Maximum Likelihood Estimation in Network Models

Alessandro Rinaldo*
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA, 15213-3890 USA

Sonja Petrović†
Department of Mathematics, Statistics, and Computer Science
University of Illinois
Chicago, IL, 60607

Stephen E. Fienberg‡
Department of Statistics
Heinz College
Machine Learning Department
Cylab
Carnegie Mellon University
Pittsburgh, PA 15213 USA

Abstract

We study maximum likelihood estimation for the statistical model for both directed and undirected random graph models in which the degree sequences are minimal sufficient statistics. In the undirected case, the model is known as the beta model. We derive necessary and sufficient conditions for the existence of the MLE that are based on the polytope of degree sequences, and wecharacterize 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 nodes increases. By reparametrizing the beta model as a log-linear model under product multinomial sampling scheme, we are able to provide usable algorithms for detecting nonexistence of the MLE and for identifying non-estimable parameters. We illustrate our approach on other random graph models for networks, such as the Rasch model, the Bradley-Terry model and the more general p_1 model of Holland and Leinhardt (1981).

Keywords: beta model, polytope of degree sequences, random graphs, Rasch model, p_1 model

1 Introduction

Virtually all models for network data rely, directly or indirectly, on the information contained in the degrees associated with the nodes of the corresponding graphs. The simplest instance of such a model is

*Email: arinaldo@cmu.edu
†Email: petrovic@math.uic.edu
‡Email: fienberg@stat.cmu.edu

the beta model (a named coined by Chatterjee et al., 2011), the exponential family of probability distributions for undirected random graphs for which the node degrees are natural sufficient statistics. Its relevance and use in the social sciences and in the physics literature is detailed and extensively reviewed by Newman et al. (2001), Newman (2003), Park and Newman (2004) and Blitzstein and Diaconis (2009), and references therein. The beta model is a simple undirected version of the p_1 class of statistical models for directed networks introduced by Holland and Leinhardt (1981), discussed later in section 6.4. In this article we address the issue of existence of maximum likelihood estimates of the probability parameters of the exponential family of probability distributions for both directed undirected random graphs for which the nodal degrees are natural sufficient statistics.

Lauritzen Lauritzen (2003, 2008) characterized beta models as the natural models for representing random binary symmetric arrays that are *weakly summarized,* i.e., random arrays whose distribution only depends on the row and column totals. The properties of these models are linked to the solutions of a certain system of functional equations of Rasch type, as well as to the properties of exchangeable and summarized doubly infinite random arrays. More recently, Chatterjee et al. (2011) have conducted an extensive analysis of the asymptotic properties of the beta model, including existence and consistency of the MLE as the dimension of the network increases, and have provided a simple algorithm for estimating the natural parameters. Furthermore, they have fully characterized the graph limits, or *graphons*, corresponding to a sequence of beta models with given degree sequence (for a connection between the theory of graphons and deFinetti's theorem for exchangeable arrays see Diaconis and Janson, 2007; Diaconis et al., 2008). Barvinok and Hartigan (2010) also explores the asymptotic behavior or sequences of random graphs with given degree sequences, proving that a different mode of convergence takes place. In their analysis, the Barvinok and Hartigan show that, as the size of the network increases, 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 following a beta model parametrized by the MLE corresponding to degree sequence. Blitzstein and Diaconis (2009) consider the problem of carrying out exact inference for the beta model and propose an algorithm for sampling from the set of graphs with given degree sequence (see also Viger and Latapay, 2005). The same problem is tackled also by Hara and Takemura (2010) and Ogawa et al. (2011) and by Petrović et al. (2010), who study Markov bases for the beta and the more general p_1 directed network model, respectively.

Here we investigate in detail the issue of existence of the MLE for the parameters of the beta model under a general sampling scheme in which each edge is observed a fixed number of times. Using the theory of exponential families, we provide necessary and sufficient conditions for existence of the MLE that are based on the polytope of degree sequences, a well-studied polytope arising in the study of threshold (see Mahadev and Peled, 1996). We show how nonexistence of the MLE is brought on by certain forbidden patterns of extremal network configurations, which we fully characterize in a combinatorial way. In particular, when the MLE does not exist, we can identify exactly which probability parameters are estimable. To illustrate our findings, we rely on the computational algebraic software `polymake` (see Gawrilow and Joswig, 2000) to compute the forbidden configurations leading to nonexistence of the MLE. Next, 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. When applied to the random graph model of fixed degree sequence, our asymptotic result sharpens the analogous result of Chatterjee et al. (2011). Our numerical experiments with `polymake` are based on re-expressing the beta model as a log-linear model under the product-multinomial sampling scheme. Though highly redundant, this reparametrization, which in polyhedral geometry is known as the Cayley embedding, has the crucial advantage of yielding geometric objects that are simpler to analyze, both computationally and theoretically. This approach 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.).

While we do not pursue a detailed treatment of the theoretical and algorithmic connections with the broader theory of log-linear models, for which the interested reader is referred to Fienberg and Rinaldo (2011), we repeatedly use the Cayley embedding device to analyze other network models that are variations on or generalizations of the beta model: the Rasch model, the Bradley-Terry model and the p_1 model. Our analysis illustrates a principled way for detecting nonexistence of the MLE and identifying non-estimable parameters that applies more generally to discrete models.

In addition to improving upon the results of Chatterjee et al. (2011) concerning the probability of an existent MLE for the beta model as the number of nodes increases, we exemplify the relevance and use of polyhedral geometry in dealing with nonexistence of the MLE and and identification of estimable parameters in discrete linear exponential families. For a more in-depth study of the geometric properties of log-linear models under general sampling schemes, we refer the reader to Fienberg and Rinaldo (2011).

We proceed as follows. In section 2 we describe a generalized version of the beta model in which we observe the edges of a graph a fixed number of times, possibly larger than one, and we express it as a natural exponential family with linear sufficient statistics. We obtain the beta model as a special case in which we observe edges only once. In section 3 we introduce the polytope of degree sequences and use it to derive necessary and sufficient conditions for the existence of the MLE. In particular, we characterize the patterns of edge counts for which the MLE does not exist, called co-facial sets. In section 3.1 we show a number of examples of co-facial sets, obtained using `polymake`. Furthermore, we use a result from Mahadev and Peled (1996) to show in section 3.2 how to construct virtually any example of random graphs for which the MLE of the beta parameters does not exist. In section 4 we once again use the polytope of degree sequences to obtain finite sample bounds on the probability that the MLE does not exist. As the number of objects to be compared increases, the MLE exists with probability approaching one. In section 5 we describe a general procedure for computing and identifying facial sets and, in section 6, we apply them to the Rasch model, a generalized beta model with no sampling restriction on the number of observed edges, the Bradley-Terry model and p_1 directed graph models.

Notation

For vectors x and y in the Euclidean space $\mathbb{R}^n$, we will denote with x_i the value of x at its i-th coordinate and with $\langle x, y\rangle := x^\top y = \sum_i x_i y_i$ their standard inner product. Operations on vector will be performed element-wise. For a matrix A, $\mathrm{convhull}(\mathrm{A})$ and $\mathrm{cone}(\mathrm{A})$ denote the set of all convex and conic combinations of the columns of A, respectively. For a polyhedron P, we denote with $\mathrm{ri}(P)$ its relative interior. 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 Beta Model

In this section we describe the beta model of Chatterjee et al. (2011) and introduce the exponential family parametrization we will be using throughout the entire article.

The beta model focuses on the occurrence of edges in a simple undirected graph, whose nodes are labeled $\{1,\ldots,n\}$, for convenience, and whose edges are $\{(i,j), i<j\}$. The associated statistical experiment consists of recording, for each pair of nodes (i,j) with $i<j$, the number of edges appearing in $N_{i,j}$ distinct observations, where the integers $\{N_{i,j}, i<j\}$ are deterministic and strictly bigger than zero (the non-randomness and positivity assumptions can in fact be relaxed, as shown in section 6.2 and 7, respectively). For $i<j$, we denote with $x_{i,j}$, the number of times edge (i,j) was observed and, accordingly, with $x_{j,i}$ the number of times object edge (i,j) was missing. Thus, for all (i,j),

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

This is the natural heterogenous version of the well-known Erdös-Rényi random graph model (Erdös and Rényi, 1959). For a more general discussion of this model and its generalizations see **?**.

The observed edge counts $\{x_{i,j}, i<j\}$ are modeled as draws from mutually independent binomial distributions, with $x_{i,j} \sim \mathrm{Bin}(N_{i,j}, p_{i,j})$, where $p_{i,j} \in (0,1)$ for each $i<j$. Accordingly, $x_{j,i} = N_{i,j} - x_{i,j}$ has a $\mathrm{Bin}(N_{i,j}, p_{j,i})$ distribution, where $p_{j,i} = 1 - p_{i,j}$, for all $i<j$.

Data arising from such an experiment can be naturally represented through 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, it is enough to consider the upper-triangular part of this contingency table. Indeed, since given $x_{i,j}$ with $i < j$, the value of $x_{j,i}$ is determined by $N_{i,j} - x_{i,j}$, the set of all possible outcomes can be represented more parsimoniously as the following subset of $\mathbb{N}^{\binom{n}{2}}$:

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

We will adopt the convention of indexing the coordinates $\{(i,j) \colon i < j\}$ of any point $x \in \mathcal{S}_n$ lexicographically.

We parametrize the beta model by points in $\mathbb{R}^n$, so that for each $\beta \in \mathbb{R}^n$, the probability parameters are represented as

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

or, equivalently, in term of odds ratios,

$$\log \frac{p_{i,j}}{p_{j,i}} = \beta_i + \beta_j, \quad \forall i \neq j.$$

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

$$\prod_{i<j} \binom{N_{i,j}}{x_{i,j}} p_{i,j}^{x_{i,j}} (1 - p_{i,j})^{N_{i,j} - x_{i,j}}, \tag{2}$$

with the probability values $p_{i,j}$ satisfying (15). Simple algebra shows that this probability can be written in exponential family form as

$$\exp\left\{\sum_{i=1}^{n} d_i \beta_i - \psi(\beta)\right\} \prod_{i<j} \binom{N_{i,j}}{x_{i,j}}, \tag{3}$$

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

$$d_i = \sum_{j<i} x_{j,i} + \sum_{j>i} x_{i,j}, \quad i = 1, \ldots, n, \tag{4}$$

and the log-partition function $\psi \colon \mathbb{R}^n \to \mathbb{R}$ by

$$\beta \mapsto \sum_{i<j} N_{i,j} \log\left(1 + e^{\beta_i + \beta_j}\right).$$

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 on $\mathcal{S}_n$ (see, e.g. Barndorff-Nielsen, 1978) with densities given by the exponential term in (3).

Random graphs with fixed degree sequence

In the special case in which $N_{i,j} = 1$ for all (i,j), the support $\mathcal{S}_n$ reduces to the set $\mathcal{G}_n := \{0,1\}^{\binom{n}{2}}$, which encodes the set of 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 beta 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 (15). Then, by (4), the i-th minimal sufficient statistic d_i is the *degree* of node i, i.e. the number of nodes adjacent to i. The vector $d(x)$ of sufficient statistics is known as the *degree sequence* of the observed graph x.

