MLE exists if and only if $\bar{n}^{(1)} \in int(A_{n-1})$.

In the case that n is even, a point in $int(A_{n-1})$ is written as $a_0 e_0 + a_1 e_1 + \cdots + a_{n-2} e_{n-2} + c\mathbf{0}_{n-2}$, where $a_k > 0$ for $0 \leq k < n-1$, $c > 0$, and $\sum_{k=0}^{n-2} a_k + c = 1$. Therefore, $\bar{n}^{(1)} \in int(A_{n-1})$ if and only if $a_k = \bar{n}_k^{(1)}$ and $c = n - \sum_{k=0}^{n-2} \bar{n}_k^{(1)}$, which implies the result.

In the case that n is odd, a point in $int(A_{n-1})$ is written as $\sum_{\substack{l=0 \\ l \text{ even}}}^{n-3} a_l e_l + \sum_{\substack{k=1 \\ k \text{ odd}}}^{n-2} \sum_{\substack{l=0 \\ l \text{ even}}}^{n-3} b_{k,l} e_{k,l} + c\mathbf{0}_{n-2}$, where $a_l > 0$ for even l, $0 \leq l < n-1$, $b_{k,l} > 0$ for odd k and even l, $0 \leq k, l < n-1$, $c > 0$, and $\sum_{\substack{l=0 \\ l \text{ even}}}^{n-3} a_l + \sum_{\substack{l=0 \\ l \text{ even}}}^{n-3} \sum_{\substack{k=1 \\ k \text{ odd}}}^{n-2} b_{k,l} + c = 1$.

If $\bar{n}^{(1)} \in int(A_{n-1})$, for every odd k, $\sum_{l=0}^{n-3} b_{k,l} = n_k/(n-1)$, which implies $n_k > 0$. In addition, we know that $\sum_{l=0}^{n-3} b_{k,l} < 1$, which implies that $n_k < n-1$. For every even l, $a_l = (1/n)(n_l - \sum_{k=1}^{n-2} b_{k,l})$, which implies $n_l > 0$. Now by using $\sum_{l=0}^{n-3} a_l + \sum_{l=0}^{n-3} \sum_{k=1}^{n-2} b_{k,l} < 1$, we obtain $\sum_{r=0}^{n-2} n_r < n$.

Conversely, if conditions (i) and (ii) hold, we let $b_{k,l} = (2n_k)/(n-1)^2$ and $a_l = (1/n)(n_l - (2/(n-1)^2 \sum_{k=1}^{n-2} n_k)$. We conclude that $a_l > 0$, $b_{k,l} > 0$, and $\sum_{l=0}^{n-3} a_l + \sum_{l=0}^{n-3} \sum_{k=1}^{n-2} b_{k,l} < 1$, which imply the result. □

Corollary 1. *For a single observation in the* $1K$ *model, the MLE does not exist.*

Proof. For an observed g, if $n_{n-1}(g) \neq 0$, then $n_0(g) = 0$ and vice vera. $n_{n-1}(g) = 0$ implies $\sum_{k=0}^{n-2} \bar{n}_k = n$. Therefore, in either case the MLE does not exist. □

Example 2. *We proceed with Example 1 for networks with* 3 *nodes, illustrated in Figure 1. Theorem 1 implies that in order for the MLE to exist (i)* $\bar{n}_0 \neq 0$*,* $\bar{n}_1 \neq 0$*, and* $\bar{n}_1 + \bar{n}_0 < 3$*; (ii) one should observe either* g_1 *or* g_4 *since both* g_2 *and* g_3 *are of form* R_{kl}*. If only* (i) *or* (ii) *holds, then only one parameter is estimable.*

Estimable parameters. By partially differentiating the log-likelihood function in (6), we conclude that the MLE $\hat{\alpha}$ should satisfy the following system of equations for $k \in \{0. \dots, n-2\}$:

$$\frac{\partial \psi(\hat{\alpha})}{\partial \hat{\alpha}_k} = \bar{n}_k. \tag{7}$$

Therefore, from (5), the MLE for 1K model with m observations must satisfy the following system of equations for $k \in \{0, \dots, n-2\}$:

$$\frac{\sum_{i=1}^{M} n_k(g_i) e^{\sum_{l=0}^{n-2} n_l(g_i)\hat{\alpha}_l}}{\sum_{i=1}^{M} e^{\sum_{l=0}^{n-2} n_l(g_i)\hat{\alpha}_l}} = \bar{n}_k. \tag{8}$$

Notice that if for a fixed k, $\bar{n}_k = 0$ then $\sum_{i=1}^{M} n_k(g_i) e^{\sum_{l=0}^{n-2} n_l(g_i)\hat{\alpha}_l} = 0$. We observe that $\hat{\alpha}_k$ is the only parameter that appears in all terms and does not appear in all terms of the denominator of the left hand side of (8). Therefore, $\hat{\alpha}_k = -\infty$. This corresponds to $\hat{\tilde{p}}_k = 0$, or equivalently $\hat{p}_k = 0$ since it is that assumed $p_{n-1} \neq 0$.

