The β-model — Maximum Likelihood, 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 study the maximum likelihood estimator 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 β-model for random graphs by replacing the constant model parameters with regression functions. Cramér-Rao bounds are derived for special cases of the undirected β-model, the directed β-model, and the covariate-based β-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 significance. Last, the applicability of the model is demonstrated using temporal social network data describing communication among healthcare workers.

Index Terms—The β-model, Cramér-Rao bounds, hypothesis testing, random graphs, 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 generalizing the already established β-model for random graphs. 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 β-model belongs to the class of exponential random graphs models (ERGMs), also known as p^* 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 β-model [15], [16][1]. The popularity of the β-model can be attributed to its flexibility in adjusting to observed degree sequences. Specifically, in a β-model (ERGM) where the model (canonical) parameters are obtained by using the method of maximum likelihood (ML), the degree of each node (the sufficient statistics) has an expectation value equal to the observed degree of the same node (the observed sufficient statistics) [16]. Since the ML estimates of the parameters in the β-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,

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: {jwahlst, 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).

---

[1]The p_1 model discussed in [9] is equivalent to the directed β-model but with the number of reciprocated edges as an added sufficient statistic.

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 two nodes is then constrained to only depend on the 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 special cases of the undirected, directed, and covariate-based β-models, and the performance of the ML estimator is illustrated by means of simulations. Further, examples are given on how to use the presented ML estimators to perform significance tests and tests of directionality. The applicability of the estimation framework is demonstrated in a case study where we use real-world data describing the interaction among ten healthcare workers. The resulting social network is regressed on categorical covariates representing the time of day and the day of the week.

II. ESTIMATION FRAMEWORK

To prepare for the covariate-based β-model and the results in later sections, we review the undirected and directed β-models in Section II-A and Section II-B, respectively. The covariate-based β-model is then introduced in Section II-C. The notation is defined independently in each subsection, and all graphs are assumed to be unweighted and without self-loops. Generalizations of the β-model to weighted random graphs have previously been considered in [25] and [26].

A. The Undirected β-model

Consider a graph with n nodes. In the undirected β-model, the probability of having an edge between the ith node and the jth node is

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

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), β_i signifies the tendency of the ith node to form edges with other nodes, i.e., its so called *differential attractiveness* [9]. Specifically, the ith node can be expected to have a large (small) number of ties if β_i is positive (negative) and of large magnitude.

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

$$Y_{ij} \sim \text{Bin}(N, p_{ij}) \tag{2}$$

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[2]. Examples of

[2]The model can easily be extended to the case where the number of measurements N depends on the considered nodes i and j [14], [27].

applications where one might be able to repeatedly sample a given network include studies of resting-state brain networks, as well as temporal social networks (see Section V).

The likelihood function of $\boldsymbol{\theta} \triangleq [\beta_1 \ \ldots \ \beta_n]^\mathsf{T}$, conditioned on N observations associated with each unordered pair of nodes (i,j), is proportional to [16]

$$L(\boldsymbol{\theta}|\mathbf{Y}) = \frac{e^{\sum_{i=1}^{n} N\beta_i \bar{d}_i}}{\prod_{i=1,j=i+1}^{n}(1+e^{\beta_i+\beta_j})^N}. \tag{3}$$

Here, $\mathbf{Y} \triangleq \{Y_{ij}\}_{i=1,j=i+1}^{n}$, while the degree of node i, averaged over all observations, has been defined as $\bar{d}_i \triangleq \sum_{j \neq i} Y_{ij}/N$. Since the Hessian of $L(\boldsymbol{\theta}|\mathbf{Y})$ with respect to $\boldsymbol{\theta}$ can easily be shown to be negative semidefinite, the ML estimates $\hat{\boldsymbol{\theta}} \triangleq [\hat{\beta}_1 \ \ldots \ \hat{\beta}_n]^\mathsf{T}$ must necessarily satisfy $\partial L(\hat{\boldsymbol{\theta}}|\mathbf{Y})/\partial \beta_i = 0$ for $i = 1,\ldots,n$. This is equivalent to saying that [16]

$$\bar{d}_i = \sum_{j \neq i} \frac{e^{\hat{\beta}_i+\hat{\beta}_j}}{1+e^{\hat{\beta}_i+\hat{\beta}_j}} \tag{4}$$

for $i = 1,\ldots,n$. The interpretation of (4) is that the observed average degree of each node must be equal to the expectation value of the corresponding average 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 Np_{ij}. To find the ML estimates, [16] introduced the function $\boldsymbol{\varphi} : \mathbb{R}^n \to \mathbb{R}^n$ (when $N = 1$), whose outputs are defined according to

$$\varphi_i(\mathbf{z}) \triangleq \log(\bar{d}_i) - \log \sum_{j \neq i} \frac{1}{e^{-z_j}+e^{z_i}} \tag{5}$$

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

$$\mathbf{z}^{(m+1)} = \boldsymbol{\varphi}(\mathbf{z}^{(m)}) \tag{6}$$

until convergence, starting from some initial value $\mathbf{z}^{(0)}$. Assuming that a unique ML solution exists, geometrically fast convergence to the fixed point was shown in [16] and [28]. Conditions for the asymptotic normality of the ML estimator when n tends to infinity were presented in [29]. Necessary and sufficient conditions for the existence of a finite ML estimate were presented in [14], [28], and [30]. As an example, it should be clear that there is no finite $\hat{\boldsymbol{\theta}}$ satisfying (4) whenever $\bar{d}_i = n-1$ or $\bar{d}_i = 0$ for some $i \in \{1,\ldots,n\}$.

B. The Directed β-model

In the directed β-model [9], the probability of having an edge directed from the ith node to the jth node is

$$p_{ij} = \frac{e^{\alpha_i+\beta_j}}{1+e^{\alpha_i+\beta_j}} \tag{7}$$

for any $i \neq j$. Assuming n nodes, this means that the $n(n-1)$ edge probabilities are parameterized by the $2n$ parameters $\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_n$. As should be obvious, α_i affects the probability of having an edge directed *from* the ith node, while β_i affects the probability of having an edge directed *to* the

ith node. The parameters α_i and β_i 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 α_i for $i = 1, \ldots, n$ and simultaneously subtracting the same constant from β_i for $i = 1, \ldots, n$), we will without loss of generality assume that β_n is known to be zero [31]. The directed β-model has been called the p_0 model [32], and is also closely related to the Rasch model [33], [34].

Now, let Y_{ij} denote the number of times that an edge directed from node i to node j is present in N independent measurements. It then follows that for any $i \neq j$

$$Y_{ij} \sim \mathrm{Bin}(N, p_{ij}). \tag{8}$$

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

$$L(\boldsymbol{\theta}|\mathbf{Y}) = \frac{e^{\sum_{i=1}^n N\alpha_i \bar{d}_i^\alpha + N\beta_i \bar{d}_i^\beta}}{\prod_{i \neq j}(1 + e^{\alpha_i + \beta_j})^N}. \tag{9}$$

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

$$\bar{d}_i^\alpha = \sum_{j \neq i} \frac{e^{\hat{\alpha}_i + \hat{\beta}_j}}{1 + e^{\hat{\alpha}_i + \hat{\beta}_j}}, \tag{10a}$$

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

$$\bar{d}_i^\beta = \sum_{j \neq i} \frac{e^{\hat{\alpha}_j + \hat{\beta}_i}}{1 + e^{\hat{\alpha}_j + \hat{\beta}_i}}, \tag{10b}$$

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

$$\varphi_i(\mathbf{z}) \triangleq \log(\bar{d}_i^\alpha) - \log \sum_{j \neq i} \frac{1}{e^{-z_{j+n}} + e^{z_i}}, \tag{11a}$$

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

$$\varphi_{i+n}(\mathbf{z}) \triangleq \log(\bar{d}_i^\beta) - \log \sum_{j \neq i} \frac{1}{e^{-z_j} + e^{z_{i+n}}}, \tag{11b}$$

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

$$\mathbf{z}^{(m+1)} = \boldsymbol{\varphi}(\mathbf{z}^{(m)}) \tag{12}$$

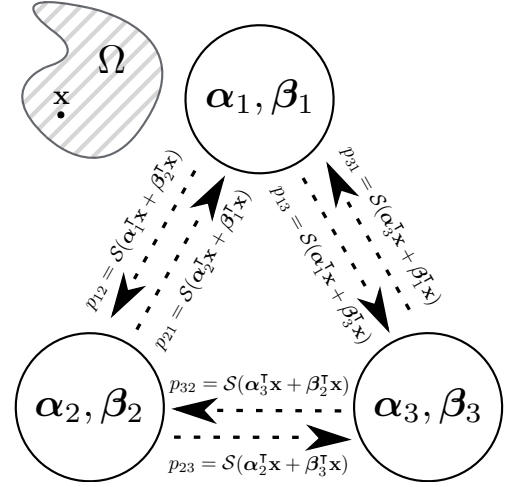until convergence, starting from some initial value $\mathbf{z}^{(0)}$. A

Fig. 1. Illustration of the covariate-based β-model with $n = 3$ nodes. Here, the circles and dashed lines represent nodes and dyads, respectively, with the probability of there being a link in between two nodes specified along the dyads. Ω is the sample space for the covariates $\mathbf{x}$.

similar iteration was proposed in [34]. Asymptotic normality of the ML estimates as n tends to infinity was established in [31].

