The $\beta$-model for Random Graphs — Regression, Cramér-Rao Bounds, and Hypothesis Testing

Johan Wahlström, Isaac Skog, Member, IEEE, Patricio S. La Rosa, Member, IEEE, Peter Händel, Senior Member, IEEE, and Arye Nehorai, Fellow, IEEE

Abstract — We develop a maximum-likelihood based method for regression in a setting where the dependent variable is a random graph and covariates are available on a graph-level. The model generalizes the well-known $\beta$-model for random graphs by replacing the constant model parameters with regression functions. Cramér-Rao bounds are derived for the undirected $\beta$-model, the directed $\beta$-model, and the generalized $\beta$-model. The corresponding maximum likelihood estimators are compared to the bounds by means of simulations. Moreover, examples are given on how to use the presented maximum likelihood estimators to test for directionality and signficance. Last, the applicability of the model is demonstrated using dynamic social network data describing communication among healthcare workers.

Index Terms — The $\beta$-model, Cramér-Rao bounds, hypothesis testing, random graphs, network analysis, dynamic social networks.

I. INTRODUCTION

A random graph is a random variable whose realizations are graphs. Typically, the number of nodes is considered to be fixed. Random graphs have been used to model e.g., interactome networks, with nodes and edges representing molecules and their biological interaction, respectively [1]; brain networks, with nodes and edges representing brain regions and their structural or functional connectivity, respectively [2], [3]; and social networks, with nodes and edges representing social actors (such as individuals or organizations) and their interaction, respectively [4]. In many of these applications, variations in the observed graphs correlate with variations in external covariates. For example, properties of the interactome relate to human disease and the function of specific proteins [5], brain connectivity is closely related to demographic and psychometric measures [6], and social interaction is related to both external events and the demographics and spatial distribution of the studied population [7]. Despite numerous examples of phenomena that can be modeled as dynamic or temporal networks [8], there is still a need to extend or generalize existing statistical methods for static graphs to enable the incorporation of dynamic side information.

The aim of the present study is to develop methods for point estimation, uncertainty estimation, and hypothesis testing, in a setting where the considered observations are graphs, and where covariate information is available on a graph-level. This is done by starting from an already established model for random graphs, namely the $\beta$-model, and then generalizing it so as to enable multivariate regression analysis. In the remainder of this section, we describe the considered graph models, and review previously employed methods for incorporating covariate information into random graphs.

The $\beta$-model belongs to the class of exponential random graphs models (ERGMs), also known as $p^\star$ models, which is the subset of the exponential family that describes random graphs [9]–[13]. Among the ERGMs studied in the literature, the sufficient statistics defining a specific ERGM have included reciprocity measures (describing the tendency of mutual connection between two nodes in a directed graph) [9], the number of k-stars (formations with one node connected to k other nodes but with no additional edges between these k nodes), and the number of triangles (three mutually connected nodes) [10]. The main appeal of ERGMs is that their probability distribution can be specified in terms of any graph attribute that may have relevance for the modeled network.

One of the most simple ERGMs is obtained by letting the sufficient statistics be the vector specifying the degree of each node in the graph [14]. This is a special case of the model for logistic regression, known as the $\beta$-model [15], [16]. The popularity of the $\beta$-model can be attributed to its flexibility in adjusting to observed degree sequences. Specifically, in a $\beta$-model where the model parameters are obtained by using the method of maximum likelihood (ML), the degree of each node has an expectation value equal to the observed degree of the same node [16]. Since the ML estimates of the parameters in the $\beta$-model cannot be written in closed form, several iterative or Monte Carlo-based estimation procedures have been proposed [17]. These have included iterative scaling, Newton’s method, Fisher’s method of scoring, iteratively reweighted least squares [9], [18], and other fixed-point iteration schemes [16].

Information from exogenous variables are often incorporated into random graphs by using nodal (associated with a node) or dyadic (associated with a pair of nodes) covariates. For example, in a social network, a nodal covariate could represent gender or status, whereas a dyadic covariate could represent absolute age difference or spatial distance. Random graphs with nodal covariates are commonly modeled as stochastic block models (SBMs) [19]–[22]. In a SBM, the nodes are divided into groups (blocks), with each group collecting all nodes with some given covariate values. The distribution of the random variable describing the relation between the nodes in the same group is determined by a set of parameters.

The work of A. Nehorai was supported by AFOSR Grant No. FA9550-11-1-0210.

J. Wahlström, I. Skog, and P. Händel are with the ACCESS Linnaeus Center, Dept. of Signal Processing, KTH Royal Institute of Technology, Stockholm, Sweden. (e-mail: {jwahlstr, skog, ph}@kth.se).

P. S. La Rosa is with Global IT Analytics, Monsanto Company, and with the School of Engineering, Washington University in St. Louis, MO 63130 USA (e-mail: pslaro@monsanto.com).

A. Nehorai is with the Preston M. Green Department of Electrical and Systems Engineering, Washington University in St. Louis, MO 63130 USA (e-mail: nehorai@ese.wustl.edu).

arXiv:1611.05699v1 [stat.ME] 14 Nov 2016
The undirected graphs are assumed to be unweighted and without self-loops. This notation is defined independently in each subsection. All attractiveness groups to which the two nodes belong. ERGMs with nodal and dyadic covariates have previously been studied in [23] and [24].

In this article, we generalize the β-model by incorporating covariates on a graph-level. This allows us to perform regression with a random graph as the dependent variable. ML estimates are obtained by generalizing the fixed-point iteration scheme that was proposed for the β-model in [16]. To characterize the considered estimation problems, Cramér-Rao bounds (CRBs) are derived for the undirected, directed, and generalized β-model, and of large magnitude. For ease of notation, we set $\phi$ introduced the function $\varphi : \mathbb{R}^n \to \mathbb{R}^n$, whose outputs are defined according to

$$\varphi_i(z) = \log(d_i) - \log \sum_{j=1}^n \frac{N_{ij}}{e^{-z_i} + e^{z_i}}$$

for $i = 1, \ldots, n$, and where we have used that $z = [z_1 \ldots z_n]^T$. Rearranging the terms in (5) and comparing with (4) it can be seen that $\theta$ is a fixed point of $\varphi$. Hence, the ML estimates may be found by iterating

$$z^{(m+1)} = \varphi(z^{(m)})$$

until convergence, starting from some initial value $z^{(0)}$. Assuming that a unique ML solution exists and that $N_{ij} = 1$ for all $i \neq j$, geometrically fast convergence to the fixed point was shown in [16] and [25]. Conditions for the asymptotic normality of the ML estimator when $n$ tends to infinity and $N_{ij} = 1$ for all $i \neq j$ were presented in [26]. Necessary and sufficient conditions for the existence of a finite ML estimate were presented in [14], [25], and [27]. As an example, it should be clear that there is no finite $\theta$ satisfying (4) whenever $d_i = \sum_{j=1}^n N_{ij}$ or $d_i = 0$ for some $i \in \{1, \ldots, n\}$.

Consider a graph with $n$ nodes. In the undirected β-model, the probability of having an edge between the $i$th node and the $j$th node is

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

for any $i \neq j$. Hence, the $n(n - 1)/2$ edge probabilities are parameterized by the $n$ parameters $\beta_1, \ldots, \beta_n$. As may be realized from (1), $\beta_i$ signifies the tendency of the $i$th node to form edges with other nodes, i.e., its so called differential attractiveness [9]. Specifically, the $i$th node can be expected to have a large (small) number of ties if $\beta_i$ is positive (negative) and of large magnitude. For ease of notation, we set $p_{ii} = 0$ for $i = 1, \ldots, n$. The model is illustrated in Fig. 1, where $S(\beta) \triangleq 1/(1 + e^{-\beta})$ denotes the sigmoid function.