Thus, based on the MLE, the model implies that the probability of a node having a specific degree is zero when no node with that degree has been observed. Hence, it is plausible to remove such parameters from the model and simply focus on the submodel of exponential family form. The corresponding model polytope would then be simply the same, embedded in the Euclidian space with remaining coordinates. By this method we can still estimate a subset of parameters whose corresponding sufficient statistic is nonzero.

Notice that even with a single observation, there are sometimes a considerable number of parameters that can be estimated. The following proposition deals with the extreme case of such graphs.

Proposition 2. *There exists a graph T_n of every size n such that $n_0(T_n) = 0$ and $n_k(T_n) \neq 0$ for $k \in \{1, \dots, n-1\}$.*

Proof. We prove the result by induction on the number of nodes. For the base, where $n = 2$, the result holds for K_2. Suppose that there exists such a graph T_n. We prove it for $n+1$: Since $n_0(T_n) = 0$ there is a $k' \in \{1, \dots, n-1\}$ such that $n_{k'} = 2$ and $n_k = 1$ if $k \neq k'$. We construct the graph T_{n+1} as follows. We add a node to T_n and start connecting it to the nodes starting from the node with the node with degree $n-1$ and stop when we reach k'. Now regardless of what degree the added node has, there are nodes of all degrees except zero in graph. □

Notice, however, that the large sum in (8) entails the common problem with ERG modeling that the computing of the MLE for this model ultimately requires MCMC methods, just like with most other ERGMs.

Asymptotics. We have shown that we can estimate a subset of parameters whose corresponding sufficient statistic is nonzero. Here we calculate the probability of a sufficient statistic to be nonzero, and discuss its behaviour asymptotically. In addition, we provide the asymptotical expected value of the number of non-zero sufficient statistics.

Under the 1K-model, for each $k \leq n-1$, we have the following:

$$\mathbb{P}(n_k(G) > 0) = \frac{\sum_{g \in \mathcal{G}_n : n_k(g) \neq 0} \prod_{k'=0}^{n-1} p_{k'}^{n_{k'}(g)}}{\sum_{g \in \mathcal{G}_n} \prod_{k'=0}^{n-1} p_{k'}^{n_{k'}(g)}}. \tag{9}$$

We will now consider the very special case where all $p_{k'}$ are equal, which implies that they are all equal to $1/n$. In this case we see that the model is the same as the Erdös-Rényi model with $p = 1/2$.

Proposition 3. *Suppose that $p_i = p$ for all i and let $k = k(n) = (n-1)/2 + a_n$, where $\{a_n\}$ is a sequence such that $0 \leq a_n \leq (n-1)/2$ for all n. It then holds that*

1. $\lim_{n\to\infty} \mathbb{P}(n_k(g) > 0) = 1$ *if* $a_n = O(\sqrt{n})$;
2. $\lim_{n\to\infty} \mathbb{P}(n_k(g) > 0) = 0$, *if* $a_n = \Omega(\sqrt{n \log n})$.

Proof. By using (9), we have that $P(n_k(g) > 0) = \frac{\sum_{g \in \mathcal{G}_n : n_k(g) \neq 0} \prod_{k'=0}^{n-1} p^n}{\sum_{g \in \mathcal{G}_n} \prod_{k'=0}^{n-1} p^n} = \frac{|\{g \in \mathcal{G}_n : n_k(g) \neq 0\}|}{|\mathcal{G}_n|}$, which is the same as the same probability in the Erdös-Rényi model with $p = 1/2$. Let $\lambda_k(n) = n\binom{n-1}{k}(1/2)^{n-1}$. From the theorem 3.1 in [3], we know that if $\lim_{n\to\infty} \lambda_k(n) = 0$, then $\lim_{n\to\infty} P(n_k(g) > 0) = 0$, and if $\lim_{n\to\infty} \lambda_k(n) = \infty$, then $\lim_{n\to\infty} P(n_k(g) > 0) = 1$. We show that the latter holds for $a_n = O(\sqrt{n})$, and the former if $a_n = \Omega(\sqrt{n \log n})$. By using Stirling's approximation,

