A Regression Framework for Studying Relationships among Attributes under Network Interference

Cornelius Fritz

Michael Schweinberger*

Subhankar Bhadra

David R. Hunter

Department of Statistics, The Pennsylvania State University

Abstract

To understand how the interconnected and interdependent world of the twenty-first century operates and make model-based predictions, joint probability models for networks and interdependent outcomes are needed. We propose a comprehensive regression framework for networks and interdependent outcomes with multiple advantages, including interpretability, scalability, and provable theoretical guarantees. The regression framework can be used for studying relationships among attributes of connected units and captures complex dependencies among connections and attributes, while retaining the virtues of linear regression, logistic regression, and other regression models by being interpretable and widely applicable. On the computational side, we show that the regression framework is amenable to scalable statistical

*Corresponding author.

computing based on convex optimization of pseudo-likelihoods using minorization-maximization methods. On the theoretical side, we establish convergence rates for pseudo-likelihood estimators based on a single observation of dependent connections and attributes. We demonstrate the regression framework using simulations and an application to hate speech on the social media platform X.

Keywords: Dependent Data, Generalized Linear Models, Minorization-Maximization, Pseudo-Likelihood

1 Introduction

In the interconnected and interdependent world of the twenty-first century, individual and collective outcomes—such as personal and public health, economic welfare, or war and peace—are affected by relationships among individual, corporate, state, and non-state actors. To understand how the world of the twenty-first century operates and make model-based predictions, it is vital to study networks of relationships and gain insight into how the structure of networks affects individual and collective outcomes.

While the structure of networks has been widely studied (see Kolaczyk, 2017, and references therein), the structure of networks is rarely of primary interest. Instead, we often wish to understand how networks affect individual or collective outcomes. For example, social, economic, and financial relationships among individual and corporate actors can affect the welfare of people, but the outcome of primary interest is the welfare of billions of people around the world. Relationships among state and non-state actors can affect war and peace, but the outcome of primary interest is the welfare of nations. Contact networks mediate the spread of infectious diseases, but the outcome of primary interest is public health. A final example is causal inference under network interference: If the outcomes

of units are affected by the treatments or outcomes of other units, the spillover effect of treatments on outcomes can be represented by an intervention network, but the target of statistical inference is the direct and indirect causal effects of treatments on outcomes.

To learn how networks are wired and how the structure of networks affects outcomes of interest, data on outcomes $\boldsymbol{Y} \coloneqq (Y_i)_{i=1}^N$ and connections $\boldsymbol{Z} \coloneqq (Z_{i,j})_{i,j}^N$ among N units are needed along with predictors $\boldsymbol{X} \coloneqq (\boldsymbol{X}_i)_{i=1}^N$. Statistical work on joint probability models for dependent outcomes and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ is scarce. Snijders et al. (2007) and Niezink and Snijders (2017) develop models for behavioral outcomes and connections using continuous-time Markov processes, assuming that the behavioral outcomes and connections are observed at two or more time points. Wang et al. (2024) combine Ising models for binary outcomes with exponential family models for binary connections, with applications to causal inference (Clark and Handcock, 2024). In a Bayesian framework, Fosdick and Hoff (2015) unite models for continuous outcomes with latent variable models that capture dependencies among connections. A common feature of these approaches is that the models and methods in these works may be useful in small populations with, say, hundreds of members, but may be less useful in large populations with, say, thousands or millions of members. For example, many of these models make dependence assumptions that are reasonable in small populations but are less reasonable in large populations. In the special case of exponential-family models, it is known that models that make unreasonable dependence assumptions can give rise to undesirable probabilistic and statistical behavior in large populations, such as model near-degeneracy (Handcock, 2003; Schweinberger, 2011; Chatterjee and Diaconis, 2013). In addition, these works rely on Monte Carlo and Markov chain Monte Carlo methods for moment- and likelihood-based inference, which limits the scalability of the mentioned approaches. Last, but not least, the theoretical properties of

statistical procedures based on dependent outcomes and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$, such as the convergence rates of estimators, are unknown.

While statistical work on joint probability models for $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ is scarce, recent progress has been made on conditional models for outcomes $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{z})$ and connections $\boldsymbol{Z} \mid \boldsymbol{X} = \boldsymbol{x}$. For example, the literature on network-aware regression uses conditional models for outcomes $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{z})$: Lei et al. (2024) assume that the dependence among outcomes decays as a function of distance in the population network, while Li et al. (2019) and Le and Li (2022) encourage outcomes of connected units to be similar. A related branch of literature, concerned with causal inference under network interference, leverages conditional models for outcomes $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{z})$ given treatment assignments $\boldsymbol{X} = \boldsymbol{x}$ and connections $\boldsymbol{Z} = \boldsymbol{z}$. Some of them consider fixed connections (e.g., Tchetgen Tchetgen et al., 2021; Ogburn et al., 2024), while others combine conditional models for $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{z})$ with marginal models for $\boldsymbol{Z}$, assuming that connections are independent (Li and Wager, 2022). Other works advance autoregressive network models for $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{z})$ (Huang et al., 2019, 2020; Zhu et al., 2020). Conditional models for connections $\boldsymbol{Z} \mid \boldsymbol{X} = \boldsymbol{x}$ include stochastic block and exponential-family models with covariates (e.g., Handcock, 2003; Huang et al., 2024; Wang et al., 2024; Stein et al., 2025).

All of the cited work is limited to special cases, such as real-valued outcomes or binary connections, rather than presenting a comprehensive regression framework for studying relationships among attributes $(\boldsymbol{X}, \boldsymbol{Y})$ under network interference $\boldsymbol{Z}$. To fill the void left by existing work, we propose a comprehensive regression framework for studying relationships among attributes $(\boldsymbol{X}, \boldsymbol{Y})$ under network interference $\boldsymbol{Z}$ based on joint probability models for $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$. The proposed regression framework has important advantages over existing work, including interpretability, scalability, and provable theoretical guarantees:

1. We show in Sections 2.1 and 2.2 that the proposed regression framework can be viewed as a generalization of linear regression, logistic regression, and other regression models for studying relationships among attributes under network interference, adding a simple and widely applicable set of tools to the toolbox of data scientists. We demonstrate the advantages of the regression framework with an application to hate speech on the social media platform X in Section 6.

2. The proposed regression framework can be applied to small and large populations by leveraging additional structure to control the dependence among outcomes and connections, facilitating the construction of models with complex dependencies among outcomes and connections in small and large populations.

3. We develop scalable methods using minorization-maximization algorithms for convex optimization of pseudo-likelihoods in Section 3. To disseminate the regression framework and its scalable methods, we provide an `R` package.

4. We establish theoretical guarantees for pseudo-likelihood estimators in Section 4. To the best of our knowledge, these are the first theoretical guarantees for joint probability models of $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ based on a single observation $(\boldsymbol{y}, \boldsymbol{z})$ of $(\boldsymbol{Y}, \boldsymbol{Z})$. The simulation results in Section 5 demonstrate that pseudo-likelihood estimators perform well as the number of units N and the number of parameters p increases.

In addition, the regression framework has conceptual and statistical advantages:

5. Compared with conditional models for outcomes $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{z})$ and connections $\boldsymbol{Z} \mid \boldsymbol{X} = \boldsymbol{x}$, the proposed regression framework for $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ provides insight into outcome-connection dependencies, in addition to outcome-outcome and connection-connection dependencies.

6. Compared with conditional models for outcomes $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{z})$, conclusions based on the proposed regression framework are not limited to a specific population network $\boldsymbol{z}$, but can be extended to the superpopulation of all possible population networks. In addition, the proposed regression framework provides insight into the probability law governing the superpopulation of all possible population networks.

7. The proposed regression framework retains the advantages of two general approaches to building joint probability models for dependent data, elucidated in the celebrated paper by Besag (1974): Specifying a joint probability distribution directly guarantees desirable mathematical properties, while specifying it indirectly via conditional probability distributions helps build complex models from simple building blocks. We show how to directly specify a joint probability model from simple building blocks. The resulting regression framework possesses desirable mathematical properties and induces conditional distributions that can be represented by regression models, facilitating interpretation. We showcase these advantages in Sections 2.2 and 6.2.

We elaborate the proposed regression framework in the remainder of the article.

2 Regression under Network Interference

Consider a population of $N \geq 2$ units $\mathcal{P}_N := \{1, \ldots, N\}$, where each unit $i \in \mathcal{P}_N$ possesses

- one or more binary, count-, or real-valued predictors $\boldsymbol{X}_i \in \mathcal{X}_i$, which may include covariates and treatment assignments;

- binary, count-, or real-valued outcomes or responses $Y_i \in \mathcal{Y}_i$;

- binary, count-, or real-valued connections $Z_{i,j} \in \mathcal{Z}_{i,j}$ to other units $j \in \mathcal{P}_N \setminus \{i\}$,

which represent indicators of connections or weights of connections (e.g., the number of interactions between i and j).

We first consider undirected connections, for which $Z_{i,j}$ equals $Z_{j,i}$, and describe extensions to directed connections in Section 6. We write $\boldsymbol{X} := (\boldsymbol{X}_i)_{1 \leq i \leq N}$, $\boldsymbol{Y} := (Y_i)_{1 \leq i \leq N}$, $\boldsymbol{Z} := (Z_{i,j})_{1 \leq i < j \leq N}$, $\mathcal{X} := \bigtimes_{i=1}^N \mathcal{X}_i$, $\mathcal{Y} := \bigtimes_{i=1}^N \mathcal{Y}_i$, and $\mathcal{Z} := \bigtimes_{i<j} \mathcal{Z}_{i,j}$, and refer to $\boldsymbol{Y}$ without Y_i and $\boldsymbol{Z}$ without $Z_{i,j}$ as $\boldsymbol{Y}_{-i} \in \mathcal{Y}_{-i}$ and $\boldsymbol{Z}_{-\{i,j\}} \in \mathcal{Z}_{-\{i,j\}}$, respectively. In line with Generalized Linear Models (GLMs), we introduce a known scale parameter $\psi \in (0, +\infty)$ and define $Y_i^\star := Y_i / \psi$ and $\boldsymbol{Y}_{-i}^\star := \boldsymbol{Y}_{-i} / \psi$. Throughout, $\mathbb{I}(\cdot)$ is an indicator function, which is 1 if its argument is true and is 0 otherwise. We write $a_N = O(b_N)$ and $a_N = o(b_N)$ to indicate that $|a_N/b_N|$ remains bounded and $\lim_{N \to \infty} |a_N/b_N| = 0$, respectively.

Following the bulk of the literature on regression models, we condition on predictors $\boldsymbol{X} = \boldsymbol{x}$. To construct joint probability models for dependent responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$, we introduce a family of probability measures $\{\mathbb{P}_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ dominated by a σ-finite measure ν, with densities of the form

$$
\begin{aligned}
f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{z} \mid \boldsymbol{x}) \;=\; & \frac{1}{\varphi(\boldsymbol{\theta})} \left[\prod_{i=1}^N a_{\mathcal{Y}}(y_i) \, \exp\left(\boldsymbol{\theta}_g^\top \, g_i(\boldsymbol{x}_i, y_i^\star)\right) \right] \\
& \times \left[\prod_{i=1}^{N-1} \prod_{j=i+1}^N a_{\mathcal{Z}}(z_{i,j}) \, \exp\left(\boldsymbol{\theta}_h^\top \, h_{i,j}(\boldsymbol{x}, y_i^\star, y_j^\star, \boldsymbol{z})\right) \right] :
\end{aligned}
\tag{1}
$$

- $a_{\mathcal{Y}} : \mathcal{Y}_i \mapsto [0, +\infty)$ and $a_{\mathcal{Z}} : \mathcal{Z}_{i,j} \mapsto [0, +\infty)$ are known functions of responses Y_i of units $i \in \mathcal{P}_i$ and connections $Z_{i,j}$ of pairs of units $\{i, j\} \subset \mathcal{P}_N$;

- $g_i : \mathcal{X}_i \times \mathcal{Y}_i \mapsto \mathbb{R}^q$ are known functions describing the relationship of predictors $\boldsymbol{x}_i$ and responses Y_i of units $i \in \mathcal{P}_N$, which can depend on ψ;

- $h_{i,j} : \mathcal{X} \times \mathcal{Y}_i \times \mathcal{Y}_j \times \mathcal{Z} \mapsto \mathbb{R}^r$ are known functions specifying how the responses and connections of pairs of units $\{i, j\} \subset \mathcal{P}_N$ depend on the predictors, responses, and connections to other units, which can depend on ψ;

- $\boldsymbol{\theta} := (\boldsymbol{\theta}_g, \boldsymbol{\theta}_h) \in \boldsymbol{\Theta}$ is a parameter vector of dimension $p := q + r$, where $\boldsymbol{\Theta} := \{\boldsymbol{\theta} \in \mathbb{R}^p : \varphi(\boldsymbol{\theta}) < \infty\}$ and $\varphi : \boldsymbol{\Theta} \mapsto (0, +\infty]$ ensures that $\int_{\mathcal{Y} \times \mathcal{Z}} f_{\boldsymbol{\theta}}(\boldsymbol{y}, \boldsymbol{z} \,|\, \boldsymbol{x}) \, \mathrm{d}\, \nu(\boldsymbol{y}, \boldsymbol{z}) = 1$, with the dependence of φ on $\boldsymbol{x}$ suppressed;

- ν is a σ-finite product measure of the form

$$\nu(\boldsymbol{y}, \boldsymbol{z}) := \left[\prod_{i=1}^{N} \nu_{\mathcal{Y}}(y_i) \right] \left[\prod_{i=1}^{N-1} \prod_{j=i+1}^{N} \nu_{\mathcal{Z}}(z_{i,j}) \right],$$

where $\nu_{\mathcal{Y}}$ and $\nu_{\mathcal{Z}}$ are σ-finite measures that depend on the support sets of responses Y_i and connections $Z_{i,j}$ (e.g., Lebesgue or counting measure).

