Statistica Sinica **35** (2025), 1671-1687
doi:https://doi.org/10.5705/ss.202022.0236

EFFICIENT ESTIMATION AND INFERENCE FOR THE SIGNED β-MODEL IN DIRECTED SIGNED NETWORKS

Haoran Zhang and Junhui Wang*

Southern University of Science and Technology and
The Chinese University of Hong Kong

Abstract: This paper proposes a novel signed β-model for directed signed network, which is frequently encountered in application domains but largely neglected in literature. The proposed signed β-model decomposes a directed signed network as the difference of two unsigned networks and embeds each node with two latent factors for in-status and out-status. The presence of negative edges leads to a non-concave log-likelihood, and a one-step estimation algorithm is developed to facilitate parameter estimation, which is efficient both theoretically and computationally. We also develop an inferential procedure for pairwise and multiple node comparisons under the signed β-model, which fills the void of lacking uncertainty quantification for node ranking. Theoretical results are established for the coverage probability of confidence interval, as well as the false discovery rate (FDR) control for multiple node comparison. The finite sample performance of the signed β-model is also examined through extensive numerical experiments on both synthetic and real-life networks.

Key words and phrases: Directed network, estimating equation, false discovery rate, node ranking, one-step estimation, status theory.

1. Introduction

Network data has attracted increasing attention from different scientific communities, due to its flexibility in describing various pairwise relations among multiple objects of interest. In literature, various network models have been developed, such as the Erdös-Rényi model (Erdös and Rényi, 1960), the stochastic block model (Holland, Laskey and Leinhardt, 1983; Zhao, Levina and Zhu, 2012), the β-model (Chatterjee, Diaconis and Sly, 2011), the latent space model (Hoff, Raftery and Handcock, 2002), and the network embedding model (Zhang, He and Wang, 2022). Among them, the β-model is one of the most popular models (Rinaldo, Petrović and Fienberg, 2013; Karwa and Slavković, 2016; Graham, 2017; Chen, Kato and Leng, 2021), which explicitly represents each node i with a numeric factor β_i to accommodate degree heterogeneity. Yet, most existing development of the β-model focuses on undirected and unsigned networks, and it is only recently that the directed β-model (Yan, Leng and Zhu, 2016; Yan et al.,

*Corresponding author. E-mail: junhuiwang@cuhk.edu.hk

2019) has been developed to analyze directed unsigned networks.

In this paper, we propose a novel signed β-model for directed signed network, which is frequently encountered in various application domains but largely neglected in literature. For instances, on many social network platforms such as Facebook or Twitter, users may send likes (positive edges) or dislikes (negative edges) to other users' posts, leading to a directed signed social network. In a citation network, authors may cite papers of other authors, where the citations can be categorized as either endorsement (positive edges) or criticism (negative edges). One interesting feature of directed signed network is the so-called status theory (Guha et al., 2004), which essentially suggests that a directed signed edge pointing from one node to another highly depends on their relative status. The status theory follows from the intuition that nodes with higher status tend to be more influential in the network and attract more attention from nodes with lower status (Leskovec, Huttenlocher and Kleinberg, 2010).

Motivated by the status theory, the proposed signed β-model aims at quantifying the bi-faceted roles of each node, including its in-status and out-status. It models the probability of a directed signed edge from node i to node j in such a way that it is determined by both the out-status factor of node i and the in-status factor of node j. More specifically, a node with lower out-status tends to send more negative edges and less positive edges to other nodes with low in-status, whereas a node with higher in-status tends to receive more positive edges and less negative edges from other nodes with high out-status. Furthermore, the signed β-model decomposes the directed signed network as the difference of two directed unsigned networks, corresponding to the positive and negative edges, respectively. The presence of negative edges leads to a non-concave log-likelihood, casting great challenges for parameter estimation. To circumvent the difficulty, a one-step estimation algorithm is developed, in which a single update is conducted from an initial estimate obtained via estimating equation. Besides computational efficiency, asymptotic estimation efficiency of the one-step estimate is also established. The signed β-model also admits some novel inferential procedures for pairwise and multiple node comparisons with respect to either in-status or out-status, with theoretical guarantees on the coverage probability of confidence interval, as well as the FDR control for multiple node comparison.

Contribution. The main contribution of this paper is three-fold. First, it proposes a novel statistical model for the under-investigated directed signed network, which decomposes it as the weighted difference of two unsigned network, and embeds each node with two latent factors for in-status and out-status. Second, it develops an efficient one-step estimation algorithm to address the non-concavity of the log-likelihood induced by the negative edges, as well as an estimation procedure for the negative sparsity parameters, which guarantees a

theoretically efficient estimate and overcomes the computational-statistical gap. Third, to the best of our limited knowledge, this paper is the first attempt to provide an inferential procedure for pairwise and multiple node comparisons in directed signed networks, which fills the void of lacking uncertainty quantification for node ranking.

Related works. In addition to the directed β-model, there have been some recent works on directed unsigned networks, including the stochastic co-block model (Rohe, Qin and Yu, 2016), and the network embedding model (Zhang, He and Wang, 2022). All these models represent each node with two sets of latent factors, but largely rely on the nature of binary networks and cannot be directly extended to accommodate negative edges. Moreover, there have also been some other works in machine learning literature on community detection in undirected signed networks, such as Chiang, Whang and Dhillon (2012), Chiang et al. (2014), Cucuringu et al. (2019), and Cucuringu et al. (2021). Most of these works focus on the balance theory for undirected signed network (Heider, 1946; Cartwright and Harary, 1956), which is substantially different from the status theory induced by the directed edges.

The proposed model is also related to the Rasch model (Haberman, 1977; Chen et al., 2023) in item response theory, which also relies on the exponential family assumption and thus cannot be applied to directed signed networks. The Bradley–Terry model (Chen et al., 2019; Gao, Shen and Zhang, 2023; Han et al., 2020; Chen, Gao and Zhang, 2022) has also been widely used for ranking problems, where the pairwise comparison is determined by the latent scores assigned to each item in the comparison. Yet, the Bradley–Terry model model is particularly designed for the "skew-symmetric" network, and it remains unclear how to extend the latent scores to incorporate both in-status and out-status in directed signed network.

Organization of the paper. The rest of the paper is organized as follows. Section 2 presents the proposed signed β-model for directed signed network as well as a one-step estimation algorithm. Section 3 establishes the uniform estimation consistency and asymptotic normality of the one-step estimate. Section 4 presents the inferential procedures for pairwise and multiple node comparisons, as well as their theoretical guarantees. Section 5 conducts numerical experiments on synthetic and real-life networks to examine the finite sample performance of the proposed model. Section 6 concludes the paper with a brief discussion, and technical proofs and necessary lemmas are provided in the Appendix.

Throughout this paper, we use c to denote a generic positive constant whose value may vary according to context. For two nonnegative sequences a_n and b_n, $a_n \lesssim b_n$ means there exists a positive constant c such that $a_n \leq cb_n$ when n is sufficiently large. For a vector $\mathbf{h}$, let $\mathbf{h}_{[1:d]}$ denote its first d entries; for a matrix

$\mathbf{H}$, let $\mathbf{H}_{[1:d,1:d]}$ denote its upper left $d \times d$ block.

