

MAXIMUM LIKELIHOOD ESTIMATION IN THE β -MODEL

BY ALESSANDRO RINALDO SONJA PETROVIĆ STEPHEN E. FIENBERG

We study maximum likelihood estimation for the statistical model for undirected random graphs, known as the β -model, in which the degree sequences are minimal sufficient statistics. We derive necessary and sufficient conditions, based on the polytope of degree sequences, for the existence of the maximum likelihood estimator (MLE) of the model parameters. We characterize in a combinatorial fashion sample points leading to a nonexistent MLE, and non-estimability of the probability parameters under a nonexistent MLE. We formulate conditions that guarantee that the MLE exists with probability tending to one as the number of nodes increases.

1. Introduction. Many statistical models for the representation and analysis of network data rely on information contained in the *degree sequence*, the vector of node degrees of the observed graph. Node degrees not only quantify the overall connectivity of the network, but also reveal other potentially more refined features of interest. The study of the degree sequences and, in particular, of the degree distributions of real networks is a classic topic in network analysis, which has received extensive treatment in the statistical literature (see, e.g., [Fienberg and Wasserman, 1981a](#); [Fienberg et al., 1985](#); [Holland and Leinhardt, 1981](#)), the physics literature (see, e.g., [Albert and Barabási, 2002](#); [Foster et al., 2007](#); [Newman et al., 2006](#); [Newman, 2003](#); [Newman et al., 2001](#); [Park and Newman, 2004](#); [Willinger et al., 2009](#)) as well as in the social network literature (see, e.g., [Goodreau, 2007](#); [Handcock and Morris, 2007](#); [Robins et al., 2008](#), and references therein). See also the monograph by [Goldenberg et al. \(2010\)](#) and the books by [Kolaczyk \(2009\)](#), [Cohen and Havlin \(2010\)](#) and [Newman \(2010\)](#).

The simplest instance of a statistical network model based exclusively on the node degrees is the exponential family of probability distributions for undirected random graphs with the degree sequence as its natural sufficient statistic. This is in fact a simpler, undirected version of the broader class of statistical models for directed networks known as the p_1 -models, introduced by [Holland and Leinhardt \(1981\)](#). We will refer to this model as the *beta* model (henceforth the β -model), a name recently coined by [Chatterjee et al.](#)

AMS 2000 subject classifications: Primary 62F99; secondary 62F99

Keywords and phrases: β -model, polytope of degree sequences, random graphs, maximum likelihood estimator

(2011), and refer to [Blitzstein and Diaconis \(2010\)](#) for details and extensive references.

Despite its apparent simplicity and popularity, the β -model, much like most network models, exhibits non-standard statistical features, since its complexity, measured by the dimension of the parameter space, increases with the size of the graph. [Lauritzen \(2003, 2008\)](#) characterized β -models as the natural models for representing exchangeable binary arrays that are *weakly summarized*, i.e., random arrays whose distribution only depends on the row and column totals. More recently, [Chatterjee et al. \(2011\)](#) conducted an analysis of the asymptotic properties of the β -model, including existence and consistency of the maximum likelihood estimator (MLE) as the dimension of the network increases, and provided a simple algorithm for estimating the natural parameters. They also characterized the graph limits, or *graphons* (see [Lovász and Szegedy, 2006](#)), corresponding to a sequence of β -models with given degree sequences (for a connection between the theory of graphons and exchangeable arrays see [Diaconis and Janson, 2007](#)). Concurrently, [Barvinok and Hartigan \(2010\)](#) explored the asymptotic behavior of sequences of random graphs with given degree sequences, and studied a different mode of stochastic convergence. Among other things, they show that, as the size of the network increases and under a “tameness” condition, the number of edges of a uniform graph with given degree sequence converges in probability to the number of edges of a random graph drawn from a β -model parametrized by the MLE corresponding to degree sequence. [Yan and Xu \(2012\)](#) and [Yan et al. \(2012\)](#) derived asymptotic conditions for uniform consistency and asymptotic normality of the MLE of the β -model, and asymptotic normality of the likelihood ratio test for homogeneity of the model parameters. [Perry and Wolfe \(2012\)](#) consider a general class of models for network data parametrized by node-specific parameters, of which the β -model is a special case. The authors derive non-asymptotic conditions under which the MLEs of model parameters exist and can be well approximated by simple estimators.

In an attempt to avoid the reliance on asymptotic methods, whose applicability to network models remains largely unclear (see, e.g., [Haberman, 1981](#)), several researchers have turned to exact inference for the β -model, which hinges upon the non-trivial task of sampling from the set of graphs with a given degree sequence. [Blitzstein and Diaconis \(2010\)](#) developed and analyzed a sequential importance sampling algorithm for generating a random graph with the prescribed degree sequence (see also [Viger and Latapay, 2005](#), for a different algorithm). [Hara and Takemura \(2010\)](#) and [Ogawa et al. \(2011\)](#) tackled the same task using more abstract algebraic methods and

Petrović et al. (2010) studied Markov bases for the more general p_1 model.

In this article we study the existence of the MLE for the parameters of the β -model under a more general sampling scheme in which each edge is observed a fixed number of times (instead of just once, as in previous works) and for increasing network sizes. We view the issue of existence of the MLE as a natural measure of the intrinsic statistical difficulty of the β -model for two reasons. First, existence of the MLE is a natural minimum requirement for feasibility of statistical inference in discrete exponential families, such as the β -model: nonexistence of the MLE is in fact equivalent to non-estimability of the model parameters, as illustrated in Fienberg and Rinaldo (2012). Thus, establishing conditions for existence of the MLE amounts to specifying the conditions under which statistical inference for these models is fully possible. Secondly, under the asymptotic scenario of growing network sizes, existence of the MLE will provide a natural measure of sample complexity of the β -model, and will indicate the asymptotic scaling of the model parameters for which statistical inference is viable.

Though Chatterjee et al. (2011) and Barvinok and Hartigan (2010)¹ also considered the existence of the MLE, our analysis differs substantially from theirs in that it is rooted in the statistical theory of discrete linear exponential families and relies in a fundamental way on the geometric properties of these families (see, in particular, Geyer, 2009; Rinaldo et al., 2009). Our contributions are as follows.

- We provide explicit necessary and sufficient conditions for existence of the MLE for the β -model that are based on the polytope of degree sequences, a well-studied polytope arising in the study of threshold graphs (see Mahadev and Peled, 1996). In contrast, the conditions of Chatterjee et al. (2011) are only sufficient. We then show that nonexistence of the MLE is brought on by certain forbidden patterns of extremal network configurations, which we characterize in a combinatorial way. Furthermore, when the MLE does not exist, we can identify exactly which probability parameters are estimable.
- We use the properties of the polytope of degree sequences to formulate geometric conditions that allow us to derive finite sample bounds on the probability that the MLE does not exist. Our asymptotic results improve analogous results of Chatterjee et al. (2011) and our proof is both simpler and more direct. Furthermore, we show that the tame-

¹In the analysis of Barvinok and Hartigan (2010), the *maximum entropy matrix* associated to a degree sequence is in fact exactly the MLE corresponding to the observed degree sequence. This is a well-known property of linear exponential families: see, e.g., Cover and Thomas (1991, Chapter 11).

ness condition of [Barvinok and Hartigan \(2010\)](#) is stronger than our conditions for existence of the MLE.

- Our analysis is not specific to the β -model but, in fact, follows a principled way for detecting nonexistence of the MLE and identifying non-estimable parameters that is based on polyhedral geometry and applies more generally to discrete models. We illustrate this point by analyzing other network models that are variations or generalizations of the β -model: the β -model with random numbers of edges, the Rasch model, the Bradley-Terry model and the p_1 model. Due to space limitations, the details of these additional analyses are contained in the supplementary material.

While this is a self-contained article, the results derived here are best understood as applications of the geometric and combinatorial properties of log-linear models under product-multinomial sampling schemes, as detailed in [Fienberg and Rinaldo \(2012\)](#) and its supplementary material, to which we refer the reader for further details as well as for practical algorithms.

The article is organized in the following way. Section 2 introduces the β -model and establishes the exponential family parametrization that is key to our analysis. In Section 3 we derive necessary and sufficient conditions for existence of the MLE of the β -model parameters and characterize parameter estimability under a nonexistent MLE. These results are further discussed with examples in Section 4. In Section 5 we provide sufficient conditions on the expected degree sequence guaranteeing that, with high probability as the network size increases, the MLE exists. Finally, in Section 6 we indicate possible extensions of our work and briefly discuss some of the computational issues directly related to detecting nonexistence of the MLE and parameter estimability.

We will assume throughout some familiarity with basic concepts from polyhedral geometry (see, e.g., [Schrijver, 1998](#)) and the theory of exponential families (see, e.g., [Barndorff-Nielsen, 1978](#); [Brown, 1986](#)).

2. The (Generalized) β -Model. In this section we describe the exponential family parametrization of a simple generalization of the β -model, which, with slight abuse of notation, we will refer to as the β -model as well.