Remark: Importance of Additional Structure. To respect real-world constraints and facilitate theoretical guarantees, joint probability models for dependent responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ should leverage additional structure, e.g., local dependence structure. For one thing, units in large populations may not be aware of most other units in the population, so it is not credible that the responses and connections of units depend on the responses and connections of all other units in the population. In addition, models permitting strong dependence among the responses and connections of all units in the population may suffer from model near-degeneracy (Handcock, 2003; Schweinberger, 2011; Chatterjee and Diaconis, 2013). By contrast, Stewart and Schweinberger (2025) demonstrate that leveraging additional structure to control dependence can lead to theoretical guarantees. Motivated by these considerations, we assume that each unit $i \in \mathcal{P}_N$ has a known set of neighbors $\mathcal{N}_i \subset \mathcal{P}_N$, which includes i and is independent of connections $\boldsymbol{Z}$, and that the dependence among responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ is local in the sense that it is limited to overlapping neighborhoods. We provide examples of joint probability models for $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ with local dependence in Sections 2.2 and 6.1.

Remark: Fixed versus Random Design. In line with the bulk of the literature on regression models, we consider a fixed design: We view predictors $\boldsymbol{X}$ and neighbor-

hoods $\mathcal{N}_1, \ldots, \mathcal{N}_N$ as exogenous and known, and we do not make assumptions about the mechanism generating them. Thus, Equation (1) specifies the joint probability density function of responses and connections $(\boldsymbol{Y}, \boldsymbol{Z})$ conditional on predictors $\boldsymbol{X} = \boldsymbol{x}$ and neighborhoods $\mathcal{N}_1, \ldots, \mathcal{N}_N$. If $\boldsymbol{X}$ and $\mathcal{N}_1, \ldots, \mathcal{N}_N$ were random, the conditional model for $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}, \mathcal{N}_1, \ldots, \mathcal{N}_N$ could be combined with marginal models for $\boldsymbol{X}$ and $\mathcal{N}_1, \ldots, \mathcal{N}_N$. In the social media application in Section 6, the neighborhoods are fixed and known: The neighborhoods of users are the sets of followees, because users choose whom to follow and hence who can influence them, and these choices are observed. If the neighborhoods were unobserved, one could view them as unobserved constants (if neighborhoods were fixed) or unobserved variables (if neighborhoods were random) and learn them from attributes $(\boldsymbol{X}, \boldsymbol{Y})$ or connections $\boldsymbol{Z}$. The problem of how to learn neighborhoods is an open problem and constitutes a promising avenue for future research.

2.1 GLM Representations

The proposed joint probability models of $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ can be viewed as generalizations of Generalized Linear Models (GLMs) (Efron, 2022). GLMs form a well-known, interpretable, and widely applicable statistical framework for univariate responses $Y_i \in \mathcal{Y}_i$ given predictors $\boldsymbol{x}_i \in \mathbb{R}^d$ ($d \geq 1$), including logistic regression ($Y_i \in \{0, 1\}$), Poisson regression ($Y_i \in \{0, 1, \ldots\}$), and linear regression ($Y_i \in \mathbb{R}$). GLMs are characterized by two properties:

1. *Conditional mean:* The conditional mean $\mu_i(\eta_i) := \mathbb{E}_{\eta_i}(Y_i \mid \boldsymbol{x}_i)$ of response $Y_i \in \mathcal{Y}_i$, conditional on predictors $\boldsymbol{x}_i \in \mathbb{R}^d$ with weights $\boldsymbol{\beta} \in \mathbb{R}^d$, is a (possibly nonlinear) function of a linear predictor $\eta_i := \boldsymbol{\beta}^\top \boldsymbol{x}_i$.

2. *Conditional distribution:* The conditional distribution of response Y_i is an exponential family distribution with a known scale parameter $\psi \in (0, +\infty)$, which admits a

density with respect to a σ-finite measure ν_y of the form

$$f_{\eta_i}(y_i \mid \boldsymbol{x}_i) \;\; := \;\; a_y(y_i)\,\exp\left(\frac{\eta_i\,y_i - b_i(\eta_i)}{\psi}\right),$$

with cumulant-generating function

$$b_i(\eta_i) \;\; := \;\; \psi\,\log\int_{y_i} a_y(y)\,\exp\left(\frac{\eta_i\,y}{\psi}\right)\mathrm{d}\,\nu_y(y).$$

The conditional mean $\mu_i(\eta_i)$ can be obtained by differentiating $b_i(\eta_i)$: $\mu_i(\eta_i) = \nabla_{\eta_i} b_i(\eta_i)$ (Corollary 2.3, Brown, 1986, pp. 35–36).

The relationship to GLMs facilitates the interpretation and dissemination of results. The following proposition clarifies the relationship to GLMs.

Proposition 1: GLM Representation of Conditionals. *Consider any pair of units $\{i, j\} \subset \mathcal{P}_N$ $(i < j)$ and assume that g_i and $h_{i,j}$ are affine functions of $y_i^\star$ for any given $(\boldsymbol{x},\,\boldsymbol{y}_{-i},\,\boldsymbol{z}) \in \mathcal{X} \times \mathcal{Y}_{-i} \times \mathcal{Z}$, in the sense that there exist known functions $g_{i,0} : \mathcal{X}_i \mapsto \mathbb{R}^q$, $g_{i,1} : \mathcal{X}_i \mapsto \mathbb{R}^q$, $h_{i,j,0} : \mathcal{X} \times \mathcal{Y}_j \times \mathcal{Z} \mapsto \mathbb{R}^r$, and $h_{i,j,1} : \mathcal{X} \times \mathcal{Y}_j \times \mathcal{Z} \mapsto \mathbb{R}^r$ such that*

$$g_i(\boldsymbol{x}_i,\, y_i^\star) \;\; := \;\; g_{i,0}(\boldsymbol{x}_i) + g_{i,1}(\boldsymbol{x}_i)\, y_i^\star$$

$$h_{i,j}(\boldsymbol{x},\, y_i^\star,\, y_j^\star,\, \boldsymbol{z}) \;\; := \;\; h_{i,j,0}(\boldsymbol{x},\, y_j^\star,\, \boldsymbol{z}) + h_{i,j,1}(\boldsymbol{x},\, y_j^\star,\, \boldsymbol{z})\, y_i^\star.$$

Then the conditional distribution of response $Y_i \mid (\boldsymbol{X},\, \boldsymbol{Y}_{-i},\, \boldsymbol{Z}) = (\boldsymbol{x},\, \boldsymbol{y}_{-i},\, \boldsymbol{z})$ by unit i can be represented by a GLM with linear predictor

$$\eta_i(\boldsymbol{\theta};\, \boldsymbol{x},\, \boldsymbol{y}_{-i}^\star,\, \boldsymbol{z}) \;\; := \;\; \boldsymbol{\theta}^\top\left(g_{i,1}(\boldsymbol{x}_i),\, \sum_{j\,\in\,\mathcal{P}_N\backslash\{i\}} h_{i,j,1}(\boldsymbol{x},\, y_j^\star,\, \boldsymbol{z})\right)$$

and cumulant-generating function

$$b_i(\eta_i(\boldsymbol{\theta};\, \boldsymbol{x},\, \boldsymbol{y}_{-i}^\star,\, \boldsymbol{z})) \;\; := \;\; \psi\,\log\int_{y_i} a_y(y)\,\exp\left(\frac{\eta_i(\boldsymbol{\theta};\, \boldsymbol{x},\, \boldsymbol{y}_{-i}^\star,\, \boldsymbol{z})\, y}{\psi}\right) d\nu_y(y).$$

To ease the notation, we henceforth write η_i instead of $\eta_i(\boldsymbol{\theta};\, \boldsymbol{x},\, \boldsymbol{y}_{-i}^\star,\, \boldsymbol{z})$.

Proposition 1 supplies a recipe for representing the conditional distribution of responses $Y_i \mid (\boldsymbol{X},\, \boldsymbol{Y}_{-i},\, \boldsymbol{Z}) = (\boldsymbol{x},\, \boldsymbol{y}_{-i},\, \boldsymbol{z})$ by a GLM:

1. *Conditional distribution:* The conditional distribution of response Y_i is an exponential family distribution, which can be represented by a GLM with conditional mean $\mu_i(\eta_i)$, linear predictor η_i, and scale parameter ψ.

2. *Conditional mean:* The conditional mean $\mu_i(\eta_i) := \mathbb{E}_{\eta_i}(Y_i \mid \boldsymbol{x},\, \boldsymbol{y}_{-i},\, \boldsymbol{z})$ can be obtained by differentiating $b_i(\eta_i)$: $\mu_i(\eta_i) = \nabla_{\eta_i} b_i(\eta_i)$. Since the map $\eta_i \mapsto \mu_i$ is one-to-one and invertible (Theorem 3.6, Brown, 1986, p. 74), η_i can be obtained by inverting $\mu_i(\eta_i)$.

Thus, the proposed regression framework for dependent responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ inherits the GLM advantages of being interpretable and widely applicable, without assuming that responses or connections are independent. As a result, the proposed regression framework can be viewed as a generalization of GLMs.

2.2 Example: Model Specification

We showcase how a joint probability model for dependent responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ with local dependence can be constructed, leveraging additional structure in the form of overlapping neighborhoods $\mathcal{N}_1, \ldots, \mathcal{N}_N$ to control the dependence among responses and connections in small and large populations.

We focus on units $i \in \mathcal{P}_N$ with binary, count-, or real-valued predictors $x_i \in \mathcal{X}_i$ and responses $Y_i \in \mathcal{Y}_i$ and binary connections $Z_{i,j} \in \{0, 1\}$. Starting with g_i, we capture the main effect of $Y_i^\star$ and the interaction effect of x_i and $Y_i^\star$ by specifying g_i as follows:

$$\boldsymbol{\theta}_g := \begin{pmatrix} \alpha_y \\ \beta_{x,y} \end{pmatrix} \in \mathbb{R}^2, \quad g_i := \begin{pmatrix} y_i^\star \\ x_i\, y_i^\star \end{pmatrix} \in \mathbb{R}^2. \tag{2}$$

Turning to $h_{i,j}$, we define neighborhood-related terms

$$c_{i,j} := \mathbb{1}(\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset), \quad d_{i,j}(\boldsymbol{z}) := \mathbb{1}(\exists\, k \in \mathcal{N}_i \cap \mathcal{N}_j : z_{i,k} = z_{k,j} = 1). \tag{3}$$

To capture unobserved heterogeneity in the propensities of units to form connections, we introduce the N-vector $\boldsymbol{\alpha}_z := (\alpha_{z,1}, \ldots, \alpha_{z,N}) \in \mathbb{R}^N$. In addition, we penalize connections among units i and j with non-overlapping neighborhoods and capture transitive closure along with treatment and outcome spillover by specifying $h_{i,j}$ as follows:

$$
\boldsymbol{\theta}_h := \begin{pmatrix} \boldsymbol{\alpha}_z \\ \lambda \\ \gamma_{z,z} \\ \gamma_{x,y,z} \\ \gamma_{y,y,z} \end{pmatrix} \in \mathbb{R}^{N+4}, \quad h_{i,j} := \begin{pmatrix} \boldsymbol{e}_{i,j}\, z_{i,j} \\ -(1 - c_{i,j})\, z_{i,j} \log N \\ d_{i,j}(\boldsymbol{z})\, z_{i,j} \\ c_{i,j}\, (x_i\, y_j^{\star} + x_j\, y_i^{\star})\, z_{i,j} \\ c_{i,j}\, y_i^{\star}\, y_j^{\star}\, z_{i,j} \end{pmatrix} \in \mathbb{R}^{N+4}, \tag{4}
$$

where $\boldsymbol{e}_{i,j}$ denotes the N-vector whose whose ith and jth coordinates are 1 and whose other coordinates are all 0. The parameters $\alpha_{z,1}, \ldots, \alpha_{z,N}$ can be interpreted as the propensities of units $1, \ldots, N$ to form connections; $\lambda > 0$ discourages connections among units with non-overlapping neighborhoods; $\gamma_{z,z}$ quantifies the tendency towards transitive closure among connections; and $\gamma_{x,y,z}$ and $\gamma_{y,y,z}$ capture treatment and outcome spillover, respectively. Sections 2.2.1 and 2.2.2 demonstrate that the interpretation of these effects is facilitated by the fact that the conditional distributions of Y_i and $Z_{i,j}$ can be represented by GLMs.

2.2.1   GLM Representation of Responses Y_i

To interpret the model specified by Equations (2) and (4), we take advantage of the fact that the conditional distribution of response $Y_i \mid (\boldsymbol{X}, \boldsymbol{Y}_{-i}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z})$ by unit i can be represented by a GLM with linear predictor

$$
\eta_i \;=\; \alpha_y + \beta_{x,y}\, x_i + \gamma_{x,y,z} \sum_{j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset} x_j\, z_{i,j} + \gamma_{y,y,z} \sum_{j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset} y_j^{\star}\, z_{i,j}. \tag{5}
$$

Figure 1 depicts the predictors, responses, and connections that affect the conditional distribution of response Y_i. We provide three specific examples, depending on the support
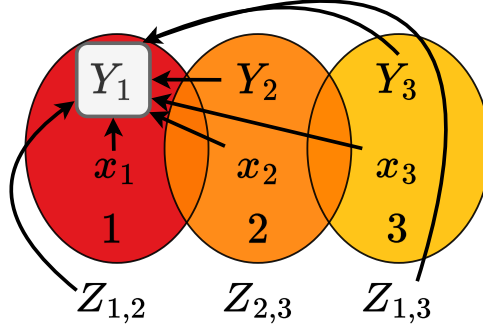
12

Figure 1: Given $N = 3$ units $1, 2, 3$ with neighborhoods $\mathcal{N}_1 := \{1, 2\}$, $\mathcal{N}_2 := \{1, 2, 3\}$, and $\mathcal{N}_3 := \{2, 3\}$, the arrows indicate which predictors, responses, and connections can affect the response Y_1 of unit 1 according to the model specified by Equations (2) and (4).

set of response Y_i.