2. Proposed Method

Suppose a directed signed network $\mathcal{G}$ is observed, with n nodes labeled by $[n] = \{1, \ldots, n\}$ and an adjacency matrix $\mathbf{Y} = (y_{ij})_{n\times n}$ with $y_{ij} \in \{-1, 0, 1\}$. Here, $y_{ij} = 1$ if there is a positive edge from node i to node j, $y_{ij} = -1$ if there is a negative edge from node i to node j, and $y_{ij} = 0$ if no edge is observed at all. Suppose no self loop is allowed, and thus $y_{ii} = 0$ for all $i \in [n]$.

2.1. Signed β-model

The proposed signed β-model first decomposes $\mathcal{G}$ as the difference of two unsigned networks. Specifically, it formulates $y_{ij} = z_{ij}^+ - z_{ij}^-$, where z_{ij}^+ and z_{ij}^- are two independent Bernoulli random variables, and

$$\Pr(z_{ij}^+ = 1) = \frac{e^{\alpha_i+\beta_j}}{1 + e^{\alpha_i+\beta_j}}, \quad \text{and} \quad \Pr(z_{ij}^- = 1) = \frac{\kappa_i}{1 + e^{\alpha_i+\beta_j}}.$$

Here, $\alpha_i + \beta_j$ measures the relative status between nodes i and j, and $\kappa_i \in \{\kappa_{00}, \kappa_{01}\}$ with $0 < \kappa_{00} < \kappa_{01} < 1$ quantifies two different patterns of sending negative edges.

It is clear that as $\alpha_i + \beta_j$ increases, node i is more likely to send a positive edge and less likely to send a negative edge to node j. The probability mass function of y_{ij} can be specified as

$$p(y \mid \alpha_i + \beta_j, \kappa_i) = \begin{cases} \dfrac{e^{2(\alpha_i+\beta_j)} + e^{\alpha_i+\beta_j}(1-\kappa_i)}{(1+e^{\alpha_i+\beta_j})^2}, & \text{if } y = 1; \\ \dfrac{e^{\alpha_i+\beta_j}(1+\kappa_i) + (1-\kappa_i)}{(1+e^{\alpha_i+\beta_j})^2}, & \text{if } y = 0; \\ \dfrac{\kappa_i}{(1+e^{\alpha_i+\beta_j})^2}, & \text{if } y = -1. \end{cases} \tag{2.1}$$

It is interesting to note that (2.1) accommodates the status theory (Guha et al., 2004; Leskovec, Huttenlocher and Kleinberg, 2010) for directed signed network, where β_j represents the in-status for node j and α_i represents the out-status for node i. It implies that a node with higher in-status tends to receive more positive edges, and a node with higher out-status tends to send more positive edges.

The signed β-model is flexible and includes the standard β-model (Chatterjee, Diaconis and Sly, 2011; Graham, 2017) and the directed β-model (Yan, Leng and Zhu, 2016) as its special cases. Particularly, the signed β-model reduces to the directed β-model if all κ_i's are set as 0, and the standard β-model if we further set $\beta_i = \alpha_i$. More interestingly, if we set $\kappa_i = 1$, the signed β-model reduce to two separate β-models, one for z^+ and the other for z^-, except that $\mathbb{E}z^+$ increases as $\alpha_i + \beta_j$ increases, while $\mathbb{E}z^-$ decreases as $\alpha_i + \beta_j$ increases. However,

as negative edges are often much less frequently observed than positive edges in signed networks (Tang et al., 2016), it is more appropriate to employ small κ_i in the signed β-model. In particular, we suppose κ_i could take two different values, $\kappa_i \in \{\kappa_{00}, \kappa_{01}\}$, to characterize two different patterns of sending negative edges, where $\kappa_{00} < \kappa_{01}$ and both of them may decay with n to accommodate sparse networks. For example, we may employ an extremely small κ_{00} for nodes who rarely send negative edges, while κ_{01} could be estimated from data for those nodes who occasionally send negative edges. An estimation procedure determining the class of each κ_i and the value of κ_{01} is provided in the supplement. The signed β-model is also closely related with the ordinal regression model (Hoff, 2021) when y_{ij} is regarded as ordinal response.

Note that the parameters are not identifiable in (2.1), as one can add a constant to α_i and subtract it from β_j without affecting the distribution of y_{ij}. We thus set $\beta_n = 0$ for identifiability, and denote $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top)^\top \in \mathbb{R}^{2n-1}$ as the unknown parameters to be estimated, with $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{n-1})^\top$. We also denote $\boldsymbol{\kappa} = (\kappa_1, \ldots, \kappa_n)^\top$ as the sparsity parameters for negative edges. The presence of negative edges in the directed signed network casts new challenges to the analysis of the signed β-model. Specifically, define $l_{ij}(\alpha_i + \beta_j; \kappa_i) = \log p(y_{ij} \mid \alpha_i + \beta_j; \kappa_i)$ as the log-likelihood for edge y_{ij}. Then, given that y_{ij}'s are mutually independent, the log-likelihood function of $\mathcal{G}$ takes the form

$$l(\boldsymbol{\theta}; \boldsymbol{\kappa}) = \sum_{i,j=1, i \neq j}^{n} l_{ij}(\alpha_i + \beta_j; \kappa_i). \tag{2.2}$$

As the positive value of κ_i leads to a non-concave l_{ij} with respect to α_i and β_j, which further leads a non-concave log-likelihood function with respect to $\boldsymbol{\theta}$, the standard maximum likelihood estimation as in Yan, Leng and Zhu (2016) is no longer feasible. To facilitate parameter estimation, we develop an efficient one-step estimation algorithm, which does not require global optimum but still achieves asymptotical efficiency. In sharp contrast to the asymptotic analysis in Yan, Leng and Zhu (2016), the signed β-model is generally not a member of exponential family, and thus it becomes substantially more challenging to quantify the asymptotic behavior of the one-step estimate.

2.2. One-step estimation

To circumvent the non-concavity issue of $l(\boldsymbol{\theta}; \boldsymbol{\kappa})$ in (2.2), the proposed one-step estimation algorithm performs a single update from an initial estimate of $\boldsymbol{\theta}$ obtained from the estimating equation approach. We assume known $\boldsymbol{\kappa}$ and conduct the asymptotic analysis in the sequel, while an estimation procedure for $\boldsymbol{\kappa}$ and its asymptotic properties are deferred to the supplement.

First, it follows from (2.1) that

$$\mathbb{E}y_{ij} = \frac{e^{\alpha_i+\beta_j} - \kappa_i}{1 + e^{\alpha_i+\beta_j}}.$$

Denote $F(\boldsymbol{\theta};\boldsymbol{\kappa}) = (F_1(\boldsymbol{\theta};\boldsymbol{\kappa}), \ldots, F_{2n-1}(\boldsymbol{\theta};\boldsymbol{\kappa}))^\top$, with

$$F_i(\boldsymbol{\theta};\boldsymbol{\kappa}) = \sum_{k=1,k\neq i}^{n} y_{ik} - \frac{e^{\alpha_i+\beta_k} - \kappa_i}{1 + e^{\alpha_i+\beta_k}}, \text{ for } i \in [n],$$
$$F_{n+j}(\boldsymbol{\theta};\boldsymbol{\kappa}) = \sum_{k=1,k\neq j}^{n} y_{kj} - \frac{e^{\alpha_k+\beta_j} - \kappa_k}{1 + e^{\alpha_k+\beta_j}}, \text{ for } j \in [n-1].$$

Then the initial estimate of $\boldsymbol{\theta}$ can be obtained by solving the following estimating equations,

$$F(\boldsymbol{\theta};\boldsymbol{\kappa}) = \mathbf{0}. \tag{2.3}$$

As will be shown in Theorem 1, (2.3) has a unique solution, denoted as $\check{\boldsymbol{\theta}} = (\check{\boldsymbol{\alpha}}^\top, \check{\boldsymbol{\beta}}^\top)^\top$. We remark that $\check{\boldsymbol{\theta}}$ is derived in the same way as in Yan, Leng and Zhu (2016), which can be further refined via a one-step estimation algorithm.