We are concerned with modeling the occurrence of edges in a simple undirected random graph with node set $\{1, \dots, n\}$. The statistical experiment consists of recording, for each pair of nodes (i, j) with $i < j$, the number of edges appearing in $N_{i,j}$ i.i.d. samples, where the integers $\{N_{i,j}, i < j\}$ are deterministic and positive (we can relax both the non-randomness and positivity assumptions). Thus, in our setting we allow for the possibility that

each edge in the network be sampled a different number of times, a realistic feature that makes the model more flexible. For $i < j$, we denote by $x_{i,j}$, the number of times we observe the edge (i, j) and, accordingly, by $x_{j,i}$ the number of times edge (i, j) is missing. Thus, for all (i, j) ,

$$x_{i,j} + x_{j,i} = N_{i,j}.$$

We model the observed edge counts $\{x_{i,j}, i < j\}$ as draws from mutually independent binomial distributions, with $x_{i,j} \sim \text{Bin}(N_{i,j}, p_{i,j})$, where $p_{i,j} \in (0, 1)$ for each $i < j$.

Data arising from such an experiment has a representation in the form of a $n \times n$ contingency table with empty diagonal cells and whose (i, j) -th cell contains the count $x_{i,j}$, $i \neq j$. For modeling purposes, however, we need only consider the upper-triangular part of this table. Indeed, since, given $x_{i,j}$, the value of $x_{j,i}$ is determined by $N_{i,j} - x_{i,j}$, we can represent the sample space more parsimoniously as the following subset of $\mathbb{N}^{\binom{n}{2}}$:

$$\mathcal{S}_n := \{x_{i,j} : i < j \text{ and } x_{i,j} \in \{0, 1, \dots, N_{i,j}\}\}.$$

We index the coordinates $\{(i, j) : i < j\}$ of any point in \mathcal{S}_n lexicographically.

In the β -model, we parametrize the $\binom{n}{2}$ edge probabilities by points $\beta \in \mathbb{R}^n$ as follows. For each $\beta \in \mathbb{R}^n$, the probability parameters are uniquely determined as

$$(1) \quad p_{i,j} = \frac{e^{\beta_i + \beta_j}}{1 + e^{\beta_i + \beta_j}} \quad \text{and} \quad p_{j,i} = 1 - p_{i,j} = \frac{1}{1 + e^{\beta_i + \beta_j}}, \quad \forall i \neq j,$$

or, equivalently, in terms of log-odds,

$$(2) \quad \log \frac{p_{i,j}}{1 - p_{i,j}} = \beta_i + \beta_j, \quad \forall i \neq j.$$

The magnitude and sign of β_i quantifies the propensity of node i to have ties: the degree of node i is expected to be large (small) if β_i is positive (negative) and of large magnitude. Thus the β -model is the natural heterogeneous version of the well-known Erdős-Rényi random graph model ([Erdős and Rényi, 1959](#)). For a discussion of this model and its generalizations see [Goldenberg et al. \(2010\)](#).

For a given choice of β , the probability of observing the vector of edge counts $x \in \mathcal{S}_n$ is

$$(3) \quad p_\beta(x) = \prod_{i < j} \binom{N_{i,j}}{x_{i,j}} p_{i,j}^{x_{i,j}} (1 - p_{i,j})^{N_{i,j} - x_{i,j}},$$

with the probability values $p_{i,j}$ satisfying (1). Simple algebra allows us to rewrite this expression in exponential family form as

$$(4) \quad p_\beta(x) = \exp \left\{ \sum_{i=1}^n d_i \beta_i - \psi(\beta) \right\} \prod_{i < j} \binom{N_{i,j}}{x_{i,j}},$$

where the coordinates of the vector of the minimal sufficient statistics $d = d(x) \in \mathbb{N}^n$ are

$$(5) \quad d_i = \sum_{j < i} x_{j,i} + \sum_{j > i} x_{i,j}, \quad i = 1, \dots, n,$$

and the log-partition function is $\psi(\beta) = \sum_{i < j} N_{i,j} \log(1 + e^{\beta_i + \beta_j})$.

Note that $e^{\psi(\beta)} < \infty$ for all $\beta \in \mathbb{R}^n$, so \mathbb{R}^n is the natural parameter space of the full and steep exponential family with support \mathcal{S}_n (see, e.g. [Barndorff-Nielsen, 1978](#)) and densities given by the exponential term in (4).

Random graphs with fixed degree sequence. When $N_{i,j} = 1$ for all (i, j) , the support \mathcal{S}_n reduces to the set $\mathcal{G}_n := \{0, 1\}^{\binom{n}{2}}$, which encodes all undirected simple graphs on n nodes: for any $x \in \mathcal{G}_n$, the corresponding graph has an edge between nodes i and j , with $i < j$, if and only if $x_{i,j} = 1$. In this case, the β -model yields a class of distributions for random undirected simple graphs on n nodes, where the edges are mutually independent Bernoulli random variables with probabilities of success $\{p_{i,j}, i < j\}$ satisfying (1). Then, by (5), the i th minimal sufficient statistic d_i is the *degree* of node i , i.e., the number of nodes adjacent to i , and the vector $d(x)$ of sufficient statistics is the *degree sequence* of the observed graph x . This is the version of the β -model studied by [Chatterjee et al. \(2011\)](#).

3. Existence of the MLE for the β -Model. We now derive a necessary and sufficient condition for the existence of the MLE of the natural parameter β or, equivalently, of the probability parameters $\{p_{i,j}, i < j\}$ as defined in (1). For a given $x \in \mathcal{S}_n$, we say that the MLE does not exist when

$$\{\beta^* : p_{\beta^*}(x) = \sup_{\beta \in \mathbb{R}^n} p_\beta(x)\} = \emptyset,$$

where $p_\beta(x)$ is given in (4). For the natural parameters, nonexistence of the MLE implies that we cannot attain the supremum of the likelihood function (4) by any finite vector in \mathbb{R}^n . For the probability parameters, nonexistence signifies that the supremum of (3) cannot be attained by any set of probability values bounded away from 0 and 1, and satisfying the equations (1).

Either way, nonexistence of the MLE implies that only a random subset of the model parameters is estimable: see [Fienberg and Rinaldo \(2012\)](#).

Our analysis on the existence of the MLE and parameter estimability for the β -model is based on a geometric object that plays a key role throughout the rest of the paper: the *polytope of degree sequences*. To this end, we note that, for each $x \in \mathcal{S}_n$, we can obtain the vector of sufficient statistics $d(x)$ for the β -model as

$$d(x) = Ax$$

where A is the $n \times \binom{n}{2}$ design matrix equal to the node-edge incidence matrix of a complete graph on n nodes. Specifically, we index the rows of A by the node labels $i \in \{1, \dots, n\}$, and the columns by the set of all pairs (i, j) with $i < j$, ordered lexicographically. The entries of A are ones along the coordinates $(i, (i, j))$ and $(j, (i, j))$ for $i < j$, and zeros otherwise. For instance, when $n = 4$

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix},$$

where we index the columns lexicographically by the pairs $(1, 2)$, $(1, 3)$, $(1, 4)$, $(2, 3)$, $(2, 4)$, and $(3, 4)$. In particular, for any undirected simple graph $x \in \mathcal{G}_n$, Ax is the associated degree sequence. The polytope of degree sequences P_n is the convex hull of all possible degree sequences, i.e.,

$$P_n := \text{convhull}(\{Ax, x \in \mathcal{G}_n\}).$$

The integral polytope P_n is a well-studied object in graph theory: e.g., see Chapter 3 in [Mahadev and Peled \(1996\)](#). In particular, when $n = 2$, P_n is just a line segment in \mathbb{R}^2 connecting the points $(0, 0)$ and $(1, 1)$, while, for all $n \geq 3$, $\dim(P_n) = n$.

We now fully characterize the existence of the MLE for the β -model using the polytope of degree sequences in the following fashion. For any $x \in \mathcal{S}_n$, let

$$\tilde{p}_{i,j} := \frac{x_{i,j}}{N_{i,j}}, \quad i < j,$$

and set $\tilde{d} = \tilde{d}(x) \in \mathbb{R}^n$ to be the vector with coordinates

$$(6) \quad \tilde{d}_i := \sum_{j < i} \tilde{p}_{j,i} + \sum_{j > i} \tilde{p}_{i,j}, \quad i = 1, \dots, n,$$

a rescaled version of the sufficient statistics (5), normalized by the number of observations. In particular, for the random graph model, $\tilde{d} = d$.

THEOREM 3.1. *Let $x \in \mathcal{S}_n$ be the observed vector of edge counts. The MLE exists if and only if $\tilde{d}(x) \in \text{int}(P_n)$.*

Theorem 3.1 verifies the conjecture contained in Addendum A in Chatterjee et al. (2011) for the random graph model: the MLE exists if and only if the degree sequence belongs to the interior of P_n . This result follows from the standard properties of exponential families: see Theorem 9.13 in Barndorff-Nielsen (1978) or Theorem 5.5 in Brown (1986). It also confirms the observation made by Chatterjee et al. (2011) that the MLE never exists if $n = 3$: indeed, since P_3 has exactly 8 vertices, as many as the possible graphs on 3 nodes, no degree sequence can be inside P_3 .