$$\lambda_k(n) \sim \frac{n(n-1)^{n-\frac{1}{2}}}{(n-1-k)^{n-k-\frac{1}{2}}2^{n-\frac{1}{2}}k^{k+\frac{1}{2}}\sqrt{\pi}} = \left(\frac{n-1}{2}\right)^{n-\frac{1}{2}} \frac{n}{(\frac{n-1}{2}-a_n(k))^{\frac{n}{2}-a_n(k)}(\frac{n-1}{2}+a_n(k))^{\frac{n}{2}+a_n(k)}\sqrt{\pi}} \asymp \frac{n^{n+1/2}}{2^n}\frac{1}{\left(\frac{n}{2}\right)^n\left(1-\frac{2a_n(k)}{n}\right)^{\frac{n}{2}-a_n(k)}\left(1+\frac{2a_n(k)}{n}\right)^{\frac{n}{2}+a_n(k)}} = \frac{\sqrt{n}}{\left(1-\frac{2a_n(k)}{n}\right)^{\frac{n}{2}-a_n(k)}\left(1+\frac{2a_n(k)}{n}\right)^{\frac{n}{2}+a_n(k)}} = \frac{\sqrt{n}}{\left(1-\frac{4a_n^2}{n}\frac{1}{n}\right)^{\frac{n}{2}}\left(1-2\frac{a_n^2}{n}\frac{1}{a_n}\right)^{-a_n}\left(1+2\frac{a_n^2}{n}\frac{1}{a_n}\right)^{a_n}}.$$

If $a_n = O(\sqrt{n})$ the previous display is $\asymp \sqrt{n}$ and hence tends to infinity. If $a_n = \Omega(\sqrt{n \log n})$ the term will vanish instead. The claim is proved. $\square$

We now consider the number N of non-zero entries in the degree distribution and show that it is of order $O\left(\sqrt{n \log n}\right)$, which corresponds to the number of estimable parameters in the model.

Proposition 4. *Suppose that $p_i = p$ for all i and let N denote the number of entries in the degree distribution that are non-zero. Then,*

$$\mathbb{E}[N] = O\left(\sqrt{n \log n}\right).$$

Proof. Let $k_n = \sqrt{cn \log n}$ for some $c > 0$ which will be set below. Assume that $(n-1)p > k_n$ and let $\mathcal{A}$ denote the event that there exists a node with degree less than $(n-1)p - k_n$ or more than $(n-1)p + k_n$. Let D_i denote the degree of node i and notice that $D_i \sim \mathrm{Bin}(n-1, p)$, for all i. Then $\mathcal{A} = \bigcup_i \{|D_i - \mathbb{E}[D_i]| > k_n\}$.

By the union bound and Hoeffding inequality we get that, for any $r > 0$, $\mathbb{P}(\mathcal{A}) \leq n^2 \exp\{-2c \log n\} = \frac{2}{n^r}$, provided that $c = r/2$. It is now easy to see that the expected number of non-zero entries in the degree distribution is of order $O(\sqrt{n \log n})$. In details, $\mathbb{E}[N] = \mathbb{E}[N|\mathcal{A}]\mathbb{P}(\mathcal{A}) + \mathbb{E}[N|\mathcal{A}^c]\mathbb{P}(\mathcal{A}^c) \leq \frac{n}{n^r} + \left(1 - \frac{2}{n^r}\right)\sqrt{n \log n} = O(\sqrt{n \log n})$. $\square$

We have seen that the case $p_i = p$ for all i and some $p \in (0, 1)$ corresponds to the Erdös-Rényi model with $p = 1/2$. Using (3), this is equivalent to having $\alpha_k = 0$, for all $k = 0, 1, \ldots, n-2$. We now consider the slightly more general case in which the vector $\alpha = (\alpha_0, \alpha_1, \ldots, \alpha_{n-2})$ belongs to a subset A of $\mathbb{R}^{n-1}$ such that $\log(c_n) \leq \alpha_i \leq \log(C_n)$, where $c_n \in (0, 1]$ and $C_n \in [1, \infty)$, for all i. Then it is easy to see that, for any $\alpha \in A$,

$$\mathbb{P}_\alpha(n_k(G) > 0) \leq \left(\frac{C_n}{c_n}\right)^n \mathbb{P}_0(n_k(G) > 0) \leq \left(\frac{C_n}{c_n}\right)^n \lambda_k(n), \tag{10}$$