Example 1: Real-valued Responses $Y_i \in \mathbb{R}$. Let $\psi \in (0, +\infty)$ and

$$a_y(y_i) \quad := \quad \frac{1}{\sqrt{2\pi\psi}} \, \exp\left(-\frac{y_i^2}{2\psi}\right) \, \mathbb{I}(y_i \in \mathbb{R}).$$

1. *Conditional distribution:* The conditional distribution of response Y_i is $N(\mu_i(\eta_i), \psi)$.

2. *Conditional mean:* The conditional mean $\mu_i(\eta_i)$ can be obtained by differentiating $b_i(\eta_i) = \eta_i^2 / 2$ with respect to η_i, giving $\mu_i(\eta_i) = \eta_i$:

$$\mu_i(\eta_i) \quad = \quad \alpha_y + \beta_{x,y} \, x_i + \gamma_{x,y,z} \sum_{j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset} x_j \, z_{i,j} + \gamma_{y,y,z} \sum_{j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset} y_j^\star \, z_{i,j}.$$

Under certain restrictions on $\gamma_{y,y,z}$, the conditional distribution of $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{z})$ is N-variate Gaussian. The restrictions on $\gamma_{y,y,z}$ depend on the neighborhoods $\mathcal{N}_i$ and $\mathcal{N}_j$ and connections $Z_{i,j}$ of pairs of units $\{i, j\} \subset \mathcal{P}_N$; see Proposition 2 in Section A of the Supplementary Materials.

Example 2: Count-valued Responses $Y_i \in \{0, 1, \dots\}$. Let $\psi := 1$ and

$$a_y(y_i) \quad := \quad \frac{1}{y_i!} \, \mathbb{I}(y_i \in \{0, 1, \dots\}).$$

13

1. *Conditional distribution:* The conditional distribution of response Y_i is $\text{Poisson}(\mu_i(\eta_i))$.

2. *Conditional mean:* The conditional mean $\mu_i(\eta_i)$ can be obtained by differentiating $b_i(\eta_i) = \exp(\eta_i)$ with respect to η_i, giving $\mu_i(\eta_i) = \exp(\eta_i)$.

Example 3: Binary Responses $Y_i \in \{0,\ 1\}$. Let $\psi := 1$ and $a_y(y_i) := \mathbb{I}(y_i \in \{0,1\})$.

1. *Conditional distribution:* The conditional distribution of response Y_i is $\text{Bernoulli}(\mu_i(\eta_i))$.

2. *Conditional mean:* The conditional mean $\mu_i(\eta_i)$ can be obtained by differentiating $b_i(\eta_i) = \log(1 + \exp(\eta_i))$ with respect to η_i, giving $\mu_i(\eta_i) = \text{logit}^{-1}(\eta_i)$.

Interpretation of Examples. According to Equations (2) and (4), regardless of the conditional distribution of response Y_i, α_y can be viewed as an intercept, while $\beta_{x,y}$ captures the relationship between predictor x_i and response Y_i. The parameters $\gamma_{x,y,z}$ and $\gamma_{y,y,z}$ capture two distinct spillover effects:

- *Treatment spillover:* $\gamma_{x,y,z} \neq 0$ allows the outcome Y_i of unit i to be affected by the treatments x_j of its neighbors $j \in \mathcal{N}_i$ and non-neighbors $j \notin \mathcal{N}_i$, provided $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and i and j are connected (see Figure 1).

- *Outcome spillover:* $\gamma_{y,y,z} \neq 0$ allows the outcome Y_i of unit i to be affected by the outcomes y_j of its neighbors $j \in \mathcal{N}_i$ and non-neighbors $j \notin \mathcal{N}_i$, provided $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and i and j are connected (see Figure 1).

The proposed regression framework can be used for causal inference under network interference, which studies treatment spillover. That said, the framework is considerably broader, because it permits outcome spillover, and spillover need not be studied in a causal setting.

2.2.2 GLM Representation of Connections $Z_{i,j}$

The conditional mean $\mu_{i,j}(\eta_{i,j}) := \mathbb{E}_{\eta_{i,j}}(Z_{i,j} \mid \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}}) = \mathrm{logit}^{-1}(\eta_{i,j})$ of connection $Z_{i,j} \in \{0, 1\}$ depends on the linear predictor

$$\eta_{i,j} = \alpha_{z,i} + \alpha_{z,j} - (1 - c_{i,j})\,\lambda \log N + c_{i,j}\left[\gamma_{z,z}\,\Delta_{i,j}(\boldsymbol{z}) + \gamma_{x,y,z}\,(x_i\,y_j^\star + x_j\,y_i^\star) + \gamma_{y,y,z}\,y_i^\star\,y_j^\star\right],$$

where $\Delta_{i,j} : \mathcal{Z} \mapsto \mathbb{R}$ is the change in $\sum_{a<b}^N d_{a,b}(\boldsymbol{z})$ due to transforming $z_{i,j}$ from 0 to 1. The logistic regression representation of $Z_{i,j} \mid (\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}_{-\{i,j\}}) = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}})$ facilitates interpretation: e.g., $\alpha_{z,i}$ captures heterogeneity among units i in forming connections. If $\lambda > 0$, the sparsity-inducing term $-(1 - c_{i,j})\,\lambda \log N$ penalizes connections between pairs of units with non-overlapping neighborhoods, where the $\log N$-term can be motivated in the special case of Bernoulli random graphs (Krivitsky et al., 2023): If $Z_{i,j} \overset{\mathrm{iid}}{\sim} \mathrm{Bernoulli}(\pi)$ and the expected degrees $\mathbb{E}\sum_{j=1}^N Z_{i,j}$ are bounded, then $\pi = O(1/N)$ and $\mathrm{logit}(\pi) = O(\log N)$. In addition, the model captures three forms of dependencies. First, the model encourages i and j to be connected when i and j are both connected to some $k \in \mathcal{N}_i \cap \mathcal{N}_j$, provided $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and $\gamma_{z,z} > 0$. Second, the model encourages i and j to be connected when $x_i\,y_j^\star > 0$ or $x_j\,y_i^\star > 0$, provided $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and $\gamma_{x,y,z} > 0$. Third, the model encourages i and j to be connected when $y_i^\star\,y_j^\star > 0$, provided $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and $\gamma_{y,y,z} > 0$.

3 Scalable Statistical Computing

To learn the regression framework from a single observation $(\boldsymbol{y}, \boldsymbol{z})$ of dependent responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$, we develop scalable methods based on convex optimization of pseudo-likelihoods using minorization-maximization methods.

3.1 Pseudo-Loglikelihood

Let

$$\ell(\boldsymbol{\theta};\, \boldsymbol{y},\, \boldsymbol{z}) \;\; \coloneqq \;\; \sum_{i=1}^{N} \ell_i(\boldsymbol{\theta};\, \boldsymbol{y},\, \boldsymbol{z}) + \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \ell_{i,j}(\boldsymbol{\theta};\, \boldsymbol{y},\, \boldsymbol{z}), \tag{6}$$

where the dependence on predictors $\boldsymbol{x} \in \mathcal{X}$ is suppressed and ℓ_i and $\ell_{i,j}$ are defined by

$$\ell_i(\boldsymbol{\theta};\, \boldsymbol{y},\, \boldsymbol{z}) \;\; \coloneqq \;\; \log f_{\boldsymbol{\theta}}(y_i \mid \boldsymbol{y}_{-i},\, \boldsymbol{z}) \;\; \text{and} \;\; \ell_{i,j}(\boldsymbol{\theta};\, \boldsymbol{y},\, \boldsymbol{z}) \;\; \coloneqq \;\; \log f_{\boldsymbol{\theta}}(z_{i,j} \mid \boldsymbol{y},\, \boldsymbol{z}_{-\{i,j\}}).$$

The pseudo-loglikelihood ℓ is based on full conditional densities of responses Y_i and connections $Z_{i,j}$ and is hence tractable. In addition, ℓ is a sum of exponential family loglikelihood functions ℓ_i and $\ell_{i,j}$, each of which is concave and twice differentiable on the convex set $\boldsymbol{\Theta}$ (Brown, 1986, Theorem 1.13, p. 19 and Lemma 5.3, p. 146), proving Lemma 1:

Lemma 1: Convexity and Smoothness. *The set $\boldsymbol{\Theta}$ is convex and the pseudo-loglikelihood function $\ell : \boldsymbol{\Theta} \mapsto \mathbb{R}$, considered as a function of $\boldsymbol{\theta}$ for fixed $(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{Y} \times \mathcal{Z}$, is twice differentiable with a negative semidefinite Hessian matrix on $\boldsymbol{\Theta}$.*

In light of the tractability and concavity of ℓ, it makes sense to base statistical learning on pseudo-likelihood estimators of the form

$$\widehat{\boldsymbol{\Theta}}(\delta_N) \;\; \coloneqq \;\; \left\{ \boldsymbol{\theta} \in \boldsymbol{\Theta} : \; \| \nabla_{\boldsymbol{\theta}} \, \ell(\boldsymbol{\theta};\, \boldsymbol{y},\, \boldsymbol{z}) \|_{\infty} \;\; \leq \;\; \delta_N \right\}, \tag{7}$$

where $\nabla_{\boldsymbol{\theta}}$ denotes the gradient with respect to $\boldsymbol{\theta}$ while $\|\boldsymbol{v}\|_{\infty} \coloneqq \max_{1 \leq k \leq p} |v_k|$ denotes the ℓ_{∞}-norm of vectors $\boldsymbol{v} \in \mathbb{R}^p$. The quantity $\delta_N \in [0, +\infty)$ can be viewed as a convergence criterion of a root-finding algorithm and can depend on N. The set $\widehat{\boldsymbol{\Theta}}(\delta_N)$ consists of maximizers of ℓ when $\delta_N = 0$, and maximizers and near-maximizers when $\delta_N > 0$.

3.2 Minorization-Maximization (MM)

While pseudo-likelihood estimators $\widehat{\boldsymbol{\theta}} \in \widehat{\boldsymbol{\Theta}}(\delta_N)$ can be obtained by standard root-finding algorithms, inverting the $p \times p$ negative Hessian of ℓ at each iteration is time-consuming,

because inversions require $O(p^3)$ operations and p can increase with N. We thus divide the task of estimating p parameters into two subtasks using MM methods (Hunter and Lange, 2004). In the example model specified by Equations (2) and (4) for binary, count-, or real-valued predictors and responses (X_i, Y_i) and binary connections $Z_{i,j}$, we partition $\boldsymbol{\theta} \in \mathbb{R}^{N+6}$ into N nuisance parameters, $\boldsymbol{\theta}_1 := (\alpha_{\mathsf{z},1}, \ldots, \alpha_{\mathsf{z},N}) \in \mathbb{R}^N$, and 6 parameters of primary interest, $\boldsymbol{\theta}_2 := (\lambda, \alpha_{\mathsf{y}}, \beta_{\mathsf{x},\mathsf{y}}, \gamma_{\mathsf{z},\mathsf{z}}, \gamma_{\mathsf{x},\mathsf{y},\mathsf{z}}, \gamma_{\mathsf{y},\mathsf{y},\mathsf{z}}) \in \mathbb{R}^6$. We then partition the negative Hessian of ℓ accordingly:

$$
-\nabla_{\boldsymbol{\theta}}^2 \, \ell(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) \;\; := \;\; \begin{pmatrix} \mathbf{A}(\boldsymbol{\theta}) & \mathbf{B}(\boldsymbol{\theta}) \\[4pt] \mathbf{B}(\boldsymbol{\theta})^\top & \mathbf{C}(\boldsymbol{\theta}) \end{pmatrix}, \tag{8}
$$

where $\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{R}^{N \times N}$, $\mathbf{B}(\boldsymbol{\theta}) \in \mathbb{R}^{N \times 6}$, and $\mathbf{C}(\boldsymbol{\theta}) \in \mathbb{R}^{6 \times 6}$. We suppress the dependence of $\mathbf{A}(\boldsymbol{\theta})$, $\mathbf{B}(\boldsymbol{\theta})$, and $\mathbf{C}(\boldsymbol{\theta})$ on $(\boldsymbol{y}, \boldsymbol{z})$ and henceforth write $\ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2; \boldsymbol{y}, \boldsymbol{z})$ instead of $\ell(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})$.

Iteration $t + 1$ then consists of two steps:

Step 1: Find $\boldsymbol{\theta}_1^{(t+1)}$ satisfying $\ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z}) \geq \ell(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z})$.

Step 2: Find $\boldsymbol{\theta}_2^{(t+1)}$ satisfying $\ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t+1)}; \boldsymbol{y}, \boldsymbol{z}) \geq \ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z})$.

In Step 1, it is inconvenient to invert the high-dimensional $N \times N$ matrix

$$
\boldsymbol{A}(\boldsymbol{\theta}^{(t)}) \;\; := \;\; -\sum_{i<j}^{N} \nabla_{\boldsymbol{\theta}_1}^2 \, \ell_{i,j}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z}) \Big|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(t)}} \;\; = \;\; \sum_{i<j}^{N} \pi_{i,j}^{(t)} \, (1 - \pi_{i,j}^{(t)}) \, \boldsymbol{e}_{i,j} \, \boldsymbol{e}_{i,j}^\top, \tag{9}
$$

where $\pi_{i,j}^{(t)} := \mathbb{P}_{\boldsymbol{\theta}^{(t)}}(Z_{i,j} = 1 \mid \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}})$. We thus increase ℓ by maximizing a minorizer of ℓ, replacing $\boldsymbol{A}(\boldsymbol{\theta}^{(t)})$ by a constant matrix $\boldsymbol{A}^\star$ that only needs to be inverted once.