C. The Covariate-based β-model

We will now demonstrate how to generalize the β-model for the purpose of including covariates. 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 α_i and β_j with the regression functions $\boldsymbol{\alpha}_i^\mathsf{T}\mathbf{x}$ and $\boldsymbol{\beta}_j^\mathsf{T}\mathbf{x}$, respectively. Hence, the probability of having an edge directed from the ith node to the jth node becomes

$$p_{ij}(\mathbf{x}) = \frac{e^{\boldsymbol{\alpha}_i^\mathsf{T}\mathbf{x} + \boldsymbol{\beta}_j^\mathsf{T}\mathbf{x}}}{1 + e^{\boldsymbol{\alpha}_i^\mathsf{T}\mathbf{x} + \boldsymbol{\beta}_j^\mathsf{T}\mathbf{x}}} \tag{13}$$

for any $i \neq j$. The probability depends on both the $2K$ regression coefficients $\boldsymbol{\alpha}_i \triangleq [\alpha_{i,1} \ \ldots \ \alpha_{i,K}]^\mathsf{T}$ and $\boldsymbol{\beta}_j \triangleq [\beta_{j,1} \ \ldots \ \beta_{j,K}]^\mathsf{T}$, and the K covariates $\mathbf{x} = [x_1 \ \ldots \ x_K]^\mathsf{T} \in \Omega$, representing e.g., time, space, or other variables describing the state of the network. Here, $\alpha_{i,k}$ and $\beta_{i,k}$ describe the effect that the kth covariate has on the tendency of the ith 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 ith node increases (decreases) as x_k increases. In analogy with the preceding subsection, we let $\boldsymbol{\beta}_n = \mathbf{0}_{K,1}$ where $\mathbf{0}_{\ell_1, \ell_2}$ is the zero matrix of dimension $\ell_1 \times \ell_2$. The model is illustrated in Fig. 1, where $\mathcal{S}(\beta) \triangleq 1/(1 + e^{-\beta})$ denotes the sigmoid function. Although this model bears some similarity to the model discussed in [35] and [36], the fundamental difference is that the model in [35] and [36] uses covariates that are unique to each edge, and associated model parameters that are constant over the graph.

Now, assume that we know the covariate values $\{\mathbf{x}_\ell\}_{\ell=1}^L$ and have made the associated observations $\mathbf{Y} \triangleq \{\{Y_{ij,\ell}\}_{i \neq j}\}_{\ell=1}^L$ 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_ℓ independent measurements from random graph ℓ. It then holds that for any $i \neq j$

$$Y_{ij,\ell} \sim \text{Bin}(N_\ell, p_{ij}(\mathbf{x}_\ell)). \tag{14}$$

Furthermore, it can be seen that the likelihood function of $\boldsymbol{\theta} \triangleq [\boldsymbol{\alpha}_1^\mathsf{T} \ \ldots \ \boldsymbol{\alpha}_n^\mathsf{T} \boldsymbol{\beta}_1^\mathsf{T} \ \ldots \ \boldsymbol{\beta}_{n-1}^\mathsf{T}]^\mathsf{T}$ is proportional to

$$L(\boldsymbol{\theta}|\mathbf{Y}) = \prod_{\ell=1}^{L} \frac{e^{\sum_{i=1}^{n} \boldsymbol{\alpha}_i^\mathsf{T} \mathbf{x}_\ell N_\ell \bar{d}_{i,\ell}^\alpha + \boldsymbol{\beta}_i^\mathsf{T} \mathbf{x}_\ell N_\ell \bar{d}_{i,\ell}^\beta}}{\prod_{i \neq j}(1 + e^{\boldsymbol{\alpha}_i^\mathsf{T} \mathbf{x}_\ell + \boldsymbol{\beta}_j^\mathsf{T} \mathbf{x}_\ell})^{N_\ell}} \tag{15}$$

where the average outdegree and indegree of node i in random graph ℓ have been defined as $\bar{d}_{i,\ell}^\alpha \triangleq \sum_{j \neq i} Y_{ij,\ell}/N_\ell$ and $\bar{d}_{i,\ell}^\beta \triangleq \sum_{j \neq i} Y_{ji,\ell}/N_\ell$, respectively. In analogy with the derivation of (10), the ML estimates $\hat{\boldsymbol{\theta}} \triangleq [\hat{\boldsymbol{\alpha}}_1^\mathsf{T} \ \ldots \ \hat{\boldsymbol{\alpha}}_n^\mathsf{T} \hat{\boldsymbol{\beta}}_1^\mathsf{T} \ \ldots \ \hat{\boldsymbol{\beta}}_{n-1}^\mathsf{T}]^\mathsf{T}$ must satisfy $\partial L(\hat{\boldsymbol{\theta}}|\mathbf{Y})/\partial \alpha_{i,k} = 0$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$, and $\partial L(\hat{\boldsymbol{\theta}}|\mathbf{Y})/\partial \beta_{i,k} = 0$ for $i = 1, \ldots, n-1$ and $k = 1, \ldots, K$. This further means that

$$\sum_{\ell=1}^{L} x_{\ell,k} N_\ell \bar{d}_{i,\ell}^\alpha = \sum_{\ell=1}^{L} \sum_{j \neq i} x_{\ell,k} N_\ell \frac{e^{\hat{\boldsymbol{\alpha}}_i^\mathsf{T} \mathbf{x}_\ell + \hat{\boldsymbol{\beta}}_j^\mathsf{T} \mathbf{x}_\ell}}{1 + e^{\hat{\boldsymbol{\alpha}}_i^\mathsf{T} \mathbf{x}_\ell + \hat{\boldsymbol{\beta}}_j^\mathsf{T} \mathbf{x}_\ell}}, \tag{16a}$$

for $i = 1, \ldots, n$ and $k = 1, \ldots, K$, where we have used that $\mathbf{x}_\ell \triangleq [x_{\ell,1} \ \ldots \ x_{\ell,K}]^\mathsf{T}$ and $\hat{\boldsymbol{\beta}}_n \triangleq \mathbf{0}_{K,1}$, while

$$\sum_{\ell=1}^{L} x_{\ell,k} N_\ell \bar{d}_{i,\ell}^\beta = \sum_{\ell=1}^{L} \sum_{j \neq i} x_{\ell,k} N_\ell \frac{e^{\hat{\boldsymbol{\alpha}}_j^\mathsf{T} \mathbf{x}_\ell + \hat{\boldsymbol{\beta}}_i^\mathsf{T} \mathbf{x}_\ell}}{1 + e^{\hat{\boldsymbol{\alpha}}_j^\mathsf{T} \mathbf{x}_\ell + \hat{\boldsymbol{\beta}}_i^\mathsf{T} \mathbf{x}_\ell}}, \tag{16b}$$

for $i = 1, \ldots, n-1$ and $k = 1, \ldots, K$.

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

$$\psi_{K(i-1)+k}(\mathbf{z}, \boldsymbol{\gamma})$$
$$\triangleq \sum_{\ell=1}^{L} x_{\ell,k} N_\ell \left(\bar{d}_{i,\ell}^\alpha - \sum_{j \neq i} \frac{e^{\mathbf{z}_i^\mathsf{T} \mathbf{x}_\ell + x_{\ell,k}(\gamma_{i,k} - z_{i,k})}}{e^{-\mathbf{z}_{j+n}^\mathsf{T} \mathbf{x}_\ell} + e^{\mathbf{z}_i^\mathsf{T} \mathbf{x}_\ell}} \right), \tag{17a}$$

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

$$\psi_{K(i+n-1)+k}(\mathbf{z}, \boldsymbol{\gamma}) \tag{17b}$$
$$\triangleq \sum_{\ell=1}^{L} x_{\ell,k} N_\ell \left(\bar{d}_{i,\ell}^\beta - \sum_{j \neq i} \frac{e^{\mathbf{z}_{i+n}^\mathsf{T} \mathbf{x}_\ell + x_{\ell,k}(\gamma_{i+n,k} - z_{i+n,k})}}{e^{-\mathbf{z}_j^\mathsf{T} \mathbf{x}_\ell} + e^{\mathbf{z}_{i+n}^\mathsf{T} \mathbf{x}_\ell}} \right),$$

for $i = 1, \ldots, n-1$ and $k = 1, \ldots, K$. Here, we have used that $\mathbf{z} \triangleq [\mathbf{z}_1^\mathsf{T} \ \ldots \ \mathbf{z}_{2n-1}^\mathsf{T}]^\mathsf{T}$, $\mathbf{z}_{2n} \triangleq \mathbf{0}_{K,1}$, and $\boldsymbol{\gamma} \triangleq [\boldsymbol{\gamma}_1^\mathsf{T} \ \ldots \ \boldsymbol{\gamma}_{2n-1}^\mathsf{T}]^\mathsf{T}$, while $\mathbf{z}_i \triangleq [z_{i,1} \ \ldots \ z_{i,K}]^\mathsf{T}$ and $\boldsymbol{\gamma}_i \triangleq [\gamma_{i,1} \ \ldots \ \gamma_{i,K}]^\mathsf{T}$ for $i = 1, \ldots, 2n-1$. Finally, we define $\boldsymbol{\varphi} : \mathbb{R}^{(2n-1)K} \to \mathbb{R}^{(2n-1)K}$ as

$$\boldsymbol{\varphi}(\mathbf{z}) \triangleq \{\boldsymbol{\gamma} : \ \boldsymbol{\psi}(\mathbf{z}, \boldsymbol{\gamma}) = \mathbf{0}_{(2n-1)K,1}\}. \tag{18}$$

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

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

$$\mathbf{z}^{(m+1)} = \boldsymbol{\varphi}(\mathbf{z}^{(m)}) \tag{19}$$

until convergence, starting from some initial value $\mathbf{z}^{(0)}$. In general, there is no closed-form expression for the elements in $\boldsymbol{\varphi}(\mathbf{z}^{(m)})$. The exceptions are the elements providing updates

of eventual intercepts $\alpha_{i,1}$ and $\beta_{i,1}$, i.e., when $x_{\ell,1} = x_1$ 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 $\boldsymbol{\psi}(\mathbf{z}, \boldsymbol{\gamma})$ only depends on one matching element in $\boldsymbol{\gamma}$. Therefore, it is easy to numerically find the corresponding elements of $\boldsymbol{\varphi}(\mathbf{z})$ by applying some one-dimensional root-finding algorithm.