The Rasch model

The Rasch model (see, e.g., Rasch, 1960; Andersen, 1980), one of the most popular statistical models used in item response theory and in educational tests, is concerned with modeling the joint probabilities that k subjects provide correct answers to a set of l items. The Rasch model can be recast as a random bipartite graph model with sufficient statistics given by the node degrees, where, without loss of generality, the bipartition of the nodes consists of the sets $I := \{1, \ldots, k\}$ and $J := \{k+1, n-1, n\}$, with $k \geq 2$ and $l := n - k \geq 2$. In this model, the set I represents the subjects and the set J the items, and edges can only be of the form (i, j), with $i \in I$ and $j \in J$. The sample space is given by the set $\mathcal{R}_n = \{0, 1\}^{kl}$, and the vector $x \in \{x_{i,j}, i \in I, j \in J\} \in \mathcal{R}_n$ encodes the bipartite graphs in which the edge (i, j) is present if and only if $x_{i,j} = 1$ if and only if subject i answered correctly to item j.

3 Existence of the MLE

We derive a necessary and sufficient condition for the existence of the MLE of the natural parameter $\beta \in \mathbb{R}^n$ and, equivalently, of the probability parameters $\{p_{i,j}, i < j\}$. Notice that nonexistence of the MLE entrails, in the case of the natural parameters, that the supremum of the likelihood function (3) cannot be attained by any finite vector in $\mathbb{R}^n$, and, in the case of the probability parameters, that the supremum of (2) cannot be attained by any set of probability values bounded away from 0 and 1.

To determine when the MLE exists, we first introduce a geometric object that will play a key role throughout the rest of the paper. First, note that the vector of sufficient statistics $d(x)$ for the beta model, for each $x \in \mathcal{S}_n$, can be obtained as

$$d(x) = \mathrm{A}x$$

where A is the $n \times \binom{n}{2}$ design matrix consisting of the node-edge incidence matrix of a complete graph on n nodes. Specifically, the rows of A are indexed by the object labels $i \in \{1, \ldots, n\}$, and the columns are indexed by the set of all pairs (i, j) with $i < j$, ordered lexicographically. The entries of A are ones along the coordinates $(i, (i, j))$, when $i < j$ and $(i, (j, i))$ when $j < i$, 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 the columns are indexed by the pairs $(1, 2)$, $(1, 3)$, $(1, 4)$, $(2, 3)$, $(2, 4)$, and $(3, 4)$. In particular, as pointed out above, for any undirected simple graph $x \in \mathcal{G}_n$, $\mathrm{A}x$ is the associated degree sequence.

The *polytope of degree sequences* P_n is the convex hull of all possible degree sequences, i.e.

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

The integral polytope P_n is a well-studied object: see Chapter 3 in Mahadev and Peled (1996). In the language of algebraic statistics, P_n is called the *model polytope* (see Sturmfels and Welker, 2011). 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 $n \geq 3$, $\dim(P_n) = n$.

The main result in this section is to show that existence of the MLE for the beta model can be fully characterized using the polytope of degree sequences in the following fashion. For any $x \in \mathcal{S}_n$, let

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

denote the frequency of wins of i over j and set $\tilde{d} = \tilde{d}(x) \in \mathbb{R}^n$ to be the vector with coordinates

$$\tilde{d}_i := \sum_{j<i} \tilde{p}_{j,i} + \sum_{j>i} \tilde{p}_{i,j}, \quad i = 1, \ldots, n. \tag{5}$$

Notice that, $\tilde{d}$ is a just a rescaled version of the sufficient statistics (4), normalized by the number of observations. It is also clear that, 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 $\tilde{d}(x) \in \mathrm{int}(P_n)$.*

Remark
Theorem 3.1 verifies the conjecture contained in Addenda 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). The theorem also confirms the observation made by Chatterjee et al. (2011) that the MLE never exists if $n = 3$: since P_3 has exactly 8 vertices, as many as possible graphs, no degree sequence can be inside P_3.

A significant consequence of the geometric nature of Theorem 3.1 is the possibility of characterizing the patterns of observed edge counts that cause nonexistence of the MLE. This is done in the next result.

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*

$$y = \mathrm{A}p, \tag{6}$$

where $p = \{p_{i,j} \colon i < j, p_{i,j} \in [0,1]\}$ is such that $p_{i,j} \in \{0,1\}$ if $(i,j) \not\in \mathcal{F}$ 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 a maximal set for which (6) holds.

Following Geiger et al. (2006) and Fienberg and Rinaldo (2011), we call any such set $\mathcal{F}$ a facial set of S_n and its complement, $\mathcal{F}^c = \{(i,j) \colon i < j\} \setminus \mathcal{F}$, a co-facial set. Facial sets form a lattice that is isomorphic to the face lattice of P_n as shown by Fienberg and Rinaldo (2011, Lemma 3.4). This means that the faces of S_n are in one-to-one correspondence with the facial sets of S_n and, for any pair of faces F and F' of S_n with associated facial sets $\mathcal{F}$ and $\mathcal{F}'$, $F \cap F'$ if and only if $\mathcal{F} \cap \mathcal{F}' = \emptyset$ and $F \subset F'$ if and only if $\mathcal{F} \subset \mathcal{F}'$. In particular, the facial set corresponding to S_n is the set $\{(i,j) \colon i < j\}$.

Facial sets are combinatorial objects that 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 entries on the contingency table that are either 0 or $N_{i,j}$. In particular, the MLE does not exist if and only if the set $\{(i,j) \colon i < j, x_{i,j} = 0 \text{ or } N_{i,j}\}$ contains a co-facial set. Secondly, 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 (2) reveals that, for any observable set of counts $\{x_{i,j} \colon i < j\}$, there always exists a unique set of maximizers $\widehat{p} = \{\widehat{p}_{i,j}, i < j\}$ which, by strict concavity, are uniquely determined by the first order optimality conditions

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

also known as the moment equations. Existence of the MLE is then equivalent to $0 < \widehat{p}_{i,j} < 1$ 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 $\{\widehat{p}_{i,j}, i < j\}$, known as the extended MLE, are no longer strictly between 0 and 1. Instead, by Lemma (3.2), the extended MLE is such that $\widehat{p}_{i,j} = \tilde{p}_{i,j} \in \{0,1\}$ for all $(i,j) \in \mathcal{F}^c$. Furthermore, it is possible to show (see, e.g., Morton, 2008) that $\widehat{p}_{i,j} \in (0,1)$ 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.

Therefore, 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 (2011).

Below, we further investigate the properties of P_n and provide several examples of co-facial sets associated to the facets of P_n.

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

Table 1: Example of a co-facial set leading to a nonexistent MLE.

$\times$	0	1	2
3	$\times$	2	1
2	1	$\times$	3
1	2	0	$\times$

$\times$	0	0.5	0.5
1	$\times$	0.5	0.5
0.5	0.5	$\times$	1
0.5	0.5	0	$\times$

Table 2: Left: data exhibiting the pattern reported in Table 1, when $N_{i,j} = 3$ for all $i \neq j$. Right: table of the extended MLE of the estimated probabilities. Under natural parametrization, the supremum of the log-likelihood is achieved in the limit for any sequence of natural parameters $\{\beta^{(k)}\}$ of the form $\beta^{(k)} = (-c_k, -c_k, c_k, c_k)$, where $c_k \to \infty$ as $k \to \infty$.

$\times$	2	1	2
1	$\times$	0	1
2	3	$\times$	3
1	2	0	$\times$

$\times$	0.225	0.384	0.725
0.775	$\times$	0.225	0.551
0.616	0.775	$\times$	0.725
0.275	0.449	0.275	$\times$

Table 3: Left: same data as in Table 2, but with the values for the cells $(1,2)$ and $(2,3)$ switched with the values in the cells $(2,1)$ and $(3,2)$, respectively. Right: table of probabilities at which the log-likelihood is optimal. The MLE of the natural parameters are $\beta = (-0.237, -1.002, -0.237, 1.205)$.

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

Table 4: Example of a co-facial set leading to a nonexistent MLE. In this case $\tilde{d}_2 = 0$.

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

Table 5: Example of a co-facial set leading to a nonexistent MLE. In this case the second row sum is 0.

3.1 The Co-facial Sets of P_n

Theorem 3.1 and Lemma 3.2 both show that the boundary of the polytope P_n plays a fundamental role in determining the existence of the MLE for beta models and in specifying which parameters are estimable.

Mahadev and Peled (1996) derived the facet-defining inequalities of P_n, for all $n \geq 4$. For reader's convenience, we report this result below. Let $\mathcal{P}$ be the set of all pairs (S, T) of disjoint non-empty subsets of

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

Table 6: Example of a co-facial set leading to a nonexistent MLE.

$\{1, \ldots, n\}$, such that $|S \cup T| \in \{2, \ldots, n-3, n\}$. For any $(S,T) \in \mathcal{P}$ and $y \in P_n$, let

$$g(S,T,y,n) := |S|(n-1-|T|) - \sum_{i \in S} y_i + \sum_{i \in T} y_i. \tag{7}$$

Theorem 3.3 (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* $i = 1, \ldots, n$;

(ii) $y_i \leq n-1$, *for* $i = 1, \ldots, n$;

(iii) $g(S,T,y,n) \geq 0$, *for all* $(S,T) \in \mathcal{P}$.

The combinatorial complexity of the face lattice of an n-dimensional polytope can be summarized by its f-vector, a vector of length $n+1$ whose i-th entry is the number of i-dimensional faces, $i = 0, \ldots, n$. Stanley (1991) studies the number faces of the polytope of degree sequences P_n and derives an expression for computing the entries of the f-vector of P_n. For example, the f-vector of P_8 is the 9-dimensional vector

$$(334982, 1726648, 3529344, 3679872, 2074660, 610288, 81144, 3322, 1),$$

so 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 `polymake`, using the methods described in section 5).

Despite the fact that much is known about P_n, the number of facet-defining inequalities appears to be exponential in n and, consequently, the tasks of identifying points on the boundary of P_n and the associated facial set remain computationally challenging. In section 5, we discuss these difficulties and propose a solution for detecting boundary points and the associated facial sets that is based on a log-linear model reparametrization. Using the methods and computations described in that section, we were able to identify few interesting cases in which the MLE is nonexistent, some of which seem to be unaccounted for in the statistical literature. Below we describe some of those cases.

Recall that the data can be represented as a $n \times n$ table of counts, in which the diagonal elements are expunged and where the $(i,j)-th$ entry of the table indicates the number of times, out of $N_{i,j}$, in which the edges (i,j) was observed. In our examples, empty cells correspond to facial set and may contain any count values, in contrast to the cells in the co-facial sets that contain either a zero value or a maximal value, namely $N_{i,j}$. As we say in Lemma 3.2, extreme count values of this nature are precisely what leads to a nonexistent MLE.

Table 1 provides an instance of a co-facial set, which corresponds to a facet of P_4. Assume for simplicity that each of 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 2, which shows, on the left, an instance of data with $N_{i,j} = 3$, for all $i \neq j$ satisfying the pattern indicated in Table 1 and, on the right, the probability values maximizing the log-likelihood function. Since the MLE does not exist, some of these probability values are 0 and 1. The order of the pattern is crucial. Indeed, Table 3 shows, on the left, data containing precisely the same counts as in Table 2, but with the values in cells $(1,2)$ and $(2,3)$ switched with the values in cell $(2,1)$ and $(3,2)$, respectively. On the left of Table 3 the MLE of the cell probabilities are shown; as the MLE exists, they are bounded away from 0 and 1.