Lemma 2: Minorizer. *Define*

$$
\boldsymbol{A}^\star \;\; := \;\; \frac{1}{4} \sum_{i<j}^{N} \boldsymbol{e}_{i,j} \, \boldsymbol{e}_{i,j}^\top \;\; = \;\; \frac{1}{4} \left[(N - 2)\, \boldsymbol{I} + \mathbf{1} \mathbf{1}^\top \right] \;\; = \;\; \left[\frac{4}{N-2} \left(\boldsymbol{I} - \frac{1}{2\,N - 2} \mathbf{1} \mathbf{1}^\top \right) \right]^{-1},
$$

where $\boldsymbol{I}$ is the $N \times N$ identity matrix and $\mathbf{1}$ is the N-vector of ones. Then the function

$$
\begin{aligned}
m(\boldsymbol{\theta}_1; \boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}, \boldsymbol{y}, \boldsymbol{z}) \ := \ & \ell(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z}) \\
& + \ \left(\nabla_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z}) \Big|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(t)}} \right)^{\top} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_1^{(t)}) \\
& + \ \frac{1}{2} (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_1^{(t)})^{\top} (-\boldsymbol{A}^{\star}) (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_1^{(t)})
\end{aligned}
$$

is a minorizer of ℓ at $\boldsymbol{\theta}_1^{(t)}$ for fixed $\boldsymbol{\theta}_2^{(t)}$, in the sense that

$$
m(\boldsymbol{\theta}_1; \boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}, \boldsymbol{y}, \boldsymbol{z}) \ \leq \ \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z}) \ \text{ for all } \ \boldsymbol{\theta}_1 \in \mathbb{R}^N,
$$

$$
m(\boldsymbol{\theta}_1^{(t)}; \boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}, \boldsymbol{y}, \boldsymbol{z}) \ = \ \ell(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z}).
$$

Lemma 2 is proved in Section B of the Supplementary Materials. Step 1 may be implemented by an MM algorithm, as the closed-form maximizer of the minorizer m is

$$
\boldsymbol{\theta}_1^{(t+1)} \ := \ \boldsymbol{\theta}_1^{(t)} + (\mathbf{A}^{\star})^{-1} \left(\nabla_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z}) \Big|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(t)}} \right). \tag{10}
$$

We accelerate the MM step in Equation (10) with quasi-Newton methods. Details can be found in Section E of the Supplementary Materials. Compared to a Newton-Raphson algorithm, the accelerated MM-step reduces the per-iteration computational complexity from $O(N^3)$ to $O(N^2)$. Step 2 updates the low-dimensional parameter vector of interest $\boldsymbol{\theta}_2^{(t+1)} \in \mathbb{R}^6$ given the high-dimensional nuisance parameter vector $\boldsymbol{\theta}_1^{(t+1)} \in \mathbb{R}^N$ by a Newton-Raphson step. The concavity of ℓ, established in Lemma 1, guarantees that

$$
\ell(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z}) \ \leq \ \ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t)}; \boldsymbol{y}, \boldsymbol{z}) \ \leq \ \ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t+1)}; \boldsymbol{y}, \boldsymbol{z}).
$$

3.3 Quantifying Uncertainty

The uncertainty about the maximum pseudo-likelihood estimator $\widehat{\boldsymbol{\theta}}$ of the data-generating parameter vector $\boldsymbol{\theta}^{\star}$ can be quantified based on the covariance matrix of the sampling distribution of $\widehat{\boldsymbol{\theta}}$, which we derive as follows: The mean-value theorem for vector-valued

functions (Ortega and Rheinboldt, 2000, Equations (2) and (3), pp. 68–69) implies that there exist real numbers $t_1, \ldots, t_p \in (0, 1)$ such that

$$\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} - \nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\star}} = \boldsymbol{H}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}^{\star}; \boldsymbol{y}, \boldsymbol{z})\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\star}), \qquad (11)$$

where

$$\boldsymbol{H}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}^{\star}; \boldsymbol{y}, \boldsymbol{z}) \;\; := \;\; \begin{pmatrix} g_1'(\boldsymbol{\theta}^{\star} + t_1\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\star}); \boldsymbol{y}, \boldsymbol{z}) \\ \vdots \\ g_p'(\boldsymbol{\theta}^{\star} + t_p\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\star}); \boldsymbol{y}, \boldsymbol{z}) \end{pmatrix}.$$

Here, $g_k(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})$ is the kth coordinate of $\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})$ and $g_k'(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})$ is the row vector of partial derivatives of $g_k(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})$ with respect to $\boldsymbol{\theta}$ ($k = 1, \ldots, p$). Leveraging Equation (11) along with $\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = \boldsymbol{0}$ gives the exact covariance matrix $\widehat{\boldsymbol{\theta}}$:

$$\mathbb{V}_{\boldsymbol{\theta}^{\star}}(\widehat{\boldsymbol{\theta}}) \;\; = \;\; \mathbb{V}_{\boldsymbol{\theta}^{\star}} \left[-\boldsymbol{H}(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}^{\star}; \boldsymbol{Y}, \boldsymbol{Z})^{-1} \, \nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{Z})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{\star}} \right].$$

If N is large, $\boldsymbol{\theta}^{\star}$ can be replaced by $\widehat{\boldsymbol{\theta}}$, because $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^{\star}\|_{\infty}$ is small with high probability according to Theorem 1 in Section 4. The resulting approximation of $\mathbb{V}_{\boldsymbol{\theta}^{\star}}(\widehat{\boldsymbol{\theta}})$ is

$$\mathbb{V}_{\widehat{\boldsymbol{\theta}}}(\widehat{\boldsymbol{\theta}}) \;\; = \;\; \mathbb{V}_{\widehat{\boldsymbol{\theta}}} \left[-\nabla_{\boldsymbol{\theta}}^2 \ell(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{Z})\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}^{-1} \, \nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{Z})\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \right], \qquad (12)$$

which can be estimated by simulating responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ from $\mathbb{P}_{\widehat{\boldsymbol{\theta}}}$ using Markov chain Monte Carlo methods.

Remark: Asymptotic Distribution. Establishing asymptotic normality for pseudo-likelihood estimators based on a single observation of dependent responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ in scenarios with $p \to \infty$ parameters is an open problem. Asymptotic normality results in the most closely related literature—the literature in applied probability concerned with Ising models, Gibbs measures, and Markov random fields in single-observation scenarios (e.g., Jensen and Künsch, 1994; Comets and Janzura, 1998)—assume the presence of lattice structure and a fixed number of parameters p, in addition to other

assumptions motivated by applications in physics. In the current setting, none of these assumptions holds, although simulation results in Section 5 suggest that the sampling distribution of $\widehat{\boldsymbol{\theta}}$ is approximately normal and that normal-based confidence intervals based on Equation (12) achieve close-to-nominal coverage probabilities.

4 Theoretical Guarantees

We establish convergence rates for pseudo-likelihood estimators $\widehat{\boldsymbol{\Theta}}(\delta_N)$ based on a single observation of dependent responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$. To cover a wide range of models for binary, count-, and real-valued responses and connections, we introduce a general theoretical framework and showcase convergence rates in a specific example.

We denote by $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$ the data-generating parameter vector and by $\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \rho) := \{\boldsymbol{\theta} \in \mathbb{R}^p : \|\boldsymbol{\theta} - \boldsymbol{\theta}^\star\|_\infty < \rho\}$ a hypercube with center $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta}$ and width $2\,\rho \in (0, +\infty)$. Let

$$\mathcal{I}(S) := \left\{ (\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{Y} \times \mathcal{Z} : -\nabla_{\boldsymbol{\theta}}^2\, \ell(\boldsymbol{\theta};\, \boldsymbol{y},\, \boldsymbol{z}) \text{ is invertible for all } \boldsymbol{\theta} \in S \right\}$$

and, for some $\epsilon^\star \in (0, +\infty)$ and $\mathcal{H} \subseteq \mathcal{I}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star))$, let

$$\Lambda_N(\boldsymbol{\theta}^\star) := \sup_{(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{H}}\ \sup_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)} \|(-\nabla_{\boldsymbol{\theta}}^2\, \ell(\boldsymbol{\theta};\, \boldsymbol{y},\, \boldsymbol{z}))^{-1}\|_\infty,$$

where $\|.\|_\infty$ is the ℓ_∞-induced matrix norm. The set $\mathcal{H}$ can be a proper subset of $\mathcal{I}(\mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star))$, provided $\mathcal{H}$ is a high probability subset of $\mathcal{Y} \times \mathcal{Z}$. The definition of $\mathcal{H}$ is motivated by the fact that characterizing the set of all $(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{Y} \times \mathcal{Z}$ for which the Hessian is invertible can be challenging, but finding a sufficient condition for invertibility is often possible.

Theorem 1: Convergence Rate. *Consider a single observation of $(\boldsymbol{Y}, \boldsymbol{Z}) \in \mathcal{Y} \times \mathcal{Z}$ generated by model (1) with parameter vector $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta} \subseteq \mathbb{R}^p$, where $\mathcal{Y} \times \mathcal{Z}$ is a finite, countably infinite, or uncountable set. Assume that there exists a sequence $\rho_1, \rho_2, \ldots \in [0, +\infty)$ satisfying $\rho_N = o(1)$ so that the events $\|\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta};\, \boldsymbol{Y}, \boldsymbol{Z})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} - \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta};\, \boldsymbol{Y}, \boldsymbol{Z})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star}\|_\infty < \delta_N$*

and $(\boldsymbol{Y}, \boldsymbol{Z}) \in \mathcal{H}$ occur with probability $1 - o(1)$, where $\delta_N \coloneqq \rho_N/(2\,\Lambda_N(\boldsymbol{\theta}^\star))$. Then there exists a positive integer N_0 such that, for all $N > N_0$, the random set $\widehat{\boldsymbol{\Theta}}(\delta_N)$ is non-empty and, with probability $1 - o(1)$, satisfies

$$\widehat{\boldsymbol{\Theta}}(\delta_N) \quad \subseteq \quad \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \, \rho_N).$$

Theorem 1 is proved in Section C of the Supplementary Materials. The requirement $\delta_N \coloneqq \rho_N/(2\,\Lambda_N(\boldsymbol{\theta}^\star))$ implies $\rho_N \propto \delta_N\,\Lambda_N(\boldsymbol{\theta}^\star)$, so the convergence rate depends on

- the strength of concentration of the gradient $\nabla_{\boldsymbol{\theta}}\,\ell(\boldsymbol{\theta};\,\boldsymbol{Y},\boldsymbol{Z})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star}$ around its expectation $\mathbb{E}\,\nabla_{\boldsymbol{\theta}}\,\ell(\boldsymbol{\theta};\,\boldsymbol{Y},\boldsymbol{Z})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star}$ via δ_N;

- the inverse negative Hessian $(-\nabla_{\boldsymbol{\theta}}^2\,\ell(\boldsymbol{\theta};\,\boldsymbol{y},\boldsymbol{z}))^{-1}$ in a neighborhood $\mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\epsilon^\star)$ of $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta}$ and a high probability subset $(\boldsymbol{y},\boldsymbol{z}) \in \mathcal{H}$ of $\mathcal{Y} \times \mathcal{Z}$ via $\Lambda_N(\boldsymbol{\theta}^\star)$.

The strength of concentration of $\nabla_{\boldsymbol{\theta}}\,\ell(\boldsymbol{\theta};\,\boldsymbol{Y},\boldsymbol{Z})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star}$ can be quantified by concentration inequalities for dependent random variables. In general, the strength of concentration depends on the sample space and the tails of the distribution, the smoothness of the functions g_i and $h_{i,j}$, and the dependence induced by model (1). To control the dependence among responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$, one can take advantage of additional structure (e.g., one or more neighborhood structures, non-overlapping or overlapping subpopulations, or a metric space in which units are embedded). For example, each unit can have one or more neighborhoods (e.g., geographical neighbors and colleagues in the workplace), and the responses and connections of the unit can be affected by any geographical neighbor and any colleague. Theoretical guarantees can be obtained as long as the neighborhoods are not too large and do not overlap too much.

Specific convergence rates depend on the model. To demonstrate, consider predictors $x_i \in \mathbb{R}$, responses $Y_i \in \{0, 1\}$, and connections $Z_{i,j} \in \{0, 1\}$ generated by a model captur-

ing heterogeneity in the propensities $\alpha_{z,1}, \ldots, \alpha_{z,N}$ of units $1, \ldots, N$ to form connections, transitive closure among connections with weight $\gamma_{z,z}$, and treatment spillover with weight $\gamma_{x,y,z}$; compare Equations (2) and (4) in Section 2.2. Since $Y_i \in \{0,1\}$ and $Z_{i,j} \in \{0,1\}$, it is reasonable to specify $a_y(y_i) := \mathbb{I}(y_i \in \{0,1\})$ and $a_z(z_{i,j}) := \mathbb{I}(z_{i,j} \in \{0,1\})$. Convergence rates can be obtained under the following conditions.