D. Implementation

The initial parameter estimate $\mathbf{z}^{(0)}$ was defined as the zero vector in all three models. The estimates were then iteratively updated as $\mathbf{z}^{(m+1)} = \boldsymbol{\varphi}(\mathbf{z}^{(m)})$ until $\|\mathbf{z}^{(m+1)} - \mathbf{z}^{(m)}\| < \epsilon$, 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 [28].

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 [37]

$$\text{Cov}(\hat{\boldsymbol{\theta}}) \succeq \mathbf{P} \tag{20}$$

where $\mathbf{P}$ is the inverse of the Fisher information matrix (FIM), and we have used $\mathbf{A} \succeq \mathbf{B}$ to denote that $\mathbf{A} - \mathbf{B}$ is positive semidefinite. The FIM is defined as $\mathcal{I} \triangleq \mathbb{E}[\mathbf{s}\mathbf{s}^\mathsf{T}]$ where $\mathbf{s}$ is the gradient of the log-likelihood function (the score function) with respect to $\boldsymbol{\theta}$. In this section, we will derive FIMs for the three models discussed in Section II, CRBs for special cases of these models, and use simulations to compare the performance of the ML estimators to the CRBs. Each subsection uses the same notation as the corresponding subsection in Section II. Similarly, the notation for FIMs and inverse FIMs is defined independently in each subsection. Although the CRBs are only derived for special cases of the considered β-models, our hope is that the derived bounds will provide practitioners with some intuition for how the accuracy of the ML estimates varies with the studied parameters also in the general cases.

A. The Undirected β-model

In the undirected β-model, the element in the ith row and jth column of the FIM takes the form

$$\mathcal{I}_{ij} = \begin{cases} \sum_{q \neq i} N p_{iq}(1 - p_{iq}), & i = j \\ N p_{ij}(1 - p_{ij}), & i \neq j. \end{cases} \tag{21}$$
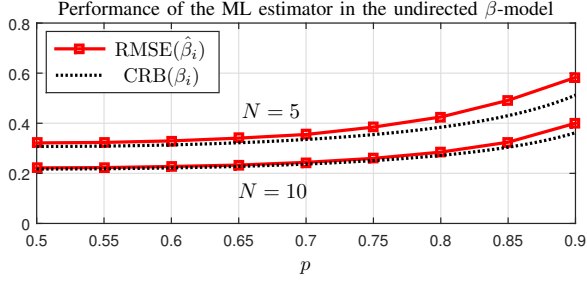
Fig. 2. Estimation performance with $n = 10$ as dependent on the edge probability p.

Expressed in words, the Fisher information for β_i is the sum of variances of the associated edge observations $Y_{i1}, \ldots, Y_{ii-1}, Y_{ii+1}, \ldots, Y_{in}$. Similarly, the Fisher information shared between β_i and β_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 $n > 2$, $\beta_i = \beta$ for $i = 1, \ldots, n$, and hence also $p_{ij} = p \triangleq e^{2\beta}/(1 + e^{2\beta})$ for all $i \neq j$. As shown in Appendix A, the inverse FIM then becomes

$$\mathbf{P} = \frac{1}{Np(1-p)} \frac{1}{(n-2)} \left(\mathbf{I}_n - \frac{1}{2(n-1)} \mathbb{1}_n \right). \quad (22)$$

Here, $\mathbf{I}_n$ denotes the identity matrix of dimension n, while $\mathbb{1}_n$ 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

$$\begin{aligned} \mathrm{Var}(\hat{\beta}_i) &\geq \frac{1}{Np(1-p)} \frac{2n-3}{2(n-1)(n-2)} \\ &\simeq \frac{1}{Np(1-p)} \frac{1}{n} \end{aligned} \quad (23)$$

where $\simeq$ 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 $p \to 0$ ($\beta \to -\infty$).

Now, continuing with the same the parameter restrictions, Fig. 2 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[3] with $n = 10$, and $\mathrm{CRB}(\beta_i)$ 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. 2, 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.

It should be stressed 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

[3]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.

Erdős-Rényi model [38]. In this case, the CRB for β_{ER}, where the edge probability is $p_{\mathrm{ER}} \triangleq e^{2\beta_{\mathrm{ER}}}/(1 + e^{2\beta_{\mathrm{ER}}})$, becomes

$$\mathrm{Var}(\hat{\beta}_{\mathrm{ER}}) \geq \frac{1}{Np(1-p)} \frac{1}{2n(n-1)}. \quad (24)$$

For a derivation of (24), see Appendix A. Since we here only have to estimate one parameter using $Nn(n-1)/2$ measurements, it should come as no surprise that the bound is, in the limit of $n \to \infty$, inversely proportional to the square of the number of nodes n (rather than inversely proportional to n as in the undirected β-model).

B. The Directed β-model

As shown in Appendix B, the FIM for the directed β-model is

$$\boldsymbol{\mathcal{I}} = \begin{bmatrix} \boldsymbol{\mathcal{I}}_\alpha & \boldsymbol{\mathcal{I}}_{\alpha,\beta} \\ \boldsymbol{\mathcal{I}}_{\alpha,\beta}^\intercal & \boldsymbol{\mathcal{I}}_\beta \end{bmatrix} \quad (25)$$

where $\boldsymbol{\mathcal{I}}_\alpha$ is the diagonal matrix of dimension n whose ith diagonal element is $\sum_{j=1}^n Np_{ij}(1 - p_{ij})$, $\boldsymbol{\mathcal{I}}_{\alpha,\beta}$ is the matrix of dimension $n \times n - 1$ whose element in the ith row and jth column is $Np_{ij}(1-p_{ij})$, and $\boldsymbol{\mathcal{I}}_\beta$ is the diagonal matrix of dimension $n-1$ whose ith diagonal element is $\sum_{j=1}^n Np_{ji}(1-p_{ji})$.

Now, consider the special case when $n > 2$, $\alpha_i = \alpha$ 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 $p_{ij} = p \triangleq e^\alpha/(1 + e^\alpha)$ for all $i \neq j$. The inverse FIM then becomes

$$\mathbf{P} = \frac{1}{Np(1-p)} \begin{bmatrix} \mathbf{P}_\alpha & \mathbf{P}_{\alpha,\beta} \\ \mathbf{P}_{\alpha,\beta}^\intercal & \mathbf{P}_\beta \end{bmatrix} \quad (26)$$

where

$$\mathbf{P}_\alpha \triangleq \begin{bmatrix} \mathbf{P}_{\alpha,1} & \mathbf{P}_{\alpha,2} \\ \mathbf{P}_{\alpha,2}^\intercal & \mathbf{P}_{\alpha,3} \end{bmatrix} \quad (27)$$

with

$$\mathbf{P}_{\alpha,1} \triangleq \frac{n-1}{n(n-2)} \left(\mathbf{I}_{n-1} + \frac{n^2 - 3n + 1}{(n-1)^2} \mathbb{1}_{n-1} \right), \quad (28a)$$

$$\mathbf{P}_{\alpha,2} \triangleq \frac{1}{n-1} \mathbb{1}_{n-1,1}, \quad (28b)$$

$$\mathbf{P}_{\alpha,3} \triangleq \frac{2n-3}{(n-1)(n-2)}, \quad (28c)$$

and

$$\mathbf{P}_{\alpha,\beta} \triangleq -\frac{1}{n(n-2)} \begin{bmatrix} (n-1)\mathbb{1}_{n-1} - \mathbf{I}_{n-1} \\ n\mathbb{1}_{1,n-1} \end{bmatrix}, \quad (29a)$$

$$\mathbf{P}_\beta \triangleq \frac{n-1}{n(n-2)} (\mathbf{I}_{n-1} + \mathbb{1}_{n-1}). \quad (29b)$$

Here, we have used $\mathbb{1}_{n,m}$ 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

$$\mathrm{Var}(\hat{\alpha}_i) \geq \frac{1}{Np(1-p)} \frac{2n-1}{n(n-1)}, \quad (30a)$$

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

$$\mathrm{Var}(\hat{\alpha}_n) \geq \frac{1}{Np(1-p)} \frac{2n-3}{(n-1)(n-2)}, \quad (30b)$$
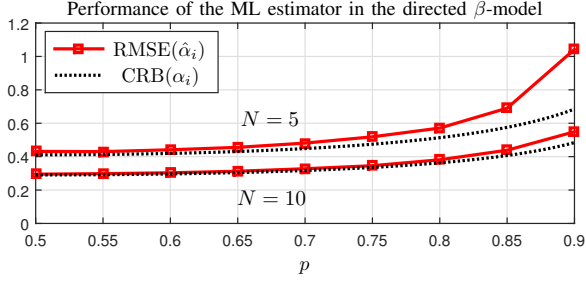
Fig. 3. Estimation performance with $n = 10$ and $i \neq n$ as dependent on the edge probability p.