In Table 4 we show another example of 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 3.3, the MLE does not exist if $\tilde{d}_i = 0$ or $\tilde{d}_i = n-1$, for some i. Table 5 shows yet one more example of a co-facial set that is easy to detect, as it leads to a zero row margin for the second row. Finally, Table 6 provides one more example of a co-facial set, which unlike the ones in Tables 4 and 5, has normalized row sums and the normalized sufficient statistics bounded away from 0 and $n-1$. In Table 7 we list all 22 co-facial sets associated with the facets of P_n, including the cases already shown in Tables 1, 4, 5 and 6.

In general, there are $2n$ facets of P_n that are determined by $\tilde{d}_i$ equal to 0 or $n-1$ and $2n$ other facets associated to values of the normalized row sums equal to 0 or $n-1$. Thus, just by inspecting the row sums or the observed sufficient statistics, one can detect $4n$ co-facial sets associated to as many facets of P_n. However, comparing this number to the entries of the f-vector calculated in Stanley (1991) 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 appear to grow exponentially in n, we conclude that most of the co-facial sets do not appear to arise in this fashion, and methods for detecting them are called for. We discuss them in section 5.

3.2 Random Graphs with Nonexistent MLEs

When dealing with the special case of $N_{i,j} = 1$ for all $i < j$, which we showed to be equivalent to a model for random graphs with independent edges and node degrees as minimal sufficient statistics, points on the boundary of P_n are, by construction, degree sequences and have a direct graph-theoretical interpretation, as shown in the next result.

Lemma 3.4 (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 $\{1,\ldots,n\}$ 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 $\mathcal{G}$.*

A direct consequence of lemma 3.4 is that the MLE does not exists if the observed network is a *split graph,* i.e. a graph whose node sets can be partitioned into a clique S and a stable set T. More generally. Lemma 3.4 can be used to create virtually any example of random graphs with fixed degree sequences for which the MLE does not exist. Notice that, in particular, having node degrees bounded away from 0 and $n-1$ is not a sufficient condition for the existence of the MLE (though its violation implies nonexistence of the MLE). We point out that, in order to detect boundary points and the associated co-facial sets, Lemma 3.4 is, however, of little help. Instead, one can use the procedures described in section 5.

Below, we provide some examples of co-facial sets for random graphs with fixed degree sequences for which the MLE does not exist, yet the node degrees are bounded from 0 and $n-1$.

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 8, which we computed numerically using the procedure described in section 5. 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 7. 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 7 that any co-facial set containing three zero counts and three maximal

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Table 7: All possible co-facial sets for P_4 (empty cells indicate any entry values).

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.

Table 9 shows an observed graph with degrees all larger than 0 and less than 3 but for which the MLE does not exist. Notice that the co-facial set corresponds to the one shown in the upper left corner of Table 8. Finally, Tables 9 and 10 show two more examples of random graphs on $n = 5$ and $n = 6$ nodes, respectively, for which the MLE does not exist (by Lemma 3.4), and yet the degrees are such that $0 < d_i < n - 1$ for all i.

$\times$	0		
1	$\times$		
		$\times$	1
		0	$\times$

$\times$		0	
	$\times$		
1		$\times$	1
		0	$\times$

$\times$			1
	$\times$	0	
	1	$\times$	
0			$\times$

$\times$	1		
0	$\times$		
		$\times$	0
		1	$\times$

$\times$		1	
	$\times$		
0		$\times$	0
		1	$\times$

$\times$			0
	$\times$	1	
	0	$\times$	
1			$\times$

Table 8: 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).

$\times$	0	1	0
1	$\times$	0	1
0	1	$\times$	1
1	0	0	$\times$

Table 9: Random graph with node degrees larger than 0 and smaller than 3 exhibiting the same co-facial set show in the upper left corner of Table 8. In this case, lemma 3.4 applies with $S = \{3, 4\}$ and $T = \{1, 2\}$.

$\times$	1	0	0	0
0	$\times$	1	1	0
1	0	$\times$	1	0
1	0	0	$\times$	1
1	1	1	0	$\times$

Table 10: Network with $n = 5$ for which the MLE does not exist and the degrees are bounded away from 0 and 4. In this case, lemma 3.4 applies with $S = \{2, 3, 4\}$ and $T = \{1, 5\}$.

$\times$	1	0	1	1	1
0	$\times$	1	0	0	1
1	0	$\times$	0	0	0
0	1	1	$\times$	0	0
0	1	1	1	$\times$	0
0	0	1	1	1	$\times$

Table 11: Network with $n = 6$ for which the MLE does not exist and the degrees are bounded away from 0 and 5. In this case, lemma 3.4 applies with $S = \{1, 2, 6\}$ and $T = \{3, 4, 5\}$.

4 Existence of the MLE: Asymptotics

In this section we derive sufficient conditions that imply existence of the MLE with large probability as n grows. We will make the simplifying assumption that $N_{i,j} = N_n$, for all i and j, where $N_n \geq 1$ could itself depend on n. In studying Bradley-Terry models, Simons and Yao (1999), for instance, study the asymptotic scenario of a fixed number $N_n = N$ of pairwise comparisons.

Recall the random vector $\tilde{d}$, whose coordinate are given in (5) and set $\overline{d} = \mathbb{E}[\tilde{d}] \in \mathbb{R}^n$. Then

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

We formulate our sufficient conditions in terms of the entries of the vector $\overline{d}$.

Theorem 4.1. *Assume that, for all $n \geq \max\{4, 2\sqrt{c\frac{n \log n}{N}} + 1\}$, the vector $\overline{d}$ satisfies the conditions*

(i) $\min_i \min\left\{\overline{d}_i, n-1-\overline{d}_i\right\} \geq 2\sqrt{c\frac{n\log n}{N}} + C$*,*

(ii) $\min_{(S,T)\in\mathcal{P}} g(S,T,\overline{d},n) > |S \cup T|\sqrt{c\frac{n\log n}{N}} + C$*,*

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_n is constant, for instance when $N_n = 1$, as in the random graph case, the conditions of Theorem 4.1 can be relaxed 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 $n \geq N_n$, we do not expect it to be sharp in general when N_n grows with n.

Corollary 4.2. *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 $\overline{d} = \mathbb{E}[\tilde{d}] \in \mathbb{R}^n$ satisfies the conditions*

(i') $\min_i \min\left\{\overline{d}_i, n-1-\overline{d}_i\right\} \geq 2\sqrt{cn\log n} + C$*;*

(ii') $\min_{(S,T)\in\mathcal{P}_n} g(S,T,\overline{d},n) > |S \cup T|\sqrt{cn\log n} + C$*,*

where

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

where the set $\mathcal{P}$ was defined before Theorem 3.3. Then, the MLE exists with probability at least $1 - \frac{2}{n^{2c-2}}$. If $N = 1$, it is enough to have $c > 1/2$, and the MLE exists with probaiblity larger than $1 - \frac{2}{n^{2c-1}}$

Remarks

1. It is clear that, asymptotically, the value of the constant C becomes irrelevant, as the constraints on its range will be satisfied by any positive C, for all n large enough.

2. Since $|S \cup T| \leq n$, one could replace assumption *(ii)* of Theorem 4.1 with the simpler but stronger condition

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

   Then, assuming for simplicity that N is a constant, turning Theorem 4.1 into asymptotic statement, the MLE exists with probability tending to one at a rate that is polynomial in n whenever

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

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

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

3. For the case $N_n = 1$, corollary 4.2 should be compared with theorem3.1 in Chatterjee et al. (2011), which also provides sufficient conditions guaranteeing the existence of the MLE with probability no smaller than $1 - \frac{1}{n^{2c-1}}$ (for all n large enough), but appear to be stronger than ours. Explicitly, those 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
$$|S|(|S|-1) - \sum_{i \in S} d_i + \sum_{i \notin S} \min\{d_i, |S|\} > c_3 n^2 \tag{8}$$
for all sets S such that $|S| > (c_1)^2 n^2$. It is easy to see that, for any non-empty subsets $S \subset \{1,\ldots,n\}$ and $T \subset \{1,\ldots,n\} \setminus S$,
$$\sum_{i \notin S} \min\{d_i, |S|\} \leq \sum_{i \in T} d_i + |S||(S \cup T)^c|,$$
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 (8) holds for some non-empty $S \subset \{1,\ldots,n\}$, it satisfies the facet conditions implied by all the pairs (S,T), for any non-empty set $T \subset \{1,\ldots,n\} \setminus S$. As a result, for any subset S, (8) is a stronger condition than any of the facet conditions of P_n specified by S. In addition, we weakened significantly their requirements that $|S| > (c_1)^2 n^2$ and $c_1(n-1) < d_i < c_2(n-1)$ for all i to $|S| > \sqrt{cn \log n} + C$ and $\min_i \min\left\{\overline{d}_i, n-1-\overline{d}_i\right\} \geq 2\sqrt{cn\log n} + C$, respectively.

4. Theorem 1.3 in Chatterjee et al. (2011) shows that, when the MLE exists, $\max_i |\widehat{\beta}_i - \beta_i| = O(\sqrt{n \log n})$, with probability at least $1 - \frac{2}{n^{2c}-1}$.

5 Computations

In this section, we describe the procedure we used to compute the facial sets of P_n. The main difficulties with working directly with P_n is that this polytope arises a Minkowksi sum and, even though the system of defining inequalities is given explicitly, its combinatorial complexity grows exponentially in n. Furthermore, we do not have available a set of vertices for P_n. Algorithms for obtaining the vertices of P_n, such as `minksum` (see Weibel, 2005), are computationally expensive and require generating all the points $\{\mathrm{A}x, x \in \mathcal{G}_n\}$, where $|\mathcal{G}_n| = 2^{\binom{n}{2}}$. In general, when n is as small as 10, this is not feasible.

Our basic strategy to overcome these problems is quite simple, and entails representing the beta model as a log-linear model with $\binom{n}{2}$ product-multinomial sampling constraints. 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 its use in the analysis of log-linear models is described in Fienberg and Rinaldo (2011). The advantages of this re-parametrization are two-fold. First, it allows us to use the highly optimized algorithms available in `polymke` for listing explicitly all the facial sets of P_n, which is the strategy we used. Secondly, the general algorithms for detecting nonexistence of the MLE and identifying facial sets proposed in Fienberg and Rinaldo (2011), which can handle larger dimensional models, can be directly applied to this problem. This reference is also relevant for dealing with inference under a non-existent MLE.

In the interest of space, we do not provide all the details, and instead only sketch the two main steps of our procedure.

- **Step 1: Enlarging the space**
  In the first step, we switch to a redundant representation of the data by considering all the observed

counts $\{x_{i,j}, i \neq j\}$ and not just $\{x_{i,j}, i < j\}$. We index the points of this enlarged set of $n(n-1)$ numbers as pairs $\mathcal{S}'_n = \{(x_{i,j}, x_{j,i})\colon i < j\} \subset \mathbb{N}^{n(n-1)}$, with the pairs ordered lexicographically based on (i,j). For instance, when $n = 4$, any point $x' \in \mathcal{S}'_4$ has coordinates indexed by

$$(1,2),(2,1),(1,3),(3,1),(1,4),(4,1),(2,3),(3,2),(2,4),(4,2),(3,4),(4,3).$$

It is clear that the sets $\mathcal{S}_n$ and $\mathcal{S}'_n$ are in one-to-one correspondence with each other and that, for each corresponding pair $x \in \mathcal{S}_n$ and $x' \in \mathcal{S}'_n$, $x'_{i,j} = x_{i,j}$ for all $i < j$ and $x'_{j,i} = N_{i,j} - x_{i,j}$ for all $j > i$.