We conclude by taking note that, by representing the sufficient statistics as a linear mapping $d = Ax$, we can recast the β -model as a log-linear model with design matrix A^\top and product-multinomial scheme, with $\binom{n}{2}$ sampling constraints, one for each edge. This simple yet far reaching observation allows us, among the other things, to design algorithms for detecting nonexistence of the MLE and identifying estimable parameters under a nonexistent MLE, as explained in the supplementary material to this article.

3.1. Parameter Estimability under a Nonexistent MLE. The geometric nature of Theorem 3.1 has important consequences. First, it allows to identify the patterns of observed edge counts that cause nonexistence of the MLE, i.e., the sample points for which the MLE is undefined. Secondly, it yields a complete description of estimability of the edge probability parameters under a nonexistent MLE, a key issue for correct evaluation of degrees of freedom of the model. The next result addresses the last two points.

LEMMA 3.2. *A point y belongs to the interior of some face F of P_n if and only if there exists a set $\mathcal{F} \subset \{(i, j), i < j\}$ such that*

$$(7) \quad y = Ap,$$

where $p = \{p_{i,j} : i < j, p_{i,j} \in [0, 1]\} \in \mathbb{R}^{\binom{n}{2}}$ is such that $p_{i,j} \in \{0, 1\}$ if $(i, j) \notin \mathcal{F}$ and $p_{i,j} \in (0, 1)$ if $(i, j) \in \mathcal{F}$. The set \mathcal{F} is uniquely determined by the face F and is the maximal set for which (7) holds.

Following Geiger et al. (2006) and Fienberg and Rinaldo (2012), we refer to any such set \mathcal{F} a *facial set* of P_n and its complement, $\mathcal{F}^c = \{(i, j) : i < j\} \setminus \mathcal{F}$, a *co-facial set*. Facial sets form a lattice that is isomorphic to the face lattice of P_n (Fienberg and Rinaldo, 2012, Lemma 5). Thus the faces of P_n are in one-to-one correspondence with the facial sets of P_n and, for any

pair of faces F and F' of P_n with associated facial sets \mathcal{F} and \mathcal{F}' , $F \cap F' = \emptyset$ if and only if $\mathcal{F} \cap \mathcal{F}' = \emptyset$ and $F \subset F'$ if and only if $\mathcal{F} \subset \mathcal{F}'$. In details, for a point $x \in \mathcal{S}_n$, $d(x) = Ax$ belongs to the interior of a face F of P_n if and only if there exists a non-negative p such that $d(x) = Ap$, where $\mathcal{F} = \{(i, j) : 0 < p_{i,j} < 1\}$ is the facial set corresponding to F . By the same token, $y \in \text{int}(P_n)$ if and only if $y = Ap$ for a vector p with coordinates strictly between 0 and 1.

Facial sets have statistical relevance for two reasons. First, non-existence of the MLE can be described combinatorially in terms of co-facial sets, i.e. patterns of edge counts that are either 0 or $N_{i,j}$. In particular, the MLE does not exist if and only if the set $\{(i, j) : i < j, x_{i,j} = 0 \text{ or } N_{i,j}\}$ contains a co-facial set. Second, apart from exhausting all possible patterns of forbidden entries in the table leading to a nonexistent MLE, facial sets specify which probability parameters are estimable. In fact, inspection of the likelihood function (3) reveals that, for any observable set of counts $\{x_{i,j} : i < j\}$, there always exists a unique maximizer $\hat{p} = \{\hat{p}_{i,j}, i < j\}$ which, by strict concavity, is uniquely determined by the first order optimality conditions

$$\tilde{d}(x) = A\hat{p},$$

also known as the moment equations. Existence of the MLE is then equivalent to $0 < \hat{p}_{i,j} < 1$ for all $i < j$. When the MLE does not exist, i.e. when \tilde{d} is on the boundary of P_n , the moment equations still hold, but the entries of the optimizer $\{\hat{p}_{i,j}, i < j\}$, known as the *extended MLE*, are no longer strictly between 0 and 1. Instead, by Lemma (3.2), the extended MLE is such that $\hat{p}_{i,j} = \tilde{p}_{i,j} \in \{0, 1\}$ for all $(i, j) \in \mathcal{F}^c$. Furthermore, it is possible to show (see, e.g., Morton, 2008) that $\hat{p}_{i,j} \in (0, 1)$ for all $(i, j) \in \mathcal{F}$. Therefore, when the MLE does not exist, only the probabilities $\{p_{i,j}, (i, j) \in \mathcal{F}\}$ are estimable by the extended MLE. We refer the reader to Barndorff-Nielsen (1978), Brown (1986), Fienberg and Rinaldo (2012) and references therein, for details about the theory of extended exponential families and extended maximum likelihood estimation in log-linear models.

To summarize, while co-facial sets encode the patterns of table entries leading to a non-existent MLE, facial sets indicate which probability parameters are estimable. A similar, though more involved interpretation holds for the estimability of the natural parameters, for which the reader is referred to Fienberg and Rinaldo (2012). Further, for a given sample point x , the realized facial set and its cardinality are both random, as they depend on the actual value of the observed sufficient statistics Ax . This implies that, with a non-existent MLE, the set of estimable parameters is itself random.

4. The Boundary of P_n . Theorem 3.1 and Lemma 3.2 show that the boundary of the polytope P_n plays a fundamental role in determining the existence of the MLE for the β -model and in specifying which parameters are estimable. In particular, the larger the number of faces (i.e. facial sets) of P_n the higher the complexity of the β -model as measured by the numbers of possible patterns of edge counts for which the MLE does not exist. Therefore, gaining an even basic understanding of the number and of the types of co-facial patterns will provide valuable insights into the behavior of the β -model. Below we further elaborate on the consequences of the results established in Section 3 and present a small selection of examples of co-facial sets associated to the facets of P_n .

Though the discussion and examples of this section will reveal a number of subtle issues, we believe that the key message is two-fold. First, the combinatorial complexity of P_n , measured by both the number of the types of co-facial sets, grows very fast with n , with the co-facial sets associated to node degrees bounded away from 0 and $n - 1$ vastly outnumbering the easily detectable cases of minimal or maximal degree. Secondly, since complete enumeration of the faces of P_n is impractical, it is important to devise algorithms for detecting a nonexistent MLE and identifying the facial sets of estimable parameters. Both these issues become more severe in large and sparse networks, where it is expected that the exploding number of possible non-trivial co-facial set renders estimation of the model parameters more difficult. Later in Section 5, we will derive conditions, based on the geometry of P_n that prevents this from happening, with large probability for large n .

4.1. *The combinatorial Complexity of P_n .* Mahadev and Peled (1996) describe the facet-defining inequalities of P_n , for all $n \geq 4$ (when $n \leq 3$ the problem is of little interest), a result we use later in Section 5. Let \mathcal{P} be the set of all pairs (S, T) of disjoint non-empty subsets of $\{1, \dots, n\}$, such that $|S \cup T| \in \{2, \dots, n - 3, n\}$. For any $(S, T) \in \mathcal{P}$ and $y \in P_n$, let

$$(8) \quad g(S, T, y, n) := |S|(n - 1 - |T|) - \sum_{i \in S} y_i + \sum_{i \in T} y_i.$$

THEOREM 4.1 (Theorem 3.3.17 in Mahadev and Peled (1996)). *Let $n \geq 4$ and $y \in P_n$. The facet-defining inequalities of P_n are*

- (i) $y_i \geq 0$, for $i = 1, \dots, n$;
- (ii) $y_i \leq n - 1$, for $i = 1, \dots, n$;
- (iii) $g(S, T, y, n) \geq 0$, for all $(S, T) \in \mathcal{P}$.

×	0		
$N_{1,2}$	×		
		×	$N_{3,4}$
		0	×

×	0	1	2
3	×	2	1
2	1	×	3
1	2	0	×

×	0	0.5	0.5
1	×	0.5	0.5
0.5	0.5	×	1
0.5	0.5	0	×

TABLE 1

Left: Co-facial set leading to a nonexistent MLE. Center: An example of data exhibiting the pattern of counts consistent with the co-facial set on the left when $N_{i,j} = 3$ for all $i \neq j$. Right: Table of the extended MLE of the estimated probabilities.

×	0		
$N_{1,2}$	×	0	0
	$N_{3,2}$	×	
	$N_{4,2}$		×

×	0		0
$N_{1,2}$	×		0
		×	
$N_{4,1}$	$N_{4,2}$		×

TABLE 2

Examples of a co-facial set leading to a nonexistent MLE. Left: $\tilde{d}_2 = 0$. Right: Example where the degrees are all bounded away from 0 and 3, the MLE does not exist.