Now, let $Y_{ij} = Y_{ji}$ denote the number of times that an edge between node $i$ and node $j$ is present in $N_{ij} = N_{ji}$ independent measurements. It then follows that

$$Y_{ij} \sim \text{Bin}(N_{ij}, p_{ij})$$

where $\text{Bin}(\cdot, \cdot)$ denotes the binomial distribution with the first and second parameter indicating the number of observations and the probability of success, respectively. Here, we let $N_{ii} = 0$ for $i = 1, \ldots, n$. Moreover, the likelihood function of $\theta \triangleq [\beta_1 \ldots \beta_n]^T$, conditioned on $N_{ij}$ observations associated with each unordered pair of nodes $(i, j)$, is proportional to [16]

$$L(\theta | Y) = \frac{e^{\sum_{i=1}^n \beta_i d_i}}{\prod_{i=1}^n (1 + e^{\beta_i})^{N_{ii}}}$$

(3)

Here, $Y \triangleq \{Y_{ij}\}_{i=1,j=i+1}^n$, while the degree of node $i$, summed over all observations, have been defined as $d_i = \sum_{j=1}^n Y_{ij}$. We will, with some abuse of terminology, continue to refer to $d_i$ as a degree. Since the Hessian of $L(\theta | Y)$ with respect to $\theta$ can easily be shown to be negative semidefinite, the ML estimates $\hat{\theta} \triangleq [\hat{\beta}_1 \ldots \hat{\beta}_n]^T$ must necessarily satisfy

$$\frac{\partial L(\theta | Y)}{\partial \beta_i} = 0 \quad \text{for} \quad i = 1, \ldots, n.$$ 

This is equivalent to saying that [16]

$$d_i = \sum_{j=1}^n N_{ij} \frac{e^{\hat{\beta}_i + \hat{\beta}_j}}{1 + e^{\hat{\beta}_i + \hat{\beta}_j}}$$

for $i = 1, \ldots, n$. The interpretation of (4) is that the observed degree of each node must be equal to the expectation value of the corresponding degree under the model that is implied by the ML estimates. This interpretation makes use of the fact that expected value of $Y_{ij}$ is $N_{ij} p_{ij}$. To find the ML estimates, [16] introduced the function $\varphi : \mathbb{R}^n \to \mathbb{R}^n$, whose outputs are defined according to

$$\varphi(z) = \log(d_i) - \sum_{j=1}^n \frac{N_{ij}}{e^{-z_i} + e^{z_i}}$$

for $i = 1, \ldots, n$, and where we have used that $z = [z_1 \ldots z_n]^T$. Rearranging the terms in (5) and comparing with (4) it can be seen that $\theta$ is a fixed point of $\varphi$. Hence, the ML estimates may be found by iterating

$$z^{(m+1)} = \varphi(z^{(m)})$$

until convergence, starting from some initial value $z^{(0)}$. Assuming that a unique ML solution exists and that $N_{ij} = 1$ for all $i \neq j$, geometrically fast convergence to the fixed point was shown in [16] and [25]. Conditions for the asymptotic normality of the ML estimator when $n$ tends to infinity and $N_{ij} = 1$ for all $i \neq j$ were presented in [26]. Necessary and sufficient conditions for the existence of a finite ML estimate were presented in [14], [25], and [27]. As an example, it should be clear that there is no finite $\theta$ satisfying (4) whenever $d_i = \sum_{j=1}^n N_{ij}$ or $d_i = 0$ for some $i \in \{1, \ldots, n\}$.

![Fig. 1. Illustration of the undirected β-model with $n = 3$ nodes.](image-url)
B. The Directed $\beta$-model

In the directed $\beta$-model [9], the probability of having an edge directed from the $i$th node to the $j$th node is

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

(7)

for any $i \neq j$, while $p_{ii} = 0$ for $i = 1, \ldots, n$. Assuming $n$ nodes, this means that the $n(n-1)$ edge probabilities are parameterized by the $2n$ parameters $\alpha_1, \alpha_2, \ldots, \alpha_n, \beta_1, \ldots, \beta_n$. As should be obvious, $\alpha_i$ affects the probability of having an edge directed from the $i$th node, while $\beta_j$ affects the probability of having an edge directed to the $j$th node. The parameters $\alpha_i$ and $\beta_j$ are called productivity and attractiveness parameters, respectively [9]. To resolve the additive ambiguity of the parameters ($p_{ij}$ is not altered when adding a constant to $\alpha_i$ for $i = 1, \ldots, n$ and simultaneously subtracting the same constant from $\beta_j$ for $i = 1, \ldots, n$), we will without loss of generality assume that $\beta_n$ is known to be zero. The model is illustrated in Fig. 2.

Now, let $Y_{ij}$ denote the number of times that an edge directed from node $i$ to node $j$ is present in $N_{ij}$ independent measurements. Once again, we let $N_{ii} = 0$ for $i = 1, \ldots, n$. It then follows that

$$ Y_{ij} \sim \text{Bin}(N_{ij}, p_{ij}). $$

(8)

Further, assuming that we have $N_{ij}$ observations associated with each ordered pair of nodes $(i, j)$, the likelihood function of $\theta \triangleq [\alpha_1 \ldots \alpha_n \beta_1 \ldots \beta_{n-1}]^\top$ is proportional to

$$ L(\theta|Y) = \frac{e^{\sum_{i=1}^n\alpha_i d_i^\alpha + \beta_i d_i^\beta}}{\prod_{i,j=1}^n (1 + e^{\alpha_i + \beta_j})^{N_{ij}}}.$$

(9)

Here, $Y \triangleq \{Y_{ij}\}_{i,j=1}^n$, while the outdegree and indegree of node $i$, summed over all observations, have been defined as $d_i^\alpha \triangleq \sum_{j=1}^n Y_{ij}$ and $d_i^\beta \triangleq \sum_{j=1}^n Y_{ji}$, respectively. The ML estimates $\hat{\theta} \triangleq [\hat{\alpha}_1 \ldots \hat{\alpha}_n \hat{\beta}_1 \ldots \hat{\beta}_{n-1}]^\top$ must necessarily satisfy $\partial L(\hat{\theta}|Y)/\partial \alpha_i = 0$ for $i = 1, \ldots, n$ and $\partial L(\hat{\theta}|Y)/\partial \beta_i = 0$ for $i = 1, \ldots, n-1$. This is equivalent to saying that

$$ d_i^\alpha = \sum_{j=1}^n N_{ij} \frac{e^\hat{\alpha}_i + \hat{\beta}_j}{1 + e^{\hat{\alpha}_i + \hat{\beta}_j}} $$

(10a)

for $i = 1, \ldots, n$, where we have used that $\hat{\beta}_n = 0$, and

$$ d_i^\beta = \sum_{j=1}^n N_{ji} \frac{e^\hat{\alpha}_j + \hat{\beta}_i}{1 + e^{\hat{\alpha}_j + \hat{\beta}_i}} $$

(10b)

for $i = 1, \ldots, n-1$. In similarity with (4), equation (10) can be interpreted as saying that the observed outdegree and indegree of each node must be equal to the expectation values of the corresponding degrees under the model that is implied by the ML estimates. To find the ML estimates, we introduce the function $\varphi: \mathbb{R}^{2n-1} \to \mathbb{R}^{2n-1}$ whose outputs are defined according to

$$ \varphi_i(z) \triangleq \log(d_i^\alpha) - \log \sum_{j=1}^n \frac{N_{ij}}{e^{-z_{i+n}} + e^{z_i}}, $$

(11a)

for $i = 1, \ldots, n$, and

$$ \varphi_{i+n}(z) \triangleq \log(d_i^\beta) - \log \sum_{j=1}^n \frac{N_{ji}}{e^{-z_j} + e^{z_{i+n}}}, $$

(11b)

for $i = 1, \ldots, n-1$. Here, we have used that $z \triangleq [z_1 \ldots z_{2n-1}]^\top$ and $z_2n \triangleq 0$. Rearranging the terms in (11) and comparing with (10) it can be seen that $\hat{\theta}$ is a fixed point of $\varphi(z)$. Hence, we may, just as for the undirected model, find the ML estimates by iterating

$$ z^{(m+1)} = \varphi(z^{(m)}) $$

(12)

until convergence, starting from some initial value $z^{(0)}$.

C. The Generalized $\beta$-model

We will now demonstrate how to generalize the $\beta$-model for the purpose of performing regression. Using this generalization, the estimated model parameters will not only provide us with a static description of the random graph, but also reveal...
how the random graph varies with the chosen covariates. The core idea is to replace the parameters \( \alpha_i \) and \( \beta_j \) with the regression functions \( \alpha_i^T x \) and \( \beta_j^T x \), respectively. Hence, the probability of having an edge directed from the \( i \)th node to the \( j \)th node becomes

\[
p_{ij}(x) = \frac{e^{\alpha_i^T x + \beta_j^T x}}{1 + e^{\alpha_i^T x + \beta_j^T x}}
\]

for any \( i \neq j \), while \( p_{ii}(x) = 0 \) for \( i = 1, \ldots, n \). The probability depends on both the 2K regression coefficients \( \alpha_i \triangleq [\alpha_{i,1} \ldots \alpha_{i,K}]^T \) and \( \beta_j \triangleq [\beta_{j,1} \ldots \beta_{j,K}]^T \), and the \( K \) covariates \( x = [x_1 \ldots x_K]^T \in \Omega \). Here, \( \alpha_{i,k} \) and \( \beta_{j,k} \) describe the effect that the \( k \)th covariate has on the tendency of the \( i \)th node to form edges with other nodes. For example, if \( \alpha_{i,k} \) is positive (negative), the probability of having an edge directed from the \( i \)th node increases (decreases) as \( x_k \) increases. In analogy with the preceding subsection, we let \( \beta_n = 0_{K,1} \) where \( 0_{i,j} \) is the zero matrix of dimension \( \ell_1 \times \ell_2 \). The model is illustrated in Fig. 3.