Fig. 4. Estimation performance with $n = 5$, $L = 2$, $\mathbf{x}_1 = [1 \ 0]^\mathsf{T}$, $\mathbf{x}_2 = [1 \ 1]^\mathsf{T}$, $\boldsymbol{\alpha} = \alpha[1 \ 1]^\mathsf{T}$, $\boldsymbol{\beta} = \mathbf{0}_{2,1}$, and $i \neq n$, as dependent on the average edge probability $\bar{p}$.

and

$$\mathrm{Var}(\hat{\beta}_i) \geq \frac{1}{Np(1-p)} \frac{2(n-1)}{n(n-2)}, \tag{30c}$$

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 (30) 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 β-model as in the undirected β-model (assuming the same N, n, and p).

Given the considered parameter restrictions, Fig. 3 shows how the ML estimator of α_i with $i \neq n$ compares to the CRB in (30a) 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. 2.

C. The Covariate-based β-model

For the covariate-based β-model, we consider the special case when $n > 2$, $\boldsymbol{\alpha}_i = \boldsymbol{\alpha}$ for $i = 1, \ldots, n$, and $\boldsymbol{\beta}_i = \boldsymbol{\beta}$ for $i = 1, \ldots, n-1$. Without loss of generality, we assume that $\boldsymbol{\beta} = \mathbf{0}_{K,1}$. Further, this also means that $p_{ij}(\mathbf{x}) = p(\mathbf{x}) \triangleq e^{\boldsymbol{\alpha}^\mathsf{T}\mathbf{x}}/(1 + e^{\boldsymbol{\alpha}^\mathsf{T}\mathbf{x}})$ for all $i \neq j$. Using $\otimes$ to denote the Kronecker product, the inverse FIM becomes

$$\mathbf{P} = \frac{1}{N} \begin{bmatrix} \mathbf{P}_\alpha & \mathbf{P}_{\alpha,\beta} \\ \mathbf{P}_{\alpha,\beta}^\mathsf{T} & \mathbf{P}_\beta \end{bmatrix} \otimes \boldsymbol{\mathcal{I}}_x^{-1} \tag{31}$$

where $\boldsymbol{\mathcal{I}}_x \triangleq \sum_{\ell=1}^L p(\mathbf{x}_\ell)(1 - p(\mathbf{x}_\ell))\mathbf{x}_\ell \mathbf{x}_\ell^\mathsf{T}$, while $\mathbf{P}_\alpha$, $\mathbf{P}_{\alpha,\beta}$, and $\mathbf{P}_\beta$ are as defined in equations (27), (28), and (29). 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 β-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 model presented in Section III-A and the CRB for the undirected covariate-based 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 $\boldsymbol{\beta}_n = \mathbf{0}_{K,1}$.) The CRB for the different regression coefficients associated with a given node are then weighted based on $\boldsymbol{\mathcal{I}}_x^{-1}$. Since the rank of $\boldsymbol{\mathcal{I}}_x$ is $\mathrm{rank}(\boldsymbol{\mathcal{I}}_x) \leq \sum_{\ell=1}^L \mathrm{rank}\big(p(\mathbf{x}_\ell)(1 - p(\mathbf{x}_\ell))\mathbf{x}_\ell \mathbf{x}_\ell^\mathsf{T}\big) \leq L$, we must have $L \geq K$ for $\boldsymbol{\mathcal{I}}_x$ to be invertible (this requirement

could have been expected since the dimension of the observations $\{\{\bar{d}_{i,\ell}^\alpha, \bar{d}_{i,\ell}^\beta\}_{i=1}^n\}_{\ell=1}^L$ is $2nL$ and the dimension of $\boldsymbol{\theta}$ 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 $n = 5$, $L = 2$, $\mathbf{x}_1 = [1 \ 0]^\mathsf{T}$, $\mathbf{x}_2 = [1 \ 1]^\mathsf{T}$, and $\boldsymbol{\alpha} = \alpha[1 \ 1]^\mathsf{T}$. Fig. 4 then shows how the ML estimator of α_{i1} with $i \neq n$ compares to the CRB as α varies. Here, the average edge probability has been defined as $\bar{p} = (p(\mathbf{x}_1) + p(\mathbf{x}_2))/2$. 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 special cases of all three studied β-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 (the ML estimator approaches the CRB as $N \to \infty$ [37]).

IV. HYPOTHESIS TESTING

The problem of choosing between two simple 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 threshold. As stated in the Neyman-Pearson lemma, the LRT is the most powerful test for a given significance level [39]. 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) [40]. If the two hypotheses can be expressed as $\mathcal{H}_0 : \boldsymbol{\theta} \in \Omega_0$ and $\mathcal{H}_1 : \boldsymbol{\theta} \in \Omega_1$, the generalized likelihood ratio becomes

$$\Lambda(\mathbf{Y}) = \frac{\sup_{\boldsymbol{\theta} \in \Omega_0} L(\boldsymbol{\theta}|\mathbf{Y})}{\sup_{\boldsymbol{\theta} \in \Omega_1} L(\boldsymbol{\theta}|\mathbf{Y})} \tag{32}$$

and we decide on $\mathcal{H}_0$ whenever $\Lambda(\mathbf{Y}) > \eta$ for some chosen threshold η.

The threshold η is often chosen so that the decision boundary becomes the set of measurements giving some predetermined p-value. However, to be able to compute p-values, we need to know the distribution of $\Lambda(\mathbf{Y})$ under $\mathcal{H}_0$. While the exact distribution is seldom known, approximate p-values can often be obtained from Wilks' theorem. Specifically, assume that Ω_0 is a subset of Ω_1 and that Ω_1 has κ more free parameters than Ω_0. Then, given certain mild regularity conditions[4], Wilks' theorem states that the asymptotic distribution of $-2\log\Lambda(\mathbf{Y})$ under $\mathcal{H}_0$, is the χ^2-distribution with κ degrees of freedom [41]. The asymptotic considered in this article is when n is held fixed while N tends to infinity. The opposite case, i.e., when $N = 1$ while n tends to infinity, considered in e.g., [29] and [42], is more delicate and will not be pursued here.

Next, we demonstrate how to use the ML estimates presented in Section II to perform GLRTs 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 [43], the undirected β-model [27], and the directed β-model [9], [44].

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 covariate-based model from Section II-C, assuming knowledge of $\{\mathbf{x}_\ell\}_{\ell=1}^L$ and $\mathbf{Y} = \{\{Y_{ij,\ell}\}_{i\neq j}\}_{\ell=1}^L$. Now, to test the significance of the kth covariate x_k, we formulate the null hypothesis

$$\mathcal{H}_0 : \alpha_{i,k} = \beta_{i,k} = 0 \text{ for } i = 1, \ldots, n. \tag{33}$$

Assuming that the alternative hypothesis $\mathcal{H}_1$ does not impose any constraints on the parameters $\boldsymbol{\theta}$, it follows that the maximizing parameter in the denominator of (32) is the ML estimate for the covariate-based β-model. Similarly, the maximizing parameter in the numerator of (32) is the ML estimate for the covariate-based β-model where the employed regression coefficients can be written as $\boldsymbol{\alpha}_i \triangleq [\alpha_{i,1} \ldots \alpha_{i,k-1} \quad \alpha_{i,k+1} \ldots \alpha_{i,K}]^\mathsf{T}$ and $\boldsymbol{\beta}_i \triangleq [\beta_{i,1} \ldots \beta_{i,k-1} \quad \beta_{i,k+1} \ldots \beta_{i,K}]^\mathsf{T}$ for $i = 1, \ldots, n$, while the covariates are $\mathbf{x}_\ell \triangleq [x_{\ell,1} \ldots x_{\ell,k-1} \; x_{\ell,k+1} \ldots x_{\ell,K}]^\mathsf{T}$ for $\ell = 1, \ldots, L$.

B. Testing for Directionality

Second, we consider the problem of deciding whether some given set of directed network data $\{Y_{ij}\}_{i\neq j}$ has originated from a directed or undirected (symmetric) random graph[5]. 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 (the undirected β-model), as

$$\mathcal{H}_0 : \alpha_i = \beta_i \text{ for } i = 1, \ldots, n - 1. \tag{34}$$

---

[4]The regularity conditions relate to 1) the existence of the first and second-order derivatives of the log-likelihood function; and 2) that the expectation of the score function is the zero vector [37] and [41]. It can easily be verified that both of these conditions hold for the models discussed in Section II.

[5]A nonparametric test for this problem was developed in [45].