Let $\check{\beta}_n = 0$, and define

$$\begin{aligned} \check{u}_i &= -\frac{\partial^2 l(\check{\boldsymbol{\theta}};\boldsymbol{\kappa})}{\partial \alpha_i^2} = -\sum_{k=1,k\neq i}^{n} l''_{ik}(\check{\alpha}_i + \check{\beta}_k;\kappa_i), \text{ for } i \in [n], \\ \check{u}_{n+j} &= -\frac{\partial^2 l(\check{\boldsymbol{\theta}};\boldsymbol{\kappa})}{\partial \beta_j^2} = -\sum_{k=1,k\neq j}^{n} l''_{kj}(\check{\alpha}_k + \check{\beta}_j;\kappa_k), \text{ for } j \in [n-1], \end{aligned} \tag{2.4}$$

and let $\check{u}_{2n} = \sum_{i=1}^{n} \check{u}_i - \sum_{j=1}^{n-1} \check{u}_{n+j}$. As will be shown in the proof of Theorem 2, the inverse Fisher information matrix $\{-\partial^2 l(\check{\boldsymbol{\theta}};\boldsymbol{\kappa})/\partial\boldsymbol{\theta}^2\}^{-1}$ can be approximated by

$$\check{\mathbf{H}} = \begin{pmatrix} \check{\mathbf{H}}_{11} & \check{\mathbf{H}}_{12} \\ \check{\mathbf{H}}_{12}^\top & \check{\mathbf{H}}_{22} \end{pmatrix}, \tag{2.5}$$

where $\check{\mathbf{H}}_{11} = \mathrm{diag}(\check{u}_1^{-1}, \ldots, \check{u}_n^{-1}) + \check{u}_{2n}^{-1}\mathbf{1}_n\mathbf{1}_n^\top$, $\check{\mathbf{H}}_{22} = \mathrm{diag}(\check{u}_{n+1}^{-1}, \ldots, \check{u}_{2n-1}^{-1}) + \check{u}_{2n}^{-1}\mathbf{1}_{n-1}\mathbf{1}_{n-1}^\top$, and $\check{\mathbf{H}}_{12} = -\check{u}_{2n}^{-1}\mathbf{1}_n\mathbf{1}_{n-1}^\top$. Here $\mathbf{1}_{2n-1}$ denotes a vector with all ones. Then the one-step estimate is given as

$$\widehat{\boldsymbol{\theta}} = \check{\boldsymbol{\theta}} + \check{\mathbf{H}} \left\{ \frac{\partial l(\check{\boldsymbol{\theta}};\boldsymbol{\kappa})}{\partial \boldsymbol{\theta}} \right\}, \tag{2.6}$$

which is equivalent to

$$\widehat{\alpha}_i = \check{\alpha}_i + \check{u}_i^{-1}\frac{\partial l(\check{\boldsymbol{\theta}};\boldsymbol{\kappa})}{\partial \alpha_i} + \check{u}_{2n}^{-1}\sum_{k=1}^{n}\frac{\partial l(\check{\boldsymbol{\theta}};\boldsymbol{\kappa})}{\partial \alpha_k} - \check{u}_{2n}^{-1}\sum_{l=1}^{n-1}\frac{\partial l(\check{\boldsymbol{\theta}};\boldsymbol{\kappa})}{\partial \beta_l}, \text{ for } i \in [n],$$

$$\widehat{\beta}_j = \check{\beta}_j + \check{u}_{n+j}^{-1}\frac{\partial l(\check{\boldsymbol{\theta}};\boldsymbol{\kappa})}{\partial \beta_j} - \check{u}_{2n}^{-1}\sum_{k=1}^{n}\frac{\partial l(\check{\boldsymbol{\theta}};\boldsymbol{\kappa})}{\partial \alpha_k} + \check{u}_{2n}^{-1}\sum_{l=1}^{n-1}\frac{\partial l(\check{\boldsymbol{\theta}};\boldsymbol{\kappa})}{\partial \beta_l},\ \text{for } j\in[n-1].$$

The final estimate is denoted as $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\alpha}}^\top, \widehat{\boldsymbol{\beta}}^\top)^\top = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_n, \widehat{\beta}_1, \ldots, \widehat{\beta}_{n-1})$ and $\widehat{\beta}_n = 0$. It is worthy pointing out that the one-step estimation in (2.6) needs not to calculate the inverse Hessian matrix as standard Newton-Raphson update, and thus is computationally more efficient. More importantly, this one-step estimation also attains asymptotic estimation efficiency without assuming the intractable global optimum, as will be shown in Theorem 2.

3. Asymptotic Theory

This section establishes the uniform consistency and asymptotic normality of the one-step estimate $\widehat{\boldsymbol{\theta}}$ in Section 2.2. Let $\boldsymbol{\theta}^* = (\alpha_1^*, \ldots \alpha_n^*, \beta_1^*, \ldots, \beta_{n-1}^*)^\top$ denote the true parameters, and $\|\boldsymbol{\theta}^*\|_\infty = \max\{|\alpha_1^*|, \ldots, |\alpha_n^*|, |\beta_1^*|, \ldots, |\beta_{n-1}^*|\}$.

For $i, j \in [n]$, let

$$\begin{aligned}
u_i &= \mathbb{E}\left\{-\frac{\partial^2 l(\boldsymbol{\theta}^*;\boldsymbol{\kappa})}{\partial \alpha_i^2}\right\},\ v_i = \sum_{k=1,k\neq i}^{n}\frac{(1+\kappa_i)e^{\alpha_i^*+\beta_k^*}}{(1+e^{\alpha_i^*+\beta_k^*})^2},\ w_i = \sum_{k=1,k\neq i}^{n}\operatorname{var}(y_{ik}),\\
u_{n+j} &= \mathbb{E}\left\{-\frac{\partial^2 l(\boldsymbol{\theta}^*;\boldsymbol{\kappa})}{\partial \beta_j^2}\right\},\ v_{n+j} = \sum_{k=1,k\neq j}^{n}\frac{(1+\kappa_k)e^{\alpha_k^*+\beta_j^*}}{(1+e^{\alpha_k^*+\beta_j^*})^2},\ w_{n+j} = \sum_{k=1,k\neq j}^{n}\operatorname{var}(y_{kj}),
\end{aligned} \tag{3.1}$$

where $u_{2n} = \mathbb{E}\{-\partial^2 l(\boldsymbol{\theta}^*;\boldsymbol{\kappa})/\partial\beta_n^2\}$ is defined by $u_{2n} = \sum_{i=1}^{n} u_i - \sum_{j=1}^{n-1} u_{n+j}$. We define two matrices as following, which will be shown as the asymptotic covariance matrices for $\check{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}$ in Theorems 1 and 2,