Now, assume that we know the covariate values \( \{x_i\}_{i=1}^L \) and have made the associated observations \( Y_i \triangleq \{Y_{ij,\ell}\}_{j=1}^n \) from a set of \( L \) random graphs, each with \( n \) nodes. Here, \( Y_{ij,\ell} \) is the number of times that an edge directed from node \( i \) to node \( j \) is present in \( N_{ij,\ell} \) independent measurements from random graph \( \ell \), where \( N_{ii,\ell} = 0 \) for \( i = 1, \ldots, n \) and \( \ell = 1, \ldots, L \). It then holds that

\[
Y_{ij,\ell} \sim \text{Bin}(N_{ij,\ell}, p_{ij}(x_\ell)).
\]

Furthermore, it can be seen that the likelihood function of \( \theta \triangleq [\alpha_1^T \ldots \alpha_n^T \beta_1^T \ldots \beta_n^T]^T \) is proportional to

\[
L(\theta|Y) = \prod_{\ell=1}^L \frac{\prod_{i=1}^n e^{\sum_{j=1}^n \alpha_i^T x_j d_{i,j,\ell} + \beta_j^T x_j d_{j,\ell}}}{\prod_{i=1}^n (1 + e^{\sum_{j=1}^n \alpha_i^T x_j + \beta_j^T x_j})^{N_{i,\ell}}}.
\]

where the outdegree and indegree of node \( i \) in random graph \( \ell \) have been defined as \( d_{i,j,\ell} \triangleq \sum_{j=1}^n Y_{ij,\ell} \) and \( d_{j,\ell} \triangleq \sum_{j=1}^n Y_{ij,\ell} \), respectively. In analogy with the derivation of (10), the ML estimates \( \hat{\theta} \triangleq [\hat{\alpha}_1^T \ldots \hat{\alpha}_n^T \hat{\beta}_1^T \ldots \hat{\beta}_n^T]^T \) must satisfy

\[
\frac{\partial L(\theta|Y)}{\partial \alpha_{i,k}} = 0 \quad \text{for} \quad i = 1, \ldots, n \quad \text{and} \quad k = 1, \ldots, K, \quad \text{and} \quad \frac{\partial L(\theta|Y)}{\partial \beta_{j,k}} = 0 \quad \text{for} \quad i = 1, \ldots, n \quad \text{and} \quad k = 1, \ldots, K.
\]

This further means that

\[
\sum_{\ell=1}^L x_{k,\ell} d_{i,j,\ell} = \sum_{\ell=1}^L n \cdot x_{k,\ell} N_{ij,\ell} \frac{e^{\alpha_i^T x_j + \beta_j^T x_j}}{1 + e^{\alpha_i^T x_j + \beta_j^T x_j}},
\]

for \( i = 1, \ldots, n \) and \( k = 1, \ldots, K \), where we have used that \( x_{k,\ell} \triangleq [x_{k,1} \ldots x_{k,L}]^T \) and \( \beta_n \triangleq 0_{K,1} \), while

\[
\sum_{\ell=1}^L x_{k,\ell} d_{j,\ell} = \sum_{\ell=1}^L n \cdot x_{k,\ell} N_{ij,\ell} \frac{e^{\alpha_i^T x_j + \beta_j^T x_j}}{1 + e^{\alpha_i^T x_j + \beta_j^T x_j}},
\]

for \( i = 1, \ldots, n \) and \( k = 1, \ldots, K \).

We now introduce the function \( \psi : \mathbb{R}^{(2n-1)K} \times \mathbb{R}^{(2n-1)K} \rightarrow \mathbb{R}^{(2n-1)K} \) whose outputs are defined according to

\[
\psi_K(i-1) + k(x, y) = \sum_{\ell=1}^L x_{\ell,k} \left( \frac{e^{\alpha_i^T x_{\ell,\ell} + \beta_j^T x_{\ell,\ell}}}{e^{\alpha_i^T x_{\ell,\ell} + \beta_j^T x_{\ell,\ell}} + e^{\gamma^T x_{\ell,\ell}}}, \right),
\]

for \( i = 1, \ldots, n \) and \( k = 1, \ldots, K \), and

\[
\psi_K(i+n-1) + k(x, y) = \sum_{\ell=1}^L x_{\ell,k} \left( \frac{e^{\alpha_i^T x_{\ell,\ell} + \beta_j^T x_{\ell,\ell} + \gamma^T y_{\ell,k}}}{e^{\alpha_i^T x_{\ell,\ell} + \beta_j^T x_{\ell,\ell} + \gamma^T y_{\ell,k}} + e^{\gamma^T x_{\ell,\ell}}}, \right),
\]

for \( i = 1, \ldots, n-1 \) and \( k = 1, \ldots, K \). Here, we have used that \( x \triangleq [x_1 \ldots x_{2n-1}]^T \), \( y \triangleq [y_{K,1} \ldots y_{K,n}]^T \), and \( \gamma \triangleq [\gamma_1 \ldots \gamma_{2n-1}]^T \), while \( x_i \triangleq [z_{i,1} \ldots z_{i,K}]^T \) and \( \gamma_i \triangleq [\gamma_{i,1} \ldots \gamma_{i,K}]^T \) for \( i = 1, \ldots, 2n-1 \). Finally, we define \( \varphi : \mathbb{R}^{(2n-1)K} \rightarrow \mathbb{R}^{(2n-1)K} \) as

\[
\varphi(z) = \{z : \psi(z, \gamma) = 0\}.
\]

In the degenerate case of \( L = 1, K = 1 \), and \( x_1 = 1, \varphi(z) \) reduces to the form given in equation (11).

Rearranging the terms in (17) and comparing with (16) it can be seen that \( \theta \) is a fixed point of \( \varphi(z) \). Hence, we will attempt to find the ML estimates by iterating

\[
z^{(m+1)} = \varphi(z^{(m)})
\]

until convergence, starting from some initial value \( z^{(0)} \). In general, there is no closed-form expression for the elements in \( \varphi(z^{(m)}) \). The exceptions are the elements providing updates of eventual intercepts \( \alpha_i \) and \( \beta_i \), i.e., when \( x_{k,\ell} = x_i \) for \( \ell = 1, \ldots, L \). For these parameters, closed-form expressions similar to those in (11) are always available. For the remaining parameters, we can use that each individual element of \( \psi(z, \gamma) \) only depends on one matching element in \( \gamma \). Therefore, it is easy to numerically find the corresponding elements of \( \varphi(z) \) by applying some one-dimensional root-finding algorithm.

D. Initial parameter estimate \( z^{(0)} \) was defined as the zero vector in all three models. The estimates were then iteratively updated as \( z^{(m+1)} = \varphi(z^{(m)}) \) until \( \|z^{(m+1)} - z^{(m)}\| < \epsilon \), where the threshold was set to \( \epsilon = 10^{-4} \). Here, \( \| \cdot \| \) denotes the Euclidean norm.

Several alternative methods, including Newton’s method, may be used to find the ML estimates. However, as was hinted in [9], a standard implementation of Newton’s method will often lead to divergence in low signal-to-noise ratio environments, i.e., where the edge probabilities are close to 0 or 1 and the number of observations is small. According to the authors’ experience, the fixed-point method presented in this section has more favorable convergence properties. This could also have been expected given the convergence theorems, earlier mentioned in Section II-A, that were presented for iteration (5) in [16] and [25].

III. CRAMÉR-RAO BOUNDS

A common way to assess the performance of an estimator is to compare its mean square error (MSE) to the CRB. The CRB
provides a lower bound on the MSE of any unbiased estimator, and can be computed directly from the likelihood function. What is more, the CRB can often be used to characterize the estimation problem in terms of its underlying parameters. We write the bound as [28]

\[ \text{Cov}(\hat{\theta}) \geq \mathbf{P} \]  

(20)

where \( \mathbf{P} \) is the inverse of the Fisher information matrix (FIM), and we have used \( \mathbf{A} \preceq \mathbf{B} \) to denote that \( \mathbf{A} \) is positive semidefinite. The FIM is defined as \( \mathcal{I} = \mathbb{E}[s s^\top] \) where \( s \) is the gradient of the log-likelihood function (the score function) with respect to \( \theta \). In this section, we will derive CRBs for the three models discussed in Section II, and use simulations to compare the performance of the ML estimators to the CRBs. Each subsection uses the same notation as in the corresponding subsection of Section II. Similarly, the notation for FIMs and inverse FIMs is defined independently in each subsection.