Assuming that the alternative hypothesis $\mathcal{H}_1$ does not impose any constraints on the parameters $\boldsymbol{\theta}$, it follows that the maximizing parameter in the denominator of (32) is the ML estimate for the directed β-model. Similarly, the maximizing parameter in the numerator of (32) is the ML estimate for the undirected β-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 $2N$ independent measurements. To extend the test to the covariate-based β-model considered in Section II-C, one would simply test the hypothesis that $\boldsymbol{\alpha}_i = \boldsymbol{\beta}_i$ for $i = 1, \ldots, n - 1$.

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 $\mathcal{H}_0$) we independently simulated α_i for $i = 1, \ldots, n$ and β_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 $\Lambda(\mathbf{Y})$ and, for any given η, make the decision to reject or accept $\mathcal{H}_0$. Finally, the true positive rate was obtained as the percentage of rejections. To compute the false positive rate (the probability of incorrectly rejecting $\mathcal{H}_0$), we simulated α_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 $\Lambda(\mathbf{Y})$, the false positive rate was obtained as the percentage of rejections of $\mathcal{H}_0$. Fig. 5 displays the ROC curve for $\rho = 0.1, 0.2, 0.3, 0.4$, with $N = 10$, $n = 10$, 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 $\{\beta_1, \ldots, \beta_n\}$ less alike under the alternative hypothesis $\mathcal{H}_1$.

Fig. 6 displays the histogram of $-2\log\Lambda(\mathbf{Y})$ under $\mathcal{H}_0$ when testing for directionality with $N = 10$, $n = 10$, and $\rho = 0.4$, using 10^4 simulations. The histogram is overlaid with the probability density function (pdf) of the χ^2-distribution with $n-1$ degrees of freedom (note that $\mathcal{H}_0$ is specified by n parameters while $\mathcal{H}_1$ is specified by $2n - 1$ parameters). As can be seen from Fig. 6, the test statistic seems to follow the χ^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. To numerically test whether $-2\log\Lambda(\mathbf{Y})$ follows a χ^2-distribution, the Kolmogorov-Smirnov test was separately applied to 10 disjoint sets, each with 10^3 samples of $-2\log\Lambda(\mathbf{Y})$ [46]. Using $\chi^2(n-1)$ as the null distribution, we obtained the (ordered) p-values 0.0267, 0.0858, 0.1225, 0.3481 0.3971, 0.4158, 0.4645, 0.5319, 0.8278, and 0.8649. Evidently, only one of the ten sets would have been rejected at a significance level of 0.05.
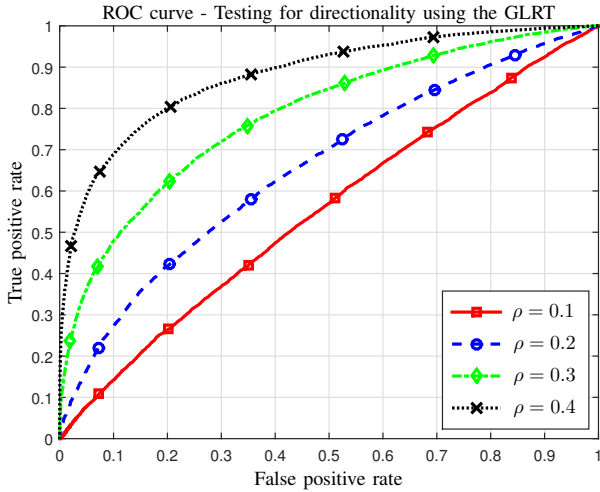
Fig. 5. ROC curve when testing for directionality in the directed β-model with $N_{ij} = 10$ for all $i \neq j$ and $n = 10$.
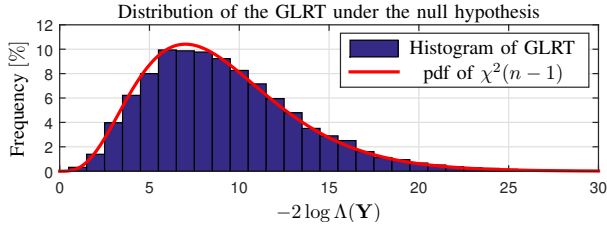

Fig. 6. The distribution of the GLRT under the null hypothesis compared to the χ^2-distribution when testing for symmetry with $N_{ij} = 10$ for all $i \neq j$, $n = 10$, and $\rho = 0.4$.

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 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 [47]. 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. 7 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
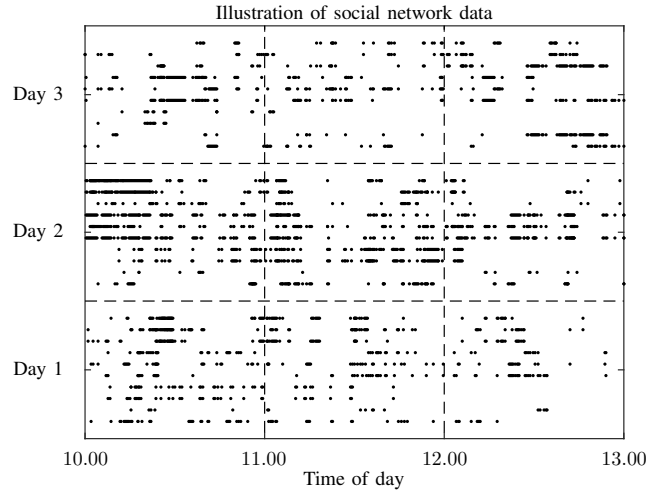

Fig. 7. Illustration of the studied social network data. Each dot represents a contact made between two individuals.

half of the observations indicate that a link is present (here, we had $\sum_{i<j,\ell} Y_{ij,\ell} / \sum_{i<j,\ell} N_\ell = 0.484$).

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 covariate-based 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 covariate-based β-model with $\mathbf{x}_1 = [1\ 1]^\mathsf{T}$ and $\mathbf{x}_2 = [1\ 0]^\mathsf{T}$. The contacts displayed in the three leftmost boxes in Fig. 7 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 $N_\ell = 3\ell$ for $\ell = 1, 2$. The ML estimates under the null hypothesis are obtained by using the undirected β-model and merging all graph data so that $N = 9$. 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 $-2\log\Lambda(\mathbf{Y})$ is χ^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-

TABLE I
p-VALUES FOR SOCIAL NETWORK DATA.

	Time of day		
	$10.00-11.00$	$11.00-12.00$	$12.00-13.00$
Simulated *p*-value	0.0051	0.2950	0.0001
Theoretical *p*-value	0.0037	0.2557	$5.7 \cdot 10^{-6}$

TABLE II
p-VALUES FOR SOCIAL NETWORK DATA.

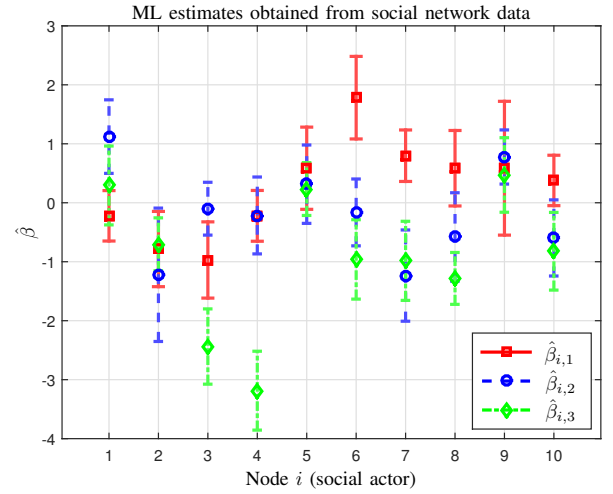|	Day		
	1	2	3
Simulated *p*-value	0.9361	0.1121	0.3125
Theoretical *p*-value	0.9238	0.0943	0.2693

Fig. 8. ML estimates obtained when regressing social network data on time of day. The vertical lines represent uncertainty intervals of $\pm$ one standard deviation as estimated from the CRB.

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 measurements of the first, second, and third graph, respectively. The associated covariates where $\mathbf{x}_1 = [1 \ 0 \ 0]^\mathsf{T}$, $\mathbf{x}_2 = [1 \ 1 \ 0]^\mathsf{T}$, and $\mathbf{x}_3 = [1 \ 0 \ 1]^\mathsf{T}$. The resulting estimates are illustrated in Fig. 8, where the vertical lines represent uncertainty intervals $[\hat{\beta}_{i,k} - \mathrm{CRB}(\beta_{i,k}), \hat{\beta}_{i,k} + \mathrm{CRB}(\beta_{i,k})]$, and $\mathrm{CRB}(\beta_{i,k})$ denotes the lower bound on the standard deviation of $\hat{\beta}_{i,k}$ as estimated from the undirected version of the FIM presented in (57) and (58). 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. 7 (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, the β-model was generalized to enable studies of how a random graph depends on a given set of covariates. Further, ML estimates and CRBs were derived for special cases of the undirected, directed, and covariate-based β-models. In the limit where the number of nodes approaches infinity, the CRB on the variance of the β-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 covariate-based β-model was applied to social network data describing contacts among healthcare workers was worked out

in detail. In summary, the article presented tractable methods for point estimation, model design, and hypothesis testing for networks that depend on external covariates. Future research may focus on deriving estimation bounds or approximate estimation bounds for the general β-models.