where $\lambda_k(n)$ is given in the proof or proposition 3 and $\mathbb{P}_\alpha$ denotes a probability of the random graph G when sampled from the degree distribution model with parameter α. Next, for a sequence $\{a_n\}_{n=1,2,\ldots}$ such that $0 \leq a_n \leq \frac{n}{2}$, let $h_n = \left(1 - \frac{2a_n}{n}\right)^{\frac{1}{2} - \frac{a_n}{n}} \left(1 + \frac{2a_n}{n}\right)^{\frac{1}{2} + \frac{a_n}{n}}$. Then, by proposition 3, (10) yields that, for $k = k(n) = (n-1)/2 + a_n$, $\mathbb{P}_\alpha(n_k(G) > 0)$ is bounded by a term that is asymptotically of order $\left(\frac{C_n}{c_n}\right)^n \frac{\sqrt{n}}{(h_n)^n}$.

Now let $b = \mathrm{liminf}_n h_n$. If $a_n = O(\sqrt{n})$, then $(h_n)^n = O(1)$ from the proof of proposition 3, which yields a trivial bound. Thus assume that $b \in (1, 2]$ since $h_n \in [1, 2]$ for all n. Then, provided that $\frac{C_n}{c_n} < b$, the probability $\mathbb{P}_\alpha(n_k(G) > 0)$ vanishes. (Notice that this implies that $1 < C_n/c_n < 2$.)

Relation with the Erdös-Rényi model. Recently and somewhat surprisingly, [5] have shown that some specifications of the ERGM family lead to models which asymptotically behave like Erdös-Rényi (ER) models for appropriate choices of p.

Here we provide some results illustrating the relationships between the ER model and the degree distribution model. Below we parameterize the degree distribution model using the natural parameter vector $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^{n-1}$. Notice that we have expunged n_0 from the vector of sufficient statistics, a choice that entails no loss of generality.

First off, we make the easy observation that the ER model with probability p can be represented by setting $\alpha_i = i\theta, \quad i = 1, \ldots, n-1$, where $\theta = \log\left(\frac{p}{1-p}\right)$. Next, we show that the degree distribution model is dramatically different from the Erdös-Rényi model in the sense of being almost singular to it whenever the natural parameters are uniformly bounded in absolute value.

Below, we will denote with P_p the probability distribution of the Erdös-Rényi model with parameter $p \in (0,1)$ and with P_α the probability distribution of the degree distribution model with natural parameter vector $\alpha = (\alpha_1, \ldots, \alpha_{n-1}) \in \mathbb{R}^{n-1}$.

Proposition 5. *Let $p \in (0,1)$ such that $\min\{|p - 1/2|, 1-p, p\} > \epsilon$, for a fixed, arbitrarily small $\epsilon > 0$. Then, there exists a sequence of subsets $\mathcal{G}_n(p)$ of $\mathcal{G}_n$ such that, for any sequence $\{\alpha_n\}$ such that $\alpha_n \in \mathbb{R}^{n-1}$ for all n and $\|\alpha_n\|_\infty = o(n)$,*

$$\lim_n P_p(\mathcal{G}_n(p)) = 1 \quad \textit{and} \quad \lim_n P_{\alpha_n}(\mathcal{G}_n(p)) = 0.$$

Proof. Let $\mathcal{G}_n(p)$ be the set of graphs such that $(n-1)p - C\sqrt{n \log n} \leq d_i(g) \leq (n-1)p + C\sqrt{n \log n}$, for all $i = 1, \ldots, n$, where $d_i(g)$ denote the degree of the ith node of graph g and C is a positive constant to be specified.

By Heoffding's inequality, for any $c > 0$, there exists a $C = C(c)$ such that $P_p(\mathcal{G}_n(p)) \geq 1 - \frac{1}{n^c}$, for n large enough that $C\sqrt{\frac{\log n}{n}} < \epsilon$. Next, the probability $P_\alpha(\mathcal{G}_p) = \frac{\sum_{g \in \mathcal{G}_n(p)} e^{\sum_{i=1}^{n-1} \alpha_i n_i(g)}}{\sum_{g \in \mathcal{G}_n} e^{\sum_{i=1}^{n-1} \alpha_i n_i(g)}}$ can be bounded from above by $\frac{e^{Mn}|\mathcal{G}_n(p)|}{e^{-Mn}|\mathcal{G}_n|} = e^{2Mn}\mathbb{P}_{1/2}(\mathcal{G}_n(p))$, a calculation that follows from simple algebra and the fact that, for each $g \in \mathcal{G}_n$, $\sum_{i=1}^{n-1} n_i(g) = n - n_0(g) \leq n$.