Even with the exhaustive characterization of P_n provided by Theorem 4.1, understanding the combinatorial complexity of P_n (i.e. the collection of all its faces and their inclusion relations) is far from trivial. Stanley (1991) studied the number faces of the polytope of degree sequences P_n and derived an expression for computing the entries of the f -vector of P_n . The f -vector of an n -dimensional polytope is the vector of length n whose i th entry contains the number of i -dimensional faces, $i = 0, \dots, n-1$. For example, the f -vector of P_8 is the 8-dimensional vector

$$(334982, 1726648, 3529344, 3679872, 2074660, 610288, 81144, 3322).$$

Thus, P_8 is an 8-dimensional polytope with 334982 vertices, 1726648 edges, and so on, up to 3322 facets. Also, according to Stanley's formula, the number of facets of P_4 , P_5 , P_6 and P_7 are 22, 60, 224 and 882, respectively (these numbers correspond to the numbers we obtained with the software `polymake`, using the methods described in the supplementary material to this article). Stanley's analysis showed that the combinatorial complexity of P_n is extraordinarily large, with both the number of vertices and the number of facets growing at least exponentially in n , and, consequently, the tasks of identifying points on the boundary of P_n and the associated facial set are far from trivial. For instance, computing directly the number of vertices of P_{10} is prohibitively expensive, even using one of the best known algorithms, such as the one implemented in the software `minksum` (see Weibel, 2005). To overcome these problems we have devised an algorithm for detecting boundary points and the associated facial sets that can handle networks with up

to hundreds of nodes. We report on this algorithm, which is based on a log-linear model reparametrization and is equivalent to what is known in computational geometry as the ‘‘Cayley trick,’’ in the supplementary material. Using the methods described there, we were able to identify a few interesting cases in which the MLE does not exist, most of which have gone unrecognized in the statistical literature. Below we describe some of our computations for the purpose of elucidating the results derived in Section 3.

4.2. *Some examples of co-facial sets.* Recall that we can represent the data as a $n \times n$ table of counts with structural zero diagonal elements and where the (i, j) – *th* entry of the table indicates the number of times, out of $N_{i,j}$, in which we observed the edges (i, j) . In our examples, empty cells correspond to facial sets and may contain arbitrary count values, in contrast to the cells in the co-facial sets that contain either a zero value or a maximal value, namely $N_{i,j}$. Lemma 3.2 implies that extreme count values of this nature are precisely what leads to the nonexistence of the MLE. The pattern shown on the left of Table 1 provides an instance of a co-facial set, which corresponds to a facet of P_4 . Assume for simplicity that the empty cells contain counts bounded away from 0 and $N_{i,j}$. Then the sufficient statistics \tilde{d} are also bounded away from 0 and $n - 1$ and, and so are the row and column sums of the normalized counts $\{\frac{x_{i,j}}{N_{i,j}} : i \neq j\}$, yet the MLE does not exist. This is further illustrated in Table 1, center, which shows an instance of data with $N_{i,j} = 3$ for all $i \neq j$, satisfying the above pattern and, on the right, the probability values maximizing the log-likelihood function. Notice that, because the MLE does not exist, the supremum of the log-likelihood under the natural parametrization is attained in the limit by any sequence of natural parameters $\{\beta^{(k)}\}$ of the form $\beta^{(k)} = (-c_k, -c_k, c_k, c_k)$, where $c_k \rightarrow \infty$ as $k \rightarrow \infty$. As a result, some of these probability values are 0 and 1. The order of the pattern is crucial. In Table 2 we show, on the left, another example of a co-facial set that is easy to detect, since it corresponds to a value of 0 for the normalized sufficient statistic \tilde{d}_2 . Indeed, from cases (i) and (ii) of Theorem 4.1, the MLE does not exist if $\tilde{d}_i = 0$ or $\tilde{d}_i = n - 1$, for some i . On the right, we show a co-facial set that is instead compatible with normalized sufficient statistics being bounded away from 0 and $n - 1$. Finally, in Table 3 we list all 22 co-facial sets associated with the facets of P_4 , including the cases already shown.

In general, there are $2n$ facets of P_n that are determined by one \tilde{d}_i equal to 0 or $n - 1$. Thus, just by inspecting the row sums or the observed sufficient statistics, we can detect only $2n$ co-facial sets associated to as many facets of P_n . Comparing this number to the entries of the f -vector calculated in

\times	0			\times	0			\times	0		0
$N_{1,2}$	\times			$N_{1,2}$	\times	0	0	$N_{1,2}$	\times		0
		\times	$N_{3,4}$		$N_{3,2}$	\times				\times	
		0	\times		$N_{4,2}$		\times	$N_{4,1}$	$N_{4,2}$		\times

\times	0	0	0	\times	0	0		\times		0	0
$N_{1,2}$	\times			$N_{1,2}$	\times	0			\times		
$N_{1,3}$		\times		$N_{1,3}$	$N_{2,3}$	\times		$N_{1,3}$		\times	0
$N_{4,1}$			\times				\times	$N_{1,4}$		$N_{3,4}$	\times

\times		0		\times	$N_{1,2}$	$N_{1,3}$	$N_{1,4}$	\times		$N_{1,3}$	$N_{1,4}$
	\times	0		0	\times				\times		
$N_{1,3}$	$N_{2,3}$	\times	0	0		\times		0		\times	$N_{3,4}$
		$N_{3,4}$	\times	0			\times	0		0	\times

\times	$N_{1,2}$	$N_{1,3}$		\times		$N_{1,3}$		\times	$N_{1,2}$		$N_{1,4}$
0	\times	$N_{2,3}$			\times	$N_{2,3}$		0	\times		$N_{2,4}$
0	0	\times		0	0	\times	$N_{3,4}$			\times	
			\times			0	\times	0	0		\times

\times			$N_{1,4}$	\times		0		\times	$N_{1,2}$		
	\times		$N_{2,4}$		\times	0		0	\times	0	0
		\times	$N_{3,4}$	$N_{1,3}$	$N_{2,3}$	\times	0		$N_{2,3}$	\times	
0	0	0	\times			$N_{3,4}$	\times		$N_{2,4}$		\times

\times				\times			0	\times	$N_{1,2}$		
	\times	0	0		\times		0	0	\times		
	$N_{2,3}$	\times	0			\times	0			\times	0
	$N_{2,4}$	$N_{3,4}$	\times	$N_{1,4}$	$N_{2,4}$	$N_{3,4}$	\times			$N_{3,4}$	\times

\times		0		\times		$N_{1,3}$		\times			$N_{1,4}$
	\times		$N_{2,4}$		\times		0		\times	0	
$N_{1,3}$		\times		0		\times			$N_{2,3}$	\times	
	0		\times		$N_{2,4}$		\times	0			\times

\times			0	\times			0
	\times	$N_{2,3}$			\times	$N_{2,3}$	
	0	\times			0	\times	
$N_{1,4}$			\times	$N_{1,4}$			\times

TABLE 3

All possible co-facial sets for P_4 corresponding to the facets of P_4 (empty cells indicate arbitrary entry values).

Stanley (1991), however, and as our computations confirm, most of the facets of P_n do not yield co-facial sets of this form. Since the number of facets

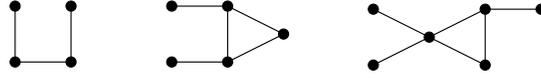


FIG 1. Examples of random graphs on 4 (left), 5 (center) and 6 (right) nodes with node degrees bounded away from 0 and $n - 1$ and for which the MLE is not defined. Lemma 4.2 applies with $S = \{3, 4\}$ and $T = \{1, 2\}$ (left), with $S = \{2, 3, 4\}$ and $T = \{1, 5\}$ (center) and with $S = \{1, 2, 6\}$ and $T = \{3, 4, 5\}$ (right).

appear to grow exponentially in n , we conclude that most of the co-facial sets do not appear to arise in this fashion. Thus, at least combinatorially, patterns of data counts leading to the nonexistence of MLEs but with the normalized degree bounded away from 0 and $n - 1$ are much more frequent, especially in larger networks.

4.3. *The Random Graph Case.* In the special case of $N_{i,j} = 1$ for all $i < j$, which is equivalent to a model for random undirected graphs, points on the boundary of P_n are, by construction, degree sequences and have a direct graph-theoretical interpretation. We say that a subset of a set of nodes of a given graph is *stable* if it induces a subgraph with no edges and a *clique* if it induces a complete subgraph.

LEMMA 4.2 (Lemma 3.3.13 in Mahadev and Peled (1996)). *Let d be a degree sequence of a graph \mathcal{G} that lies on the boundary of P_n . Then either $d_i = 0$, or $d_i = n - 1$ for some i , or there exist non-empty and disjoint subsets S and T of $\{1, \dots, n\}$ such that*

1. S is clique of \mathcal{G} ;
2. T is a stable set of \mathcal{G} ;
3. every vertex in S is adjacent to every vertex in $(S \cup T)^c$ in \mathcal{G} ;
4. no vertex of T is adjacent to any vertex of $(S \cup T)^c$ in \mathcal{G} .

Using Lemma 4.2, we can create virtually any example of a random graph whose node degree sequence lies on the boundary of P_n . In particular, we note that having node degrees bounded away from 0 and $n - 1$ is not a sufficient condition for the existence of the MLE, although its violation implies nonexistence of the MLE (see the examples of Figure 1). Nonetheless, Lemma 4.2 is of little or no practical use when it comes to detecting boundary points and the associated co-facial sets, since checking for the existence of a pair (S, T) of subsets of nodes satisfying conditions 1. through 4. is algorithmically impractical. In the supplementary material to this article,

×	0		
1	×		
		×	1
		0	×

×		0	
	×		
1		×	1
		0	×

×			1
	×	0	
	1	×	
0			×

×	1		
0	×		
		×	0
		1	×

×		1	
	×		
0		×	0
		1	×

×			0
	×	1	
	0	×	
1			×

TABLE 4

Patterns of zeros and ones yielding random graphs with non-existent MLE (empty cells indicate that the entry could be a 0 or a 1).

we describe alternative procedures that can be used in large networks.