$$\boldsymbol{\Sigma} = \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{12}^\top & \boldsymbol{\Sigma}_{22}\end{pmatrix} \quad \text{and} \quad \mathbf{H} = \begin{pmatrix}\mathbf{H}_{11} & \mathbf{H}_{12}\\ \mathbf{H}_{12}^\top & \mathbf{H}_{22}\end{pmatrix},$$

where $\boldsymbol{\Sigma}_{12} = -w_{2n}v_{2n}^{-2}\mathbf{1}_n\mathbf{1}_{n-1}^\top$, $\mathbf{H}_{12} = -u_{2n}^{-1}\mathbf{1}_n\mathbf{1}_{n-1}^\top$, and

$$\begin{aligned}
\boldsymbol{\Sigma}_{11} &= \operatorname{diag}(w_1v_1^{-2}, \ldots, w_nv_n^{-2}) + w_{2n}v_{2n}^{-2}\mathbf{1}_n\mathbf{1}_n^\top,\\
\boldsymbol{\Sigma}_{22} &= \operatorname{diag}(w_{n+1}v_{n+1}^{-2}, \ldots, w_{2n-1}v_{2n-1}^{-2}) + w_{2n}v_{2n}^{-2}\mathbf{1}_{n-1}\mathbf{1}_{n-1}^\top,\\
\mathbf{H}_{11} &= \operatorname{diag}(u_1^{-1}, \ldots, u_n^{-1}) + u_{2n}^{-1}\mathbf{1}_n\mathbf{1}_n^\top,\\
\mathbf{H}_{22} &= \operatorname{diag}(u_{n+1}^{-1}, \ldots, u_{2n-1}^{-1}) + u_{2n}^{-1}\mathbf{1}_{n-1}\mathbf{1}_{n-1}^\top.
\end{aligned}$$

We first establish the uniform consistency and asymptotic normality of the initial estimate $\check{\boldsymbol{\theta}}$ from the first step.

Theorem 1. *Suppose $\|\boldsymbol{\theta}^*\|_\infty \le c\log n$ with $0 < c < 1/40$ and $\|\boldsymbol{\kappa} - \boldsymbol{\kappa}^*\|_\infty \lesssim e^{12\|\boldsymbol{\theta}^*\|_\infty}\log n/n$. Then as n goes to infinity, with probability at least $1 - c_1/n$ for a constant c_1, it holds true that (2.3) has a unique solution $\check{\boldsymbol{\theta}}$, which satisfies that*

$$\|\check{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty \lesssim e^{6\|\boldsymbol{\theta}^*\|_\infty} \sqrt{\frac{\log n}{n}}. \tag{3.2}$$

Further, for any fixed d, $(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)_{[1:d]}$ is asymptotically multivariate normal with mean $\mathbf{0}$ and covariance matrix given by the upper $d \times d$ block of $\boldsymbol{\Sigma}$.

Theorem 1 shows that the initial estimate $\check{\boldsymbol{\theta}}$ is a fairly good estimate and converges to $\boldsymbol{\theta}^*$ at a fast rate. The asymptotic variance of $\check{\theta}_i$ is given as

$$\operatorname{avar}(\check{\theta}_i) = w_i v_i^{-2} + w_{2n} v_{2n}^{-2},$$

where the term $w_{2n} v_{2n}^{-2}$ is due to the identifiability constraint that $\beta_n = 0$. Further, if $\mathcal{G}$ is an unsigned network with $\kappa_1 = \cdots = \kappa_n = 0$, then $w_i = v_i$ and $\operatorname{avar}(\check{\theta}_i) = v_i^{-1} + v_{2n}^{-1}$, which coincides with the result in Theorem 2 of Yan, Leng and Zhu (2016). More interestingly, Theorem 1 holds true for sparse signed networks by allowing $\|\boldsymbol{\theta}^*\|_\infty \le c \log n$, which matches up with the existing sparsity results for unsigned β-model (Yan, Leng and Zhu, 2016).

We are now ready to establish the consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}$.

Theorem 2. *Under the same condition of Theorem 1, with probability at least $1 - c_2/n$ for a constant c_2, we have*

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty \lesssim e^{2\|\boldsymbol{\theta}^*\|_\infty} \sqrt{\frac{\log n}{n}}. \tag{3.3}$$

Further, for any fixed d, $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)_{[1:d]}$ is asymptotically multivariate normal with mean $\mathbf{0}$ and covariance matrix given by the upper $d \times d$ block of $\mathbf{H}$.

Theorem 2 shows that the one-step estimate $\widehat{\boldsymbol{\theta}}$ also converges to $\boldsymbol{\theta}^*$ at a fast rate, and its asymptotic variance is given as

$$\operatorname{avar}(\widehat{\theta}_i) = u_i^{-1} + u_{2n}^{-1},$$

where the term u_{2n}^{-1} is also due to the identifiability constraint that $\beta_n = 0$. It is important to remark that the one-step estimate $\widehat{\boldsymbol{\theta}}$ is as efficient as the global maximizer of the non-concave log-likelihood function of (2.2), which is difficult to obtain in directed signed networks, if not impossible. Proposition 1 shows that $\widehat{\boldsymbol{\theta}}$ outperforms the initial estimate $\check{\boldsymbol{\theta}}$ asymptotically by reducing the estimation variance, confirming the advantage of the one-step estimate.

Proposition 1. *Under the same conditions of Theorem 1, we have $avar(\check{\theta}_i) \ge avar(\widehat{\theta}_i)$.*

We should point out that the efficiency gain in Proposition 1 is non-negligible, even in the sparse regime. For example, suppose $\beta_1^* = \cdots = \beta_n^* = 0$, and let $\alpha_i^* \to -\infty$ and $\kappa_i \to 0$, and then both positive and negative edges are sparse. It

can be verified that $w_i v_i^{-2}/u_i^{-1} \to 1 + \kappa_i/e^{\alpha_i^*}$, and thus $\text{avar}(\check{\theta}_i)/\text{avar}(\widehat{\theta}_i) > 1$ as long as $\kappa_i = \omega(e^{\alpha_i^*})$, which implies a non-negligible gain in terms of the asymptotic variance.

4. Inference for Node Ranking

Node ranking has been an important task in network data analysis (Wasserman and Faust, 1994), which aims to rank nodes based on their importance or centrality. In literature, many ranking algorithms have been developed for node ranking in directed unsigned networks, including Freeman (1978), Latora and Marchiori (2007), Page et al. (1999), and Kleinberg (1999). These algorithms have also been extended to directed signed network, such as Bonacich and Lloyd (2004), Zolfaghar and Aghaie (2010), and Shahriari and Jalili (2014); readers may refer to Tang et al. (2016) for a complete literature review on node ranking in directed signed networks. Despite the rich literature, most aforementioned algorithms are intuition driven and lack of theoretical justification, not to mention developing an inferential framework to conduct uncertainty quantification for node ranking. Based on the signed β-model, this section develops some inferential procedures for pairwise and multiple node comparisons with respect to either in-status or out-status.