Condition 1: Predictors. *There exist finite constants $0 < c < C$ such that, for each $i \in \mathcal{P}_N$, $x_i \in [0, C]$ and there exists $j \in \mathcal{P}_N \setminus \{i\}$ such that $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and $x_j \in [c, C]$.*

Condition 2: Parameters. *The parameter space is $\boldsymbol{\Theta} = \mathbb{R}^{N+2}$ and there exists a constant $A \in (0, +\infty)$, not depending on N, such that $\|\boldsymbol{\theta}^\star\|_\infty < A$.*

Condition 3: Dependence. *The population $\mathcal{P}$ consists of overlapping subpopulations $\mathcal{A}_1, \mathcal{A}_2, \ldots$, which can be represented as vertices of a subpopulation graph $\mathcal{G}_\mathcal{A}$ with an edge connecting $\mathcal{A}_k$ and $\mathcal{A}_l$ if $\mathcal{A}_k \cap \mathcal{A}_l \neq \emptyset$ $(k < l)$. For each $\mathcal{A}_k$, the number of subpopulations at geodesic distance K in $\mathcal{G}_\mathcal{A}$ is $O(\log K)$. For each $i \in \mathcal{P}_i$, the neighborhood is*

$$\mathcal{N}_i := \{j \in \mathcal{P}_N : \text{ there exists } k \in \{1, 2, \ldots\} \text{ such that } i \in \mathcal{A}_k \text{ and } j \in \mathcal{A}_k\}.$$

There exists a constant $B \in (0, +\infty)$ such that $\max_{1 \leq i \leq N} |\mathcal{N}_i| < B$.

Condition 1 imposes restrictions on $\boldsymbol{x} \in \mathbb{R}^N$. Condition 2 requires that the data-generating parameter vector $\boldsymbol{\theta}^\star$ be contained in a compact subset of $\boldsymbol{\Theta} = \mathbb{R}^{N+2}$. The set of estimators $\widehat{\boldsymbol{\Theta}}(\delta_N)$ is not restricted by Condition 2 and consists of all $\boldsymbol{\theta} \in \mathbb{R}^{N+2}$ such that $\|\nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{Z})\|_\infty \leq \delta_N$. Condition 2 can be weakened in special cases, allowing $\|\boldsymbol{\theta}^\star\|_\infty = O(\log N)$; see Section D.3 of the Supplementary Materials. Condition 3 controls the dependence among responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ and can be weakened to $\max_{1 \leq i \leq N} |\mathcal{N}_i| = O(\log N)$, as demonstrated by Stewart and Schweinberger (2025) in the special case of connections $\boldsymbol{Z}$ (without predictors $\boldsymbol{X}$ and responses $\boldsymbol{Y}$).

Corollary 1: Example of Convergence Rate. *Consider a single observation of dependent responses and connections $(\boldsymbol{Y}, \boldsymbol{Z})$ generated by the model with parameter vector $\boldsymbol{\theta}^\star := (\alpha^\star_{Z,1}, \ldots, \alpha^\star_{Z,N}, \gamma^\star_{Z,Z}, \gamma^\star_{X,Y,Z}) \in \mathbb{R}^{N+2}$. If Conditions 1–3 hold, there exist constants $K \in (0, +\infty)$ and $0 < L \leq U < +\infty$ along with an integer $N_0 \in \{3, 4, \ldots\}$ such that, for all $N > N_0$, the quantity δ_N satisfies*

$$L \sqrt{N \log N} \quad \leq \quad \delta_N \quad \leq \quad U \sqrt{N \log N},$$

and the random set $\widehat{\boldsymbol{\Theta}}(\delta_N)$ is non-empty and satisfies

$$\widehat{\boldsymbol{\Theta}}(\delta_N) \quad \subseteq \quad \mathcal{B}_\infty \left(\boldsymbol{\theta}^\star, \ K \sqrt{\frac{\log N}{N}} \right)$$

with probability at least $1 - 6/N^2$.

Corollary 1 is proved in Section D of the Supplementary Materials. The same method of proof can be used to establish convergence rates for pseudo-likelihood estimators $\widehat{\boldsymbol{\Theta}}(\delta_N)$ based on other models for dependent responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$, provided there is additional structure to control the dependence among responses and connections.

5   Simulation Results

To evaluate the performance of pseudo-likelihood estimators $\widehat{\boldsymbol{\theta}} \in \widehat{\boldsymbol{\Theta}}(\delta_N)$ and the accompanying uncertainty quantification, we simulate data from the example model specified by Equations (2) and (4). The coordinates of the nuisance parameter vector, $\boldsymbol{\theta}^\star_1 := (\alpha^\star_{Z,1}, \ldots, \alpha^\star_{Z,N}) \in \mathbb{R}^N$, are independent Gaussian draws with mean $-3/2$ and standard deviation $3/10$. The parameter vector of primary interest, $\boldsymbol{\theta}^\star_2 := (\lambda^\star, \alpha^\star_Y, \beta^\star_{X,Y}, \gamma^\star_{Z,Z}, \gamma^\star_{X,Y,Z}, \gamma^\star_{Y,Y,Z}) \in \mathbb{R}^6$, is specified as $(1/5, -1, 3, 4/5, 1/2, -1/2)$. The sparsity parameter $\lambda^\star = 1/5$ ensures that each unit has on average approximately 30 connections, regardless of the value
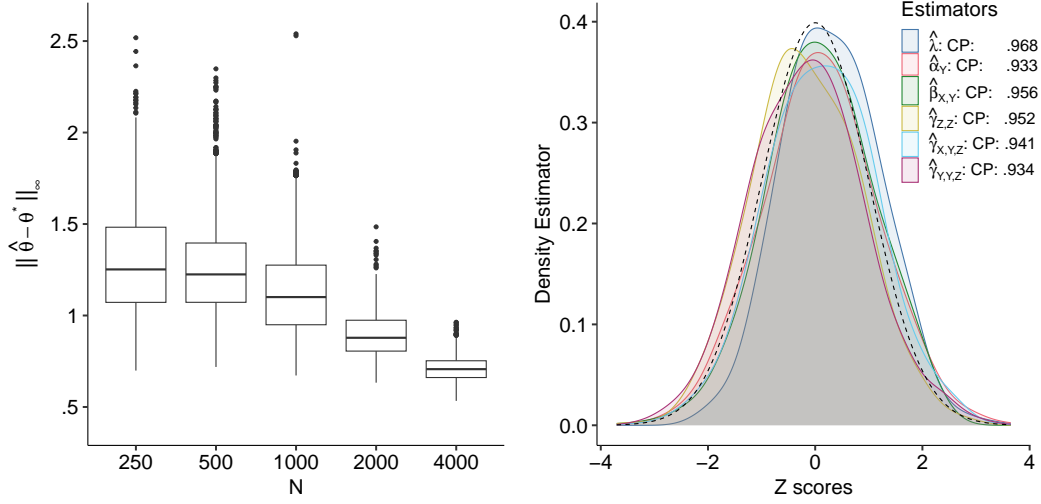
Figure 2: Simulation results based on 1,000 replications. Left: Statistical error $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty$ of maximum pseudo-likelihood estimator $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^{N+6}$ as a function of N. Right: Kernel density estimators of Z-scores $(\widehat{\lambda} - \lambda^\star)/\mathrm{S.E.}(\widehat{\lambda})$, $(\widehat{\alpha}_y - \alpha_y^\star)/\mathrm{S.E.}(\widehat{\alpha}_y)$, $(\widehat{\beta}_{x,y} - \beta_{x,y}^\star)/\mathrm{S.E.}(\widehat{\beta}_{x,y})$, $(\widehat{\gamma}_{z,z} - \gamma_{z,z}^\star)/\mathrm{S.E.}(\widehat{\gamma}_{z,z})$, $(\widehat{\gamma}_{x,y,z} - \gamma_{x,y,z}^\star)/\mathrm{S.E.}(\widehat{\gamma}_{x,y,z})$, and $(\widehat{\gamma}_{y,y,z} - \gamma_{y,y,z}^\star)/\mathrm{S.E.}(\widehat{\gamma}_{y,y,z})$ based on $N = 500$ units, where the dashed line corresponds to the standard normal density and CP is the coverage probability of interval estimators with nominal coverage probability .95.

of N. The neighborhood structure is based on $L = (N - 25)/25$ intersecting subpopulations $\mathcal{A}_1, \ldots, \mathcal{A}_L$, where $\mathcal{A}_l$ consists of the 50 units $1 + 25\,(l-1), \ldots, 25\,(l+1)$ $(l = 1, \ldots, L-1)$. For each $i \in \mathcal{P}_N$, we define the neighborhood $\mathcal{N}_i \subset \mathcal{P}_N$ to be the 50- or 75-unit union of all subpopulations $\mathcal{A}_l$ containing i.

In Figure 2, the left panel shows that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^\star\|_\infty$ decreases as N increases. The right panel depicts the empirical distributions of the standardized univariate estimators and the empirical coverage probabilities, demonstrating that the covariance estimator in Equation (12) appears accurate and normal-based inference seems reasonable. As discussed in Section 1, comparisons to other estimation approaches are infeasible due to their lack of scalability.

6 Hate Speech on X

We analyze posts of U.S. state legislators on the social media platform X in the six months preceding the insurrection at the U.S. Capitol on January 6, 2021 (Kim et al., 2022), with a view to studying how hate speech depends on the attributes of legislators and connections among them. Using Large Language Models (LLMs), we classify the contents of 109,974 posts by $N = 2{,}191$ legislators as "non-hate speech" or "hate speech," as explained in Section G of the Supplementary Materials. The response Y_i of legislator i indicates whether i released at least one post classified as hate speech. We use four covariates: $x_{i,1}$ indicates that legislator i's party affiliation is Republican, $x_{i,2}$ indicates that legislator i is female, $x_{i,3}$ indicates that legislator i is white, and $x_{i,4}$ is the state legislature that legislator i is a member of (e.g., New York). The directed connections $Z_{i,j}$ are based on the mentions and reposts exchanged between January 6, 2020 and January 6, 2021: $Z_{i,j} = 1$ if legislator i mentioned or reposted posts by legislator j in a post. To construct the neighborhoods $\mathcal{N}_i$ of legislators i, we exploit the fact that users of X choose whom to follow and that these choices are known, so $\mathcal{N}_i$ is defined as the union of $\{i\}$ and the set of users followed by i.

6.1 Model Specification

To accommodate binary responses $Y_i \in \{0, 1\}$ and connections $Z_{i,j} \in \{0, 1\}$ that are directed, i.e., $Z_{i,j}$ may not be equal to $Z_{j,i}$, we consider a model of the form

$$
\begin{aligned}
f_{\boldsymbol{\theta}}(\boldsymbol{y},\, \boldsymbol{z} \mid \boldsymbol{x}) \;\; \propto \;\; & \left[\prod_{i=1}^{N} a_y(y_i) \, \exp(\boldsymbol{\theta}_g^{\top}\, g_i(\boldsymbol{x}_i,\, y_i^{\star})) \right] \\
\times \;\; & \left[\prod_{i=1}^{N} \prod_{j=1,\, j \neq i}^{N} a_z(z_{i,j}) \, \exp(\boldsymbol{\theta}_h^{\top}\, h_{i,j}(\boldsymbol{x},\, y_i^{\star},\, y_j^{\star},\, \boldsymbol{z})) \right],
\end{aligned}
\tag{13}
$$

where $y_i^{\star} := y_i/\psi = y_i$ because $\psi := 1$ when $Y_i \in \{0, 1\}$; see Example 3 in Section 2.2.1. Since $y_i^{\star} = y_i$, we henceforth write y_i instead of $y_i^{\star}$.

Using the definitions of $c_{i,j}$ and $d_{i,j}$ in Equation (3), we specify g_i and $h_{i,j}$ as follows:

$$\boldsymbol{\theta}_g := \begin{pmatrix} \alpha_{\mathcal{Y}} \\ \beta_{\mathcal{X},\mathcal{Y},m}, \ m=1,2,3 \end{pmatrix}, \quad g_i := \begin{pmatrix} y_i \\ x_{i,m}\, y_i, \ m=1,2,3 \end{pmatrix}, \tag{14}$$

$$\boldsymbol{\theta}_h := \begin{pmatrix} \boldsymbol{\alpha}_{\mathcal{Z},O} \\ \boldsymbol{\alpha}_{\mathcal{Z},I} \\ \lambda \\ \gamma_{\mathcal{X},\mathcal{Z},1} \\ \gamma_{\mathcal{X},\mathcal{Z},m}, \ m=2,3,4 \\ \gamma_{\mathcal{Y},\mathcal{Z}} \\ \gamma_{\mathcal{Z},\mathcal{Z},1} \\ \gamma_{\mathcal{Z},\mathcal{Z},2} \\ \gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}} \end{pmatrix}, \quad h_{i,j} := \begin{pmatrix} \boldsymbol{e}_i\, z_{i,j} \\ \boldsymbol{e}_j\, z_{i,j} \\ -(1 - c_{i,j})\, z_{i,j} \log N \\ c_{i,j}\, x_{i,1}\, z_{i,j} \\ c_{i,j}\, \mathbb{I}(x_{i,m} = x_{j,m})\, z_{i,j}, \ m=2,3,4 \\ c_{i,j}\, y_j\, z_{i,j} \\ \dfrac{1}{2}\, z_{i,j}\, z_{j,i} \\ d_{i,j}(\boldsymbol{z})\, z_{i,j} \\ c_{i,j}\, x_{i,1}\, y_j\, z_{i,j} \end{pmatrix}, \tag{15}$$

where the ith coordinate of N-vector $\boldsymbol{e}_i \in \{0,1\}^N$ is 1 and all other coordinates are 0. Here, $\boldsymbol{\alpha}_{\mathcal{Z},O} := (\alpha_{\mathcal{Z},O,1}, \ldots, \alpha_{\mathcal{Z},O,N}) \in \mathbb{R}^N$ quantifies the activity of legislators $1, \ldots, N$, i.e., their tendency to mention or repost posts of other legislators; $\boldsymbol{\alpha}_{\mathcal{Z},I} := (\alpha_{\mathcal{Z},I,1}, \ldots, \alpha_{\mathcal{Z},I,N}) \in \mathbb{R}^N$ quantifies the attractiveness of legislators $1, \ldots, N$, i.e., the tendency for other legislators to mention or repost posts by them; $\lambda > 0$ discourages connections between legislators with non-overlapping neighborhoods; $\gamma_{\mathcal{X},\mathcal{Z},1}, \ldots, \gamma_{\mathcal{X},\mathcal{Z},4} \in \mathbb{R}$ capture the effects of covariates $x_{i,1}, \ldots, x_{i,4}$ on connections $Z_{i,j}$; $\gamma_{\mathcal{Y},\mathcal{Z}} \in \mathbb{R}$ is the weight of the interaction of Y_j and $Z_{i,j}$; $\gamma_{\mathcal{Z},\mathcal{Z},1} \in \mathbb{R}$ quantifies the tendency to reciprocate connections; $\gamma_{\mathcal{Z},\mathcal{Z},2} \in \mathbb{R}$ quantifies the tendency to form transitive connections; and $\gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}}$ captures spillover from covariate $x_{i,1}$ on response Y_j through connection $Z_{i,j}$; note that the spillover effect should not be interpreted as a causal effect, because the party affiliations $x_{i,1}$ of legislators i are not under the control of investigators (Kim et al., 2022). Since $\sum_{i=1}^{N} Z_{i,j} = \sum_{j=1}^{N} Z_{i,j}$ with probability 1, we set

Table 1: Maximum pseudo-likelihood estimates and standard errors based on the model specified by Equations (14) and (15).