In this new setting, we construct a new polytope $P'_n \subset \mathbb{R}^{2n}$ that is combinatorially equivalent to P_n but whose facial sets are easier to interpret. This is achieved by first constructing a new design matrix B of dimension $(2n) \times n(n-1)$, with the columns indexed according to the order described above. The matrix B has the form

$$\mathrm{B} = \begin{pmatrix} \mathrm{B}_1 \\ \mathrm{B}_2 \end{pmatrix} \tag{9}$$

where both B_1 and B_2 have n rows. For all $i < j$, the columns of B_1 corresponding to the coordinate (i,j) and the columns of B_2 corresponding to the coordinate (j,i) are both equal to $a_{i,j}$, and all the other columns are zeros. For instance, when $n = 4$,

$$B = \begin{bmatrix} 1&0&1&0&1&0&0&0&0&0&0&0 \\ 1&0&0&0&0&0&1&0&1&0&0&0 \\ 0&0&1&0&0&0&1&0&0&0&1&0 \\ 0&0&0&0&1&0&0&0&1&0&1&0 \\ 0&1&0&1&0&1&0&0&0&0&0&0 \\ 0&1&0&0&0&0&0&1&0&1&0&0 \\ 0&0&0&1&0&0&0&1&0&0&0&1 \\ 0&0&0&0&0&1&0&0&0&1&0&1 \end{bmatrix}.$$

By construction, $d = \mathrm{A}x = \mathrm{B}_1 x'$ for any corresponding pair $x \in \mathcal{S}_n$ and $x' \in \mathcal{S}'_n$. Furthermore, if we let $d' = \mathrm{B}_2 x'$, it is easy to see that d' and d are in one-to-one correspondence with each other. Indeed, recalling that $N_{i,j} = N_{j,i}$,

$$\begin{aligned} d'_i &= \textstyle\sum_{j<i} x'_{i,j} + \sum_{j>i} x'_{j,i} \\ &= \textstyle\sum_{j<i} (N_{i,j} - x_{j,i}) + \sum_{j>i} (N_{i,j} - x_{i,j}) \\ &= \textstyle\sum_{j\neq i} N_{i,j} - \left(\sum_{j<i} x_{j,i} + \sum_{j>i} x_{i,j} \right) \\ &= \textstyle\sum_{j\neq i} N_{i,j} - d_i, \end{aligned}$$

where we used (4) in the last step. Thus, $\mathrm{B}x'$ is also a sufficient statistic, though highly redundant due to linear dependencies. Next, for any $i < j$, let

$$B_{i,j} = \mathrm{convhull}(\{b_{i,j}, b_{j,i}\})$$

where $b_{i,j}$ is the column of B indexed by (i,j), and set

$$P'_n = \sum_{i<j} B_{i,j}.$$

The polytopes P_n and P'_n are combinatorially equivalent, even though their ambient dimensions are different. In fact, using arguments similar to the ones used in the proof of corollary 3.2, one can characterize the facial sets of P'_n as follows.

Lemma 5.1. *A point y' belongs to the interior of some face F' of P'_n if and only if there exists a set $\mathcal{F}' \subset \{(i,j), i \neq j\}$ such that*

$$y' = \mathrm{B}p', \tag{10}$$

where $p' = \{p'_{i,j}\colon i \neq j, p'_{i,j} \in [0,1], p'_{i,j} = 1 - p'_{j,i}\}$ is such that $p'_{i,j} = 0$ for all $(i,j) \notin \mathcal{F}'$ and $p'_{i,j} > 0$ for all $(i,j) \in \mathcal{F}$. The set $\mathcal{F}$ is uniquely determined by the face F and is a maximal set for which (10) holds.

Because P_n and P'_n are combinatorially equivalent, their co-facial sets are also in one-to-one correspondence. The advantage of using P'_n instead of P_n is that its co-facial sets arise by entries of p' that are all zeros, as opposed to the more complicated co-facial sets of P_n, which are obtained from entries of $p = \{p_{i,j} : i < j\}$ which are both ones and zeros. For instance, the co-facial set of P_n corresponding to the counts reported in Table 1 is $\{(1,2),(3,4)\}$ with $p_{1,2} = 0$ and $p_{3,4} = 1$. In contrast, the corresponding co-facial set for P'_n is $\{(1,2),(4,3)\}$, with $p'_{1,2} = 0$ and $p'_{4,3} = 0$. Clearly, they convey the same information.

- **Step 2: Lifting**
As we saw, the advantage of the larger polytope P'_n derived in the first step is that, when searching for co-facial sets, it is enough to consider points of the form $p' = \{p'_{i,j} : i \neq j, p'_{i,j} \in [0,1]\}$ with zero coordinates only. However, P'_n is still a hard object to deal with computationally, since it is prescribed as a Minkowski sum of $\binom{n}{2}$ polytopes. In this second step, we lift P'_n to a polyhedral cone in dimension $2n + \binom{n}{2}$ which is simpler to analyze (in fact, as remarked below, this polyhedral cone has smaller dimension: $n + \binom{n}{2}$). This cone is spanned by the columns of a matrix C of dimension $(2n + \binom{n}{2}) \times n(n-1)$ which has the form

$$\mathrm{C} = \begin{pmatrix} \mathrm{C}_1 \\ \mathrm{B} \end{pmatrix},$$

where the rows of C_1 are indexed by the pairs $\{(i,j) : i < j\}$ ordered lexicographically. Each row (i,j) of C_1 contains all zeroes, except for two ones in the coordinates (i,j) and (j,i). In fact for any $x' \in \mathcal{S}_n$, the vector $\mathrm{C}_1 x'$ is constant, and its (i,j)-the entry is

$$x'_{i,j} + x'_{j,i} = N_{i,j}.$$

For instance, when $n = 4$,

$$\mathrm{C} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Let $D_n = \mathrm{cone}(\mathrm{C})$ be the polyhedral cone of spanned by the columns of C. The facial sets of D_n are defined as follows (see, e.g., Geiger et al., 2006). The subset $\mathcal{F} \subset \{(i,j) : i \neq j\}$ is a facial set of D_n when there exists a $v \in \mathbb{R}^{2n+\binom{n}{2}}$ such that

$$\langle v, c_{i,j} \rangle = 0, \quad \forall i \in \mathcal{F} \quad \text{and} \quad \langle v, c_{i,j} \rangle < 0, \quad \forall i \notin \mathcal{F},$$

where $c_{i,j}$ indicates the column of C indexed by the pair (i,j). It follows that F is face of D_n if and only if $F = \mathrm{cone}(\{c_i : i \in \mathcal{F}\})$, for some facial set $\mathcal{F}$ of D_n, and that there is a one-to-one correspondence between the facial sets and the faces of D_n. Thus, as before, facial sets form a lattice isomorphic to the face lattice of D_n. Following Eriksson et al. (2006), we will call D_n the *marginal cone*.

The following result shows how one can obtain the facial sets of P_n from the facial set of D_n through the facial sets of P'_n (see also section 3 in Fienberg and Rinaldo (2011)).

Theorem 5.2. *Let* $p' = \{p'_{i,j} \colon i \neq j, p'_{i,j} \in [0,1], p'_{i,j} = 1 - p'_{j,i}\}$. *Then* $\mathrm{B}p' \in \mathrm{ri}(P'_n)$ *if and only if* $\mathrm{C}p' \in \mathrm{ri}(D_n)$. *Furthermore, if* $\mathcal{F}'$ *is a facial set of* P'_n*, then* $\mathcal{F}'$ *is a facial set of* D_n.

In particular, the only facial sets of D_n that are not facial sets of P'_n are the ones corresponding to the supports of the first $\binom{n}{2}$ rows of C, so that D_n has $\binom{n}{2}$ more facets than P_n (and P'_n). Since, by construction $x'_{i,j} + x_{j,i} = N_{i,j}$, $\mathrm{C}x'$ will never be a point in the interior of the $\binom{n}{2}$ facets of D whose facial sets are the supports of the first $\binom{n}{2}$ rows of C.

Theorem 5.2 can be used as follows. The MLE exists if and only if $\mathrm{C}x' \in \mathrm{ri}(D_n)$. When the MLE does not exist, the corresponding facial set of D_n gives the required facial set for P'_n and, therefore, for P_n.

Finally, it is clear to see that C is rank-deficient due to linear dependencies among the rows, so one could instead consider the marginal cone spanned by the columns of the matrix

$$\begin{pmatrix} \mathrm{C}_1 \\ \mathrm{B}_1 \end{pmatrix}, \tag{11}$$

which has full dimension $\binom{n}{2} + n$ and is combinatorially equivalent to D_n.

The final result of the two-step procedure just outlined is a reparametrization of the beta model in the form of a log-linear model with full-rank design matrix given in (11) and Poisson sampling scheme. The constrains on the number of observed edges translate into $\binom{n}{2}$ product-multinomial sampling restrictions for this log-linear model. However, it is well known that the conditions for existence of the MLE are the same under Poisson and product-multinomial scheme, so whether we incorporate these constraints or not has no bearing on parameter estimability. See Haberman (1974, Chapter 2) and Fienberg and Rinaldo (2011, section 3.4).

The examples of co-facial sets were obtained by first computing the matrix (11) and then using `polymake` to compute the facial sets of the resulting marginal cone[1] For a detailed description of the connection with log-linear models, and for algorithms to compute the facial sets of this cone that can be used in higher dimensions, the reader is referred to Fienberg and Rinaldo (2011).

Finally, to deal with the Rasch model, the procedure can be trivially modified by eliminating the columns of A and, in particular, of C' corresponding to all the edges between the sets I and J comprising the bipartition of the node set. In particular, the resulting matrix C' has dimension $(kl + k + l) \times 2kl$ and has rank $kl + k + l - 1$, where k and l are the cardinalities of I and J, respectively.

Algorithms

We first indicate how the non-existence of the MLE and the determination of the appropriate facial set can be addressed using simple linear programming. While checking for the existence of the MLE is immediate, the second task is more demanding.In order to decide whether the MLE exists it is sufficient to establish whether the observed sufficient statistics $\mathrm{A}x$ belong to the relative interior of P_n, which, by Theorem 5.2, happens if and only if $t := \mathrm{C}x'$ belongs to the relative interior of D_n, where for convenience the matrix C can be taken to be as in (11) (so it has dimension $n + \binom{n}{2} \times n(n-1)$ and is of full rank). In turn, we can decide this by solving the following simple linear program

$$\begin{aligned} & \max s \\ \text{s.t.} \quad & \mathrm{C}x' = t \\ & x'_{i,j} - s \geq 0 \\ & s \geq 0, \end{aligned}$$

where the scalar s and vector $x' = \{x'_{i,j}, i \neq j\} \in \mathbb{R}^{n(n-1)}$ are the variables. At the optimum (s^*, x^*), the MLE exists if and only if $s^* > 0$. Though very simple, the previous algorithm may not be sufficient to compute

[1] The `R` code we used to perform the numerical calculations is available at `http://www.stat.cmu.edu/~arinaldo/Rinaldo_Petrovic_Fienberg_Rcode.`

the support of $\hat{p}$ if the MLE does not exist. To this end, we need to resort to a more sophisticated algorithm. Consider the following $n(n-1)$ programs, one for each column of C:

$$\begin{array}{rc} & \max\langle c_{i,j}, y\rangle \\ \text{s.t.} & y^\top t = 0 \\ & \mathrm{C}^\top y \geq 0 \\ & -1 \leq y \leq 1, \end{array}$$

where the last inequalities are taken element-wise. Let $y^*_{i,j} \in \mathbb{R}^{n+\binom{n}{2}}$ denote the solution to the linear program corresponding to the (i,j)-th column of C.