A. The Undirected \( \beta \)-model

In the undirected \( \beta \)-model, the element in the \( i \)th row and \( j \)th column of the FIM takes the form

\[ \mathcal{I}_{ij} = \begin{cases} \sum_{q=1}^{n} N_{iq} p_i (1 - p_q), & i = j \\ N_{ij} p_i (1 - p_j), & i \neq j \end{cases} \]  

(21)

Expressed in words, the Fisher information for \( \beta_i \) is the sum of variances of the associated edge observations \( Y_{i1}, \ldots, Y_{in} \). Similarly, the Fisher information shared between \( \beta_i \) and \( \beta_j \) is equal to the variance of \( Y_{ij} \). A derivation of (21) is provided in Appendix A.

To demonstrate the effect that the parameters have on the estimation performance, let us consider the special case of \( N_{ij} = N \) for all \( i \neq j \), \( n > 2 \), \( \beta_i = \beta \) for \( i = 1, \ldots, n \), and hence also \( p_{ij} = p \equiv e^{2\beta} / (1 + e^{2\beta}) \) for all \( i \neq j \). As shown in Appendix A, the inverse FIM then becomes

\[ \mathbf{P} = \frac{1}{Np(1-p)} \frac{1}{(n-2)} \left( \mathbf{I}_n - \frac{1}{2(n-1)} \mathbf{1}_n \right). \]  

(22)

Here, \( \mathbf{I}_n \) denotes the identity matrix of dimension \( n \), while \( \mathbf{1}_n \) denotes the \( n \times n \)-dimensional matrix with all elements equal to one. Studying the diagonal elements of \( \mathbf{P} \), it can be seen that the variance of the individual parameter estimates are bounded from below according to

\[ \text{Var}(\hat{\beta}_i) \geq \frac{1}{Np(1-p)} \frac{2n-3}{2(n-1)(n-2)} \approx \frac{1}{Np(1-p)} \frac{1}{n} \]  

(23)

where \( \approx \) denotes asymptotic equality in the limit of \( n \to \infty \). Hence, we see that the CRB is, in the limit of \( n \to \infty \), inversely proportional to both the number of observations \( N \) and the number of nodes \( n \). Further, the CRB attains its minimum for \( p = 1/2 \) (\( \beta = 0 \)), and approaches infinity both as \( p \to 1 \) (\( \beta \to \infty \)) and \( p \to 0 \) (\( \beta \to -\infty \)). Note that we do not assume knowledge of the fact that all parameters \( \beta_1, \ldots, \beta_n \) have the same value. When assuming that this is known, the model reduces to the Erdős-Rényi model [29].

Now, continuing with the same the parameter restrictions, Fig. 4 shows how the root-mean-square error (RMSE) of the ML estimator compares to the CRB for \( N = 5 \) and \( N = 10 \). Here, the RMSE was computed from \( 10^4 \) simulations\(^1\) with \( n = 10 \), and \( \text{CRB}(\beta_i) \) is used to denote the square root of the right-hand side of equation (23). Due to symmetry, we only considered \( p \geq 0.5 \). As can be seen from Fig. 4, the ML estimator slightly exceeds the CRB when \( N = 5 \), and closely follows the CRB when \( N = 10 \). In both cases, the discrepancy between the RMSE and the CRB increases as \( p \) approaches 1.

B. The Directed \( \beta \)-model

As shown in Appendix B, the FIM for the directed \( \beta \)-model is

\[ \mathcal{I} = \begin{bmatrix} \mathcal{I}_\alpha & \mathcal{I}_{\alpha,\beta} \\ \mathcal{I}_{\alpha,\beta}^\top & \mathcal{I}_{\beta} \end{bmatrix} \]  

(24)

where \( \mathcal{I}_\alpha \) is the diagonal matrix of dimension \( n \) whose \( i \)th diagonal element is \( \sum_{j=1}^{n} N_{ij} p_i (1 - p_j) \), \( \mathcal{I}_{\alpha,\beta} \) is the matrix of dimension \( n \times n - 1 \) whose element in the \( i \)th row and \( j \)th column is \( N_{ij} p_i (1 - p_j) \), and \( \mathcal{I}_{\beta} \) is the diagonal matrix of dimension \( n - 1 \) whose \( i \)th diagonal element is \( \sum_{j=1}^{n-1} N_{ji} p_j (1 - p_i) \).

Now, consider the special case when \( N_{ij} = N \) for all \( i \neq j \), \( n > 2 \), \( \alpha_i = \alpha \) for \( i = 1, \ldots, n \), and \( \beta_i = \beta \) for \( i = 1, \ldots, n-1 \). Without loss of generality, we assume that \( \beta = 0 \), so that \( p_{ij} = p \equiv e^\alpha / (1 + e^\alpha) \) for all \( i \neq j \). The inverse FIM then becomes

\[ \mathbf{P} = \frac{1}{Np(1-p)} \begin{bmatrix} \mathbf{P}_\alpha \\ \mathbf{P}_{\alpha,\beta} \end{bmatrix} \]  

(25)

where

\[ \mathbf{P}_\alpha = \begin{bmatrix} \mathbf{P}_{\alpha,1} & \mathbf{P}_{\alpha,2} & \mathbf{P}_{\alpha,3} \end{bmatrix} \]  

(26)

with

\[ \mathbf{P}_{\alpha,1} = \frac{n-1}{n(n-2)} \begin{bmatrix} \mathbf{I}_{n-1} + \frac{n^2 - 3n + 1}{(n-1)^2} \mathbf{1}_{n-1} \end{bmatrix}, \]  

(27a)

\[ \mathbf{P}_{\alpha,2} = \frac{1}{n-1} \mathbf{1}_{n-1}, \]  

(27b)

\[ \mathbf{P}_{\alpha,3} = \frac{2n-3}{(n-1)(n-2)}, \]  

(27c)

\(^1\)To avoid non-existence of finite ML estimates, we disregarded any simulation where some node degree was equal to zero or equal to its maximum possible value.
and
\[
\mathbf{P}_{\alpha,\beta} \triangleq - \frac{1}{n(n-2)} \left[ (n-1) \mathbf{I}_{n-1} - \mathbf{I}_{n-1} \right],
\]
\[
\mathbf{P}_\beta \triangleq \frac{n-1}{n(n-2)} (\mathbf{I}_{n-1} + \mathbf{I}_{n-1}).
\]
(28a)
(28b)

Here, we have used \( \mathbf{I}_{n,m} \) to denote the \( n \times m \) dimensional matrix with all elements equal to one. Further, studying the diagonal elements of \( \mathbf{P} \) we arrive at
\[
\text{Var}(\hat{\alpha}_i) \geq \frac{1}{Np(1-p)} \frac{2n-1}{n(n-1)},
\]
(29a)
for \( i = 1, \ldots, n-1 \),
\[
\text{Var}(\hat{\alpha}_n) \geq \frac{1}{Np(1-p)} \frac{2n-3}{(n-1)(n-2)},
\]
(29b)
and
\[
\text{Var}(\hat{\beta}_i) \geq \frac{1}{Np(1-p)} \frac{2(n-1)}{n(n-2)},
\]
(29c)
for \( i = 1, \ldots, n-1 \). Just as for the undirected model, the CRB is, in the limit of \( n \to \infty \), inversely proportional to both \( N \) and \( n \), while reaching its minimum for \( p = 1/2 \) (\( \beta = 0 \)). Noting that the CRBs in (29) are asymptotically equivalent to \( 2/(Np(1-p)n) \) and comparing with (23), the asymptotic CRBs on the variance can be seen to be twice as large in the directed \( \beta \)-model as in the undirected \( \beta \)-model (assuming the same \( N, n, \) and \( p \)).

Given the considered parameter settings, Fig. 5 shows how the ML estimator of \( \alpha_i \) with \( i \neq n \) compares to the CRB in (29a) for \( N = 5 \) and \( N = 10 \). The RMSE was computed from \( 10^4 \) simulations and the number of nodes was \( n = 10 \). Overall, the estimator displays the same qualitative behavior as for the undirected model in Fig. 4.

C. The Generalized \( \beta \)-model

For the generalized \( \beta \)-model, we consider the special case when \( N_{i,\ell} = N \) for all \( i \neq j \) and \( \ell = 1, \ldots, L, n > 2, \alpha_i = \alpha \) for \( i = 1, \ldots, n, \) and \( \beta_\ell = \beta \) for \( i = 1, \ldots, n-1 \). Without loss of generality, we assume that \( \beta = 0 \). Further, this also means that \( p_{ij}(x) = p(x) \triangleq e^{\alpha x}/(1 + e^{\alpha x}) \) for all \( i \neq j \). Using \( \otimes \) to denote the Kronecker product, the inverse FIM becomes
\[
\mathbf{P} = \frac{1}{N} \left[ \mathbf{P}_\alpha \mathbf{P}_\alpha^\top \mathbf{P}_\alpha,\beta \right] \otimes \mathbf{I}_{\mathbf{x}}^{-1}
\]
(30)
where \( \mathbf{I}_{\mathbf{x}} \triangleq \sum_{\ell=1}^L p(x)(1-p(x))x_i x_j^\top \), while \( \mathbf{P}_\alpha, \mathbf{P}_{\alpha,\beta} \), and \( \mathbf{P}_\beta \) are as defined in equations (26), (27), and (28).