Weight	Estimate	Standard Error	Weight	Estimate	Standard Error
$\alpha_{\mathcal{Y}}$	$-.893$	.134	$\gamma_{\mathcal{Z},\mathcal{Z},1}$	2.57	.033
$\beta_{\mathcal{X},\mathcal{Y},1}$	$-.257$	.105	$\gamma_{\mathcal{Z},\mathcal{Z},2}$	.604	.037
$\beta_{\mathcal{X},\mathcal{Y},2}$	.069	.094	$\gamma_{\mathcal{X},\mathcal{Z},1}$	$-.007$	.07
$\beta_{\mathcal{X},\mathcal{Y},3}$	$-.034$	.127	$\gamma_{\mathcal{X},\mathcal{Z},2}$	.236	.016
$\gamma_{\mathcal{Y},\mathcal{Z}}$	.035	.005	$\gamma_{\mathcal{X},\mathcal{Z},3}$	.756	.025
$\gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}}$	.038	.013	$\gamma_{\mathcal{X},\mathcal{Z},4}$	4.729	.049
			λ	.184	.006

$\alpha_{\mathcal{Z},I,N} := 0$ to address the identifiability problem that would result if all $\alpha_{\mathcal{Z},O,i}$ and $\alpha_{\mathcal{Z},I,j}$ were allowed to vary freely. These model terms were chosen based on domain knowledge, because model selection is an open problem: For instance, the statistic $c_{i,j}\, x_{i,1}\, y_j\, z_{i,j}$ with weight $\gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}}$ is included to assess whether the party affiliation $x_{i,1}$ of state legislators i affects posts y_j of state legislators j who are connected ($z_{i,j} = 1$) and whose neighborhoods overlap ($c_{i,j} = 1$). In practice, data scientists can consult domain experts to make informed choices regarding model specifications.

The specified model is estimated by an extension of the algorithm in Section 3.2 to directed connections; see Section F of the Supplementary Materials.

6.2 Results

To interpret the results, we exploit the fact that the conditional distributions of responses Y_i and connections $Z_{i,j}$ can be represented by logistic regression models, with log odds

$$\log \frac{\mathbb{P}_{\boldsymbol{\theta}}(Y_i = 1 \mid \text{others})}{1 - \mathbb{P}_{\boldsymbol{\theta}}(Y_i = 1 \mid \text{others})} = \alpha_y + \sum_{m=1}^{3} \beta_{x,y,m}\, x_{i,m} + \sum_{j:\, \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset} (\gamma_{y,z} + \gamma_{x,y,z}\, x_{j,1})\, z_{j,i}$$

and

$$\log \frac{\mathbb{P}_{\boldsymbol{\theta}}(Z_{i,j} = 1 \mid \text{others})}{1 - \mathbb{P}_{\boldsymbol{\theta}}(Z_{i,j} = 1 \mid \text{others})} = \alpha_{z,O,i} + \alpha_{z,I,j} + \frac{1}{2}\, \gamma_{z,z,1}\, z_{j,i} - (1 - c_{i,j})\, \lambda \log N$$

$$+ \; c_{i,j} \left(\gamma_{z,z,2}\, \Delta_{i,j}(\boldsymbol{z}) + \gamma_{x,z,1}\, x_{i,1} + \sum_{m=2}^{4} \gamma_{x,z,m}\, \mathbb{I}(x_{i,m} = x_{j,m}) + \gamma_{y,z}\, y_j + \gamma_{x,y,z}\, x_{i,1}\, y_j \right).$$

For instance, the positive sign of $\widehat{\gamma}_{x,y,z} = .038$ suggests that the more Republicans interact with legislator i, the higher is the conditional probability that legislator i uses offensive text in a post, holding everything else constant. Alternatively, one can interpret $\widehat{\gamma}_{x,y,z}$ in terms of the conditional probability of observing a connection: The positive sign of $\widehat{\gamma}_{x,y,z} = .038$ indicates that Republican legislators are more likely to interact with legislators who post harmful language. Other estimates align with expectations. For example, serving for the same state is the strongest predictor for reposting and mentioning activities ($\widehat{\gamma}_{x,z,4} = 4.729$), while matching gender ($\widehat{\gamma}_{x,z,2} = .236$) and race ($\widehat{\gamma}_{x,z,3} = .756$) likewise increase the conditional probability to interact. At the same time, connections affect other connections: For example, forming a connection that leads to a transitive connection is observed more often than expected under the model with $\gamma_{z,z,2} = 0$, holding everything else constant.

6.3 Model Assessment

We assess the model using model-based predictions of $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$. First, we focus on the subnetwork of all pairs of legislators $\{i, j\} \subset \mathcal{P}_N$ with $x_{i,1} = Y_j = 1$ and $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$, with a view to assessing how well the interplay of $x_{i,1}$, Y_j, and $Z_{i,j}$ can be represented by
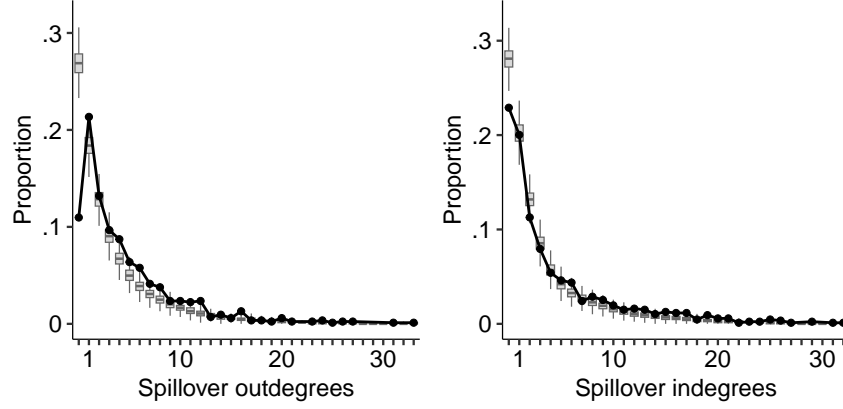
28

Figure 3: Model-based predictions of spillover in- and out-degrees of U.S. state legislators in the subnetwork with i being Republican, j using offensive language, and the neighborhoods of i and j overlapping, i.e., $x_{i,1} = Y_j = 1$ and $\mathcal{N}_i \cap \mathcal{N}_i \neq \emptyset$. By construction, the possible connections in the subnetwork act as potential channels of spillover. The spillover in- and out-degrees are defined as the respective degree of a unit in the subnetwork with $x_{i,1} = Y_j = 1$ and $\mathcal{N}_i \cap \mathcal{N}_i \neq \emptyset$. The observed spillover in- and out-degrees are shown by solid lines with overlaid points.

Table 2: Comparison of maximum pseudo-likelihood estimates based on the logistic regression model for $Y_i \mid \boldsymbol{X}_i = \boldsymbol{x}_i$ without network interference and the joint probability model for $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ with network interference.

	$\alpha_{\mathcal{Y}}$	$\beta_{\mathcal{X},\mathcal{Y},1}$	$\beta_{\mathcal{X},\mathcal{Y},2}$	$\beta_{\mathcal{X},\mathcal{Y},3}$
Without network interference	$-.101$ (.103)	$-.235$ (.097)	.032 (.093)	$-.169$ (.113)
With network interference	$-.893$ (.134)	$-.257$ (.105)	.069 (.094)	$-.034$ (.127)

the model. Figure 3 shows that the model captures the effect of $x_{i,1}$ on Y_j among pairs of legislators $\{i, j\} \subset \mathcal{P}_N$ with $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$. Second, we compare predictions of responses Y_i based on models with and without network interference. Predictions without network
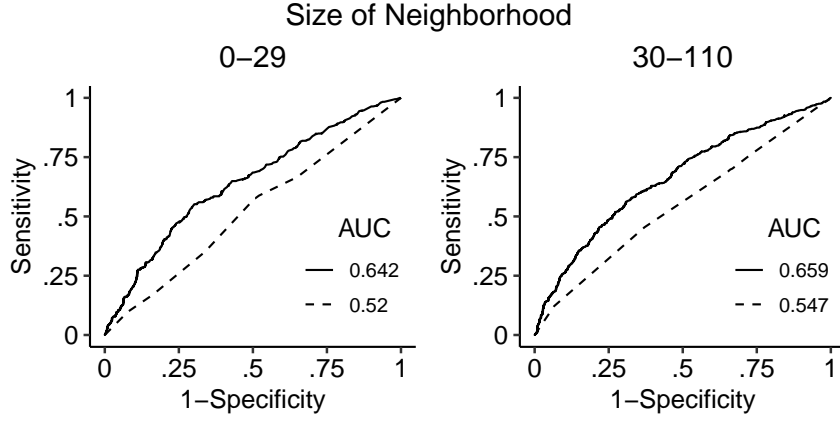
Figure 4: Comparing model-based predictions of $Y_i \mid \boldsymbol{X}_i = \boldsymbol{x}_i$ without network interference (solid) to model-based predictions of $Y_i \mid (\boldsymbol{X}, \boldsymbol{Y}_{-i}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z})$ with network interference (dashed) based on the area under the curve (AUC). Sensitivity is the true positive rate, the fraction among legislators predicted to make offensive posts who actually do so. Specificity is the true negative rate, the fraction among legislators predicted not to make offensive posts who actually do not do so.

interference are based on the logistic regression model for $Y_i \mid \boldsymbol{X}_i = \boldsymbol{x}_i$ with weights $\alpha_{\mathcal{Y}}$, $\beta_{\mathcal{X},\mathcal{Y},1}$, $\beta_{\mathcal{X},\mathcal{Y},2}$, and $\beta_{\mathcal{X},\mathcal{Y},3}$, which assumes that the posts Y_i of legislators i are independent and do not depend on the connections $\boldsymbol{Z}$ among the legislators. By contrast, predictions with network interference are based on the joint probability model for $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ specified by Equation (13), which allows the posts Y_i of state legislators i to be affected by the posts Y_j of connected legislators j whose sets of followees overlap. Figure 4 demonstrates that predictions based on models with network interference outperform those without network interference, suggesting that posts of connected state legislators with overlapping sets of followees are interdependent. Table 2 compares estimates of $\alpha_{\mathcal{Y}}$, $\beta_{\mathcal{X},\mathcal{Y},1}$, $\beta_{\mathcal{X},\mathcal{Y},2}$, and $\beta_{\mathcal{X},\mathcal{Y},3}$ based on models with and without network interference. While both models agree on the signs of parameter estimates, the estimates of $\beta_{\mathcal{X},\mathcal{Y},2}$ and $\beta_{\mathcal{X},\mathcal{Y},3}$ differ by a factor of 2 and

30

5, respectively, suggesting that network interference affects parameter estimates. Third, we demonstrate in Section G.2 of the Supplementary Materials that the model preserves salient features of connections $\boldsymbol{Z}$.

7   Discussion

The proposed regression framework is flexible, allowing data scientists to specify a wide range of models for dependent responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$.

The large set of possible models raises the question of how data scientists can select a model from a set of candidate models. While model selection for independent responses $Y_i \mid \boldsymbol{X}_i = \boldsymbol{x}_i$ is well-established and model selection for independent connections $\boldsymbol{Z} \mid \boldsymbol{X} = \boldsymbol{x}$ is an active area of research (e.g., Wang and Bickel, 2017; Wang et al., 2024; Stein et al., 2025), model selection for dependent responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$ with $p \to \infty$ parameters is an open problem. Two model selection ideas that hold promise are those of Ravikumar et al. (2010) for high-dimensional graphical models for dependent responses $\boldsymbol{Y}$ and those of Chen and Chen (2012) for high-dimensional generalized linear models for independent responses $Y_i \mid \boldsymbol{X}_i = \boldsymbol{x}_i$ based on the extended BIC.

Likewise, open questions remain in the realm of uncertainty quantification, as discussed in Section 3.3. For example, a proof of asymptotic normality based on dependent data remains elusive. A related avenue for future research is Godambe information: If $-\nabla_{\boldsymbol{\theta}}^2 \, \ell(\boldsymbol{\theta}; \boldsymbol{Y}, \boldsymbol{Z})|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}^{-1}$ were constant, it could be pulled out of the approximate covariance matrix in Equation (12), giving rise to Godambe information. Simulations suggest that uncertainty quantification based on Godambe information achieves comparable accuracy to the method reported here, while avoiding multiple matrix inversions.