Figure 1 shows three examples of graphs on 4, 5 and 6 nodes for which the MLE of the β -model is undefined even though the node degrees are bounded away from 0 and $n - 1$ in all cases. All the examples were constructed using directly Lemma 4.2, as explained in the caption. To the best of our knowledge, even these very small examples of nonexistent MLEs are unknown to practitioners and no available software for fitting the MLE is able to detect nonexistence, much less identify the relevant facial set.

For the case $n = 4$, our computations show that there are 14 distinct co-facial sets associated to the facets of P_n . Eight of them correspond to degree sequences containing a 0 or a 3, and the remaining six are shown in Table 4, which we computed numerically using the procedure described in the supplementary material. Notice that the three tables on the second row are obtained from the first three tables by switching zeros with ones. Furthermore, the number of the co-facial sets we found is smaller than the number of facets of P_n , which is 22, as shown in Table 3. This is a consequence of the fact that the only observed counts in the random graph model are 0's or 1's: it is in fact easy to see in Table 3 that any co-facial set containing three zero counts and three maximal counts $N_{i,j}$ is equivalent, in the random graph case, to a node having degree zero or 3. However, as soon as $N_{i,j} \geq 2$, the number of possible co-facial sets matches the number of faces of P_n . Therefore, the condition $N_{i,j} = 1$ is not inconsequential, as it appears to reduce the numbers of observable patterns leading to a nonexistent MLE, though we do not know the extent of the impact of such reduction in general.

5. Existence of the MLE: Finite Sample Bounds. In this section we exploit the geometry of the boundary of P_n from Lemma 4.2 to derive sufficient conditions that imply the existence of the MLE with large probability

as the size of the network n grows. These conditions essentially guarantee that the probability of observing any of the super-exponentially many (in n) co-facial sets of \mathcal{P}_n is polynomially small in n . Unlike in previous analyses, our result does not require the network to be dense.

We make the simplifying assumption that $N_{i,j} = N$, for all i and j , where $N = N(n) \geq 1$ could itself depend on n . Recall the random vector \tilde{d} , whose coordinates are given in (6) and let $\bar{d} = \mathbb{E}[\tilde{d}] \in \mathbb{R}^n$ be its expected value under the β -model. Then

$$\bar{d}_i = \sum_{j < i} p_{j,i} + \sum_{j > i} p_{i,j}, \quad i = 1 \dots, n.$$

We formulate sufficient conditions for the existence of the MLE in terms of the entries of the vector \bar{d} .

THEOREM 5.1. *Assume that, for all $n \geq \max\left\{4, 2\sqrt{c\frac{n \log n}{N}} + 1\right\}$, the vector \bar{d} satisfies the conditions*

- (i) $\min_i \min\{\bar{d}_i, n - 1 - \bar{d}_i\} \geq 2\sqrt{c\frac{n \log n}{N}} + C,$
- (ii) $\min_{(S,T) \in \mathcal{P}} g(S, T, \bar{d}, n) > |S \cup T| \sqrt{c\frac{n \log n}{N}} + C,$

where $c > 1/2$ and $C \in \left(0, \frac{n-1}{2} - \sqrt{c\frac{n \log n}{N}}\right)$. Then, with probability at least $1 - \frac{2}{n^{2c-1}}$, the MLE exists.

When N is constant, e.g., when $N = 1$ as in the random graph case, we can relax the conditions of Theorem 5.1 by requiring condition (ii) to hold only over subsets S and T of cardinality of order $\Omega(\sqrt{n \log n})$. While we present this result in greater generality by assuming only that $n \geq N$, we do not expect it to be sharp in general when N grows with n .

COROLLARY 5.2. *Let $n \geq \max\{N, 4, 2\sqrt{cn \log n} + 1\}$, $c > 1$ and $C \in \left(0, \frac{n-1}{2} - \sqrt{cn \log n}\right)$. Assume the vector \bar{d} satisfies the conditions*

- (i') $\min_i \min\{\bar{d}_i, n - 1 - \bar{d}_i\} \geq 2\sqrt{cn \log n} + C,$
- (ii') $\min_{(S,T) \in \mathcal{P}_n} g(S, T, \bar{d}, n) > |S \cup T| \sqrt{cn \log n} + C,$

where

$$\mathcal{P}_n := \{(S, T) \in \mathcal{P} : \min\{|S|, |T|\} > \sqrt{cn \log n} + C\},$$

where the set \mathcal{P} was defined before Theorem 4.1. Then, the MLE exists with probability at least $1 - \frac{2}{n^{2c-2}}$. If $N = 1$, it is sufficient to have $c > 1/2$, and the MLE exists with probability larger than $1 - \frac{2}{n^{2c-1}}$.

Discussion and Comparison with Previous Work. Since $|S \cup T| \leq n$, one could replace assumption (ii) of Theorem 5.1 with the simpler but stronger condition

$$\min_{(S,T) \in \mathcal{P}_n} g(S, T, \bar{d}, n) > n^{3/2} \sqrt{c \log n} + C_n.$$

Then, if we assume for simplicity that N is a constant, as in Corollary 5.2, the MLE exists with probability tending to one at a rate that is polynomial in n whenever

$$\min_i \min \{ \bar{d}_i, n - 1 - \bar{d}_i \} = \Omega \left(\sqrt{n \log n} \right)$$

and, for all pairs $(S, T) \in \mathcal{P}$,

$$g(S, T, \bar{d}, n) > \Omega \left(n^{3/2} \sqrt{\log n} \right).$$

For the case $N = 1$, we can compare Corollary 5.2 with Theorem 3.1 in Chatterjee et al. (2011), which also provides sufficient conditions for the existence of the MLE with probability no smaller than $1 - \frac{1}{n^{2c-1}}$ (for all n large enough). Their result appears to be stronger than ours, but that is actually not the case as we now explain. In fact, their conditions require that, for some constant c_1, c_2 and c_3 in $(0, 1)$, $c_1(n-1) < d_i < c_2(n-1)$ for all i and

$$(9) \quad |S|(|S| - 1) - \sum_{i \in S} d_i + \sum_{i \notin S} \min\{d_i, |S|\} > c_3 n^2,$$

for all sets S such that $|S| > (c_1)^2 n^2$. For any non-empty subsets $S \subset \{1, \dots, n\}$ and $T \subset \{1, \dots, n\} \setminus S$,

$$\sum_{i \notin S} \min\{d_i, |S|\} \leq \sum_{i \in T} d_i + |S| |(S \cup T)^c|,$$

which implies that

$$|S|(n - 1 - |T|) - \sum_{i \in S} d_i + \sum_{i \in T} d_i > |S|(|S| - 1) - \sum_{i \in S} d_i + \sum_{i \notin S} \min\{d_i, |S|\},$$

where we have used the equality $n = |S| + |T| + |(S \cup T)^c|$. Thus if (9) holds for some non-empty $S \subset \{1, \dots, n\}$, it satisfies the facet conditions implied by all the pairs (S, T) , for any non-empty set $T \subset \{1, \dots, n\} \setminus S$. As a result, for any subset S , condition (9) is stronger than any of the facet conditions of P_n specified by S . In addition, we weakened significantly the

requirements in [Chatterjee et al. \(2011\)](#) that $c_1(n-1) < d_i < c_2(n-1)$ for all i to $\min_i \min \{\bar{d}_i, n-1-\bar{d}_i\} \geq 2\sqrt{cn \log n} + C$. As a direct consequence of this weakening, we only need $|S| > \sqrt{cn \log n} + C$ as opposed to $|S| > (c_1)^2 n^2$. Overall, in our setting, the vector of expected degrees of the sequence of networks is allowed to lie much closer to the boundary of P_n . As we explain next, such weakening is significant, since the setting of [Chatterjee et al. \(2011\)](#) only allows us to estimate an increasing number of probability parameters (the edge probabilities) that are uniformly bounded away from 0 and 1, while our assumptions allow for these probabilities to become degenerate as the network size grows, and therefore hold even in non-dense network settings.

The non-degenerate case. We now briefly discuss the case of sequences of networks for which $N = 1$ and the edge probabilities are uniformly bounded away from 0 and 1, i.e.

$$(10) \quad \delta < p_{i,j} < 1 - \delta, \quad \forall i, j,$$

for some $\delta \in (0, 1)$ independent of n . In this scenario, the number of probability parameters to be estimated grows with n , but their values are guaranteed to be non-degenerate. It immediately follows from the non-degenerate assumption (10) that $\bar{d} \in \text{int}(P_n)$ and

$$(11) \quad \delta(n-1) < \bar{d}_i < (1-\delta)(n-1), \quad i = 1, \dots, n.$$

Then, the same arguments we used in the proof of [Corollary 5.2](#) imply that the MLE exists with high probability. We provide a sketch of the proof. First, we note that, with high probability, $g(S, T, \tilde{d}, n) \geq g(S, T, \bar{d}, n) - |S \cup T| \Omega(\sqrt{n \log n})$, for each pair $(S, T) \in \mathcal{P}$. Furthermore, because of (11), it is enough to consider only pairs (S, T) of disjoint subsets of $\{1, \dots, n\}$ of sizes of order $\Omega(n)$. For each such pair, the condition on \bar{d}_i further yields that $g(S, T, \bar{d}, n)$ is of order $\Omega(n^2)$, and, by [Theorem 8](#) the MLE exists with high probability.