Lemma 5.3. *The MLE does not exist if and only if $\langle c_{i,j}, y^* i, j\rangle > 0$ for some (i,j), in which case the co-facial set associated with t is given by*

$$\{(i,j) \colon \langle c_{i,j}, y^*_{i,j}\rangle > 0\}.$$

See Fienberg and Rinaldo (2011, section 4.1) for a more refined and efficient implementation of the above algorithms.

6 Applications and Extensions

The main arguments and the algorithmic procedures that we have used to explore nonexistence of the MLE and parameter estimability in the beta model are rather general, as they pertain to all log-linear models (see, e.g., Fienberg and Rinaldo, 2011). In this section we extend them to different models for networks.

6.1 The Rasch model

Just like in section 3.2, necessary and sufficient conditions for the existence of the MLE of the Rasch model parameters can also be formulated in geometric terms based on the polytope of degree sequence. In detail, for a bipartition of the n nodes of the form $I = \{1,\ldots,k\}$ and $J = \{k+1, n-1, n\}$, where $l = n-k$, let $P_{k,l} \subset \mathbb{R}^n$ denote the associated *polytope of bipartite degree sequences*, i.e. the convex hull of all degree sequences of bipartite undirected simple graphs on n nodes, with tge bipartition specified by I and J. Let $d(x)$ denote the degree sequence associated with the observed bipartite graph $x \in \mathcal{R}_n$. Then, a straightforward application of Theorem 9.13 in Barndorff-Nielsen (1978) yields the following result.

Theorem 6.1. *The MLE of the Rasch model parameters exists if and only if $d(x) \in \mathrm{ri}(P_{p,q})$.*

The polytope of bipartite degree sequences was introduced by Hammer et al. (1990). We briefly recall its properties (see Mahadev and Peled, 1996, section 3.4 for more details). Let

$$F_{I,J} := \{y \in P_n \colon g(y, I, J, n) = 0\}$$

be the facet of P_n specified by I and J, where g is given in (7) (the sets I and J can be interchanged). Also, let $c \in \mathbb{R}^n$ be the vector with coordinates

$$c_i = \left\{ \begin{array}{ll} k-1 & i = 1,\ldots,k \\ 0 & i = k+1,\ldots,n. \end{array} \right.$$

The polytope of bipartite degree sequences $P_{k,l}$ is just the translate by c of the facet $F_{I,J}$, which implies, in particular, that $\dim(P_{p,q}) = n-1$ (this explains why, in Theorem 6.1, we used the correct notation $\mathrm{ri}(P_{l,k})$ instead of $\mathrm{int}(P_{p,q})$).

Theorem 6.2 (Theorem 3.4.4 in Mahadev and Peled (1996))**.** $P_{k,l} = \{y - c, y \in F_{I,J}\}$.

The previous result is rather useful: in order to determine whether the MLE fails to exist, i.e. whether the degree sequence of the observed bipartite graph is on the relative boundary of $P_{k,l}$, one can use Lemma 3.4 as follows. First add an edge between each pair of nodes in I (so, the graph is no longer bipartite). Then, check whether there is a pair of sets S and T, different from I and J, for which the conditions of Lemma 3.4 apply. Thus, the MLE does not exists if and only if there exists a partition of the nodes into three non-empty sets S, T and $(S \cup T)^c$, such that, with respect to this enlarged graph,

1. $S \subseteq I$ (hence S is complete);
2. $T \subseteq J$ (hence, T is stable);
3. every vertex of S is adjacent to every vertex in $(S \cup T)^c$;
4. no vertex in T is adjacent to any vertex in $(S \cup T)^c$.

In fact, the above conditions are equivalent to the conditions for existence of the MLE in the Rasch model found independently by Haberman (1977) and Fischer (1981). Indeed, recall that Haberman's condition are as follows: the MLE does not exists if there there exists sets A, B , C and D such that

1. $A \cup B = I$ and $C \cup D = J$, with $A \cap B \cap C \cap D = \emptyset$;
2. $A \neq \emptyset$ and $C \neq \emptyset$ or $B \neq \emptyset$ and $D \neq \emptyset$;
3. $x_{i,j} = 0$ for all $i \in A$ and $j \in C$;
4. $x_{i,j} = 10$ for all $i \in B$ and $j \in D$,

were $x \in \mathcal{R}_n$ is the observed graph. Then, to see the equivalence, take $S = B$, $T = C$ and $(S \cup T)^c = A \cup D$.

6.2 Removing the Sampling Constraint in the Beta Model

We first consider a slightly modified form of the beta model, in which the number of observed edges $\{x_{i,j} \colon i \neq j\}$ are assumed to be realizations of $n(n-1)$ independent Poisson random variables with means $\{m_{i,j} \colon i \neq j\}$. As a result, the quantities $\{N_{i,j}, i \neq j\}$ are now random and can be zero with positive probabilities. In this case, the natural generalization of the beta model is to consider a parametrization of the mean values by points $\alpha \in \mathbb{R}^n$ and $\gamma \in \mathbb{R}^n$ so that

$$\log m_{i,j} = \alpha_i + \gamma_j, \quad \forall i \neq j. \tag{12}$$

As usual, we index the coordinates of a sample point x as $\{(x_{i,j}, x_{j,i}), i < j\}$, with the pairs of coordinates ordered lexicographically based on (i, j). Some algebra then shows that the probability of observing a point $x \in \mathbb{N}^{n(n-1)}$ is

$$p_{\alpha,\gamma}(x) = \exp\left\{ \sum_i \alpha_i d_i^{\text{out}} + \sum_j \gamma_j d_j^{\text{in}} - \phi(\alpha, \gamma) \right\} \prod_{i \neq j} \frac{1}{x_{i,j}!}, \quad x \in \mathbb{N}^{n(n-1)}$$

where the coordinates of the vectors of minimal sufficient statistics $d^{\text{out}} = d^{\text{out}}(x)$ and $d^{\text{in}} = d^{\text{in}}(x)$ are

$$d_i^{\text{out}} := \sum_{j \neq i} x_{i,j}, \quad i = 1, \ldots, n,$$

and

$$d_j^{\text{in}} := \sum_{i \neq j} x_{i,j}, \quad j = 1, \ldots, n,$$

respectively, and the log-partition function $\phi \colon \mathbb{R}^{2n} \to \mathbb{R}$ is given by

$$(\alpha, \gamma) \mapsto \sum_{i \neq j} \exp\{\alpha_i + \gamma_j\}.$$

×	0		0
0	×		0
		×	
0	0		×

×	0	0	
0	×	0	
0	0	×	
			×

×		0	0
	×		
0		×	0
0		0	×

×			
	×	0	0
	0	×	0
	0	0	×

Table 12: Co-facial sets of the second kind, as specified in theorem 6.3, for the case $n = 4$. Empty cells refer to arbitrary entries.

The sufficient statistics $d = d(x)$ can be obtained as

$$d = \begin{pmatrix} d^{\text{out}} \\ d^{\text{in}} \end{pmatrix} = \mathrm{A}x,$$

where A is the $2n \times n(n-1)$ whose columns are indexed in the same way as the columns B of equation (9), while the rows are indexed by the parameters $\{\alpha_1, \ldots, \alpha_n, \gamma_1, \ldots, \gamma_n\}$. The entries of the row corresponding to α_i are all zeros, except for the coordinates corresponding the columns (i,j) with $i < j$ and (j,i) with $i > j$, which are ones. Similarly, the rows corresponding to γ_j are all zeros, except for the coordinates corresponding the columns (j,i) with $i < j$ and (i,j) with $i > j$, which are ones. For instance, when $n = 4$,

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

We remark that A is rank-deficient, as its rank is $2n - 1$, which reflect the fact that the parametrization in (12) is non-identifiable (this can be easily fixed by imposing, for instance the constraint $\sum_i \alpha_i = 0$).

Notice that if the entries of $x \in \{0,1\}^{n(n-1)}$ are all zeros and ones, then x encodes a directed graph on n nodes, with an arrow going from node i to node j if and only if $x_{i,j} = 1$ (thus, there may be two edges connecting any pair of nodes, directed in opposite ways). In this case, the sufficient statistics d^{out} and d^{in} correspond to the in-degrees and out-degrees of the nodes.

If C_n denotes the polyhedral cone spanned by the columns of A, then, for a given sample point x, the MLE of (α, γ) or, equivalently, of $\{m_{i,j} : i \neq j\}$ exists if and only if $d(x) \in \mathrm{int}(C_{\mathrm{A}})$. It turned out that C_n has small combinatorial complexity, as shown in the next results.

Theorem 6.3. *The polyhedral cone C_n has $3n$ facets. The co-facial sets corresponding to the facets of C_n can be classified as follows:*

1. *the $2n$ support sets of the columns of A, each corresponding to a zero entry in the vectors of in-degree or out-degree statistics;*

2. *n co-facial sets of the form $\{(i,j) \colon i \neq j \neq k\}$, one for each $k = 1, \ldots, n$.*

For instance, when $n = 4$, there are 12 facial sets, 8 of them associated to a zero value in the 8 dimensional vector of sufficient statistics. The remaining 4 co-facial sets are shown in table 12.

The previous Theorem implies that the number of facets of C_n grows only linearly in n, unlike the number of facets of the polytope of degree sequences P_n. Thus, for this model, nonexistence of the MLE is a much less frequent phenomenon, at least combinatorially. Note in particular, that the MLE exists even if $x_{i,j} + x_{j,i} = 0$

for some (in fact many) pairs. Theorem 6.3 can be used to easily show that the MLE exists with probability tending to one as n increases. Indeed, the probability of a nonexistent MLE is no larger than

$$\sum_{i=1}^{n} e^{-\sum_{j \neq i} m_{i,j}} + \sum_{j=1}^{n} e^{-\sum_{i \neq j} m_{i,j}} + \sum_{k=1}^{n} e^{-\sum_{i \neq j \neq k} m_{i,j}}. \tag{13}$$

Then, assuming $n \geq 7$ and letting

$$m^* := \min_{i \neq j} m_{i,j},$$

the first two terms in equation (13) are each smaller than $ne^{-(n-1)m^*}$, while the last term is bounded from above by

$$ne^{-\binom{n}{2}+2(n-1)m^*} \leq ne^{-(n-1)m^*},$$

where the last inequality is due to the fact that $\binom{n}{2} - 2(n-1) \geq n-1$ for all $n \geq 7$. Thus, (13) is bounded from above by $3ne^{-(n-1)m^*}$, which implies that, if $m^* = m^*(n) = \frac{c \log n}{n-1}$, the MLE exists with probability at least $1 - \frac{3}{n^c}$. This simple calculation then shows that the MLE exists with overwhelming probability even if the expected cell counts all tend to zero, as long as these values decay at a rate $\Omega\left(\frac{\log n}{n}\right)$.