The question of neighborhood recovery is another important direction for future re-

search, as discussed in Section 2.

Supplementary Materials

The supplementary materials contain proofs of all theoretical results.

Acknowledgements

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

References

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B 36*, 192–225.

Brown, L. (1986). *Fundamentals of Statistical Exponential Families: With Applications in Statistical Decision Theory.* Hayworth, CA, USA: Institute of Mathematical Statistics.

Chatterjee, S. and P. Diaconis (2013). Estimating and understanding exponential random graph models. *The Annals of Statistics 41*, 2428–2461.

Chen, J. and Z. Chen (2012). Extended BIC for small-n-large-p sparse GLM. *Statistica Sinica 22*, 555–574.

Clark, D. A. and M. S. Handcock (2024). Causal inference over stochastic networks. *Journal of the Royal Statistical Society Series A: Statistics in Society 187*, 772–795.

Comets, F. and M. Janzura (1998). A central limit theorem for conditionally centred random fields with an application to Markov fields. *Journal of Applied Probability 35*, 608–621.

Efron, B. (2022). *Exponential families in theory and practice.* Cambridge, MA: Cambridge University Press.

Fosdick, B. K. and P. D. Hoff (2015). Testing and modeling dependencies between a network and nodal attributes. *Journal of the American Statistical Association 110*, 1047–1056.

Handcock, M. S. (2003). Statistical models for social networks: Inference and degeneracy. In R. Breiger, K. Carley, and P. Pattison (Eds.), *Dynamic Social Network Modeling and Analysis*, pp. 1–12. Washington, D.C.: National Academies Press.

Huang, D., W. Lan, H. H. Zhang, and H. Wang (2019). Least squares estimation of spatial autoregressive models for large-scale social networks. *Electronic Journal of Statistics 13*, 1135–1165.

Huang, D., F. Wang, X. Zhu, and H. Wang (2020). Two-mode network autoregressive model for large-scale networks. *Journal of Econometrics 216*, 203–219.

Huang, S., J. Sun, and Y. Feng (2024). PCABM: Pairwise covariates-adjusted block model for community detection. *Journal of the American Statistical Association 119*, 2092–2104.

Hunter, D. R. and K. Lange (2004). A tutorial on MM algorithms. *The American Statistician 58*, 30–37.

Jensen, J. L. and H. R. Künsch (1994). On asymptotic normality of pseudo likelihood estimates for pairwise interaction processes. *Annals of the Institute of Mathematical Statistics 46*, 475–486.

Kim, T., N. Nakka, I. Gopal, B. A. Desmarais, A. Mancinelli, J. J. Harden, H. Ko, and F. J. Boehmke (2022). Attention to the COVID-19 pandemic on Twitter: Partisan differences among U.S. state legislators. *Legislative Studies Quarterly 47*, 1023–1041.

Kolaczyk, E. D. (2017). *Topics at the Frontier of Statistics and Network Analysis: (Re)Visiting the Foundations.* Cambridge University Press.

Krivitsky, P. N., P. Coletti, and N. Hens (2023). A tale of two datasets: Representativeness and generalisability of inference for samples of networks. *Journal of the American Statistical Association 118*, 2213–2224.

Le, C. M. and T. Li (2022). Linear regression and its inference on noisy network-linked data. *Journal of the Royal Statistical Society Series B: Statistical Methodology 84*, 1851–1885.

Lei, J., K. Chen, and H. Moon (2024). Least squares inference for data with network dependency. Available at arXiv:2404.01977.

Li, S. and S. Wager (2022). Random graph asymptotics for treatment effect estimation under network interference. *The Annals of Statistics 50*, 2334 – 2358.

Li, T., E. Levina, and J. Zhu (2019). Prediction models for network-linked data. *The Annals of Applied Statistics 13*(1), 132–164.

Niezink, N. M. D. and T. A. B. Snijders (2017). Co-evolution of social networks and continuous actor attributes. *The Annals of Applied Statistics 11*, 1948–1973.

Ogburn, E. L., O. Sofrygin, I. Diaz, and M. J. Van der Laan (2024). Causal inference for social network data. *Journal of the American Statistical Association 119*, 597–611.

Ortega, J. M. and W. C. Rheinboldt (2000). *Iterative solution of nonlinear equations in several variables.* Society for Industrial and Applied Mathematics.

Ravikumar, P., M. J. Wainwright, and J. Lafferty (2010). High-dimensional Ising model selection using ℓ_1-regularized logistic regression. *The Annals of Statistics 38*, 1287–1319.

Schweinberger, M. (2011). Instability, sensitivity, and degeneracy of discrete exponential families. *Journal of the American Statistical Association 106*, 1361–1370.

Snijders, T. A. B., C. E. G. Steglich, and M. Schweinberger (2007). Modeling the co-evolution of networks and behavior. In K. van Montfort, H. Oud, and A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences*, pp. 41–71. Lawrence Erlbaum.

Stein, S., R. Feng, and C. Leng (2025). A sparse beta regression model for network analysis. *Journal of the American Statistical Association*. To appear.

Stewart, J. R. and M. Schweinberger (2025). Pseudo-likelihood-based M-estimators for random graphs with dependent edges and parameter vectors of increasing dimension. *The Annals of Statistics*. To appear.

Tchetgen Tchetgen, E. J., I. R. Fulcher, and I. Shpitser (2021). Auto-G-computation of causal effects on a network. *Journal of the American Statistical Association 116*, 833–844.

Wang, J., X. Cai, X. Niu, and R. Li (2024). Variable selection for high-dimensional nodal

attributes in social networks with degree heterogeneity. *Journal of the American Statistical Association 119*, 1322–1335.

Wang, Y. X. and P. J. Bickel (2017). Likelihood-based model selection for stochastic block models. *The Annals of Statistics 45*, 500–528.

Wang, Z., I. E. Fellows, and M. S. Handcock (2024). Understanding networks with exponential-family random network models. *Social Networks 78*, 81–91.

Zhu, X., D. Huang, R. Pan, and H. Wang (2020). Multivariate spatial autoregressive model for large scale social networks. *Journal of Econometrics 215*, 591–606.

Supplementary Materials: A Regression Framework for Studying Relationships among Attributes under Network Interference

A   Proofs of Propositions 1 and 2

PROOF OF PROPOSITION 1. The joint probability density function of $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$

stated in Equation (1) in Section 2 implies that the conditional probability density function

1

of $Y_i \mid (\boldsymbol{X}, \boldsymbol{Y}_{-i}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z})$ can be written as

$$f_{\boldsymbol{\theta}}(y_i \mid \boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z}) = \frac{f_{\boldsymbol{\theta}}(y_i, \boldsymbol{y}_{-i}, \boldsymbol{z} \mid \boldsymbol{x})}{\displaystyle\int_{\mathcal{Y}_i} f_{\boldsymbol{\theta}}(y, \boldsymbol{y}_{-i}, \boldsymbol{z} \mid \boldsymbol{x}) \, \mathrm{d}\,\nu_{\mathcal{Y}}(y)}$$

$$= \frac{a_{\mathcal{Y}}(y_i) \exp\left(\boldsymbol{\theta}_g^\top \, g_{i,1}(\boldsymbol{x}_i) \, y_i^\star + \left(\displaystyle\sum_{j \in \mathcal{P}_N \setminus \{i\}} \boldsymbol{\theta}_h^\top \, h_{i,j,1}(\boldsymbol{x}, \, y_j^\star, \, \boldsymbol{z})\right) y_i^\star\right)}{\displaystyle\int_{\mathcal{Y}_i} a_{\mathcal{Y}}(y) \exp\left(\boldsymbol{\theta}_g^\top \, g_{i,1}(\boldsymbol{x}_i) \, y^\star + \left(\displaystyle\sum_{j \in \mathcal{P}_N \setminus \{i\}} \boldsymbol{\theta}_h^\top \, h_{i,j,1}(\boldsymbol{x}, \, y_j^\star, \, \boldsymbol{z})\right) y^\star\right) \mathrm{d}\,\nu_{\mathcal{Y}}(y)}$$

$$= a_{\mathcal{Y}}(y_i) \exp\left(\frac{\eta_i(\boldsymbol{\theta}; \, \boldsymbol{x}, \, \boldsymbol{y}_{-i}^\star, \, \boldsymbol{z}) \, y_i - b_i(\eta_i(\boldsymbol{\theta}; \, \boldsymbol{x}, \, \boldsymbol{y}_{-i}^\star, \, \boldsymbol{z}))}{\psi}\right),$$

where $y^\star := y/\psi$, $y_i^\star := y_i/\psi$, and $\boldsymbol{y}_{-i}^\star := \boldsymbol{y}_{-i}/\psi$, while

$$\eta_i(\boldsymbol{\theta}; \, \boldsymbol{x}, \, \boldsymbol{y}_{-i}^\star, \, \boldsymbol{z}) := \boldsymbol{\theta}^\top \left(g_{i,1}(\boldsymbol{x}_i), \, \sum_{j \in \mathcal{P}_N \setminus \{i\}} h_{i,j,1}(\boldsymbol{x}, \, y_j^\star, \, \boldsymbol{z})\right)$$

$$b_i(\eta_i(\boldsymbol{\theta}; \, \boldsymbol{x}, \, \boldsymbol{y}_{-i}^\star, \, \boldsymbol{z})) := \psi \log \int_{\mathcal{Y}_i} a_{\mathcal{Y}}(y) \exp\left(\frac{\eta_i(\boldsymbol{\theta}; \, \boldsymbol{x}, \, \boldsymbol{y}_{-i}^\star, \, \boldsymbol{z}) \, y}{\psi}\right) \mathrm{d}\,\nu_{\mathcal{Y}}(y).$$

Proposition 2. *Consider Example 1 in Section 2.2.1. Let $\boldsymbol{U} \in \{0, 1\}^{N \times N}$ be the $N \times N$ matrix with elements*

$$u_{i,j} := c_{i,j} \, z_{i,j} = \mathbb{1}(\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset) \, z_{i,j}, \tag{A.1}$$

and let $\boldsymbol{v} \in \mathbb{R}^N$ be the N-vector with coordinates

$$v_i := \alpha_{\mathcal{Y}} + \beta_{\mathcal{X},\mathcal{Y}} \, x_i + \gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}} \sum_{j \in \mathcal{P}_N \setminus \{i\}} u_{i,j} \, x_j. \tag{A.2}$$

Denote by $\boldsymbol{I}$ the $N \times N$ identity matrix and define $\xi_{\mathcal{Y},\mathcal{Y},\mathcal{Z}} := \gamma_{\mathcal{Y},\mathcal{Y},\mathcal{Z}}/\psi$. If $(\boldsymbol{I} - \xi_{\mathcal{Y},\mathcal{Y},\mathcal{Z}} \, \boldsymbol{U})$ is positive definite, the conditional distribution of $\boldsymbol{Y} \mid (\boldsymbol{X}, \, \boldsymbol{Z}) = (\boldsymbol{x}, \, \boldsymbol{z})$ is N-variate Gaussian with mean vector $(\boldsymbol{I} - \xi_{\mathcal{Y},\mathcal{Y},\mathcal{Z}} \, \boldsymbol{U})^{-1} \, \boldsymbol{v}$ and covariance matrix $\psi \, (\boldsymbol{I} - \xi_{\mathcal{Y},\mathcal{Y},\mathcal{Z}} \, \boldsymbol{U})^{-1}$.

Remark. The requirement that $(\boldsymbol{I} - \xi_{\mathcal{Y},\mathcal{Y},\mathcal{Z}} \, \boldsymbol{U})$ be positive definite imposes restrictions on $\gamma_{\mathcal{Y},\mathcal{Y},\mathcal{Z}}$. The restrictions on $\gamma_{\mathcal{Y},\mathcal{Y},\mathcal{Z}}$ depend on the neighborhoods $\mathcal{N}_i$ of units $i \in \mathcal{P}_N$ and connections $Z_{i,j}$ among pairs of units $\{i, j\} \subset \mathcal{P}_N$.