4.1. Pairwise comparison

The statistical inference for $\alpha_i^* - \alpha_j^*$ and $\beta_i^* - \beta_j^*$ represents the relative differences between nodes i and j in terms of their in-status and out-status, respectively. For $i \in [n]$ and $j \in [n-1]$, define

$$\widehat{u}_i = -\frac{\partial^2 l(\widehat{\boldsymbol{\theta}}; \boldsymbol{\kappa})}{\partial \alpha_i^2}, \quad \text{and} \quad \widehat{u}_{n+j} = -\frac{\partial^2 l(\widehat{\boldsymbol{\theta}}; \boldsymbol{\kappa})}{\partial \beta_j^2}. \tag{4.1}$$

Further, for any $i \neq j \in [2n-1]$, define

$$\widehat{\delta}_{ij}^2 = \widehat{u}_i^{-1} + \widehat{u}_j^{-1}, \quad \text{and} \quad (\delta_{ij}^*)^2 = u_i^{-1} + (u_j^*)^{-1}. \tag{4.2}$$

We first establish the asymptotic normality for both $(\widehat{\alpha}_i - \widehat{\alpha}_j)$ and $(\widehat{\beta}_i - \widehat{\beta}_j)$.

Theorem 3. *Under the same condition of Theorem* 1*, for any $i \neq j \in [n]$, it holds true that*

$$\begin{aligned} \delta_{ij}^{*-1} \left\{ (\widehat{\alpha}_i - \widehat{\alpha}_j) - (\alpha_i^* - \alpha_j^*) \right\} &\to N(0,1), \\ \delta_{n+i,n+j}^{*-1} \left\{ (\widehat{\beta}_i - \widehat{\beta}_j) - (\beta_i^* - \beta_j^*) \right\} &\to N(0,1) \end{aligned} \tag{4.3}$$

in distribution. Furthermore, we also have

$$\widehat{\delta}_{ij}^{-1} \left\{ (\widehat{\alpha}_i - \widehat{\alpha}_j) - (\alpha_i^* - \alpha_j^*) \right\} \to N(0,1),$$

$$\widehat{\delta}_{n+i,n+j}^{-1}\left\{(\widehat{\beta}_i-\widehat{\beta}_j)-(\beta_i^*-\beta_j^*)\right\}\to N(0,1) \tag{4.4}$$

in distribution.

According to (4.3), the asymptotic variance for $\widehat{\alpha}_i-\widehat{\alpha}_j$ is δ_{ij}^* which is defined in (4.2), suggesting that the proposed estimate for $\alpha_i^*-\alpha_j^*$ is oracle, in the sense that its asymptotic distribution is the same as the maximum likelihood estimate with known $\{\alpha_k\}_{k\neq i,j}$ and $\{\beta_j\}_{j=1}^{n-1}$.

Furthermore, given the asymptotic normality results in Theorem 2, we can construct a confidence interval for $\alpha_i^*-\alpha_j^*$ as

$$\mathrm{CI}(\alpha_i^*-\alpha_j^*)=\left[(\widehat{\alpha}_i-\widehat{\alpha}_j)-Z_{\alpha/2}\widehat{\delta}_{ij},(\widehat{\alpha}_i-\widehat{\alpha}_j)+Z_{\alpha/2}\widehat{\delta}_{ij}\right],$$

and a confidence interval for $\beta_i-\beta_j$ as

$$\mathrm{CI}(\beta_i^*-\beta_j^*)=\left[(\widehat{\beta}_i-\widehat{\beta}_j)-Z_{\alpha/2}\widehat{\delta}_{n+i,n+j},(\widehat{\beta}_i-\widehat{\beta}_j)+Z_{\alpha/2}\widehat{\delta}_{n+i,n+j}\right],$$

where Z_α denotes the α-th upper percentile of the standard normal distribution. Furthermore, we can also test whether $\alpha_i^*>\alpha_j^*$ based on the indicator $1_{\{(\widehat{\alpha}_i-\widehat{\alpha}_j)-Z_{\alpha/2}\widehat{\delta}_{ij}>0\}}$, and similarly for testing $\beta_i^*>\beta_j^*$.

4.2. Multiple comparison

We now focus on some particular node $i\in[n]$, and find its relative rank within a subgroup of nodes. We take out-status for illustration, and similar procedure can be developed for in-status. Particularly, let $\mathcal{S}\subseteq[n]/\{i\}$ be the subgroup of nodes, and $K=|\mathcal{S}|$. For each $k\in\mathcal{S}$, we want to test

$$H_0^{(k)}:\alpha_i^*=\alpha_k^*\ \text{ v.s. }\ H_a^{(k)}:\alpha_i^*\neq\alpha_k^*.$$

We employ the Benjamini–Hochberg procedure (Benjamini and Yekutieli, 2001) to control the false discovery rate for this multiple testing problem. Let p_k denote the p-value for testing $H_0^{(k)}$, which takes the form

$$p_k=2\left\{1-\Phi\left(\widehat{\delta}_{ik}^{-1}|\widehat{\alpha}_i-\widehat{\alpha}_k|\right)\right\},$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. We order the K p-values as $p^{(1)}\leq\cdots\leq p^{(K)}$, and define

$$r=\max_{1\leq l\leq K}\left\{l:p^{(l)}\leq\frac{\alpha l}{KL}\right\}, \tag{4.5}$$

which is a modification of the traditional Benjamini–Hochberg procedure (Benjamini and Yekutieli, 2001, Thm. 1.3). Here $\alpha>0$ is a given significance level, and $L=\sum_{l=1}^K 1/l$. Then, for any $k\in\mathcal{S}$, we reject $H_0^{(k)}$ if $p_k\leq p^{(r)}$. If the set in

(4.5) is empty, we reject all null hypotheses.

Let $\mathcal{S}_0 = \{k \in \mathcal{S} : \alpha_i^* = \alpha_k^*\}$ be the set of true null hypotheses, then the false discovery rate for the Benjamini–Hochberg procedure takes the form

$$\mathrm{FDR} = \mathbb{E}\left(\frac{\sum_{k\in\mathcal{S}_0} 1_{\{p_k \le p^{(r)}\}}}{\max\left\{\sum_{k\in\mathcal{S}} 1_{\{p_k \le p^{(r)}\}}, 1\right\}}\right) = \mathbb{E}\left(\frac{\sum_{k\in\mathcal{S}_0} 1_{\{p_k \le \alpha r/KL\}}}{\max\{r, 1\}}\right).$$

Theorem 4. *Under the same condition of Theorem* 1*, further suppose* $e^{20\|\boldsymbol{\theta}^*\|_\infty} K_0 n^{-1/2} (\log n)^2 = o(1)$*, with* $K_0 = |\mathcal{S}_0|$*. Then, we have*

$$FDR \le \frac{\alpha K_0}{K}\left(1 + \frac{1}{L}\right) + o(1).$$

Theorem 4 immediately implies that when K_0 is dominated by $e^{-20\|\boldsymbol{\theta}^*\|_\infty} n^{1/2} (\log n)^{-2}$, the false discovery rate of the Benjamini–Hochberg procedure is upper bounded by α asymptotically as long as $(K_0/K)(1+1/L) \le 1$, which is a fairly mild condition since $K_0 \le K$ and L is roughly $\log K$.

5. Numerical Experiments

This section examines the finite sample performance of the proposed one-step estimate as well as the inferential procedures, where the sparse factor $\boldsymbol{\kappa}$ is estimated as described in the supplement and the estimating equation in (2.3) is solved by Newton's method.