APPENDIX A

This appendix derives the FIM in (21), given for the undirected β-model. Further, we derive the inverse FIM in the special case considered in (22).

Let us denote the score function as $\mathbf{s} \triangleq [s_1 \ \dots \ s_n]^\mathsf{T}$. As follows directly from the likelihood function in (3), it holds that

$$s_i = \sum_{j \neq i}(Y_{ij} - Np_{ij}). \tag{35}$$

Further, using (2) and that the measurements from different edges are independent yields

$$\mathbb{E}[(s_i)^2] = \sum_{j \neq i} Np_{ij}(1 - p_{ij}) \tag{36a}$$

and

$$\begin{aligned}
\mathbb{E}[s_i s_j] &= \mathbb{E}[(\textstyle\sum_{q \neq i} Y_{iq} - Np_{iq})(\textstyle\sum_{r \neq j} Y_{rj} - Np_{rj})] \\
&= \mathbb{E}[(Y_{ij} - Np_{ij} + \textstyle\sum_{q \neq i, q \neq j} Y_{iq} - Np_{iq}) \\
&\quad \cdot (Y_{ij} - Np_{ij} + \textstyle\sum_{r \neq j, r \neq i} Y_{rj} - Np_{rj})] \\
&= Np_{ij}(1 - p_{ij}) \tag{36b}
\end{aligned}$$

for any $i \neq j$. The FIM in (21) can now be obtained from (36).

Under the assumptions preceding (22), it follows that

$$\boldsymbol{\mathcal{I}} = Np(1 - p)\big((n - 2)\mathbf{I}_n + \mathbb{1}_n\big). \tag{37}$$

Moreover, the Sherman-Morrison formula gives that [48]

$$(a\mathbf{I}_n + b\mathbb{1}_n)^{-1} = \frac{1}{a}\mathbf{I}_n - \frac{b}{a}\frac{1}{a + bn}\mathbb{1}_n \tag{38}$$

for any constants $a \neq 0$ and $b \neq -a/n$. Now, applying the special case of the Sherman-Morrison formula in (38) to the FIM in (37) while assuming $n > 2$ immediately gives the inverse FIM in (22).

To derive the CRB in (24), we apply the chain rule $\partial \log L(\boldsymbol{\theta})/\partial \beta_{\mathrm{ER}} = (\partial \log L(\boldsymbol{\theta})/\partial \boldsymbol{\theta}) \cdot \partial \boldsymbol{\theta}/\partial \beta_{\mathrm{ER}}$. Since the Erdős-Rényi model is equivalent to the undirected β-model with $\beta_{\mathrm{ER}} = \beta_1 = \cdots = \beta_n$ (and where this is known a priori), it holds that $\partial \boldsymbol{\theta}/\partial \beta_{\mathrm{ER}} = \mathbb{1}_{n,1}$, and hence, the Fisher information for β_{ER} is

$$
\begin{aligned}
\mathcal{I}_{\mathrm{ER}} &= \mathbb{1}_{1,n} \boldsymbol{\mathcal{I}} \mathbb{1}_{n,1} \\
&= 2Np(1-p)(n-1)\mathbb{1}_{1,n}\mathbb{1}_{n,1} \\
&= 2Np(1-p)n(n-1).
\end{aligned}
\tag{39}
$$

The same result can be derived by noting that the likelihood function for the Erdős-Rényi model is

$$
L(\boldsymbol{\theta}|\mathbf{Y}) = p_{\mathrm{ER}}^{\sum_{i=1}^{n}\sum_{j=i+1}^{n}Y_{ij}}(1-p_{\mathrm{ER}})^{\sum_{i=1}^{n}\sum_{j=i+1}^{n}(N-Y_{ij})}.
\tag{40}
$$

APPENDIX B

Here, we derive the FIM in (25), given for the directed β-model. In addition, we also derive the inverse FIM in the special case considered in (26).

Let us denote the score function as $\mathbf{s} \triangleq [(\mathbf{s}^\alpha)^\mathsf{T}\ (\mathbf{s}^\beta)^\mathsf{T}]^\mathsf{T}$, where $\mathbf{s}^\alpha \triangleq [s_1^\alpha\ \ldots\ s_n^\alpha]^\mathsf{T}$ and $\mathbf{s}^\beta \triangleq [s_1^\beta\ \ldots\ s_{n-1}^\beta]^\mathsf{T}$. It then follows from (9) that

$$
s_i^\alpha = \sum_{j\neq i}(Y_{ij} - Np_{ij}),
\tag{41a}
$$

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

$$
s_i^\beta = \sum_{j\neq i}(Y_{ji} - Np_{ji}).
\tag{41b}
$$

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

$$
\mathbb{E}[(s_i^\alpha)^2] = \sum_{j\neq i}Np_{ij}(1-p_{ij}),
\tag{42a}
$$

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

$$
\mathbb{E}[(s_i^\beta)^2] = \sum_{j\neq i}Np_{ji}(1-p_{ji}),
\tag{42b}
$$

$$
\mathbb{E}[s_i^\alpha s_i^\beta] = 0,
\tag{42c}
$$

for $i = 1, \ldots, n-1$, while

$$
\mathbb{E}[s_i^\alpha s_j^\alpha] = 0,
\tag{43a}
$$

$$
\mathbb{E}[s_i^\beta s_j^\beta] = 0,
\tag{43b}
$$

and

$$
\mathbb{E}[s_i^\alpha s_j^\beta] = Np_{ij}(1-p_{ij}),
\tag{43c}
$$

for any $i \neq j$. The FIM in (25) now follows directly from (42) and (43). Approximations of the inverse FIM has previously been discussed in [36].

Given the assumptions preceding (26), it follows that the submatrices in (25) are

$$
\mathcal{I}_\alpha = Np(1-p)(n-1)\,\mathbf{I}_n,
\tag{44a}
$$

$$
\mathcal{I}_\beta = Np(1-p)(n-1)\,\mathbf{I}_{n-1},
\tag{44b}
$$

$$
\mathcal{I}_{\alpha,\beta} = Np(1-p)[\mathbb{1}_{n-1} - \mathbf{I}_{n-1}\ \ \mathbb{1}_{n-1,1}]^\mathsf{T}.
\tag{44c}
$$

Further, applying the formula for block matrix inversion [49] and assuming $n > 2$, we have that

$$
\begin{aligned}
\mathbf{P}_\alpha &= Np(1-p)(\mathcal{I}_\alpha - \mathcal{I}_{\alpha,\beta}\mathcal{I}_\beta^{-1}\mathcal{I}_{\alpha,\beta}^\mathsf{T})^{-1} \\
&= Np(1-p)\left(\begin{bmatrix}\mathcal{I}_{\alpha,1} & \mathcal{I}_{\alpha,2}\\ \mathcal{I}_{\alpha,2}^\mathsf{T} & \mathcal{I}_{\alpha,3}\end{bmatrix}\right)^{-1}
\end{aligned}
\tag{45}
$$

where

$$
\mathcal{I}_{\alpha,1} \triangleq c\left(n(n-2)\mathbf{I}_{n-1} - (n-3)\mathbb{1}_{n-1}\right),
\tag{46a}
$$

$$
\mathcal{I}_{\alpha,2} \triangleq -c(n-2)\mathbb{1}_{n-1,1},
\tag{46b}
$$

$$
\mathcal{I}_{\alpha,3} \triangleq c(n-1)(n-2),
\tag{46c}
$$

with $c \triangleq Np(1-p)/(n-1)$. Here, we have used that

$$
\begin{aligned}
&\mathcal{I}_{\alpha,\beta}\mathcal{I}_\beta^{-1}\mathcal{I}_{\alpha,\beta}^\mathsf{T} \\
&= c\begin{bmatrix}\mathbb{1}_{n-1} - \mathbf{I}_{n-1}\\ \mathbb{1}_{1,n-1}\end{bmatrix}\begin{bmatrix}\mathbb{1}_{n-1} - \mathbf{I}_{n-1}\\ \mathbb{1}_{1,n-1}\end{bmatrix}^\mathsf{T} \\
&= c\begin{bmatrix}(\mathbb{1}_{n-1} - \mathbf{I}_{n-1})^2 & (\mathbb{1}_{n-1} - \mathbf{I}_{n-1})\mathbb{1}_{n-1,1}\\ \mathbb{1}_{1,n-1}(\mathbb{1}_{n-1} - \mathbf{I}_{n-1}) & \mathbb{1}_{1,n-1}\mathbb{1}_{n-1,1}\end{bmatrix} \\
&= c\begin{bmatrix}\mathbf{I}_{n-1} + (n-3)\mathbb{1}_{n-1} & (n-2)\mathbb{1}_{n-1,1}\\ (n-2)\mathbb{1}_{1,n-1} & n-1\end{bmatrix}.
\end{aligned}
\tag{47}
$$