The results just obtained can be specialized to the Rasch model, in which the nodes are partitioned into two sets I and J of cardinality k and $l = n - k$, and edges can only occur between a node $i \in I$ and a node $j \in J$, though the number of edges between any pair of nodes (i, j) is random. The observed set edge counts takes the form of a $k \times l$ contingency table and the sufficient statistics are the k row sums and the l column sums. As noted by Haberman (1977), in this case the MLE exists if and only if the row and column sums are all positive.

6.3 The Bradley-Terry Model

We can specialize the model described in section 6.2 to a directed graph without multiple edges, thus obtaining the Bradley-Terry model for pairwise comparisons. See Bradley and Terry (1952), David (1988), Hunter (2004) and references therein. In detail, let $p_{i,j}$ denote the probability of a directed edge from i to j and $p_{j,i}$ the probability of a directed edge from j to i. According to the Bradley-Terry model, the probabilities of directed edges can be parametrized by vectors $\beta \in \mathbb{R}^n$ so that

$$p_{i,j} = \frac{e^{\beta_i}}{e^{\beta} + e^{\beta_j}}, \quad \forall i \neq j, \tag{14}$$

or, equivalently, in terms of log-odd ratios,

$$\log \frac{p_{i,j}}{p_{j,i}} = \beta_i - \beta_j, \quad \forall i < j.$$

Notice that this parametrization is redundant, and identifiability is typically enforced by requiring that $\sum_{i=1}^{n} e^{\beta_i} = 1$. Data are obtained by recording, for each pair of nodes (i, j) the outcomes of $N_{i,j}$ pairwise comparisons, where $N_{i,j}$ are fixed positive integers, resulting in $x_{i,j}$ instances of node i being preferred to node j and $x_{j,i}$ instances of node j being preferred to node i, with $x_{i,j} + x_{j,j} = N_{i,j}$. The outcomes of the pairwise comparison are assumed mutually independent. Thus, for $i < j$, the Bradley-Terry model treats the $n(n-1)$ observed counts $\{x_{i,j} \colon i \neq j\}$ as a realization of mutually independent $\mathrm{Bin}(N_{i,j}, p_{i,j})$ distributions, where the probability parameters $\{p_{i,j} \colon i \neq j\}$ satisfy (14),

Despite the apparent similarity between equations (15) and (14) the beta model and the Bradley-Terry model are radically different. Indeed, for the Bradley-Terry model, it is well known that the minimal sufficient statistics are the row sums (or the column sums) of the observed table. Indeed, this model can be alternatively prescribed as a model of quasi-symmetry and quasi-independence (see, e.g. Fienberg and Larntz, 1976). Necessary and sufficient conditions for the existence of the MLE, due to Zermelo (1929) and Ford

(1957), are as follows: *In every possible partition of the objects into two nonempty subsets, some object in the second set has been preferred at least once over some object in the first set* (see Ford, 1957, page 29). We can express the Zermelo-Ford condition equivalently in a graph theoretic form as follows: the MLE exists if and only if the observed directed graph is strongly connected, a property which we can easily check by a depth-first search. According to this condition, the number of facial sets corresponding to the facets of the associated convex support is

$$\sum_{i=1}^{n-1} \binom{n}{i} = 2^n - 2.$$

See Simons and Yao (1999) for an analysis of the existence and asymptotic normality of the MLE for the Bradley-Terry model under the condition that all the terms $N_{i,j}$ are constant and the number of objects n increases.

We conclude this section by noting the arguments and algorithms for facial set identification discussed in section 5 apply to this model as well. In this case, the marginal cone is spanned by a matrix of dimension $\left(\binom{n}{2} + n\right) \times n(n-1)$, the first $\binom{n}{2}$ rows corresponding to the sampling constraints $\{x_{i,j} + x_{j,i} = N_{i,j} \colon i < j\}$, and the remaining n rows to the row sums.

6.4 p_1 Models

Both the beta model and the Bradley-Terry model can be obtained as a special cases of the class of p_1 models for directed graphs proposed by Holland and Leinhardt (1981). In fact, existence of the MLE and the identification of the facial sets for p_1 models can be treated using the very same arguments we have presented in the first part of the article. In this final section we detail these arguments for the more general and challenging class of p_1 models.

In p_1 models, the occurrence of a random edge between any pair of nodes i and j, or *dyad*, is modeled independently from all the others edges. We keep track of four possible edge configurations within each dyad: node i has an outgoing edge into node j: $i \rightarrow j$; node i as an incoming edge originating from node j: $i \leftarrow j$; nodes i and j are linked by a bi-directed edge: $i \longleftrightarrow j$; and node i and j are not adjacent in the network. Following the notation we established in Petrović et al. (2010), which is slightly different than the original notation of Holland and Leinhardt (1981), for every pair of nodes (i, j) we define the probability vector

$$p_{i,j} = (p_{i,j}(0,0), p_{i,j}(1,0), p_{i,j}(0,1), p_{i,j}(1,1)) \in \Delta_3 \tag{15}$$

containing the probabilities of the four possible edge types, where Δ_3 is the standard simplex in $\mathbb{R}^4$. The numbers $p_{i,j}(1,0)$, $p_{i,j}(0,1)$ and $p_{ij}(1,1)$ denote the probabilities of the edge configurations $i \rightarrow j$, $i \leftarrow j$ and $i \longleftrightarrow j$, respectively, and $p_{i,j}(0,0)$ is the probability that there is no edge between i and j (thus, 1 denotes the outgoing side of the edge). Notice that, by symmetry $p_{i,j}(a,b) = p_{j,i}(b,a)$, for all $a, b \in \{0,1\}$ and that

$$p_{i,j}(0,0) + p_{i,j}(1,0) + p_{i,j}(0,1) + p_{i,j}(1,1) = 1. \tag{16}$$

The fundamental assumption underlying p_1 models is that the dyads are independent. This is formalized by modeling each of the $\binom{n}{2}$ dyads as mutually independent draws from multinomial distributions with class probabilities $p_{i,j}$, $i < j$. Specifically, the Holland-Leinhardt p_1 model specifies the multinomial probabilities of each dyad (i, j) in logarithmic form as follows (see Holland and Leinhardt, 1981):

$$\begin{array}{rcl} \log p_{i,j}(0,0) & = & \lambda_{ij} \\ \log p_{i,j}(1,0) & = & \lambda_{ij} + \alpha_i + \beta_j + \theta \\ \log p_{i,j}(0,1) & = & \lambda_{ij} + \alpha_j + \beta_i + \theta \\ \log p_{i,j}(1,1) & = & \lambda_{ij} + \alpha_i + \beta_j + \alpha_j + \beta_i + 2\theta + \rho_{i,j}. \end{array} \tag{17}$$

The parameter α_i quantifies the effect of an outgoing edge from node i, the parameter β_j instead measures the effect of an incoming edge into node j, while $\rho_{i,j}$ controls the added effect of reciprocated edges (in both directions). The parameter θ measures the propensity of the network to have edges and, therefore, controls the "density" of the graph. The parameters $\{\lambda_{i,j} : i < j\}$ are normalizing constants to ensure that (16) holds for each each dyad (i, j) and need not be estimated. Note that, in order for the model to be identifiable, additional linear constraints need to be imposed on its parameters. We refer the interested readers to the original paper on p_1 model by Holland and Leinhardt (1981) for an extensive interpretation of the model parameters.

As noted in Fienberg and Wasserman (1981a,b), different variants of the p_1 model can be obtained by constraining the model parameters. In Petrović et al. (2010) we consider three special cases of the basic p_1 model, which differ in the way the reciprocity parameter is modeled:

1. $\rho_{ij} = 0$, *no reciprocal effect*;
2. $\rho_{ij} = \rho$, *constant reciprocation*;
3. $\rho_{ij} = \rho + \rho_i + \rho_j$, *edge-dependent reciprocation*.

As it is often the case with network data, we assume that data become available in the form of one observed network. Thus, each dyad (i, j) is observed in only one of its four possible states and this one observation is a random vector in $\mathbb{R}^4$ with a $\mathrm{Multinomial}(1, p_{i,j})$ distribution. As a result, data are sparse and, even though the dyadic probabilities are strictly positive according to the defining equations (17), only some of the model parameters may be estimated from the data. Extension to the case in which the dyads are observed multiple times are straightforward.

For a network on n nodes, we represent the vector of $2n(n-1)$ dyadic probabilities as

$$p = (p_{12}, p_{13}, \ldots, p_{n-1,n}) \in \mathbb{R}^{2n(n-1)},$$

where, for each $i < j$, p_{ij} is given as in (15). The p_1 *model* is the set of all probability distributions that satisfy the Holland-Leinhardt equations (17). The design matrix associated with a given p_1 model can be constructed as follows (this construction is by no means unique and leads to rank-deficient matrices, though it is rather simple). The columns of A are indexed by the entries of the vectors $p_{i,j}$, $i < j$, where the $p_{i,j}$'s are ordered lexicographically, and its rows by the model parameters, ordered arbitrarily. The (r, c) entry of A is equal to the coefficient of the c-th parameter in the logarithmic expansion of the r-the probability as indicated in (17). In particular, notice that the entries of A can only be 0, 1 or 2. For example, in the case $\rho_{ij} = \rho + \rho_i + \rho_j$, the matrix A has $\binom{n}{2} + 3n + 2$ rows. When $n = 3$, the design matrix corresponding to this model is

$$
\begin{array}{c|cccc|cccc|cccc}
 & \multicolumn{4}{c|}{p_{1,2}} & \multicolumn{4}{c|}{p_{1,3}} & \multicolumn{4}{c}{p_{2,3}} \\
\hline
\lambda_{12} & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\lambda_{13} & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
\lambda_{23} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
\theta & 0 & 1 & 1 & 2 & 0 & 1 & 1 & 2 & 0 & 1 & 1 & 2 \\
\alpha_1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
\alpha_2 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
\alpha_3 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
\beta_1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
\beta_2 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\beta_3 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
\rho & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
\rho_1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
\rho_2 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
\rho_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{array}
$$

Let $\mathcal{S}_n = \{x_{i,j}, i \neq j\} \subset \{0, 1\}^{2n(n-1)}$ denote the sample space, i.e. the set of all observable networks on n nodes. Then, every point x in the sample space $\mathcal{X}$ can be written as

$$x = (x_{1,2}, x_{1,3}, \ldots, x_{n-1,n}),$$

n	$\rho_{i,j}=0$	$\rho_{i,j}=\rho$	$\rho=\rho_i+\rho_j$	$2n(n-1)$
3	62	62	62	12
4	1,862	2,415	3,086	24
5	88,232	158,072	347,032	40

Table 13: Number of vertices for the polytopes P_{A} for different specifications of the p_1 model and different network sizes. Computations carried out using `minksum` Weibel (2005). The last column indicates the number of columns of the design matrix A, which correspond to the number of generators of C_{A}.

where each of the $\binom{n}{2}$ subvectors $x_{i,j}$ is a vertex of Δ_3. Notice that $|\mathcal{X}_n| = 4^{n(n-1)}$. This way of representing a network on n nodes with a highly-constrained $0/1$ vector of dimension $2n(n-1)$ may appear cumbersome and redundant. Indeed, as in Holland and Leinhardt (1981), we could more naturally represent an n-node network using the $n \times n$ incidence matrix with $0/1$ off-diagonal entries, where the (i,j) entry is 1 is there is an edge from i to j and 0 otherwise. While this representation is more intuitive and parsimonious (as it only requires $\frac{n(n-1)}{2}$ bits), whenever $\rho \neq 0$, the sufficient statistics for the reciprocity parameter are not linear functions of the observed network. As a consequence, the adjacency matrix representation does not lead directly to a linear exponential family.