Hence, the CRBs for the regression coefficients pertaining to a given node are subject to the same dependence on \( n \) as the regression coefficients pertaining to the corresponding node in the special case of the directed \( \beta \)-model considered in the preceding subsection. (As a side note, we remark that the analogous statement can be made regarding the CRB for the undirected generalized model in Section III-A and the CRB for the undirected model presented in Section III-A and the CRB for the undirected generalized model, i.e., the model where \( \alpha_{i,k} = \beta_{i,k} \) for \( i = 1, \ldots, n \) and \( k = 1, \ldots, K \), while discarding the assumption \( \beta_\ell = 0 \).) The CRB for the different regression coefficients associated with a given node are then weighted based on \( \mathbf{I}_{\mathbf{x}}^{-1} \). Since the rank of \( \mathbf{I}_{\mathbf{x}} \) is \( \text{rank}(\mathbf{I}_{\mathbf{x}}) \leq \sum_{\ell=1}^L \text{rank}(p(x)(1-p(x))x_i x_j^\top) \leq L, \) we must have \( L \geq K \) for \( \mathbf{I}_{\mathbf{x}} \) to be invertible (this condition was expected since the dimension of the observations \( \{(\mathbf{d}_{\alpha,i}^\ell, \mathbf{d}_{\beta,i}^\ell)\}_{i=1}^n \) is \( 2nL \) and the dimension of \( \theta \) is \( 2nK \)). Appendix C first presents the FIM in the general case and then derives the inverse FIM in the special case considered here.

Now, to illustrate the performance of the ML estimator, we set \( n = 5, L = 2, x_1 = [1 0]^\top, x_2 = [1 1]^\top, \) and \( \alpha = [1 1]^\top. \) Fig. 6 then shows how the ML estimator of \( \alpha_{i,1} \) with \( i \neq n \) compares to the CRB as \( \alpha \) varies. Here, the average edge probability has been defined as \( \bar{p} = (p(x_1) + p(x_2))/2 \). Both the cases \( N = 5 \) and \( N = 10 \) were considered, and the RMSE was computed from \( 10^4 \) simulations. Once again, the RMSE exceeds the CRB for small \( N \) and then approaches the CRB as \( N \) is modestly increased.

To summarize, this section has presented CRBs for all three studied \( \beta \)-models. The dependence of the CRBs on the model parameters (number of observations, number of nodes in the graph, and edge probabilities) was shown to be similar across the models. Further, simulations indicated that the ML estimators presented in Section II have a RMSE in the vicinity of the CRB already for a comparatively small number of measurements.

IV. HYPOTHESIS TESTING

The problem of choosing between two possible hypotheses is often approached by performing a likelihood ratio test (LRT). This means that the decision is taken based on whether the ratio of the likelihoods of the two hypotheses exceed or fall below some given constant. As stated in the Neyman-Pearson lemma, the LRT is the most powerful of all possible
hypothesis tests [30]. Put differently, the LRT minimizes the probability of a missed detection (type II error) for a given probability of a false alarm (type I error). When the likelihoods of the hypotheses are dependent on some unknown parameters, it is common to apply the LRT with the parameter values that maximize the respective likelihoods. This is referred to as the generalized likelihood ratio test (GLRT) [31]. If the two hypotheses can be expressed as $H_0 : \theta \in \Omega_0$ and $H_1 : \theta \in \Omega_1$, the generalized likelihood ratio becomes

$$\Lambda(Y) = \frac{\sup_{\theta \in \Omega_0} L(\theta | Y)}{\sup_{\theta \in \Omega_1} L(\theta | Y)} \quad (31)$$

and we decide on $H_0$ whenever $\Lambda(Y) > \eta$ for some chosen threshold $\eta$.

The threshold $\eta$ is often chosen so that the decision boundary becomes the set of measurements giving some pre-determined $p$-value. However, to be able to compute $p$-values, we need to know the distribution of $\Lambda(Y)$ under $H_0$. While the exact distribution is seldom known, approximate $p$-values can often be obtained from Wilks’ theorem. Specifically, assume that $\Omega_0$ is a subset of $\Omega_1$ and that $\Omega_1$ has $k$ more free parameters than $\Omega_0$. Then, given certain regularity conditions, Wilks’ theorem states that the asymptotic distribution of $-2 \log \Lambda_{L,R}$ under $H_0$, where $\Lambda_{L,R}$ is the likelihood ratio, is the $\chi^2$-distribution with $k$ degrees of freedom [32]. Analogues to Wilks’ theorem, specified for the generalized likelihood ratio, are discussed in [33].

Next, we demonstrate how to use the ML estimates presented in Section II to perform GLRT’s for two simple null hypotheses of practical importance. Goodness-of-fit tests are omitted from the discussion, and we instead refer the reader to earlier studies on ERGMs [34], the undirected $\beta$-model [35], and the directed $\beta$-model [9], [36].

A. Significance Tests

First, we consider the problem of testing whether a specific covariate has any significant impact on the studied graph. To this end, we use the generalized model from Section II-C, assuming knowledge of $\{x_{i\ell} \}_{i=1}^L$ and $Y = \{Y_{ij,i'=1}^{\tiny{\forall \ell \in L}}\}_{i=1}^L$. Now, to test the significance of the $k$th covariate $x_{i,k}$, we formulate the null hypothesis

$$H_0 : \alpha_{i,k} = \beta_{i,k} = 0 \quad (32)$$

Assuming that the alternative hypothesis $H_1$ does not impose any constraints on the parameters $\theta$, it follows that the maximizing parameter in the denominator of (31) is the ML estimate for the generalized $\beta$-model. Similarly, the maximizing parameter in the numerator of (31) is the ML estimate for the generalized $\beta$-model where the employed regression coefficients can be written as $\alpha_i \triangleq [\alpha_{i,1} \ldots \alpha_{i,k-1} \alpha_{i,k+1} \ldots \alpha_{i,L}]^T$ and $\beta_i \triangleq [\beta_{i,1} \ldots \beta_{i,k-1} \beta_{i,k+1} \ldots \beta_{i,L}]^T$ for $i = 1, \ldots, n$, while the covariates are $x_{\ell,i} \triangleq [x_{\ell,1} \ldots x_{\ell,k-1} x_{\ell,k+1} \ldots x_{\ell,L}]^T$ for $\ell = 1, \ldots, L$.

B. Testing for Directionality

Second, we consider the problem of deciding whether some given set of directed network data $\{Y_{ij}\}_{i,j=1}^n$ has originated from a directed or undirected (symmetric) random graph\(^2\). Using the notation of the directed model presented in Section II-B, we can write the null hypothesis, indicating that the data originates from an undirected graph, as

$$H_0 : \alpha_i = \beta_i \quad \text{for } i = 1, \ldots, n. \quad (33)$$

Assuming that the alternative hypothesis $H_1$ does not impose any constraints on the parameters $\theta$, it follows that the maximizing parameter in the denominator of (31) is the ML estimate for the directed $\beta$-model. Similarly, the maximizing parameter in the numerator of (31) is the ML estimate for the undirected $\beta$-model, where $Y_{ij} + Y_{ji}$ is considered to be the number of times that an edge between node $i$ and node $j$ is present in $N_{ij} + N_{ji}$ independent measurements. It is trivial to extend the test to the generalized $\beta$-model considered in Section II-C.

C. Testing for Directionality - Numerical Example

We now study the receiver operating characteristics (ROC) of the GLRT detector, for the problem of testing for directionality, by means of simulations. For the computation of the true positive rate (the probability of correctly rejecting $H_0$) we independently simulated $\alpha_i$ for $i = 1, \ldots, n$ and $\beta_i$ for $i = 1, \ldots, n-1$ from the uniform distribution over the interval $(-\rho, \rho)$. These parameters were in turn used to simulate $Y_{ij}$ for all $i \neq j$, which then enabled us to compute $\Lambda(Y)$ and, for any given $\eta$, make the decision to reject or accept $H_0$. Finally, the true positive rate was obtained as the percentage of rejections. To compute the false positive rate (the probability of incorrectly rejecting $H_0$), we simulated $\alpha_i$ for $i = 1, \ldots, n$ from the uniform distribution over the interval $(-\rho, \rho)$, while letting $\beta_i = \alpha_i$ for $i = 1, \ldots, n$. After simulating $Y_{ij}$ for all $i \neq j$ (although the distribution of $Y_{ij}$ and $Y_{ji}$ is the same, their simulated values need not be) and computing $\Lambda(Y)$, the false positive rate was obtained as the percentage of rejections of $H_0$. Fig. 7 displays the ROC curve for $\rho = 0.1, 0.2, 0.3, 0.4$, with $N_{ij} = 10$ for all $i \neq j$, $n = 10$, when computing each true positive rate and false positive rate from $10^4$ simulations.