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

$$
\mathbf{P}_\alpha \triangleq \begin{bmatrix}\mathbf{P}_{\alpha,1} & \mathbf{P}_{\alpha,2}\\ \mathbf{P}_{\alpha,2}^\mathsf{T} & \mathbf{P}_{\alpha,3}\end{bmatrix}
\tag{48}
$$

where

$$
\begin{aligned}
&\mathbf{P}_{\alpha,1} \\
&= Np(1-p)(\mathcal{I}_{\alpha,1} - \mathcal{I}_{\alpha,2}\mathcal{I}_{\alpha,3}^{-1}\mathcal{I}_{\alpha,2}^\mathsf{T})^{-1} \\
&= (n-1)\big(n(n-2)\mathbf{I}_{n-1} - (n-3)\mathbb{1}_{n-1} \\
&\quad - (n-2)/(n-1)\mathbb{1}_{n-1}\big)^{-1} \\
&= (n-1)\big(n(n-2)\mathbf{I}_{n-1} - (n^2-3n+1)/(n-1)\mathbb{1}_{n-1}\big)^{-1} \\
&= \frac{n-1}{n(n-2)}\left(\mathbf{I}_{n-1} + \frac{n^2-3n+1}{(n-1)^2}\mathbb{1}_{n-1}\right).
\end{aligned}
\tag{49}
$$

Here, the last equality follows from (38). Further, we have

$$
\begin{aligned}
\mathbf{P}_{\alpha,3} &= Np(1-p)(\mathcal{I}_{\alpha,3} - \mathcal{I}_{\alpha,2}^\mathsf{T}\mathcal{I}_{\alpha,1}^{-1}\mathcal{I}_{\alpha,2})^{-1} \\
&= \frac{n-1}{n-2}\Big((n-1) - \frac{1}{n}\mathbb{1}_{1,n-1}\mathbf{I}_{n-1}\mathbb{1}_{n-1,1} \\
&\quad - \frac{(n-3)}{n(2n-3)}\mathbb{1}_{1,n-1}\mathbb{1}_{n-1}\mathbb{1}_{n-1,1}\Big)^{-1} \\
&= \frac{2n-3}{(n-1)(n-2)}
\end{aligned}
\tag{50}
$$

where we have used (38) to arrive at

$$
\mathcal{I}_{\alpha,1}^{-1} = \frac{1}{cn(n-2)}\left(\mathbf{I}_{n-1} + \frac{n-3}{2n-3}\mathbb{1}_{n-1}\right).
\tag{51}
$$

Finally, it holds that

$$\mathbf{P}_{\alpha,2} = -\boldsymbol{\mathcal{I}}_{\alpha,1}^{-1}\boldsymbol{\mathcal{I}}_{\alpha,2}\mathbf{P}_{\alpha,3} \tag{52}$$
$$= \frac{1}{n}\Big(\mathbf{I}_{n-1} + \frac{n-3}{2n-3}\mathbb{1}_{n-1}\Big)\mathbb{1}_{n-1,1}\mathbf{P}_{\alpha,3}$$
$$= \frac{1}{n-1}\mathbb{1}_{n-1,1}$$

which concludes the derivation of $\mathbf{P}_\alpha$ as given in equations (27) and (28).

To derive (29), we begin by showing that

$$\mathbf{P}_\beta = Np(1-p)(\boldsymbol{\mathcal{I}}_\beta - \boldsymbol{\mathcal{I}}_{\alpha,\beta}^{\mathsf{T}}\boldsymbol{\mathcal{I}}_\alpha^{-1}\boldsymbol{\mathcal{I}}_{\alpha,\beta})^{-1}$$
$$= \big((n-1)\mathbf{I}_{n-1} - 1/(n-1)(\mathbf{I}_{n-1} + (n-2)\mathbb{1}_{n-1})\big)^{-1}$$
$$= \frac{n-1}{n-2}\big(n\mathbf{I}_{n-1} - \mathbb{1}_{n-1}\big)^{-1} \tag{53}$$
$$= \frac{n-1}{n(n-2)}(\mathbf{I}_{n-1} + \mathbb{1}_{n-1})$$

where the second equality uses that

$$\boldsymbol{\mathcal{I}}_{\alpha,\beta}^{\mathsf{T}}\boldsymbol{\mathcal{I}}_\alpha^{-1}\boldsymbol{\mathcal{I}}_{\alpha,\beta} = c\begin{bmatrix}\mathbb{1}_{n-1} - \mathbf{I}_{n-1} \\ \mathbb{1}_{1,n-1}\end{bmatrix}^{\mathsf{T}}\begin{bmatrix}\mathbb{1}_{n-1} - \mathbf{I}_{n-1} \\ \mathbb{1}_{1,n-1}\end{bmatrix}$$
$$= c\big((\mathbb{1}_{n-1} - \mathbf{I}_{n-1})^2 + \mathbb{1}_{n-1,1}\mathbb{1}_{1,n-1}\big) \tag{54}$$
$$= c\big(\mathbf{I}_{n-1} + (n-2)\mathbb{1}_{n-1}\big)$$

and the last equality follows from (38). Finally, we have

$$\mathbf{P}_{\alpha,\beta} = -\boldsymbol{\mathcal{I}}_\alpha^{-1}\boldsymbol{\mathcal{I}}_{\alpha,\beta}\mathbf{P}_\beta \tag{55}$$
$$= -\frac{1}{n(n-2)}\begin{bmatrix}\mathbb{1}_{n-1} - \mathbf{I}_{n-1} \\ \mathbb{1}_{1,n-1}\end{bmatrix}(\mathbf{I}_{n-1} + \mathbb{1}_{n-1})$$
$$= -\frac{1}{n(n-2)}\begin{bmatrix}(n-1)\mathbb{1}_{n-1} - \mathbf{I}_{n-1} \\ n\mathbb{1}_{1,n-1}\end{bmatrix}$$

and we are done.

APPENDIX C

We now show how to obtain the FIM for the covariate-based β-model, and then go on to derive the inverse FIM for the special case considered in (31).

To begin with, let us denote the score function as $\mathbf{s} \triangleq [(\mathbf{s}_1^\alpha)^{\mathsf{T}} \ \dots \ (\mathbf{s}_n^\alpha)^{\mathsf{T}} \ (\mathbf{s}_1^\beta)^{\mathsf{T}} \ \dots \ (\mathbf{s}_{n-1}^\beta)^{\mathsf{T}}]^{\mathsf{T}}$ where $\mathbf{s}_i^\alpha \triangleq [s_{i,1}^\alpha \ \dots \ s_{i,K}^\alpha]^{\mathsf{T}}$ for $i = 1, \dots, n$ and $\mathbf{s}_i^\beta \triangleq [s_{i,1}^\beta \ \dots \ s_{i,K}^\beta]^{\mathsf{T}}$ for $i = 1, \dots, n-1$. It then follows from (15) that

$$s_{i,k}^\alpha = \sum_{\ell=1}^{L}\sum_{j\neq i} x_{\ell,k}(Y_{ij,\ell} - N_\ell p_{ij}(\mathbf{x}_\ell)) \tag{56a}$$

for $i = 1, \dots, n$ and $k = 1, \dots, K$, while

$$s_{i,k}^\beta = \sum_{\ell=1}^{L}\sum_{j\neq i} x_{\ell,k}(Y_{ji,\ell} - N_\ell p_{ji}(\mathbf{x}_\ell)) \tag{56b}$$

for $i = 1, \dots, n-1$ and $k = 1, \dots, K$. Further, using (14) and the independence of measurements from different edges and different graphs we have that

$$\mathbb{E}[s_{i,k}^\alpha s_{i,k'}^\alpha] = \sum_{\ell=1}^{L}\sum_{j\neq i} x_{\ell,k}x_{\ell,k'} N_\ell p_{ij}(\mathbf{x}_\ell)(1 - p_{ij}(\mathbf{x}_\ell)) \tag{57a}$$

for $i = 1, \dots, n$, $k = 1, \dots, K$, and $k' = 1, \dots, K$, while

$$\mathbb{E}[s_{i,k}^\beta s_{i,k'}^\beta] = \sum_{\ell=1}^{L}\sum_{j\neq i} x_{\ell,k}x_{\ell,k'} N_\ell p_{ji}(\mathbf{x}_\ell)(1 - p_{ji}(\mathbf{x}_\ell)), \tag{57b}$$

and

$$\mathbb{E}[s_{i,k}^\alpha s_{i,k'}^\beta] = 0, \tag{57c}$$

for $i = 1, \dots, n-1$, $k = 1, \dots, K$, and $k' = 1, \dots, K$. Moreover, we have that

$$\mathbb{E}[s_{i,k}^\alpha s_{j,k'}^\alpha] = 0, \tag{58a}$$
$$\mathbb{E}[s_{i,k}^\beta s_{j,k'}^\beta] = 0, \tag{58b}$$

and

$$\mathbb{E}[s_{i,k}^\alpha s_{j,k'}^\beta] = \sum_{\ell=1}^{L} x_{\ell,k}x_{\ell,k'} N_\ell p_{ij}(\mathbf{x}_\ell)(1 - p_{ij}(\mathbf{x}_\ell)), \tag{58c}$$

for any $i \neq j$, $k = 1, \dots, K$, and $k' = 1, \dots, K$. The FIM for the covariate-based β-model can now be obtained from equations (57) and (58).