In fact, the boundedness assumption of [Chatterjee et al. \(2011\)](#) that $\|\beta\|_\infty < L$ with L independent of n , is equivalent to the non-degenerate assumption (10), as we see from equation (1). Unlike [Chatterjee et al. \(2011\)](#), who focus on the non-degenerate case, our results hold under weaker scaling, as we only require for instance that \bar{d}_i be of order $\Omega(\sqrt{n \log n})$ for all i . Relatedly, we note that the tameness condition of [Barvinok and Hartigan \(2010\)](#) is equivalent to $\delta < \hat{p}_{i,j} < 1 - \delta$ for all i and j and a fixed $\delta \in (0, 1)$, where $\hat{p}_{i,j}$ is the MLE of $p_{i,j}$. Therefore, the tameness condition is stronger

than the existence of the MLE. In fact, using again Theorem 1.3 in [Chatterjee et al. \(2011\)](#), for all n sufficiently large, the tameness condition is equivalent to the boundedness condition of [Chatterjee et al. \(2011\)](#).

We conclude this section with two useful remarks. First, Theorem 1.3 in [Chatterjee et al. \(2011\)](#) demonstrates that, when the MLE exists, $\max_i |\hat{\beta}_i - \beta_i| = O(\sqrt{\log n/n})$, with probability at least $1 - \frac{2}{n^{2c-1}}$. Combined with our Corollary 5.2, this implies that the MLE is a consistent estimator under a growing network size and with edge probabilities approaching the degenerate values of 0 and 1.

Secondly, after the submission of this article we learned about the interesting asymptotic results of [Yan and Xu \(2012\)](#); [Yan et al. \(2012\)](#), who claim that, based on a modification of the arguments of [Chatterjee et al. \(2011\)](#), it is possible to show the MLE of the β -model exists and is uniformly consistent if $L = o(\log n)$ and $L = o(\log \log n)$, respectively, where $L = \max_i |\beta_i|$.

6. Discussion and extensions. We have used polyhedral geometry to analyze the conditions for existence of the MLE of a generalized version of the β -model and to derive finite sample bounds for the probability associated with the existence of the MLE. Our results offer a novel and explicit characterization of the patterns of edge counts leading to non-existent MLEs. The problem of nonexistence occurs in numbers and with a complexity that was not previously known. Our results allow us to sharpen conditions for existence of the MLE. Our analysis in particular highlights the fact that requiring node degrees equal to 0 and $n - 1$ is only a sufficient condition for nonexistence of the MLE and non-estimability of the edge probabilities. We show that we need to account for many more edge patterns. We note that the use of polyhedral geometry in statistical models for discrete data is a hallmark of the theory of exponential families, but its considerable potential for use and applications in the analysis of log-linear and network models has only recently begun to be investigated (see [Fienberg and Rinaldo, 2012](#); [Rinaldo et al., 2009](#)).

Our generalization of the β -model allows for Poisson and binomial, not simply Bernoulli distributions for edges. Email databases and others involving repeated transactions among pairs of parties provides the simplest examples of situations for networks where edges can occur multiple times. These are often analyzed as weighted networks but that may not necessarily make as much sense as using a Poisson for random numbers of occurrences.

As our results indicate, the nonexistence of the MLE is equivalent to non-estimability of a subset of the parameters of the model, but by no means does

it imply that no statistical inference can take place. In fact, when the MLE does not exist, there always exists a “restricted” β -model that is specified by the appropriate facial set, and for which all parameters are estimable. Thus, for such smaller model, traditional statistical tasks such as hypothesis testing and assessment of parameter uncertainty are possible, even though it becomes necessary to adjust the number of degrees of freedom for the non-estimable parameters. A complete description of this approach, which is rooted in the theory of extended exponential families, is beyond the scope of the article. See [Fienberg and Rinaldo \(2012\)](#) for details.

We can extend our study of the β -model in a number of ways. In the supplementary material to this article, we consider various generalizations of the β -model setting, including the β -model with random numbers of edges, the Rasch model from item response theory, the Bradley-Terry paired comparisons model and the p_1 network model. For most of these models we were able to carry out a fairly explicit analysis based on the underlying geometry, but for the full p_1 model the complexity of the model polytope appears to make such a direct analysis very difficult (this is reflected in the high complexity of the Markov basis for p_1 model, of which we give full account in [Petrović et al., 2010](#)). Another interesting extension of our results of Section 5 would be to translate our conditions, which are formulated in terms of expected degree sequences, into conditions on the $p_{i,j}$ ’s themselves, for instance by establishing appropriate bounds for $\min_{i<j} p_{i,j}$, $\max_{i<j} p_{i,j}$, or $\max_{i \neq j} \frac{p_{i,j}}{1-p_{i,j}}$.

We conclude with some remarks on the computational aspects of our analysis, which constitute a non-trivial component of our work and is of key importance for detecting the nonexistence of the MLE and identifying estimable parameters. The main difficulty in applying our results is that the polytope of degree sequences P_n is difficult to handle algorithmically in general. Indeed, P_n arises a Minkowski sum and, even though the system of defining inequalities is given explicitly, its combinatorial complexity grows exponentially in n . More importantly, the vertices of P_n are not known explicitly. Algorithms for obtaining the vertices of P_n , such as `minksum` (see [Weibel, 2005](#)), are computationally expensive and require generating all the points $\{Ax, x \in \mathcal{G}_n\}$, where $|\mathcal{G}_n| = 2^{\binom{n}{2}}$, a task that, even for n as small as 10, is impractical. See, for instance, our analysis of the p_1 model included in the supplementary material. Thus, deciding whether a given degree sequence is a point in the interior of P_n and identifying the facial set corresponding to an observed degree sequence on its boundary is highly non-trivial. Our strategy to overcome these problems entails re-expressing the β -model as a log-linear model with $\binom{n}{2}$ product-multinomial sampling constraints.

This approach is not new, and it harks back to the earlier re-expression of the Holland-Leinhardt p_1 model and its natural generalizations as log-linear models (Fienberg and Wasserman, 1981a,b; Fienberg et al., 1985; Meyer, 1983). Though this re-parametrization increases the dimensionality of the problem, it nonetheless has the crucial computational advantage of reducing the determination of the facial sets of P_n to the determination of the facial sets of a pointed polyhedral cone spanned by $n(n - 1)$ vectors, which is a much simpler object to analyze, both theoretically and algorithmically. This procedure is known as the Cayley embedding in polyhedral geometry, and Fienberg and Rinaldo (2012) describe its use in the analysis of log-linear models. The advantages of this re-parametrization are two-fold. First, it allows us to use the highly optimized algorithms available in `polymake` for listing explicitly all the facial sets of P_n . This is how we computed the facial sets in all the examples presented in this article. Secondly, the general algorithms for detecting nonexistence of the MLE and identifying facial sets proposed in Fienberg and Rinaldo (2012), which can handle larger dimensional models (with n in the order of hundreds), can be directly applied to this problem. This reference is also relevant for dealing with inference under a non-existent MLE.

The details of our computations and the associated algorithms are provided in the supplementary material accompanying this article. The R routines used to carry out the computations for the results presented in the paper and for creating the input files for `polymake` are available at http://www.stat.cmu.edu/~arinaldo/Rinaldo_Petrovic_Fienberg_Rcode.txt

SUPPLEMENTARY MATERIAL

(<http://lib.stat.cmu.edu/aoas/???/???>). In the supplementary material we extend our analysis to other models for network data: the Rasch model, the β -model with no sampling constraints on the number of observed edges per dyad, the Bradley-Terry model and the p_1 model of Holland and Leinhardt (1981). We also provide details on how to determine whether a given degree sequence belongs to the interior of the polytope of degree sequences P_n and on how to compute the facial set corresponding to a degree sequence on the boundary of P_n .

7. Acknowledgments. This research was supported in part by grant FA9550-12-1-0392 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA), NSF grant DMS- 0631589, and by a grant from the Singapore National Research Foundation (NRF) under the Interactive & Digital Media Programme Office to

the Living Analytics Research Centre (LARC). A previous version of this manuscript was completed while the second author was in residence at Institut Mittag-Leffler, to which she is grateful for its hospitality.

8. Proofs.

PROOF OF THEOREM 3.1. Throughout the proof, we will use standard results and terminology from the theory of exponential families, for which standard references are [Brown \(1986\)](#) and [Barndorff-Nielsen \(1978\)](#). The polytope

$$S_n := \text{convhull}(\{Ax, x \in \mathcal{S}_n\})$$

is the convex support for the sufficient statistics of the natural exponential family described in Section 2. Furthermore, by a fundamental result in the theory of exponential families (see, e.g., Theorem 9.13 in [Barndorff-Nielsen, 1978](#)), the MLE of the natural parameter $\beta \in \mathbb{R}^n$ (or, equivalently of the set probabilities $\{p_{i,j}, i < j\} \in \mathbb{R}^{\binom{n}{2}}$ satisfying (1)) exists if and only if $d \in \text{int}(S_n)$. Thus, it is sufficient to show that $d \in \text{int}(S_n)$ if and only if $\tilde{d} \in \text{int}(P_n)$.