5.1. Simulation

The simulated directed signed networks are generated as follows. We first generate 10 groups of nodes, where nodes in the same group have the same in-status and out-status. Specifically, let $\psi_i \in \{1, \ldots, 10\}$ denote the group membership of node i, generated independently from a multinomial distribution with probabilities $(0.15 \times \mathbf{1}_5^\top, 0.05 \times \mathbf{1}_5^\top)$ to accommodate unbalanced groups. We then set $\alpha_i^* = a_{\psi_i} \sim N(-0.5, 0.5)$, $\beta_i^* = b_{\psi_i} \sim N(0, 0.5)$ and $\beta_n^* = 0$. The directed edges, y_{ij}, are then generated independently from (2.1). The κ_i are randomly generated from $\{\kappa_{00}, \kappa_{01}\}$ with $\Pr(\kappa_i = \kappa_{01}) = 0.8$. Various scenarios are considered, with $n \in \{200, 600, 1000\}$, $\kappa_{01} \in \{0.05, 0.1, 0.2, 0.25\}$ and $\kappa_{00} = 0.001$.

In each scenario, the averaged estimation errors, measured by $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty$ and $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2^2/(2n-1)$, over 500 independent replications with their standard errors are reported in Tables 1 and 2. The averaged coverage frequencies of the 95% confidence interval for 100 randomly selected $\widehat{\alpha}_i - \widehat{\alpha}_j$, together with their standard errors, are reported in Table 3. In addition, we randomly select a node from the group with the largest in-status, and compare the in-status of this node with the rest nodes in this group and the first 10 nodes from each of the other groups. The

Table 1. The averaged estimation errors over 500 independent replications and their standard errors in parenthesis.

	n	$\kappa_{01} = 0.05$	$\kappa_{01} = 0.1$	$\kappa_{01} = 0.2$	$\kappa_{01} = 0.25$
$\|\check{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty$	200	0.6087 (0.1098)	0.6210 (0.1179)	0.6873 (0.1296)	0.7483 (0.1467)
	600	0.3756 (0.0617)	0.4415 (0.0697)	0.6474 (0.0923)	0.7528 (0.1007)
	1,000	0.3057 (0.0483)	0.3929 (0.0686)	0.6352 (0.0862)	0.7445 (0.0865)
$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_\infty$	200	0.5905 (0.1003)	0.5908 (0.1064)	0.6196 (0.1103)	0.6616 (0.1237)
	600	0.3603 (0.0577)	0.3820 (0.0581)	0.5070 (0.0784)	0.5939 (0.0892)
	1,000	0.2899 (0.0453)	0.3133 (0.0511)	0.4666 (0.0749)	0.5690 (0.0759)

Table 2. The averaged mean squared errors over 500 independent replications and their standard errors in parenthesis.

	n	$\kappa_{01} = 0.05$	$\kappa_{01} = 0.1$	$\kappa_{01} = 0.2$	$\kappa_{01} = 0.25$
$\frac{\|\check{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2^2}{2n-1}$	200	0.0467 (0.0296)	0.0473 (0.0299)	0.0508 (0.0324)	0.0525 (0.0309)
	600	0.0147 (0.0103)	0.0156 (0.0093)	0.0180 (0.0085)	0.0195 (0.0090)
	1,000	0.0091 (0.0067)	0.0093 (0.0060)	0.0107 (0.0070)	0.0114 (0.0066)
$\frac{\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_2^2}{2n-1}$	200	0.0449 (0.0289)	0.0449 (0.0292)	0.0468 (0.0298)	0.0478 (0.0290)
	600	0.0141 (0.0100)	0.0144 (0.0093)	0.0157 (0.0082)	0.0168 (0.0084)
	1,000	0.0088 (0.0067)	0.0087 (0.0061)	0.0096 (0.0066)	0.0100 (0.0063)

Table 3. The averaged coverage frequencies of the 95% confidence interval over 500 independent replications for 100 randomly selected pairs (i, j), and their standard errors in parenthesis.

	n	$\kappa_{01} = 0.05$	$\kappa_{01} = 0.1$	$\kappa_{01} = 0.2$	$\kappa_{01} = 0.25$
Coverage Frequency ($\check{\boldsymbol{\theta}}$)	200	0.9371 (0.0185)	0.9341 (0.0245)	0.9051 (0.0603)	0.8886 (0.0817)
	600	0.9423 (0.0166)	0.9206 (0.0491)	0.8887 (0.0921)	0.8852 (0.0944)
	1,000	0.9431 (0.0145)	0.9319 (0.0339)	0.9248 (0.0439)	0.9211 (0.0472)
Coverage Frequency ($\widehat{\boldsymbol{\theta}}$)	200	0.9441 (0.0124)	0.9419 (0.0155)	0.9218 (0.0388)	0.9043 (0.0610)
	600	0.9477 (0.0098)	0.9353 (0.0237)	0.8979 (0.0778)	0.8895 (0.0894)
	1,000	0.9466 (0.0101)	0.9399 (0.0188)	0.9262 (0.0415)	0.9217 (0.0471)

averaged false discovery proportions (FDP) and power, as well as their standard errors, are reported in Tables 4 and 5. We also report these evaluation metrics of the initial estimate $\check{\boldsymbol{\theta}}$ for comparison.

It is evident from Tables 1 and 2 that the estimation errors for both $\check{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\theta}}$ decrease as n increases, which validates the asymptotic estimation consistency in Theorems 1 and 2. Also, the performance of $\widehat{\boldsymbol{\theta}}$ is consistently better than that of $\check{\boldsymbol{\theta}}$ in all scenarios, which is expected according to Theorem 2. It is also interesting to note that the estimation errors of $\widehat{\boldsymbol{\theta}}$ are fairly robust against the sparsity level of the negative edges. In Table 3, it is clear that the coverage frequencies for the 95% confidence interval are close to the nominal level as n grows, which supports the asymptotic normality results in Theorem 3. As for multiple testing, as shown in Tables 4 and 5, the false discovery proportion is way below 0.05 in all scenarios

Table 4. The averaged FDP over 500 independent replications and their standard errors in parenthesis.

	n	$\kappa_{01} = 0.05$	$\kappa_{01} = 0.1$	$\kappa_{01} = 0.2$	$\kappa_{01} = 0.25$
FDP ($\check{\boldsymbol{\theta}}$)	200	0.0013 (0.0062)	0.0015 (0.0063)	0.0030 (0.0094)	0.0051 (0.0122)
	600	0.0027 (0.0095)	0.0065 (0.0176)	0.0289 (0.0539)	0.0341 (0.0618)
	1,000	0.0053 (0.0233)	0.0077 (0.0301)	0.0182 (0.0572)	0.0240 (0.0676)
FDP ($\widehat{\boldsymbol{\theta}}$)	200	0.0014 (0.0069)	0.0014 (0.0060)	0.0027 (0.0089)	0.0042 (0.0112)
	600	0.0025 (0.0089)	0.0048 (0.0143)	0.0225 (0.0447)	0.0296 (0.0570)
	1,000	0.0050 (0.0229)	0.0065 (0.0265)	0.0154 (0.0512)	0.0227 (0.0652)

Table 5. The averaged power over 500 independent replications and their standard errors in parenthesis.

	n	$\kappa_{01} = 0.05$	$\kappa_{01} = 0.1$	$\kappa_{01} = 0.2$	$\kappa_{01} = 0.25$
Power ($\check{\boldsymbol{\theta}}$)	200	0.6610 (0.1068)	0.6440 (0.1220)	0.6553 (0.1355)	0.6652 (0.1410)
	600	0.7871 (0.0712)	0.7935 (0.0767)	0.8029 (0.0920)	0.8042 (0.0937)
	1,000	0.8370 (0.0608)	0.8388 (0.0609)	0.8432 (0.0634)	0.8468 (0.0624)
Power ($\widehat{\boldsymbol{\theta}}$)	200	0.6531 (0.1087)	0.6428 (0.1185)	0.6558 (0.1303)	0.6670 (0.1346)
	600	0.7863 (0.0712)	0.7935 (0.0753)	0.8041 (0.0872)	0.8071 (0.0898)
	1,000	0.8375 (0.0605)	0.8395 (0.0605)	0.8438 (0.0625)	0.8494 (0.0609)

and the power increases as n grows, justifying the false discovery rate control in Theorem 4. Similarly, the performance of $\widehat{\boldsymbol{\theta}}$ is better than $\check{\boldsymbol{\theta}}$.