As expected, the classification performance improves as the simulations of $\alpha_1, \ldots, \alpha_n, \beta_1, \ldots, \beta_{n-1}$ are drawn from larger intervals, thereby making the sequences $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_{n-1}\}$ less alike under the alternative hypothesis $H_1$.

Fig. 8 displays the histogram of $-2\log \Lambda(Y)$ under $H_0$ when testing for directionality with $N_{ij} = 10$ for all $i \neq j$, $n = 10$, and $\rho = 0.4$, using $10^4$ simulations. The histogram is overlaid with the probability density function (pdf) of the $\chi^2$-distribution with $n-1$ degrees of freedom (note that $H_0$ is specified by $n$ parameters while $H_1$ is specified by $2n-1$ parameters). As can be seen from Fig. 8, the test statistic seems to follow the $\chi^2$-distribution closely, thereby indicating that Wilks’ theorem can be of use when performing hypothesis tests under conditions similar to those in the given example.

V. A CASE STUDY USING SOCIAL NETWORK DATA

We now apply the theory presented in the preceding sections to data describing an undirected social network of healthcare

\(^2\)A nonparametric test for this problem was developed in [37].
Fig. 7. ROC curve when testing for directionality in the directed β-model with $N_{ij} = 10$ for all $i \neq j$ and $n = 10$.

Fig. 8. The distribution of the GLRT under the null hypothesis compared to the $\chi^2$-distribution when testing for symmetry with $N_{ij} = 10$ for all $i \neq j$, $n = 10$, and $\rho = 0.4$.

Fig. 9. Illustration of the studied social network data. Each dot represents a contact made between two individuals.

workers. Similar data sets have previously been studied using the β-model in [9] and [24]. As noted in [9], the β-model captures the elementary social tendency of differential attraction, while retaining tractability. The data consists of a list of contacts, i.e., events where two individuals were facing each other at a distance less than $1 - 1.5 |m|$, and specifies both the time point (a 20 second interval) at which each contact took place, and the two individuals that were in contact [38]. All in all, contacts among 75 individuals were recorded over a period of 97 hours. However, to ensure the existence of all ML estimates, we chose to study the contacts of ten individuals (one administrator, six nurses, and three medical doctors) as registered between 10.00 and 13.00 during three consecutive days. Fig. 9 displays the contacts (made within the smaller group of ten individuals) of a given individual on a specific day along a horizontal line. As can be seen, the studied individuals were all active during the considered time periods. To construct the graph observations, we divided each day into three one-hour periods, with each period generating one observation of each dyad, i.e., each pair of individuals. An edge was considered to be present if and only if there was a contact between the corresponding individuals during the associated time period. Based on the analysis in Section III, a rule of thumb may be to design the model so that roughly half of the observations indicate that a link is present (here, we had $\sum_{i,j,\ell} Y_{i,j,\ell} / \sum_{i,j,\ell} N_{i,j,\ell} = 0.484$).

Next, we will test how well the studied graphs can be explained by covariates representing the time of day and the day of the week. This is done by incorporating these features as categorical covariates in the generalized β-model. Specifically, the GLRT is used to test a null hypothesis, under which the distributions of the observations are not dependent on any external covariates, against an alternative hypothesis with one binary covariate indicating whether a given observation is made during a specific day or time of day. This test is then performed for each of the six (the three days and the three time periods) possible covariates. For example, when focusing on the period between 10.00 and 11.00, the ML estimates under the alternative hypothesis are computed by using the undirected version of the generalized β-model with $x_1 = [1 
 1]^\top$ and $x_2 = [1 
 0]^\top$. The contacts displayed in the three leftmost boxes in Fig. 9 are used to obtain the measurements of the first graph, while the contacts in the remaining boxes provide the measurements of the second graph. Obviously, this means that $N_{ij,\ell} = 3\ell$ for all $i \neq j$ and $\ell = 1, 2$. The ML estimates under the null hypothesis are obtained by using the undirected β-model and merging all graph data so that $N_{ij} = 9$ for all $i \neq j$. The $p$-value of the GLRT can then be estimated in two ways. Firstly, we can study the distribution of the generalized likelihood ratios obtained by simulating measurements from the undirected β-model. The simulations are made using the ML estimates of the undirected β-parameters obtained from the real-world data. Secondly, we can employ Wilks’ theorem and assume that $-2 \log \Lambda(Y)$ is $\chi^2$-distributed with $n = 10$ degrees of freedom. Now, Tables I and II display the simulated and theoretical (as estimated using Wilks’ theorem) $p$-values obtained when testing the significance of each of the three time periods (one at a time) and each of the three days (one at a time), respectively. The simulated $p$-values where obtained from $10^4$ simulations. As can be seen, the networks seem to vary considerably with the time of day (two out of three time intervals have a $p$-value smaller than 0.01), while the day-to-day variations are minor (no day has a $p$-value smaller than 0.05).

To further investigate how the network depends on the
studied time intervals, the network was regressed on all three
time-of-day categories simultaneously. Hence, we used the
contacts gathered between 10.00 and 11.00, between 11.00
and 12.00, and between 12.00 and 13.00, to obtain measure-
ments of the first, second, and third graph, respectively. The
associated covariates where \(x_1 = [1 0 0]^\top\), \(x_2 = [1 1 0]^\top\),
and \(x_3 = [1 0 1]^\top\). The resulting estimates are illustrated in
Fig. 10, where the vertical lines represent uncertainty intervals
\([\hat{\beta}_{i,k} - \text{CRB}(\hat{\beta}_{i,k}), \hat{\beta}_{i,k} + \text{CRB}(\hat{\beta}_{i,k})]\), and \(\text{CRB}(\hat{\beta}_{i,k})\) denotes
the lower bound on the standard deviation of \(\hat{\beta}_{i,k}\) as estimated
from the undirected version of the FIM presented in (54) and
(55). Here, \(\hat{\beta}_{i,1}\) describes the attractiveness of node \(i\) in
the first time period, while \(\hat{\beta}_{i,1} + \hat{\beta}_{i,2}\) and \(\hat{\beta}_{i,1} + \hat{\beta}_{i,3}\) describe
the attractiveness of node \(i\) in the second and third time periods,
respectively. Averaging \(\hat{\beta}_{i,2}\) and \(\hat{\beta}_{i,3}\) over the ten individuals
we obtain \(-0.19\) and \(-0.94\), respectively. In other words,
the overall attractiveness is lower in the second time period
than in the first, and is lower in the third time period than in
the second. This is also what one would have guessed given
the distribution of contacts in Fig. 9 (the total number of
observed links in the networks were 81, 71, and 44, in the
three respective time periods).

VI. Summary

In this paper, we generalized the \(\beta\)-model with the aim of
enabling regression of graphs on a given set of covariates. Fur-
ther, ML estimates and CRBs were derived for the undirected,
directed, and generalized \(\beta\)-model. In the limit where the
number of nodes approaches infinity, the CRB on the variance
of the \(\beta\)-parameters was shown to be inversely proportional
to both the number of nodes and the number of observations.
Moreover, it was demonstrated how to use the ML estimators
to perform statistical tests, with simulations indicating that
Wilks’ theorem can be of use for the computation of \(p\)-values.
A case study where the generalized \(\beta\)-model was applied
to social network data describing contacts among healthcare
workers was worked out in detail. In summary, the article pre-
sented tractable methods for point estimation, model design,
and hypothesis testing in applications studying networks that
depend on external covariates. Future research may study the

\[
\begin{array}{|c|c|c|c|}
\hline
\text{Time of day} & 10.00-11.00 & 11.00-12.00 & 12.00-13.00 \\
\hline
\text{Simulated } p\text{-value} & 0.0051 & 0.2950 & 0.0001 \\
\text{Theoretical } p\text{-value} & 0.0037 & 0.2557 & 5.7 \cdot 10^{-6} \\
\hline
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
\text{Day} & 1 & 2 & 3 \\
\hline
\text{Simulated } p\text{-value} & 0.9361 & 0.1121 & 0.3125 \\
\text{Theoretical } p\text{-value} & 0.9238 & 0.0943 & 0.2693 \\
\hline
\end{array}
\]
FIM in (36) while assuming \( n > 2 \) immediately gives the inverse FIM in (22).

**APPENDIX B**

Here, we derive the FIM in (24), given for the directed \( \beta \)-model. In addition, we also derive the inverse FIM in the special case considered in (25).