The convex support for this family is the polytope obtained as the Minkowski sum

$$P_{\mathrm{A}} := \sum_{i<j} \mathrm{A}_{i,j},$$

where $\mathrm{A}_{i,j}$ is the sub-matrix of A comprised by the four columns referring to the dyad (i,j). Given an observed network $x \in \mathcal{S}_n$ the MLE of the parameters exists if and only of $\mathrm{A}x \in \mathrm{ri}(\mathcal{S}_n)$ and, when the MLE does not exist, the associated facial set provides the non-estimable probability parameters. Like with the convex support of the beta model, the combinatorial complexity of this object is quite high and increases very rapidly with n (though, unlike the beta model, the convex supports for these models do not appear to be a known or well studied polytopes). See table 6.4 and the discussion below.

The arguments and results of section 3 extend in a straightforward way: *the MLE exists if and only if* $\mathrm{A}x \in \mathrm{ri}(C_{\mathrm{A}})$*, where* $C_{\mathrm{A}} = \mathrm{cone}(\mathrm{A})$*, and the facial sets of* P_{A} *are also facial sets of* C_{A}.

Numerical Experiments

We conclude this section by describing some numerical experiments illustrating the reduction in complexity associated to the Cayley trick from in section 5 for the genral p_1 model. Table 6.4 displays the number of vertices of the polytopes P_{A} for the three p_1 model specifications we consider and various networks sizes. The last column of the table contains the number of columns of the design matrix, which is also the number of extreme rays of the marginal cone C_{A}. In comparison, the number of vertices of P_{A}, whose determination is computationally very hard, is very large and grows extremely fast with n.

In Table 6.4 we report the number of facets, dimensions and ambient dimensions of the cones C_{A} for different values of n and for the three specification of the reciprocity parameters $\rho_{i,j}$ we consider here. Though this only provides and indirect measure of the complexity of these models and of the non-zero patterns in extended MLEs, it does show how quickly the complexity of p_1 models may scale with the network size n.

Another point of interest is the assessment of how often the existence of the MLE arises. In fact, because of the product Multinomial sampling constraint, nonexistence of the MLE is quite severe, especially for smaller networks. Below we report our findings, which are necessarily restricted to networks of small sizes:

1. $n = 3$.
   The sample space consists of $4^3 = 64$ possible networks. When $\rho_{i,j} = 0$ for all i and j, there are 63 different observable sufficient statistics, only one of which belongs to $\mathrm{ri}(P_{\mathrm{A}})$. Thus, only one of the 63

n	$\rho_{i,j} = 0$			$\rho_{i,j} = \rho$			$\rho = \rho_i + \rho_j$		
	Facets	Dim.	Ambient Dim.	Facets	Dim.	Ambient Dim.	Facets	Dim.	Ambient Dim.
3	30	7	9	56	8	10	15	10	13
4	132	12	14	348	13	15	148	16	19
5	660	18	20	3,032	19	21	1,775	23	26
6	3,181	25	27	94,337	26	28	57,527	31	34

Table 14: Number of facets, dimensions and ambient dimensions of the the cones C_{A} for different specifications of the p_1 model and different network sizes. The number of facets of C_{A} is equal to the number of facets of P_{A} plus $\binom{n}{2}$, these additional facets corresponding to the sampling constraints of one observation per dyad.

observable sufficient statistics leads to the existence of the MLE. This sufficient statistic corresponds to the two nextworks

$$\begin{bmatrix} \times & 0 & 1 \\ 1 & \times & 0 \\ 0 & 1 & \times \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \times & 1 & 0 \\ 0 & \times & 1 \\ 1 & 0 & \times \end{bmatrix}.$$

In both cases, the associated MLE is the 12-dimensional vector with entries all equal to 0.25. Incidentally, the polytope P_{A} has 62 vertices and 30 facets. When $\rho_{i,j} = \rho \neq 0$ or $\rho_{i,j} = \rho_i + \rho_j$ the MLE never exists.

2. $n = 4$.
The sample space contains 4096 observable networks. If $\rho_{i,j} = 0$, there are $2,656$ different observable sufficient statistics, only 64 of which yield existent MLEs. Overall, out of the 4096 possible networks, only 426 have MLEs. When $\rho_{i,j} = \rho \neq 0$, there are $3,150$ different observable sufficient statistics, only 48 of which yield existent MLEs. Overall, out of the $4,096$ possible networks, only 96 have MLEs. When $\rho_{i,j} = \rho_i + \rho_j$, there are $3,150$ different observable sufficient statistics and the MLE never exists.

3. $n = 5$.
The sample space consists of $4^{10} = 1,048,576$ different networks. If $\rho_{i,j} = 0$, there are $225,025$ different sufficient statistics, and the MLE exists for $7,983$. If $\rho_{i,j} = \rho \neq 0$ the number of distinct possible sufficient statistics is $349,500$, and the MLE exists in $12,684$ cases. Finally, when $\rho_{i,j} = \rho_i + \rho_j$, the number of different sufficient statistics is $583,346$ and the MLE never exists.

7 Discussion

In this paper we derived necessary and sufficient conditions for the existence of the MLE of the parameters of the beta model. These conditions are tied to the polytope of degree sequences, whose geometric properties we exploit to characterize sample points leading to a non-existent MLE and to formulate conditions that imply that the MLE exists with probability approaching one as the number of nodes increases. For the random graph model, our results improve similar results recently obtained by Chatterjee et al. (2011). Finally, we showed how representing the beta model in the form of a log-linear model with product multinomial constraints is well-suited to characterize the nonexistence of the MLE, and we extend the procedure to more general discrete models, such as the Rasch model, the Bradley-Terry model and more general p_1 models.

We can relax the assumption that we observe each edge with positive probability a straightforward way. In our framework, this corresponds to removing from the design matrix the columns corresponding to the edges that never occur. In fact, this is precisely how we can obtain the Rasch model.

For the random graph models we have considered, the nonexistence of the MLE typically has no effect on the convergence properties of algorithms for maximum likelihood estimation that maximize the log-likelihood with respect to the probability parameters, such as iterative scaling algorithms. Indeed, the likelihood function (2) is strictly concave for any observable set of counts. What is unclear, however, is the effect on the rate of convergence. On the other hand, precisely because convergence always occurs, these algorithms are not suited to detect the nonexistence of the MLE or the non-estimability of the model parameters should the MLE be undefined. The procedures we describe in section 5, or the ones we detail in Fienberg and Rinaldo (2011, section 4), are instead appropriate.

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.tx

8 Acknowledgments

This research was supported in part by 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). The authors would like to thank Christian Haase and Josephine Yu for suggesting the proof of Theorem 6.3. This manuscript was completed while the second author was in residence at Institut Mittag-Lefller.

9 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

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

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 (15)) exists if and only if $d \in \mathrm{int}(S_n)$. Thus, it is sufficient to show that $d \in \mathrm{int}(S_n)$ if and only if $\tilde{d} \in \mathrm{int}(P_n)$.

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

$$P_{i,j} = \mathrm{convhull}\{0, a_{i,j}\} \subset \mathbb{R}^n. \tag{18}$$

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 $P_{i,j}$:

$$P_n = \sum_{i<j} P_{i,j}. \tag{19}$$

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 (19) 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

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

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 $F(P;c) := \{x \in P\colon x^\top c \geq y^\top c, \forall y \in P\}$. Any face F of P can be written in this way, where is 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. ∎

Proof of Lemma 3.2. By Proposition 2.1 in Fukuda (2004),

$$F = F(P_n, c) = \sum_{i<j} F(P_{i,j}, c), \tag{20}$$

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 (18)), 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 (20), $\mathcal{F}$ 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. ∎

Proof of Theorem 4.1. Let $\tilde{d} = (\tilde{d}_1, \ldots, \tilde{d}_n)$ be the random vector defined in (5). 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 can conveniently re-express the random vector $\tilde{d}$ as an average of independent and identically distributed graphical degree sequences. In details, we can write

$$\tilde{d} = \frac{1}{N} \sum_{k=1}^{N} d^{(k)}, \tag{21}$$

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

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

$$\mathcal{O}_n := \left\{ \max_i |\tilde{d}_i - \overline{d}_i| \leq \sqrt{c \frac{n \log n}{N}} \right\} \tag{22}$$

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

By assumption *(i)*, for each i,

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

so that

$$0 < \tilde{d}_i < n - 1, \quad i = 1, \dots, n. \tag{23}$$

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

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

which yields

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

Using assumption *(ii)*, the previous displays implies that

$$\min_{(S,T) \in \mathcal{P}} g(S, T, \tilde{d}, n) > C > 0. \tag{24}$$

Thus, we have shown that (23) and (24) hold, provided that the event $\mathcal{O}_n$ is true and assuming *(i)* and *(ii)*. Therefore, by Theorem 3.3 the MLE exists.

■

Proof of Corollary 4.2. Using the same setting and notation of Theorem 4.1, we will assume throughout the proof that the event

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

holds true. Note that by Hoeffding's inequality and the union bound,

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

where we have used the inequality $\log N \leq \log n$. A simple calculation shows that, when $\mathcal{O}'_n$ is satisfied, we also have

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

Then, by the same arguments used in the proof of Theorem 4.1, assumption *(i')* yields that

$$0 < \tilde{d}_i < n - 1, \quad i = 1, \dots, n. \tag{25}$$

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

$$g(S, T, \tilde{d}, n) \geq g(S, T, \overline{d}, n) - |S \cup T| \sqrt{cn \log n}. \tag{26}$$

Now, it is easy to see that, on the event $\mathcal{O}'_n$, assumption *(i')* also yields

$$\min_k \min_i \min \left\{ d_i^{(k)}, n - 1 - d_i^{(k)} \right\} \geq \sqrt{cn \log n} + C. \tag{27}$$

We now show that, when (25) and the previous equation are satisfied, the MLE exists if

$$\min_{(S,T) \in \mathcal{P}_n} g(S, T, d, n) > C > 0. \tag{28}$$

Indeed, suppose that (25) 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 3.4 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 (25) and (27) 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 3.3, when (25) and (27) hold, (28) implies that $\tilde{d} \in \mathrm{int}(P_n)$, so the MLE exists. However, equation (26) and assumption *(ii')* implies (28), so the proof is complete. ∎

Proof of Theorem 5.2. We first define a new polytope $Q_n \subset \mathbb{R}^{2n+\binom{n}{2}}$ which is combinatorially equivalent to P'_n and, therefore, to the polytope of degree sequences P_n. Let $c_{i,j}$ be the column of C index by the pair (i,j) and, for each $i < j$, set

$$C_{i,j} := \mathrm{convhull}\left(\{c_{i,j}, c_{j,i}\}\right)$$

and

$$Q_n := \sum_{i<j} C_{i,j}.$$

By construction, $w \in P'_n$ if and only if

$$\begin{pmatrix} 1 \\ w \end{pmatrix} \in Q_n,$$

where $1 \in \mathbb{R}^{\binom{n}{2}}$ is a vector of all ones, which shows that P'_n and Q_n are combinatorially equivalent, so they have the same facial sets. We make a simple but useful observation: because the first $\binom{n}{2}$ coordinates of any point in Q_n are all ones, and given the patter of non-zero entries in the first $\binom{n}{2}$ rows of C, it must be that if $y \in Q_n$ and $y = \mathrm{C}p'$, the vector p' is of the form $\{p'_{i,j} \colon i \neq j, p'_{i,j} \in [0,1], p'_{i,j} = 1 - p'_{j,i}\}$.