5.2. Signed citation network

We now apply the proposed method to analyze a real-life signed citation network (Kumar, 2016), which is collected based on citations within the natural language processing (NLP) community during 1975–2013. Particularly, each author is represented as a node, which sends edges to other nodes by citing their papers. According to the citation sentiments, the citations can be categorized into "endorsement", "criticism" and "neural". Both "endorsement" and "neural" are treated as positive edges, while "criticism" is treated as negative edge. We remove all authors who send or receive less than 5 edges, as well as about 80 authors who send or receive more negative edges than positive edges, leading to spuriously high out-status or low in-status. This pre-processing step leads to a directed network with 1,849 nodes, 56,517 positive edges and 3,726 negative edges. We set $\kappa_i = \kappa_{00} = 0.001$ for those nodes who do not send negative edges, and $\kappa_i = \kappa_{01}$ for the rest, with κ_{01} estimated as described in the supplement.

The left panel of Figure 1 shows the in-degrees for the 10 nodes with highest estimated in-status. In particular, each bar above the x-axis shows the positive in-degree $\sum_{j=1,j\neq i}^{n} y_{ji} 1_{\{y_{ji}>0\}}$, while the bar below the x-axis shows the negative in-degree $\sum_{j=1,j\neq i}^{n} |y_{ji}| 1_{\{y_{ji}<0\}}$. The black line further shows the estimated $\widehat{\beta}_i$ for the 10 nodes after rescaling. The right panel of Figure 1 shows the out-degrees

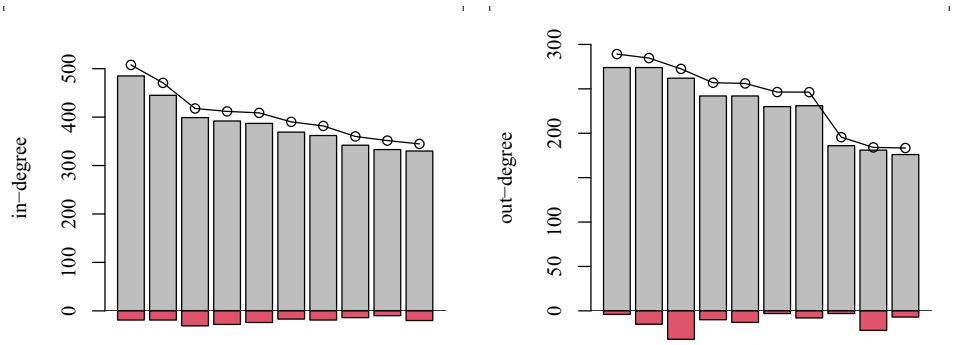

Figure 1. The left panel shows the positive in-degree (above x-axis) and negative in-degree (below x-axis) for the 10 nodes with largest $\widehat{\beta}_i$. The black line shows the corresponding $\widehat{\beta}_i$ after rescaling. The right panel shows the positive out-degree (above x-axis) and negative out-degree (below x-axis) for the 10 nodes with largest $\widehat{\alpha}_i$. The black line shows the corresponding $\widehat{\alpha}_i$ after rescaling.

for the 10 nodes with highest estimated out-status. It can be seen that the estimated in-status is highly positively correlated with the positive in-degree, reflecting the attractiveness or popularity of the node; whereas the estimated out-status is highly positively correlated with the positive out-degree, reflecting the social status of the node to some extent. It is also interesting to note that in the right panel, the 6th node actually sends out less positive edges than the 7th node, but it possesses a higher out-status as it also sends out less negative edges.

To further scrutinize the results in Figure 1, we note that the 10 nodes with highest estimated in-status are Christopher Manning, Daniel Klein, Michael Collins, Salim Roukos, Franz Josef Och, Fernando Pereira, Daniel Marcu, Philipp Koehn, Vincent J. Della Pietra and Eugene Charniak. Clearly, they are all highly cited researchers in the NLP community, with google scholar citation counts ranging from 25,000 to 189,000. We also note that the 10 nodes with highest estimated out-status are Joakim Nivre, Noah Smith, Mirella Lapata, Timothy Baldwin, Christopher Manning, Chris Callison-Burch, Junichi Tsujii, Eduard Hovy, Dan Roth and Regina Barzilay. These researchers are very active and productive, with number of publications ranging from 254 to 732, according to google scholar. It is also interesting to note that Christopher Manning is among both lists, who is well regarded as an influential and productive NLP expert.

Furthermore, we give two examples of node comparison with respect to their in-status. Specifically, in each example, we choose a node i and compare it with a randomly chosen subset $\mathcal{S} = \{i_k : k = 1, \ldots, 50\}$ with respect to in-status. Figure 2 shows the in-degree of nodes in $\{i\} \cup \mathcal{S}$ after reordering, where each bar above the x-axis shows the positive in-degree and the bar below the x-axis shows the negative in-degree. The solid line shows the estimated in-status for the 51 nodes after rescaling, where the diamond represents node i, and the items

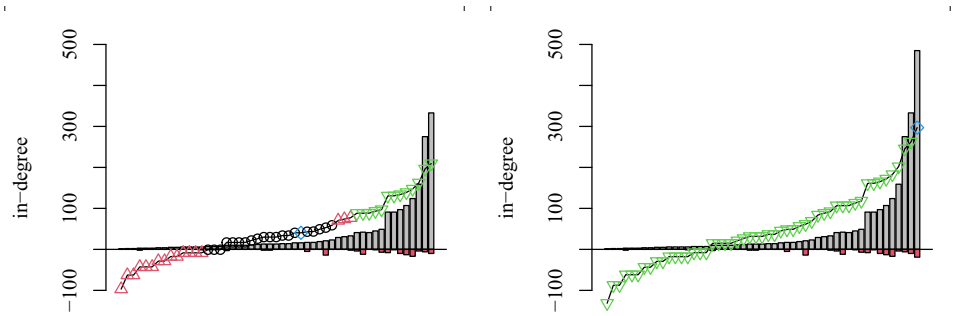

Figure 2. The estimated $\widehat{\beta}_j$ for $j \in \{i\} \cup \mathcal{S}$ after rescaling in two examples, where the diamond represents node i, and the items represent nodes in $\mathcal{S}$. Both triangles and inverted triangles indicate statistical significance at level 95% in comparison with node i, whereas inverted triangles also indicate statistical significance for multiple testing. The bars further show the positive/negative in-degrees for these nodes.

represent nodes in $\mathcal{S}$. In the left panel, it is clear that the 14 leftmost nodes are significantly less attractive than node i whereas the 16 rightmost nodes are significantly more attractive. It is further shown that the in-status of the 13 rightmost nodes are significantly different from that of node i simultaneously. In the right panel, node i is chosen to be Christopher Manning, and it is evident that his in-status is significantly higher than that of all nodes in $\mathcal{S}$ simultaneously.