Then, since p is ϵ-away from $1/2$, Theorem 2.3 in [7] (see also [5]) yields that, for all n large enough, $\mathbb{P}_{1/2}(\mathcal{G}_n(p)) \leq e^{-n^2 c}$, for an appropriate constant c, which depends on p. Thus, we have shown that, for all n large enough and for each α in an L_∞ ball of 0 in $\mathbb{R}^{n-1}$ of radius M, $P_\alpha(\mathcal{G}_n(p)) \leq e^{2Mn - cn^2}$, which vanishes provided $M = o(n)$. $\square$

The case of $p = 1/2$ was not covered by the previous result and one might wonder whether setting all the natural parameters to be equal to a constant vector i.e. $\alpha_i = \theta$ for all i, might yield a model close to an ER mode with $p = 1/2$. This in fact not the case, as we demonstrate next, unless $\theta = 0$.

Towards this end, notice first that the assumption that $\alpha_i = \theta$, for all $i \geq 1$, is equivalent to $\log\left(\frac{p_k}{p_0}\right) = \theta$. Thus, for any $\theta \in \mathbb{R}$, the probability of observing a graph g is $\frac{e^{-\theta n_0(g)}}{\sum_{j=0}^{n} e^{-\theta j}\nu_n(j)}$, where $\nu_n(j)$ is the number of n-graphs with j isolated nodes. It is clear that the above probability is rather different from the ER model with $p = 1/2$. To add more details, it is possible to show that $\nu_n(0) \equiv f(n) = \sum_{k=0}^{n} (-1)^{n-2}\binom{n}{k} 2^{\binom{n}{k}}$, and $\nu_n(j) = \binom{n}{j} f(n-j)$. This can be obtained as the solution to the recursion $f(n) = 2^{\binom{n}{2}} - \sum_{i=0}^{n-1} f(i)\binom{n}{i}$, with the initial conditions $f(0) = 1$ and $f(1) = 0$.

The term $\nu_n(0)$ dominates all the others, in the sense that $\frac{f(n)}{2^{\binom{n}{2}}} \to 1$, as $n \to \infty$. Thus, when $\theta < 0$ the model will favor networks with many isolated nodes.

References

[1] O. Bardorff-Nielsen. *Information and Exponential Families in Statistical Theory*. John Wiley and Sons, New York, NY, USA, 1978.

[2] Joseph Blitzstein and Persi Diaconis. A sequential importance sampling algorithm for generating random graphs with prescribed degrees. *Internet Math.*, 6(4):489–522, 2010.

[3] B. Bollobas. *Random Graphs*. Cambridge University Press, Cambridge, UK, 2001.

[4] L. Brown. *Fundamentals of Statistical Exponential Families*, volume 9. IMS Lecture NotesMonograph Series, Hayward, CA, USA, 1986.

[5] Sourav Chatterjee and Persi Diaconis. Estimating and understanding exponential random graph models. *submitted*, 2013. http://arxiv.org/abs/1102.2650.

[6] Sourav Chatterjee, Persi Diaconis, and Allan Sly. Random graphs with a given degree sequence. *Annals of Applied Probability*, 21(4):1400–1435, 2011.

[7] Sourav Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdös-Rényi random graph. *Eur. J. Comb.*, 32(7):1000–1017, 2011.

[8] P. Erdös and A. Rényi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.

[9] Ove Frank and David Strauss. Markov graphs. *Journal of the American Statistical Association*, 81:832–842, 1986.

[10] M. S. Handcock and M. Morris. A simple model for complex networks with arbitrary degree distribution and clustering. *Statistical Network Analysis: Models, Issues and New Directions*, 4503:103–114, 2007.

[11] Paul Holland and Samuel Leinhardt. An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, 76:33–50, 1981.

[12] B.D. McKay. Asymptotics for symmetric 0 1 matrices with prescribed row sums. *Ars Combinatoria*, 19(A):15–25, 1985.

[13] Sumit Mukherjee. Large deviation for the empirical degree distribution of an erdös-rényi graph. *submitted*, 2013.

[14] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[15] Alessandro Rinaldo, Sonja Petrović, and Stephen E. Fienberg. Maximum lilkelihood estimation in the beta model. *Annals of Statistics*, 41(3):1085–1110, 2013.

[16] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph models for social networks. *Social Networks*, 29:173–191, 2007.

[17] Tom A. B. Snijders, Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153, 2006.