Since $Q_n \subset D_n$ and both sets are closed, $y \in \mathrm{ri}(Q_n)$ implies that $y \in \mathrm{ri}(D_n)$. As for the converse statement, suppose y belongs to the interior of a proper face of Q_n with facial set $\mathcal{F}'$. Then, by Proposition 2.1 in Fukuda (2004), y can be uniquely expressed as

$$y = y_{1,2} + y_{1,3} + \ldots + y_{n-1,n} \tag{29}$$

where $y_{i,j} \in \mathrm{ri}(C_{i,j})$ if and only if (i,j) and (j,i) are in $\mathcal{F}'$. Equivalently, $y_{i,j} = c_{i,j}$ or $y_{i,j} = c_{j,i}$ if and only if $(i,j) \notin \mathcal{F}'$ or $(j,i) \notin \mathcal{F}'$, respectively. Arguing by contradiction, suppose that $y \in \mathrm{ri}(D_n)$. Then, there exists a point $p^* = \{p^*_{i,j} \colon i \neq j\}$ with strictly positive entries such that $y = \mathrm{C}p^*$. By the observation above, it must be that $p^*_{i,j} \in (0,1)$ and $p^*_{i,j} = 1 - p^*_{j,i}$, for all $i < j$. In turn, this implies that, in equation (29), $y_{i,j} \in \mathrm{ri}(C_{i,j})$ for all $i < j$, i.e. $y_{i,j} \notin \{c_{i,j}, c_{j,i}\}$ for all $i < j$. Then, using again Proposition 2.1 in Fukuda (2004), $y \in \mathrm{ri}(Q_n)$, a contradiction.

To prove the second claim, notice that, the arguments so far yield that, for every proper face F of Q_n, there exists one face G of D_n such that $\mathrm{ri}(F) \subset \mathrm{ri}(G)$, so that $\mathcal{F}' \subseteq \mathcal{G}$, where $\mathcal{F}'$ and $\mathcal{G}$ are the facial sets associated with F and G, respectively. We now show that $\mathcal{F}' = \mathcal{G}$. To see this, let $y \in \mathrm{ri}(F)$ for some face F of Q_n with facial set $\mathcal{F}'$, so that

$$y = Cp'$$

for some $p' = \{p'_{i,j} \colon i \neq j, p'_{i,j} \in [0,1], p'_{i,j} = 1 - p'_{j,i}\}$ such that $p'_{i,j} > 0$ if and only if $(i,j) \in \mathcal{F}'$. On the other hand, since $y \in \mathrm{ri}(G)$,

$$y = Cp^*,$$

where $p^* = \{p^*_{i,j} \colon p^*_{i,j} \geq 0\}$ is such that $p^*_{i,j} > 0$ if and only if $(i,j) \in \mathcal{G}$. However, using the observation above, it must be that $p^*_{i,j} \in [0,1]$ and $p^*_{i,j} = 1 - p^*_{j,i}$, for all $i < j$. By maximality of the facial sets, $\mathcal{F}' = \mathcal{G}$, as claimed.

Thus, we have shown that if $\mathcal{F}'$ is a facial set of Q_n and hence of P'_n, it is also a facial set of D_n. ∎

Proof of Lemma 5.3. Let $\widetilde{\mathcal{F}} = \{(i,j)\colon \langle c_{i,j}, y^*_{i,j}\rangle = 0\}$. If $\widetilde{\mathcal{F}} = \{1,\ldots,n\}$, then there does not exist any vector $v \in \mathbb{R}^{n+\binom{n}{2}}$ such that $\langle v, c_{i,j}\rangle \geq 0$ with strict inequality for some (i,j). Thus, the normal cone at t is the zero vector, so $t \in \mathrm{ri}(D_n)$, and the MLE exists by Theorem 5.2. We now show that the if the MLE does not exist, then $\widetilde{\mathcal{F}} = \mathcal{F}$, where $\mathcal{F}$ is the facial set associated with the face of D_n whose relative interior contains t. To see this, let $\tilde{v} = \sum_{(i,j)\in\widetilde{\mathcal{F}}} y^*_{i,j}$. It is clear that $\widetilde{\mathcal{F}} \subseteq \mathcal{F}$, for otherwise the vector $\tilde{v}$ would produce a strictly larger facial set, which violates the maximality of $\mathcal{F}$. On the other hand, if $(i,j) \in \mathcal{F}\setminus\widetilde{\mathcal{F}}$, then there does not exist any vector $y^*_{i,j}$ in the feasible set of the (i,j)-th program such that $\langle y^*_{i,j}, c_{i,j}\rangle = 0$. However, the vector v specifying $\mathcal{F}$ is clearly in that feasible set and, by definition, $\langle v, c_{i,j}\rangle = 0$, which gives a contradiction. Thus $\widetilde{\mathcal{F}} = \mathcal{F}$, as claimed. ∎

Proof of Theorem 6.3. Equivalently, since the row span of A contains the constant vectors, we study the facets of the polytope $P := \mathrm{conv}(B) \subset \mathbb{R}^n \times \mathbb{R}^n$. Denote by x_i and x'_i the coordinates of the two spaces, and by e_i and e'_i the corresponding standard unit vectors in $\mathbb{R}^n$. The polytope P is contained in the product of simplices $\Delta_{n-1} \times \Delta_{n-1} := \mathrm{conv}\{e_i \times e'_j : 1 \leq i,j \leq n\}$, where, for two vectors x and x' in $\mathbb{R}^n$,

$$x \times x' := \begin{pmatrix} x \\ x' \end{pmatrix} \in \mathbb{R}^{2n}.$$

The point $e_i \times e'_j$ corresponds to the (i,j)-entry of the $n \times n$ incidence table of the network. P is obtained from the product of simplices by removing the n vertices $\{e_i \times e'_i\colon i = 1\ldots,n\}$. To show that P has $3n$ facets, we will use the fact that $\Delta_{n-1} \times \Delta_{n-1}$ has $2n$ facets whose defining inequalities are $x_i \geq 0$, $x'_i \geq 0$, for $i = 1\ldots,n$. Note that these facets correspond to zero margins in the incidence table: for example, $x_i = 0$ refers to the zero margin corresponding to the i-th row and $x'_i = 0$ to the zero margin for the $(i+n)$-th row.

Define a new polytope, P', cut out by the following $3n$ inequalities:

$$P' := \{x_i \geq 0, x'_i \geq 0, x_i + x'_i \leq 1, \text{ for all } i\}.$$

We need to show that $P = P'$ and that the defining inequalities are all facets. For the first claim, we already see that $P \subseteq P'$. Since $\Delta_{n-1} \times \Delta_{n-1}$ is simple, every vertex has dimension many neighbors. Thus, removing the vertex $e_i \times e'_i$ introduces one new facet, namely, $x_i + x'_i \leq 1$. Since we are removing n non-adjacent vertices, $P = P'$. Next, our arguments so far already imply that the n new inequalities $\{x_i + x'_i \leq 1\colon i = 1,\ldots,n\}$ define facets, so we need to show that other $2n$ inequalities, corresponding to zero row margins, define facets as well. But this follows from the fact that the support sets of each of the rows of A are facial sets of P and that they are incomparable, in the sense that none of them is contained in any of the others. Thus, since the lattice of facial sets of P is isomorphic to the face lattice of P, the $2n$ null margins each specifies a different facet of P. ∎

References

Andersen, E. B. (1980). *Discrete Statistical Models with Social Science Applications,* Elsevier, Amsterdam.

Agresti, A. (2002). *Categorical Data Analysis*. Wiley, New York.

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. Preprint. Available at http://www.people.fas.harvard.edu/~blitz/BlitzsteinDiaconisGraphAlgorithm.pdf

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs. I. The method of paired comparisons, *Biometrika,* 39, 324–345.

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, to appear in the *Annals of Applied Probability*, available at http://arxiv.org/abs/1005.1136.

David, H. A. (1988). *The Method of Paired Comparisons,* Oxford University Press.

Diaconis, P., Holmes, S., and Janson, S. (2008). Threshold graph limits and random threshold graphs, *Internet Mathematics,* 5(3), 267–320.

Diaconis, P. and Janson, S. (2007). Graph limits and exchangeable random graphs, available at http://arxiv.org/abs/0712.2749

Erdös, P. and Rényi, A. (1959) On Random Graphs. I. *Publicationes Mathematicae*, 6, 290–297.

Eriksson, N,. Fienberg, S.E., Rinaldo, A. and Sullivant, S. (2006). Polyhedral conditions for the nonexistence of the MLE for hierarchical log-linear models, *Journal of Symbolic Computations,* 41, 222–233.

Fienberg, S.E., and Larntz, K. (1976) Loglinear representation for paired and multiple comparison models. *Biometrika,* 63, 245–254.

Fienberg, S.E. and Rinaldo, A. (2011). Maximum Likelihood Estimation in Log-Linear Models: Theory and Algorithm, http://arxiv.org/abs/1104.3618

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.

Fischer, G.H. (1981). On the existence and uniqueness of the maximum likelihood estimates in the Rasch model, *Psychometrika,* 46(1), 59–77.

Ford, L.R. Jr. (1957). Solution of a ranking problem from binary comparisons, *The American Mathematical Monthly,* 64(8), part 2, 28–33.

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.

Geiger, D., Meek, C. and Sturmfels, B. (2006). On the toric algebra of graphical models, *Annals of Statistics,* 34(3), 1463–1492.

Goldenberg, A., Zheng, A.X., Fienberg, S.E., and Edoardo M. Airoldi, E.M. (2010). A survey of statistical network models, *Foundations and Trends in Machine Learning*, 2 (2), 129–233.

Haberman, S. (1974). *The Analysis of Frequency Data,* University of Chicago Press, Chicago.

Hammer P.L., Peled, U.N. and Sun, X. (1990). Difference Graphs, *Discrete Applied Mathematics,* 28, 35–44.

Haberman, S. (1977). Maximum likelihood estimates in exponential response models, *Annals of Statistics,* 5(5), 815–841.

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.

Hunter, D.R. (2004). MM algorithms for generalized Bradley-Terry models, *Annals of Statistics,* 32(1), 384–406.

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.

Mahadev, N.V.R. and Peled, U.N. (1995). Threshold Graphs and Related Topics, North Holland.

Morton, J. (2008). "Relations among conditional probabilities," arXiv:0808.1149v1.

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. (2011). Graver basis for an undirected graph and its application to testing the beta model of random graphs, available at http://arxiv.org/pdf/1102.2583v1.

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.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests,* University of Chicago Press, Chicago.

Schrijver, A. (1998). *Theory of Linear and Integer Programming*, Wiley, New York.

Simons, G. and Yao, Y. (1995). Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons, *Annals of Statistics,* 27(3), 1041–1060.

Stanley, R.P. (1991). A zonotope associated with graphical degree sequences, *DIMACS Series in Discrete Mathematics*, 4, 555–570.

Sturmfels, B. and Welker, V. (2011). Commutative Algebra of Statistical Ranking, available at http://arxiv.org/abs/1101.1597.

Viger, F. and Latapy (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.

Zermelo, E. (1929). Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Math. Z.,* 29, 436–460.