Denote with $a_{i,j}$ the column of A corresponding to the ordered pair (i, j) , with $i < j$, and set

$$(12) \quad P_{i,j} = \text{convhull}\{0, a_{i,j}\} \subset \mathbb{R}^n.$$

Each $P_{i,j}$ is a line segment between its vertices 0 and $a_{i,j}$. Then, P_n can be expressed as the zonotope obtained as the Minkowski sum of the line segments $P_{i,j}$:

$$(13) \quad P_n = \sum_{i < j} P_{i,j}.$$

This identity can be established as follows. On one hand, P_n is the convex hull of vectors that are Boolean combinations of the columns of A. Since all such combinations are in $\sum_{i < j} P_{i,j}$, and both P_n and $\sum_{i < j} P_{i,j}$ are closed sets, we obtain $P_n \subseteq \sum_{i < j} P_{i,j}$. On the other hand, the vertices of $\sum_{i < j} P_{i,j}$ are also Boolean combinations of the columns of A (see, e.g., corollary 2.2 in [Fukuda, 2004](#)), and, therefore, $\sum_{i < j} P_{i,j} \subseteq P_n$.

Equation (13) shows, in particular, that $\tilde{d} \in P_n$. Furthermore, using the same arguments, we see that, similarly to P_n , S_n too can be expressed as a Minkowski sum:

$$S_n = \sum_{i < j} S_{i,j},$$

where

$$S_{i,j} := P_{i,j}N_{i,j} = \{xN_{i,j} : x \in P_{i,j}\}$$

is the rescaling of $P_{i,j}$ by a factor of $N_{i,j}$. In fact, we will prove that S_n and P_n are combinatorially equivalent.

For a polytope P and a vector c , we set $F(P; c) := \{x \in P : x^\top c \geq y^\top c, \forall y \in P\}$. Any face F of P can be written in this way, where c is any vector in the interior of the normal cone to F . By Proposition 2.1 in Fukuda (2004), F is a face of P_n with $F = F(P_n, c)$ if and only if it can be written uniquely as

$$F(P_n, c) = \sum_{i < j} F(P_{i,j}, c),$$

for any c in the interior of the normal cone to F . It is immediate to see that $F(P_{i,j}, c)$ is a face of $P_{i,j}$ if and only if $F(S_{i,j}, c)$ is a face of $S_{i,j}$, and that $F(S_{i,j}, c) = N_{i,j}F(P_{i,j}, c)$; in fact, $P_{i,j}$ and $S_{i,j}$ are combinatorially equivalent. Therefore, invoking again Proposition 2.1 in Fukuda (2004), we conclude that $F(P_{i,j}, c)$ is a face of P_n if and only if

$$\sum_{i < j} N_{i,j}F(P_{i,j}, c)$$

is a face of S_n (and this representation is unique). From this, we see that P_n and S_n have the same normal fan and, therefore, are combinatorially equivalent. \square

PROOF OF LEMMA 3.2. By Proposition 2.1 in Fukuda (2004),

$$(14) \quad F = F(P_n, c) = \sum_{i < j} F(P_{i,j}, c),$$

for any c in the interior of the normal cone to F , where the above representation is unique. Since $P_{i,j}$ is a line segment (see (12)), its only proper faces are the vertices 0 and $a_{i,j}$. Let the set \mathcal{F} be the complement of the set of pairs (i, j) with $i < j$ such that $F(P_{i,j}, c)$ is either the vector 0 or $a_{i,j}$. By the uniqueness of the representation (14), \mathcal{F} is unique as well and, in particular, maximal. Furthermore, as it depends on F only through the interior of its normal cone and since the interiors of the normal cones of P_n are disjoint, different faces will be associated with different facial sets. \square

PROOF OF THEOREM 5.1. Let $\tilde{d} = (\tilde{d}_1, \dots, \tilde{d}_n)$ be the random vector defined in (6). We will show that, under the stated assumptions, $\tilde{d} \in \text{int}(P_n)$ with probability no smaller than $1 - \frac{2}{n^{2c-1}}$.

Since N is constant, we conveniently re-express the random vector \tilde{d} as an average of independent and identically distributed graphical degree sequences. In details, we can write

$$(15) \quad \tilde{d} = \frac{1}{N} \sum_{k=1}^N d^{(k)},$$

where each $d^{(k)}$ is the degree sequence arising from of an independent realization of random graph with edge probabilities $\{p_{i,j} : i < j\}$, for $k = 1, \dots, N$.

Thus, each \tilde{d}_i is the sum of $N(n-1)$ independent random variables taking values in $\{0, \frac{1}{N}\}$. Then, an application of Hoeffding's inequality and of the union bound yields that the event

$$(16) \quad \mathcal{O}_n := \left\{ \max_i |\tilde{d}_i - \bar{d}_i| \leq \sqrt{c \frac{n \log n}{N}} \right\}$$

occurs with probability at least $1 - \frac{2}{n^{2c-1}}$. Throughout the rest of the proof we assume that the event \mathcal{O}_n holds.

By assumption (i), for each i ,

$$0 < C + \sqrt{c \frac{n \log n}{N}} \leq \bar{d}_i - \sqrt{c \frac{n \log n}{N}} \leq \tilde{d}_i \leq \bar{d}_i + \sqrt{c \frac{n \log n}{N}} \leq n-1 - C - \sqrt{c \frac{n \log n}{N}} < n-1,$$

so that

$$(17) \quad 0 < \tilde{d}_i < n-1, \quad i = 1, \dots, n.$$

Notice that the assumed constraint on the range of C guarantees the above inequalities are well defined. Next, for each pair $(S, T) \in \mathcal{P}$,

$$|g(S, T, \tilde{d}, n) - g(S, T, \bar{d}, n)| \leq |S \cup T| \max_i |\tilde{d}_i - \bar{d}_i|,$$

which yields

$$g(S, T, \tilde{d}, n) \geq g(S, T, \bar{d}, n) - |S \cup T| \sqrt{c \frac{n \log n}{N}}.$$

Using assumption (ii), the previous inequality implies that

$$(18) \quad \min_{(S, T) \in \mathcal{P}} g(S, T, \tilde{d}, n) > C > 0.$$

Thus, we have shown that (17) and (18) hold, provided that the event \mathcal{O}_n is true and assuming (i) and (ii). Therefore, by Theorem 4.1 the MLE exists. \square

PROOF OF COROLLARY 5.2. Using the same setting and notation of Theorem 5.1, we will assume throughout the proof that the event

$$\mathcal{O}'_n := \left\{ \max_k \max_i |d_i^{(k)} - \bar{d}_i| \leq \sqrt{cn \log n} \right\}$$

holds true. By Hoeffding's inequality, the union bound, and the inequality $\log N \leq \log n$, we have

$$\mathbb{P}(\mathcal{O}'_n) \leq 2 \exp \{-2c \log n + \log n + \log N\} \leq \frac{2}{n^{2c-2}}.$$

A simple calculation shows that, when \mathcal{O}'_n is satisfied, we also have

$$\left\{ \max_i |\tilde{d}_i - \bar{d}_i| \leq \sqrt{cn \log n} \right\}.$$

Then, by the same arguments we used in the proof of Theorem 5.1, assumption (i') yields that

$$(19) \quad 0 < \tilde{d}_i < n - 1, \quad i = 1, \dots, n.$$

and, for each pair $(S, T) \in \mathcal{P}$,

$$(20) \quad g(S, T, \tilde{d}, n) \geq g(S, T, \bar{d}, n) - |S \cup T| \sqrt{cn \log n}.$$

It is easy to see that, for the event \mathcal{O}'_n , assumption (i') also yields

$$(21) \quad \min_k \min_i \min \left\{ d_i^{(k)}, n - 1 - d_i^{(k)} \right\} \geq \sqrt{cn \log n} + C.$$

We now show that, when (19) and the previous equation are satisfied, the MLE exists if

$$(22) \quad \min_{(S, T) \in \mathcal{P}_n} g(S, T, d, n) > C > 0.$$

Indeed, suppose that (19) is true and that \tilde{d} belongs to the boundary of P_n . Then, by the integrality of the polytope P_n , there exist non-empty and disjoint subsets T and S of $\{1, \dots, n\}$ satisfying the conditions of Lemma 4.2 for each of the degree sequences $d^{(1)}, \dots, d^{(k)}$. If $\min_k \min_i d_i^{(k)} > \sqrt{cn \log n} + C$, then, necessarily, $|S| > \sqrt{cn \log n} + C$, because $|S|$ is the maximal degree of every node $i \in T$. Similarly, since each $i \in S$ has degree at least $|S| - 1 + |(S \cup T)^c|$, if $\max_k \max_i d_i^{(k)} < n - 1 - \sqrt{cn \log n} - C$, the inequality

$$|S| - 1 + |(S \cup T)^c| < n - 1 - \sqrt{cn \log n} - C$$

must hold, implying that $|T| = n - |S| - |(S \cup T)^c| > \sqrt{cn \log n} + C$. Thus, we have shown that, if (19) and (21) hold, and \tilde{d} belongs to the boundary of P_n , the cardinalities of the sets S and T defining the facet of P_n to which \tilde{d} belongs cannot be smaller than $\sqrt{cn \log n} + C$. By Theorem 4.1, when (19) and (21) hold, (22) implies that $\tilde{d} \in \text{int}(P_n)$, so the MLE exists. However, equation (20) and assumption (ii') implies (22), so the proof is complete. \square

References.