6. Conclusion

This paper proposes a general signed β-model for directed signed network, which embeds each node with two latent factors for in-status and out-status and includes the standard β-model and the directed β-model as special cases. We develop an efficient one-step estimation algorithm for parameter estimation and inferential procedure for node comparison, which leads to a computationally feasible estimation with asymptotic estimation efficiency. Note that the proposed model builds upon the mutual independence between the latent variables z_{ij}^+ and z_{ij}^-, it is thus of interest to relax this assumption and consider correlated latent variables. This will bring challenges in both computation and theoretical derivation, as the resultant likelihood function can be more complex than (2.2) and the current one-step estimation procedure is no longer applicable. Further, it is also interesting to extend the current procedure to analyze much sparser signed networks with exploitation of regularization method (Shao et al., 2021). We leave it for future investigation.

Supplementary Material

In the supplement, we provide technical proofs for all the theoretical results, as well as an estimation procedure for $\boldsymbol{\kappa}$.

Acknowledgments

This research is supported in part by HK RGC Grants GRF-11304520, GRF-11301521 and GRF-11311022.

References

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* **57**, 1165–1188.

Bonacich, P. and Lloyd, P. (2004). Calculating status with negative relations. *Social Networks* **26**, 331–338.

Cartwright, D. and Harary, F. (1956). Structural balance: A generalization of Heider's theory. *Psychological Review* **63**, 277–293.

Chatterjee, S., Diaconis, P. and Sly, A. (2011). Random graphs with a given degree sequence. *The Annals of Applied Probability* **21**, 1400–1435.

Chen, M., Kato, K. and Leng, C. (2021). Analysis of networks via the sparse β-model. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **83**, 887–910.

Chen, P., Gao, C. and Zhang, A.Y. (2022). Optimal full ranking from pairwise comparisons. *The Annals of Statistics* **50**, 1775–1805.

Chen, Y., Fan, J., Ma, C. and Wang, K. (2019). Spectral method and regularized mle are both optimal for top-k ranking. *The Annals of statistics* **47**, 2204–2235.

Chen, Y., Li, C., Ouyang, J. and Xu, G. (2023). Statistical inference for noisy incomplete binary matrix. *Journal of Machine Learning Research* **24**, 4368–4433.

Chiang, K.-Y., Hsieh, C.-J., Natarajan, N., Dhillon, I.S. and Tewari, A. (2014). Prediction and clustering in signed networks: A local to global perspective. *The Journal of Machine Learning Research* **15**, 1177–1213.

Chiang, K.-Y., Whang, J.J. and Dhillon, I.S. (2012). Scalable clustering of signed networks using balance normalized cut. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, 615–624.

Cucuringu, M., Davies, P., Glielmo, A. and Tyagi, H. (2019). SPONGE: A generalized eigenproblem for clustering signed networks. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics* **PMLR 89**, 1088–1098.

Cucuringu, M., Singh, A.V. , Sulem, D. and Tyagi, H. (2021). Regularized spectral methods for clustering signed networks. *Journal of Machine Learning Research* **22**, 1–79.

Erdös, P. and Rényi, A. (1960). On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl* **5**, 17–61.

Freeman, L.C. (1978). Centrality in social networks conceptual clarification. *Social Networks* **1**, 215–239.

Gao, C., Shen, Y. and Zhang, A.Y. (2023). Uncertainty quantification in the bradley–terry–luce model. *Information and Inference: A Journal of the IMA* **12**, 1073–1140.

Graham, B.S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica* **85**, 1033–1063.

Guha, R., Kumar, R., Raghavan, P. and Tomkins, A. (2004). Propagation of trust and distrust. In *Proceedings of the 13th International Conference on World Wide Web*, 403–412. New York, NY.

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

Han, R., Ye, R., Tan, C. and Chen, K. (2020). Asymptotic theory of sparse bradley–terry model. *The Annals of Applied Probability* **30**, 2491–2515.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology* **21**, 107–112.

Hoff, P. (2021). Additive and multiplicative effects network models. *Statistical Science* **36**, 34–50.

Hoff, P., Raftery, A. and Handcock, M. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association* **97**, 1090–1098.

Holland, P.W., Laskey, K.B. and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks* **5**, 109–137.

Karwa, V. and Slavković, A. (2016). Inference using noisy degrees: Differentially private β-model and synthetic graphs. *The Annals of Statistics* **44**, 87–112.

Kleinberg, J.M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)* **46**, 604–632.

Kumar, S. (2016). Structure and dynamics of signed citation networks. In *Proceedings of the 25th International Conference Companion on World Wide Web*, 63–64.

Latora, V. and Marchiori, M. (2007). A measure of centrality based on network efficiency. *New Journal of Physics* **9**, 188.

Leskovec, J., Huttenlocher, D. and Kleinberg, J. (2010). Signed networks in social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1361–1370. New York, NY.

Page, L., Brin, S., Motwani, R. and Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Technical report. Stanford Digital Library Technologies Project.

Rinaldo, A., Petrović, S. and Fienberg, S.E. (2013). Maximum lilkelihood estimation in the β-model. *The Annals of Statistics* **41**, 1085–1110.

Rohe, K., Qin, T. and Yu, B. (2016). Co-clustering directed graphs to discover asymmetries and directional communities. *Proceedings of the National Academy of Sciences* **113**, 12679–12684.

Shahriari, M. and Jalili, M. (2014). Ranking nodes in signed social networks. *Social Network Analysis and Mining* **4**, Article No. 172.

Shao, M., Zhang, Y., Wang, Q., Luo, J. and Yan T. (2021). L_2 regularized maximum likelihood for β-model in large and sparse networks. *arXiv:2110.11856*.

Tang, J., Chang, Y., Aggarwal, C. and Liu, H. (2016). A survey of signed network mining in social media. *ACM Computing Surveys (CSUR)* **49**, Article No. 42.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge.

Yan, T., Jiang, B., Fienberg, S. and Leng, C. (2019). Statistical inference in a directed network model with covariates. *Journal of the American Statistical Association* **114**, 857–868.

Yan, T., Leng, C. and Zhu, J. (2016). Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *The Annals of Statistics* **44**, 31–57.

Zhang, J., He, X. and Wang, J. (2022). Directed community detection with network embedding. *Journal of the American Statistical Association* **117**, 1809–1819.

Zhao, Y., Levina, E. and Zhu, J. (2012). Consistency of community detection in networks under degree-corrected stochastic block models. *The Annals of Statistics* **40**, 2266–2292.

Zolfaghar, K. and Aghaie, A. (2010). Mining trust and distrust relationships in social web applications. In *Proceedings of the 2010 IEEE 6th International Conference on Intelligent Computer Communication and Processing*, 73–80. Cluj-Napoca, Romania.

(Received July 2022; accepted August 2023)