Let us denote the score function as \( s^\alpha \triangleq [(s^\alpha_i)^T (s^\beta_j)^T]^T \), where \( s^\alpha \triangleq [s^\alpha_0 \ldots s^\alpha_n]^T \) and \( s^\beta \triangleq [s^\beta_1 \ldots s^\beta_{n-1}]^T \). It then follows from (9) that

\[
s^\alpha_i = \sum_{j=1}^{n} Y_{ij} - N_{ij}p_{ij},
\]

for \( i = 1, \ldots, n \), and

\[
s^\beta_i = \sum_{j=1}^{n} Y_{ji} - N_{ji}p_{ji},
\]

for \( i = 1, \ldots, n-1 \). Further, we use (8) and the independence of measurements from different edges to obtain

\[
\mathbb{E}[(s^\alpha_i)^2] = \sum_{j=1}^{n} N_{ij}p_{ij}(1-p_{ij}),
\]

for \( i = 1, \ldots, n \), and

\[
\mathbb{E}[(s^\beta_i)^2] = \sum_{j=1}^{n} N_{ji}p_{ji}(1-p_{ji}),
\]

for \( i = 1, \ldots, n-1 \), while

\[
\mathbb{E}[s^\alpha_is^\beta_j] = 0,
\]

and

\[
\mathbb{E}[s^\alpha_is^\beta_j] = N_{ij}p_{ij}(1-p_{ij}),
\]

for any \( i \neq j \). The FIM in (24) now follows directly from (39) and (40).

Given the assumptions preceding (25), it follows that the submatrices in (24) are

\[
\begin{align*}
\mathbf{I}_\alpha & = Np(1-p)(n-1)\mathbb{I}_n, \\
\mathbf{I}_\beta & = Np(1-p)(n-1)\mathbb{I}_{n-1}, \\
\mathbf{I}_{\alpha,\beta} & = Np(1-p)[\mathbb{I}_{n-1} - \mathbb{I}_{n-1} \mathbb{I}_{n-1}]^T.
\end{align*}
\]

Further, applying the formula for block matrix inversion [40] and assuming \( n > 2 \), we have that

\[
\begin{align*}
\mathbf{P}_\alpha & = Np(1-p)(\mathbf{I}_\alpha - \mathbf{I}_{\alpha,\beta} \mathbf{I}_\beta^{-1} \mathbf{I}_{\alpha,\beta})^{-1} \quad \text{with } \delta \triangleq Np(1-p)/(n-1). \quad \text{Here, we have used that}
\end{align*}
\]

\[
\begin{align*}
\mathbf{I}_{\alpha,\beta} \mathbf{I}_\beta^{-1} \mathbf{I}_{\alpha,\beta} & = \mathbf{c} \left[ \begin{array}{c} 
1_{n-1} - 1_{n-1} \\
1_{1,n-1}\end{array} \right] \left[ \begin{array}{c} 
1_{n-1} - 1_{n-1} \\
1_{1,n-1}\end{array} \right]^T \\
& = \mathbf{c} \left[ 
\mathbf{1}_{n-1} - 1_{n-1} \mathbf{1}_{n-1} - 1_{n-1} \\
\mathbf{1}_{1,n-1} - 1_{1,n-1} - 1_{n-1,1} \end{array} \right] \\
& = \mathbf{c} \left[ 
\mathbf{1}_{n-1} + (n-3)1_{n-1} \\
(n-2)1_{n-1} - 1_{n-1} \\
n-1 \end{array} \right].
\end{align*}
\]

Using the block matrix inversion formula, it can be seen that

\[
\mathbf{P}_\alpha = \begin{bmatrix}
\mathbf{P}_{\alpha,1} & \mathbf{P}_{\alpha,2} \\
\mathbf{P}_{\alpha,2} & \mathbf{P}_{\alpha,3}
\end{bmatrix}
\]

where

\[
\begin{align*}
\mathbf{P}_{\alpha,1} & = Np(1-p)(\mathbf{I}_\alpha - \mathbf{I}_{\alpha,2} \mathbf{I}_{\alpha,2}^{-1} \mathbf{I}_{\alpha,2})^{-1} \\
& = \frac{1}{n-1} \left( \begin{array}{c} 
(n-2)1_{n-1} - 1_{n-1} \\
(n-3)1_{n-1} - 1_{n-1} \\
2n-3
\end{array} \right),
\end{align*}
\]

Here, the last equality follows from (37). Further, we have

\[
\begin{align*}
\mathbf{P}_{\alpha,3} & = Np(1-p)(\mathbf{I}_\alpha - \mathbf{I}_{\alpha,2} \mathbf{I}_{\alpha,2}^{-1} \mathbf{I}_{\alpha,2})^{-1} \\
& = \frac{1}{n-1} \left( \begin{array}{c} 
(n-2)1_{n-1} - 1_{n-1} \\
(n-3)1_{n-1} - 1_{n-1} \\
2n-3
\end{array} \right),
\end{align*}
\]

where we have used (37) to arrive at

\[
\mathbf{I}_{\alpha,1}^{-1} = \frac{1}{cn(n-2)} \left( \mathbf{I}_{n-1} + \frac{n-3}{2n-3} \mathbb{I}_{n-1} \right).
\]

Finally, it holds that

\[
\begin{align*}
\mathbf{P}_{\alpha,2} & = -\mathbf{I}_{\alpha,1}^{-1} \mathbf{I}_{\alpha,2} \mathbf{P}_{\alpha,3} \\
& = \frac{1}{n-1} \left( \begin{array}{c} 
(n-2)1_{n-1} - 1_{n-1} \\
(n-3)1_{n-1} - 1_{n-1} \\
2n-3
\end{array} \right),
\end{align*}
\]

which concludes the derivation of \( \mathbf{P}_\alpha \) as given in equations (26) and (27).

To derive (28), we begin by showing that

\[
\begin{align*}
\mathbf{P}_\beta & = Np(1-p)(\mathbf{I}_\beta - \mathbf{I}_{\beta,\alpha} \mathbf{I}_{\alpha}^{-1} \mathbf{I}_{\beta,\alpha})^{-1} \\
& = \left( \begin{array}{c} 
(n-1)1_{n-1} - 1/(n-1)\mathbb{I}_{n-1} \\
(n-2)1_{n-1} - 1_{n-1} \\
(n-1)
\end{array} \right)^{-1} \\
& = \frac{n-1}{n-2} \left( \begin{array}{c} 
1_{n-1} \\
1_{n-1} - 1_{n-1} \\
n-1
\end{array} \right).
\end{align*}
\]
where the second equality uses that
\[
I^T_{\alpha,\beta}I^{-1}_{\alpha,\beta} = c\left[\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^T \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \right] = c\left[\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^T \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right]
\]
and the last equality follows from (37). Finally, we have
\[
P_{\alpha,\beta} = -I^T_{\alpha,\beta}I^{-1}_{\alpha,\beta}P_{\beta}
\]
and we are done.

APPENDIX C

We now show how to obtain the FIM for the generalized \(\beta\)-model, and then go on to derive the inverse FIM for the special case considered in (30).

To begin with, let us denote the score function as
\[
s_{i,k}^\alpha = \sum_{\ell=1}^{L} n \sum_{j=1}^n x_{\ell,k}(Y_{ij,\ell} - N_{ij,\ell}p_{ij}(x_\ell))
\]
for \(i = 1, \ldots, n\) and \(k = 1, \ldots, K\), while
\[
s_{i,k}^\beta = \sum_{\ell=1}^{L} n \sum_{j=1}^n x_{\ell,k}(Y_{ij,\ell} - N_{ij,\ell}p_{ij}(x_\ell))
\]
for \(i = 1, \ldots, n - 1\) and \(k = 1, \ldots, K\). Further, using (14) and the independence of measurements from different edges and different graphs we have that
\[
\mathbb{E}[s_{i,k}^\alpha s_{i,k'}^\beta] = \sum_{\ell=1}^{L} n \sum_{j=1}^n x_{\ell,k}x_{\ell,k'}N_{ij,\ell}p_{ij}(x_\ell)(1 - p_{ij}(x_\ell))
\]
for \(i = 1, \ldots, n\), \(k = 1, \ldots, K\), and \(k' = 1, \ldots, K\), and
\[
\mathbb{E}[s_{i,k}^\alpha s_{i,k'}^\beta] = 0,
\]
for \(i = 1, \ldots, n - 1\), \(k = 1, \ldots, K\), and \(k' = 1, \ldots, K\). Moreover, we have that
\[
\mathbb{E}[s_{i,k}^\alpha s_{i,k'}^\beta] = 0
\]
and
\[
\mathbb{E}[s_{i,k}^\alpha s_{i,k'}^\beta] = 0,
\]
for \(i = 1, \ldots, n - 1\), \(k = 1, \ldots, K\), and \(k' = 1, \ldots, K\). The FIM for the generalized \(\beta\)-model can now be obtained from equations (54) and (55).