- Albert, A. and Barabási, A.-L. (2002). Statistical mechanics of complex networks, *Reviews of Modern Physics*, 74, 47–97.
- Barndorff-Nielsen, O.E. (1978). *Information and Exponential Families in Statistical Theory*, Wiley, New York.
- Barvinok, A. and Hartigan, J.A. (2010). The number of graphs and a random graph with a given degree sequence, available at <http://arxiv.org/pdf/1003.0356v2>.
- Blitzstein, J. and Diaconis, P. (2009). A sequential importance sampling algorithm for generating random graphs with prescribed degrees, *Internet Mathematics*, 6 (4), 489–522.
- Brown, L. (1986). *Fundamentals of Statistical Exponential Families*, IMS Lecture Notes-Monograph Series, Vol.9, Hayward, CA.
- Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence, *Annals of Applied Probability*, 21 (4), 1400–1435.
- Cohen, R. and Havlin, S. (2010). *Complex Networks: Structure, Robustness and Function*, Cambridge University Press.
- Cover, T. M. and Thomas, J. A. (1991). *Elements of Information Theory*, Wiley, New York.
- Diaconis, P. and Janson, S. (2007). Graph limits and exchangeable random graphs, available at <http://arxiv.org/abs/0712.2749>
- Erdős, P. and Rényi, A. (1959) On random graphs. I, *Publicationes Mathematicae*, 6, 290–297.
- Fienberg, S.E. and Rinaldo, A. (2012). Maximum likelihood estimation in log-linear models, *Annals of Statistics*, 40 (2), 996–1023.
- Fienberg, S.E. and Wasserman, S.S. (1981a). Categorical data analysis of single sociometric relations, *Sociological Methodology 1981*, 156–192.
- Fienberg, S.E. and Wasserman, S.S. (1981b). An exponential family of probability distributions for directed graphs: Comment. *Journal of the American Statistical Association*, 76, 54–57.
- Fienberg, S.E., Meyer, M.M., and Wasserman, S.S. (1985). Statistical analysis of multiple sociometric relations, *Journal of the American Statistical Association*, 80, 51–67.
- Foster, J. B., Foster, D. V., Grassberger, P. and Paczuski, M. (2007). Link and sub-graph likelihoods in random undirected networks with fixed and partially fixed degree sequences, *Physical Review E*, 76, 046112.
- Fukuda, K. (2004). From the zonotope construction to the Minkowski addition of convex polytopes, *Journal of Symbolic Computation*, 38(4), 1251–1272.
- Gawrilow, E. and Joswig, M. (2000). `polymake`: A framework for analyzing convex polytopes, in *Polytopes—Combinatorics and Computation*, 43–74, G. Kalai and G. M. Ziegler (eds.), Birkhäuser, Basel.
- Geiger, D., Meek, C. and Sturmfels, B. (2006). On the toric algebra of graphical models, *Annals of Statistics*, 34(3), 1463–1492.

- Geyer, C. (2009). Likelihood inference in exponential families and directions of recession, *Electronic Journal of Statistics*, 3, 259-289.
- Goldenberg, A., Zheng, A.X., Fienberg, S.E., and Airoldi, E.M. (2010). A survey of statistical network models, *Foundations and Trends in Machine Learning*, 2 (2), 129-233.
- Goodreau, S. M. (2007). Advances in exponential random graph (p^*) models applied to a large social network, *Social Networks*, 29(2), 231-248.
- Handcock, M. S. and Morris, M. (2007). A simple model for complex networks with arbitrary degree distribution and clustering, *Statistical Network Analysis: Models, Issues and New Directions*, E. Airoldi, D. Blei, S.E. Fienberg, A. Goldenberg, E. Xing, and A. Zheng (eds.), Lecture Notes in Computer Science, Volume 4503, 103-114.
- Haberman, S. (1981). Discussion of the article by Holland, P. W. and Leinhardt, S. "An exponential family of probability distributions for directed graphs," *Journal of the American Statistical Association*, 76, 60-61.
- Hammer P.L., Peled, U.N. and Sun, X. (1990). Difference Graphs, *Discrete Applied Mathematics*, 28, 35-44.
- Hara, H. and Takemura, A. (2010). Connecting tables with zero-one entries by a subset of a Markov basis, in *Algebraic Methods in Statistics and Probability II*, M. Viana and H. Wynn (eds.), volume 516 of *Contemporary Mathematics*, American Mathematical Society, Providence RI.
- Holland, P.W. and Leinhardt, S. (1981). An an exponential family of probability distributions for random graphs, *Journal of the American Statistical Association*, 76(373), 33-50.
- Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data*, Springer.
- Lauritzen, S.L. (2003). Rasch models with exchangeable rows and columns, in *Bayesian Statistics 7*, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, (eds.), Oxford University Press, 215-232.
- Lauritzen, S.L. (2008). Exchangeable Rasch matrices, *Rendiconti di Matematica*, Series VII, 28, 83-95.
- Lovász, L. and Szegedy, B. (2006). Limits of dense graph sequences, *Journal of Combinatorial Theory Series B*, 96(6), 933-957.
- Mahadev, N.V.R. and Peled, U.N. (1995). *Threshold Graphs and Related Topics*, North Holland.
- Meyer, M.M. (1983). Transforming contingency tables, *Annals of Statistics*, 10(4), 1172-1181.
- Morton, J. (2008). "Relations among conditional probabilities," to appear in *J. Symbolic Computation*.
- Newman, MN. (2010). *Networks: An Introduction*, Oxford University Press.
- Newman, M., Barabási A.-L., and Watts, J. D. (eds.). (2006). *The Structure and Dynamics of Networks*, Princeton University Press.
- Newman, M.E.J. (2003). The structure and function of complex networks, *SIAM Review*, 45(2), 167-256.
- Newman, M.E.J., Strogatz, S. H. and Watts D. J. (2001). Random graphs with arbitrary degree distributions and their applications, *Physical Review E*, 026118.
- Ogawa, M., Hara, H. and Takemura, A. (2012). Graver basis for an undirected graph and its application to testing the beta model of random graphs, *Annals of Institute of Statistical Mathematics*, in press.
- Park, J. and Newman, M. E. (2004). Statistical mechanics of networks, *Physical Review E*, 066117.
- Petrović, S., Rinaldo, A. and Fienberg, S.E. (2010). Algebraic statistics for a directed random graph model with reciprocation, in *Algebraic Methods in Statistics and Probability*

- II, M. Viana and H. Wynn (eds.), volume 516 of *Contemporary Mathematics*, American Mathematical Society, Providence, RI.
- Perry, P.O. and Wolfe. P.J. (2012). Null models for network data, available at <http://arxiv.org/abs/1201.5871>
- Rinaldo, A., Fienberg, S. E. and Zhou, Y. (2009). On the geometry of discrete exponential families with application to exponential random graph models, *Electronic Journal of Statistics*, 3, 446–484.
- Robins, D., Pattison, P., Kalish, Y. and Lusher D. (2007). An introduction to exponential random graph (p^*) models for social networks, *Social Networks*, 29(2), 173–191.
- Schrijver, A. (1998). *Theory of Linear and Integer Programming*, Wiley, New York.
- Stanley, R.P. (1991). A zonotope associated with graphical degree sequences, *DIMACS Series in Discrete Mathematics*, 4, 555–570.
- Viger, F. and Latapy, M. (2005). Efficient and simple generation of random simple connected graphs with prescribed degree sequence, in *Lecture Notes in Computer Science*, 3595, 440–449, Springer, Berlin.
- Weibel, C. (2005). “Minksum version 1.1.” Mathematics Institute, EPF Lausanne, 2005, available at <http://roso.epfl.ch/cw/poly/public.php>.
- Willinger, W., Alderson, D. and Doyle, J. C. (2009). Mathematics and the internet: A source of enormous confusion and great potential, *Notices of the American Mathematical Society*, 56(5), 586–599.
- Yan, T. and Xu, J. (2012). A central limit theorem in the β -model for undirected random graphs with a diverging number of vertices, available at <http://arxiv.org/abs/1202.3307>
- Yan, T, Xu, J. and Yang Y. (2012). High dimensional Wilks phenomena in random graph models, available at <http://arxiv.org/abs/1201.0058>

DEPARTMENT OF STATISTICS
 CARNEGIE MELLON UNIVERSITY
 E-MAIL: arinaldo@cmu.edu

DEPARTMENT OF STATISTICS
 PENNSYLVANIA STATE UNIVERSITY
 E-MAIL: petrovic@stat.psu.edu

DEPARTMENT OF STATISTICS,
 MACHINE LEARNING DEPARTMENT,
 HEINZ SCHOOL AND CYLAB
 CARNEGIE MELLON UNIVERSITY
 E-MAIL: fienberg@stat.cmu.edu