For the special case considered in Section III-C, it is easily shown that the FIM reduces to

$$\boldsymbol{\mathcal{I}} = \begin{bmatrix}\boldsymbol{\mathcal{I}}_\alpha & \boldsymbol{\mathcal{I}}_{\alpha,\beta} \\ \boldsymbol{\mathcal{I}}_{\alpha,\beta}^{\mathsf{T}} & \boldsymbol{\mathcal{I}}_\beta\end{bmatrix} \otimes \boldsymbol{\mathcal{I}}_x \tag{59}$$

where

$$\boldsymbol{\mathcal{I}}_\alpha \triangleq N(n-1)\mathbf{I}_n, \tag{60a}$$
$$\boldsymbol{\mathcal{I}}_\beta \triangleq N(n-1)\mathbf{I}_{n-1}, \tag{60b}$$
$$\boldsymbol{\mathcal{I}}_{\alpha,\beta} \triangleq N[\mathbb{1}_{n-1} - \mathbf{I}_{n-1} \ \mathbb{1}_{n-1,1}]^{\mathsf{T}}, \tag{60c}$$

and $\boldsymbol{\mathcal{I}}_x \triangleq \sum_{\ell=1}^{L} p(\mathbf{x}_\ell)(1 - p(\mathbf{x}_\ell))\mathbf{x}_\ell\mathbf{x}_\ell^{\mathsf{T}}$. Noting the similarity of equations (44) and (60), while using that $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$ for any invertible matrices $\mathbf{A}$ and $\mathbf{B}$, the FIM in (31) follows immediately when assuming that $\boldsymbol{\mathcal{I}}_x$ is invertible and that $n > 2$.

REFERENCES

[1] M. Vidal, M. E. Cusick, and A.-L. Barabási, "Interactome networks and 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, , and W. D. Shannon, "Gibbs distribution for statistical analysis of graphical data with a sample application to fcMRI brain images," *Statistics in Med.*, vol. 35, no. 4, pp. 566–580, Feb. 2016.

[3] K. W. Andersen, M. Mørup, 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. Ulitsky, 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 behavior," *Nature Neurosci.*, vol. 18, no. 11, pp. 1565–1567, Nov. 2015.

[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. 15–24.

[8] P. Holme and J. Saramäki, "Temporal networks," *Physics Reports*, vol. 519, no. 3, pp. 97–125, Oct. 2012.

[9] P. W. Holland and S. Leinhardt, "An exponential family of probability distributions for directed graphs," *J. American Statistical Association*, vol. 76, no. 373, pp. 33–50, Mar. 1981.

[10] O. Frank and D. Strauss, "Markov graphs," *J. American Statistical Association*, vol. 81, no. 395, pp. 832–842, Sep. 1986.

[11] D. Strauss, "On a general class of models for interaction," *SIAM Rev.*, vol. 28, no. 4, pp. 513–527, Dec. 1986.

[12] M. E. J. Newman, "The structure and function of complex networks," *SIAM Rev.*, vol. 45, no. 2, pp. 167–256, May 2003.

[13] 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.

[14] A. Rinaldo, S. Petrovic, and S. E. Fienberg, "Maximum likelihood estimation in the β-model," *The Ann. Statist.*, vol. 41, no. 3, pp. 1085–1110, Jun. 2013.

[15] S. J. Cranmer and B. A. Desmarais, "Inferential network analysis with exponential random graph models," *Political Analysis*, vol. 19, no. 1, pp. 66–86, 2011.

[16] S. Chatterjee, P. Diaconis, and A. Sly, "Random graphs with a given degree sequence," *Ann. Appl. Probab.*, vol. 21, no. 4, pp. 1400–1435, Aug. 2011.

[17] T. A. B. Snijders, "Markov chain Monte Carlo estimation of exponential random graph models," *J. Social Structure*, vol. 3, no. 2, pp. 1–40, Apr. 2002.

[18] T. P. Minka, "A comparison of numerical optimizers for logistic regression," Microsoft, Tech. Rep., 2003.

[19] S. E. Fienberg and S. Wasserman, "Categorical data analysis of single sociometric relations," *Sociological Methodology*, vol. 12, pp. 156–192, Jan. 1981.

[20] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Netw.*, vol. 5, no. 2, pp. 109–137, Jun. 1983.

[21] G. Y. W. Yuchung J. Wang, "Stochastic blockmodels for directed graphs," *J. American Statistical Association*, vol. 82, no. 397, pp. 8–19, Mar. 1987.

[22] A. Celisse, J.-J. Daudin, and L. Pierre, "Consistency of maximum-likelihood and variational estimators in the stochastic block model," *Electron. J. Statistics*, vol. 6, pp. 1847–1899, Oct. 2012.

[23] T. Suesse, "Marginalized exponential random graph models," *J. Comput. Graphical Statistics*, vol. 21, no. 4, pp. 883–900, Oct. 2012.

[24] M. A. J. Van Duijn, T. A. B. Snijders, and B. J. H. Zijlstra, "p2 - A random effects model with covariates for directed graphs." *Statistica Neerlandica*, vol. 58, no. 2, pp. 234–254, May 2004.

[25] C. J. Hillar and A. Wibisono, "Maximum entropy distributions on graphs," Aug. 2013, arXiv:1301.3321v2.

[26] T. Yan, Y. Zhao, and H. Qin, "Asymptotic normality in the maximum entropy models on graphs with an increasing number of parameters," *J. Multivariate Analysis*, vol. 133, pp. 61–76, Jan. 2015.

[27] M. Ogawa, H. Hara, and A. Takemura, "Graver basis for an undirected graph and its application to testing the beta model of random graphs," *Ann. of the Institute of Statistical Mathematics*, vol. 65, no. 1, pp. 191–212, Feb. 2013.

[28] V. Csiszár, P. Hussami, J. Komlós, T. F. Móri, L. Rejtö, and G. Tusnády, "When the degree sequence is a sufficient statistic," *Acta Math. Hungar.*, vol. 134, no. 1-2, pp. 45–53, Jun. 2012.

[29] T. Yan and J. Xu, "A central limit theorem in the β-model for undirected random graphs with a diverging number of vertices," *Biometrika*, vol. 100, no. 2, pp. 519–524, Nov. 2013.

[30] V. Karwa and A. Slavkovic, "Inference using noisy degrees: Differentially private β-model and synthetic graphs," *The Ann. Statist.*, vol. 44, no. 1, pp. 87–112, Feb. 2016.

[31] T. Yan, C. Leng, and J. Zhu, "Asymptotics in directed exponential random graph models with an increasing bi-degree sequence," *The Ann. Statistics*, vol. 44, no. 1, pp. 31–57, Feb. 2016.

[32] J. Luo, H. Qin, T. Yan, and L. Zeyneb, "A note on asymptotic distributions in directed exponential random graph models with bi-degree sequences," *Accepted in Commun. Statistics - Theory and Methods*.

[33] G. Rasch, "On general laws and the meaning of measurement in psychology," in *Proc. 4th Berkeley Symp.*, vol. 4, Berkeley, CA, 1961, pp. 321–333.

[34] M. Bolla and A. Elbanna, "Estimating parameters of a probabilistic heterogeneous block model via the EM algorithm," *J. Probability and Statistics*, Nov. 2015, article ID 657965.

[35] B. S. Graham, "An econometric model of link formation with degree heterogeneity," *The National Bureau of Economic Research*, Jul. 2014, nBER Working Paper No. 20341.

[36] T. Yan, B. Jiang, S. E. Fienberg, and C. Leng, "Statistical inference in a directed network model with covariates," Sep. 2016, arXiv:1609.04558.

[37] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, Inc., 1993.

[38] R. van der Hofstad, *Random Graphs and Complex Networks. Vol. I.* Eindhoven University of Technology, 2015.

[39] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philosophical Trans. the Royal Soc. London*, vol. 231, pp. 289–337, Feb. 1933.

[40] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice-Hall, 1998.

[41] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Ann. Math. Statist.*, vol. 9, no. 1, pp. 60–62, Mar. 1938.

[42] T. Yan and Y. Zhao, "Asymptotics of score test in the generalized β-model for networks," *Statistics and Probability Lett.*, vol. 119, pp. 163–169, Dec. 2016.

[43] Y. Li and K. C. Carriere, "Assessing goodness of fit of exponential random graph models," *Int. J. Statistics and Probability*, vol. 2, no. 4, pp. 64–74, Nov. 2013.

[44] S. Petrovic, A. Rinaldo, and S. E. Fienberg, "Algebraic statistics for a directed random graph model with reciprocation," *Algebraic Methods in Statistics and Probability II*, vol. 516, pp. 261–283, Jun. 2010.

[45] G. Fagiolo, "Directed or undirected? A new index to check for directionality of relations in socio-economic networks," *Economics Bulletin*, vol. 3, no. 34, pp. 1–12, Dec. 2006.

[46] F. J. Massey, "The Kolmogorov-Smirnov test for goodness of fit," *J. American Statistical Association*, vol. 46, no. 253, pp. 68–78, Mar. 1951.

[47] P. Vanhems, A. Barrat, C. Cattuto, J.-F. Pinton, N. Khanafer, C. Régis, B. a Kim, B. Comte, and N. Voirin, "Estimating potential infection transmission routes in hospital wards using wearable proximity sensors," *PLoS ONE*, vol. 8, no. 9, p. 73970, Sep. 2013.

[48] G. H. Golub and C. F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, 1996.

[49] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

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 Pontifical 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.

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.

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 Movelo 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.