For the special case considered in Section III-C, it is easily shown that the FIM reduces to
\[
\mathcal{I} = \begin{bmatrix} I_{\alpha}^\top & I_{\alpha,\beta} \end{bmatrix} \mathbb{E} \left[ \begin{bmatrix} I_{\alpha}^\top & I_{\alpha,\beta} \end{bmatrix} \right]
\]
where
\[
I_{\alpha} \triangleq N(n-1)I_n,
\]
\[
I_{\beta} \triangleq N(n-1)I_{n-1},
\]
\[
I_{\alpha,\beta} \triangleq N[1_{n-1} - I_{n-1} 1_{n-1}^\top] \mathbb{I}
\]
and noting the similarity of equations (41) and (57), while using that \((A \otimes B)^{-1} = A^{-1} \otimes B^{-1}\) for any invertible matrices \(A\) and \(B\), the FIM in (30) follows immediately when assuming that \(\mathcal{I}_x\) is invertible and that \(n > 2\).

REFERENCES

[1] M. Vidal, M. E. Cusick, and A.-L. Barabási, “A network of human disease,” Cell., vol. 144, no. 6, pp. 986–998, Mar. 2011.
[2] P. S. L. Rosa, T. L. Brooks, E. Deych, B. Shands, F. Prior, L. J. Larson-Prior, et al., “Gibbs distribution for statistical analysis of exponential family data with a sample application to fMRI brain images,” Statist. Med., vol. 35, no. 4, pp. 566–580, Feb. 2016.
[3] K. W. Andersen, M. Morup, H. Siebner, K. H. Madsen, and L. K. Hansen, “Identifying modular relations in complex brain networks,” in Proc. IEEE Int. Workshop Mach. Learning Signal Process., Santander, Spain, Sep. 2012.
[4] G. Robins, T. Snijders, P. Wang, M. Handcock, and P. Pattison, “Recent developments in exponential random graph (p*) models for social networks,” Social Netw., vol. 29, no. 2, pp. 192–215, May 2007.
[5] R. Sharan, I. Ullity, and R. Shamir, “Network-based prediction of protein function,” Mol. Syst. Biol., vol. 3, no. 88, Mar. 2007.
[6] S. M. Smith, T. E. Nichols, D. Vidaurre, A. M. Winkler, T. E. J. Behrens, M. F. Glasser, K. Ugurbil, D. M. Barch, D. C. Van Essen, and K. L. Miller, “A positive-negative mode of population covariation links brain connectivity, demographics and social strategies in mobile social networks,” in Proc. 20th ACM Int. Conf. Knowledge Discovery and Data Mining, New York, NY, Aug. 2014, pp. 15–24.
[7] Y. Dong, Y. Yang, J. Tang, Y. Yang, and N. V. Chawla, “Inferring user demographics and social strategies in mobile social networks,” in Proc. 20th ACM Int. Conf. Knowledge Discovery and Data Mining, New York, NY, Aug. 2014, pp. 97 – 125, Oct. 2012.
[8] E. M. Frank and D. Strauss, “Markov graphs,” J. American Statistical Association, vol. 81, no. 395, pp. 832–842, Sep. 1986.
[9] D. Strauss, “On a general class of models for interaction,” SIAM Rev., vol. 28, no. 4, pp. 513–527, Dec. 1986.
[10] M. E. J. Newman, “The structure and function of complex networks,” SIAM Rev., vol. 45, no. 2, pp. 167–256, May 2003.
[11] M. N. Schmidt and M. Morup, “Nonparametric Bayesian modeling of complex networks: An introduction,” IEEE Signal Process. Mag., vol. 30, no. 3, pp. 110–128, May 2013.
[12] A. Rinaldo, S. Petrovic, and S. E. Fienberg, “Maximum likelihood estimation in the β-model,” The Ann. Statistics, vol. 41, no. 3, pp. 1085–1110, Jun. 2013.
[13] S. J. Cramer and B. A. Desmarais, “Inferential network analysis with exponential random graph models,” Political Analysis, vol. 19, no. 1, pp. 66–86, 2011.
Johan Wahlström received his MSc degree in Engineering Physics from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2014. He subsequently joined the Signal Processing Department at KTH, working towards his PhD. His main research topic is insurance telematics. In 2015, he received a scholarship from the Sweden-America foundation and spent six months at Washington University, St. Louis, USA.

Isaac Skog (S’09-M’10) received the BSc and MSc degrees in Electrical Engineering from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2003 and 2005, respectively. In 2010, he received the PhD degree in Signal Processing with a thesis on low-cost navigation systems. In 2009, he spent 5 months at the Mobile Multi-Sensor System research team, University of Calgary, Canada, as a visiting scholar and in 2011 he spent 4 months at the Indian Institute of Science (IISc), Bangalore, India, as a visiting scholar. He is currently a Researcher at KTH coordinating the KTH Insurance Telematics Lab. He was a recipient of a Best Survey Paper Award by the IEEE Intelligent Transportation Systems Society in 2013.

Patricio S. La Rosa (S’07-M’09) received the BSc degree in engineering from the Pontificial Catholic University of Chile (PUC) in 1999, the MSc degree (with maximum distinction) in electrical engineering from the University of Chile in 2003, and the MSc and the PhD degrees in electrical engineering from Washington University in St. Louis (WUSTL) in 2010, under the guidance of Prof. A. Nehorai. He was a Research Assistant in the Signal and Image Research Laboratory at the University of Illinois at Chicago, from 2003 to 2005, and in the Center for Sensor Signal and Information Processing at WUSTL, from 2006 to 2010. He has been a Postdoctoral Research Associate at the General Medical Sciences Division, Department of Internal Medicine, Washington University School of Medicine. Currently, he is a Research Data Scientist at Monsanto Company. His research interests are in statistical array signal processing and its applications to medical imaging and biosignal processing, and in biophysical modeling, including nonlinear waveform phenomena in excitable media, bioelectromagnetism, and optical mapping. Dr. La Rosa received the John Paul II Foundation scholarship between the years 1995 and 2000 for undergraduate studies in engineering sciences at PUC.

Peter Händel (S’88-M’94-SM’98) received a PhD degree from Uppsala University, Uppsala, Sweden, in 1993. From 1987 to 1993, he was with Uppsala University. From 1993 to 1997, he was with Ericsson AB, Kista, Sweden. From 1996 to 1997, he was a Visiting Scholar with the Tampere University of Technology, Tampere, Finland. Since 1997, he has been with the KTH Royal Institute of Technology, Stockholm, Sweden, where he is currently a Professor of Signal Processing and the Head of the Department of Signal Processing. From 2000 to 2006, he held an adjunct position at the Swedish Defence Research Agency. He has been a Guest Professor at the Indian Institute of Science (IISc), Bangalore, India, and at the University of Gävle, Sweden. He is a co-founder of Movelto AB. Dr. Händel has served as an associate editor for the IEEE Transactions on Signal Processing. He was a recipient of a Best Survey Paper Award by the IEEE Intelligent Transportation Systems Society in 2013.
Arye Nehorai (S’80-M’83-SM’90-F’94) is the Eugene and Martha Lohman Professor of Electrical Engineering in the Preston M. Green Department of Electrical and Systems Engineering (ESE) at Washington University in St. Louis (WUSTL). He served as chair of this department from 2006 to 2016. Under his department chair leadership, the undergraduate enrollment has more than tripled in four years and the master’s enrollment grew seven-fold in the same time period. He is also Professor in the Department of Biomedical Engineering (by courtesy), Professor in the Division of Biology and Biomedical Studies (DBBS), and Director of the Center for Sensor Signal and Information Processing at WUSTL. Prior to serving at WUSTL, he was a faculty member at Yale University and the University of Illinois at Chicago. He received the BSc and MSc degrees from the Technion, Israel and the PhD from Stanford University, California.

Dr. Nehorai served as Editor-in-Chief of the IEEE Transactions on Signal Processing from 2000 to 2002. From 2003 to 2005 he was the Vice President (Publications) of the IEEE Signal Processing Society (SPS), the Chair of the Publications Board, and a member of the Executive Committee of this Society. He was the founding editor of the special columns on Leadership Reflections in IEEE Signal Processing Magazine from 2003 to 2006.

Dr. Nehorai received the 2006 IEEE SPS Technical Achievement Award and the 2010 IEEE SPS Meritorious Service Award. He was elected Distinguished Lecturer of the IEEE SPS for a term lasting from 2004 to 2005. He received several best paper awards in IEEE journals and conferences. In 2001 he was named University Scholar of the University of Illinois. Dr. Nehorai was the Principal Investigator of the Multidisciplinary University Research Initiative (MURI) project titled Adaptive Waveform Diversity for Full Spectral Dominance from 2005 to 2010. He is a Fellow of the IEEE since 1994, Fellow of the Royal Statistical Society since 1996, and Fellow of AAAS since 2012.