PROOF OF PROPOSITION 2. Example 1 in Section 2.2.1 demonstrates that the conditional distribution of $Y_i \mid (\boldsymbol{X}, \boldsymbol{Y}_{-i}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z})$ is Gaussian with conditional mean

$$
\mathbb{E}(Y_i \mid \boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z}) = \alpha_y + \beta_{x,y} x_i + \gamma_{x,y,z} \sum_{j \in \mathcal{P}_N \setminus \{i\}} u_{i,j} x_j + \gamma_{y,y,z} \sum_{j \in \mathcal{P}_N \setminus \{i\}} u_{i,j} y_j^\star
$$

$$
= v_i + \xi_{y,y,z} \sum_{j \in \mathcal{P}_N \setminus \{i\}} u_{i,j} y_j, \tag{A.3}
$$

where

$$
v_i := \alpha_y + \beta_{x,y} x_i + \gamma_{x,y,z} \sum_{j \in \mathcal{P}_N \setminus \{i\}} u_{i,j} x_j
$$

and

$$
\xi_{y,y,z} := \frac{\gamma_{y,y,z}}{\psi}.
$$

The conditional variance of $Y_i \mid (\boldsymbol{X}, \boldsymbol{Y}_{-i}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z})$ is

$$
\mathbb{V}(Y_i \mid \boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z}) = \psi. \tag{A.4}
$$

Let $\boldsymbol{m} := (m_i) \in \mathbb{R}^N$ be the conditional mean of $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{z})$. Upon taking expectation on both sides of (A.3) conditional on $(\boldsymbol{X}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{z})$, we obtain

$$
m_i = v_i + \xi_{y,y,z} \sum_{j \in \mathcal{P}_N \setminus \{i\}} u_{i,j} m_j, \tag{A.5}
$$

which implies that

$$
v_i = m_i - \xi_{y,y,z} \sum_{j \in \mathcal{P}_N \setminus \{i\}} u_{i,j} m_j
$$

and hence

$$
\mathbb{E}(Y_i \mid \boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z}) = v_i + \xi_{y,y,z} \sum_{j \in \mathcal{P}_N \setminus \{i\}} u_{i,j} y_j
$$

$$
= m_i - \xi_{y,y,z} \sum_{j \in \mathcal{P}_N \setminus \{i\}} u_{i,j} m_j + \xi_{y,y,z} \sum_{j \in \mathcal{P}_N \setminus \{i\}} u_{i,j} y_j
$$

$$
= m_i + \xi_{y,y,z} \sum_{j \in \mathcal{P}_N \setminus \{i\}} u_{i,j} (y_j - m_j)
$$

$$
= m_i - \sum_{j \in \mathcal{P}_N \setminus \{i\}} b_{i,j} (y_j - m_j), \tag{A.6}
$$

3

where

$$b_{i,j} \; := \; - \xi_{y,y,z} \, u_{i,j}.$$

By comparing Equations (A.4) and (A.6) to Equations (2.17) and (2.18) of Rue and Held (2005) and invoking Theorem 2.6 of Rue and Held (2005), we conclude that the conditional distribution of $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{z})$ is N-variate Gaussian with mean vector $\boldsymbol{m} \in \mathbb{R}^N$ and precision matrix $\boldsymbol{P} \in \mathbb{R}^{N \times N}$ with elements

$$p_{i,j} \; := \; \begin{cases} \dfrac{1}{\psi} & \text{if } i = j \\[2ex] \dfrac{b_{i,j}}{\psi} & \text{if } i \neq j, \end{cases}$$

provided $u_{i,j} = u_{j,i}$ for all $i \neq j$ and $\boldsymbol{P}$ is positive definite; note that $u_{i,j} = u_{j,i}$ is satisfied in undirected networks with $z_{i,j} = z_{j,i}$.

To state these results in matrix form, note that (A.5) can be expressed as

$$\boldsymbol{m} \;\; = \;\; \boldsymbol{v} + \xi_{y,y,z} \, \boldsymbol{U} \, \boldsymbol{m},$$

implying

$$\boldsymbol{m} \;\; = \;\; (\boldsymbol{I} - \xi_{y,y,z} \, \boldsymbol{U})^{-1} \, \boldsymbol{v},$$

while $\boldsymbol{P}$ can be expressed as

$$\boldsymbol{P} \;\; = \;\; \frac{1}{\psi} \, (\boldsymbol{I} - \xi_{y,y,z} \, \boldsymbol{U}),$$

implying

$$\boldsymbol{P}^{-1} \;\; = \;\; \psi \, (\boldsymbol{I} - \xi_{y,y,z} \, \boldsymbol{U})^{-1}.$$

To conclude, the conditional distribution of $\boldsymbol{Y} \mid (\boldsymbol{X}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{z})$ is N-variate Gaussian with mean vector $(\boldsymbol{I} - \xi_{y,y,z} \, \boldsymbol{U})^{-1} \, \boldsymbol{v}$ and covariance matrix $\psi \, (\boldsymbol{I} - \xi_{y,y,z} \, \boldsymbol{U})^{-1}$, provided $(\boldsymbol{I} - \xi_{y,y,z} \, \boldsymbol{U})$ is positive definite.

B Proofs of Lemmas 1 and 2

PROOF OF LEMMA 1. Lemma 1 is proved in the sentence preceeding the statement of Lemma 1 in Section 3.1.

PROOF OF LEMMA 2. Letting $\boldsymbol{\Theta}_1$ denote the parameter space of $\boldsymbol{\theta}_1$, suppose that $v : \boldsymbol{\Theta}_1 \mapsto \mathbb{R}$ is any twice differentiable function and that $\nabla^2 v(\boldsymbol{\theta}) - \boldsymbol{M}$ is non-negative definite for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_1$ for some constant matrix $\boldsymbol{M} \in \mathbb{R}^{d \times d}$ $(d \geq 1)$. Then the function $u : \boldsymbol{\Theta}_1 \mapsto \mathbb{R}$ given by

$$u(\boldsymbol{\theta}_1) \;\; := \;\; v(\boldsymbol{\theta}_0) + (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^\top \nabla v(\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^\top \boldsymbol{M}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0), \quad \boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_1$$

satisfies $u(\boldsymbol{\theta}_1) \leq v(\boldsymbol{\theta}_1)$ for all $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}_1$, because Taylor's theorem (Theorem 6.11, Magnus and Neudecker, 2019, p. 124) gives

$$u(\boldsymbol{\theta}_1) - v(\boldsymbol{\theta}_1) \;\; = \;\; \frac{1}{2}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^\top \left[\nabla^2 v(\dot{\boldsymbol{\theta}}) - \boldsymbol{M} \right] (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0),$$

where $\dot{\boldsymbol{\theta}} := \phi\, \boldsymbol{\theta}_0 + (1 - \phi)\, \boldsymbol{\theta}_1 \in \boldsymbol{\Theta}_1$ $(\phi \in [0, 1])$. The inequality $1/4 \geq \pi_{i,j}\,(1 - \pi_{i,j})$ implies that

$$-[\boldsymbol{A}(\boldsymbol{\theta}_1) - \boldsymbol{A}^\star] \;\; = \;\; \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left[\frac{1}{4} - \pi_{i,j}^{(t)}\,(1 - \pi_{i,j}^{(t)}) \right] \boldsymbol{e}_{i,j}\, \boldsymbol{e}_{i,j}^\top$$

is non-negative definite. Lemma 1 proves that $\boldsymbol{\theta}_1$ is concave and that the restriction of $\ell(\boldsymbol{\theta})$ to $\boldsymbol{\theta}_1$ has the properties of $v(\boldsymbol{\theta}_1)$ stated above, proving Lemma 2.

C Proof of Theorem 1

Theorem 1 is a generalization of Theorem 2 of Stewart and Schweinberger (2025, abbreviated as S25) from exponential family models for binary connections $\boldsymbol{Z}$ to exponential family models for binary, count-, and real-valued responses and connections $(\boldsymbol{Y}, \boldsymbol{Z}) \mid \boldsymbol{X} = \boldsymbol{x}$. We henceforth suppress predictors $\boldsymbol{x} \in \mathcal{X}$.

PROOF OF THEOREM 1. Let $s(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) := \nabla_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})$ and consider events

$$\mathcal{C}(\delta_N) := \{(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{Y} \times \mathcal{Z} : \|s(\boldsymbol{\theta}^\star; \boldsymbol{y}, \boldsymbol{z})\|_\infty \leq \delta_N\}$$

$$\mathcal{H} \subseteq \{(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{Y} \times \mathcal{Z} : -\nabla_{\boldsymbol{\theta}}^2 \ell(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) \text{ is invertible for all } \boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)\}.$$

Define

$$\Lambda_{N,\boldsymbol{y},\boldsymbol{z}}(\boldsymbol{\theta}^\star) := \sup_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)} \|(-\nabla_{\boldsymbol{\theta}}^2 \ell(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}))^{-1}\|_\infty, \quad (\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{H}$$

$$\Lambda_N(\boldsymbol{\theta}^\star) := \sup_{(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{H}} \Lambda_{N,\boldsymbol{y},\boldsymbol{z}}(\boldsymbol{\theta}^\star).$$

It follows from results of S25 that $s(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})$, considered as a function of $\boldsymbol{\theta} \in \Theta$ for fixed $(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{H}$, is a homeomorphism and is continuously differentiable.

In the event $(\boldsymbol{Y}, \boldsymbol{Z}) \in \mathcal{C}(\delta_N)$, the set $\Theta(\delta_N)$ is non-empty. By construction of the sets $\mathcal{C}(\delta_N)$ and $\widehat{\Theta}(\delta_N)$, the set $\widehat{\Theta}(\delta_N)$ is non-empty for all $(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{C}(\delta_N)$, because $\widehat{\Theta}(\delta_N)$ contains the data-generating parameter vector $\boldsymbol{\theta}^\star \in \Theta$ provided $(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{C}(\delta_N)$:

$$\boldsymbol{\theta}^\star \in \widehat{\Theta}(\delta_N) := \{\boldsymbol{\theta} \in \Theta : \|s(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})\|_\infty \leq \delta_N\}.$$

In the event $(\boldsymbol{Y}, \boldsymbol{Z}) \in \mathcal{C}(\delta_N) \cap \mathcal{H}$, the set $\widehat{\Theta}(\delta_N)$ satisfies $\widehat{\Theta}(\delta_N) \subseteq \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \rho_N)$ provided $N > N_0$. By assumption, there exists a sequence $\rho_1, \rho_2, \cdots \in [0, +\infty)$ such that $\rho_N = o(1)$. Therefore, there exists an integer $N_0 \in \{1, 2, \dots\}$ such that $\rho_N < \epsilon^\star$ for all $N > N_0$. Consider any $N > N_0$ and any $(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{C}(\delta_N) \cap \mathcal{H}$. Since $s^{-1}(\cdot; \boldsymbol{y}, \boldsymbol{z})$ is continuous on Θ, there exists, for each $(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{H}$, a real number $\epsilon_N(\rho_N) \in (0, +\infty)$ (which depends on $(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{H}$) such that

$$\|s(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) - s(\boldsymbol{\theta}^\star; \boldsymbol{y}, \boldsymbol{z})\|_\infty \leq \epsilon_N(\rho_N) \quad \text{implies} \quad \|\boldsymbol{\theta} - \boldsymbol{\theta}^\star\|_\infty \leq \rho_N. \quad (\text{C.1})$$

As $s(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})$ is a homeomorphism and continuously differentiable, we can invoke Lemma 1 of S25 to conclude that $\epsilon_N(\rho_N)$ is related to ρ_N by the following inequality:

$$\frac{\rho_N}{\Lambda_{N,\boldsymbol{y},\boldsymbol{z}}(\boldsymbol{\theta}^\star)} \leq \epsilon_N(\rho_N). \quad (\text{C.2})$$

6

To take advantage of (C.2), observe that, for all $\boldsymbol{\theta} \in \widehat{\boldsymbol{\Theta}}(\delta_N)$ and all $(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{C}(\delta_N) \cap \mathcal{H}$,

$$\|\boldsymbol{s}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) - \boldsymbol{s}(\boldsymbol{\theta}^\star; \boldsymbol{y}, \boldsymbol{z})\|_\infty \leq \|\boldsymbol{s}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})\|_\infty + \|\boldsymbol{s}(\boldsymbol{\theta}^\star; \boldsymbol{y}, \boldsymbol{z})\|_\infty \leq 2\,\delta_N = \frac{\rho_N}{\Lambda_N(\boldsymbol{\theta}^\star)}, \quad \text{(C.3)}$$

because $\|\boldsymbol{s}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z})\|_\infty \leq \delta_N$ for all $\boldsymbol{\theta} \in \widehat{\boldsymbol{\Theta}}(\delta_N)$, $\|\boldsymbol{s}(\boldsymbol{\theta}^\star; \boldsymbol{y}, \boldsymbol{z})\|_\infty \leq \delta_N$ for all $(\boldsymbol{y}, \boldsymbol{z}) \in$ $\mathcal{C}(\delta_N) \cap \mathcal{H}$, and $\delta_N := \rho_N / (2\,\Lambda_N(\boldsymbol{\theta}^\star))$. Using (C.3) along with the definition of $\Lambda_N(\boldsymbol{\theta}^\star) :=$ $\sup_{(\boldsymbol{y}, \boldsymbol{z}) \in \mathcal{H}} \Lambda_{N,\boldsymbol{y},\boldsymbol{z}}(\boldsymbol{\theta}^\star) > 0$, we obtain

$$\|\boldsymbol{s}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) - \boldsymbol{s}(\boldsymbol{\theta}^\star; \boldsymbol{y}, \boldsymbol{z})\|_\infty \leq \frac{\rho_N}{\Lambda_N(\boldsymbol{\theta}^\star)} \leq \frac{\rho_N}{\Lambda_{N,\boldsymbol{y},\boldsymbol{z}}(\boldsymbol{\theta}^\star)}, \quad \text{(C.4)}$$

and, using (C.2),

$$\|\boldsymbol{s}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) - \boldsymbol{s}(\boldsymbol{\theta}^\star; \boldsymbol{y}, \boldsymbol{z})\|_\infty \leq \frac{\rho_N}{\Lambda_{N,\boldsymbol{y},\boldsymbol{z}}(\boldsymbol{\theta}^\star)} \leq \epsilon_N(\rho_N). \quad \text{(C.5)}$$

In light of the fact that

$$\|\boldsymbol{s}(\boldsymbol{\theta}; \boldsymbol{y}, \boldsymbol{z}) - \boldsymbol{s}(\boldsymbol{\theta}^\star; \boldsymbol{y}, \boldsymbol{z})\|_\infty \leq \epsilon_N(\rho_N) \quad \text{implies} \quad \|\boldsymbol{\theta} - \boldsymbol{\theta}^\star\|_\infty \leq \rho_N,$$

the set $\widehat{\boldsymbol{\Theta}}(\delta_N)$ is non-empty and satisfies

$$\widehat{\boldsymbol{\Theta}}(\delta_N) \subseteq \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \rho_N) \quad \text{(C.6)}$$

in the event $(\boldsymbol{Y}, \boldsymbol{Z}) \in \mathcal{C}(\delta_N) \cap \mathcal{H}$, provided $N > N_0$.

The event $(\boldsymbol{Y}, \boldsymbol{Z}) \in \mathcal{C}(\delta_N) \cap \mathcal{H}$ occurs with probability $1 - o(1)$. The probability of event $(\boldsymbol{Y}, \boldsymbol{Z}) \in \mathcal{C}(\delta_N) \cap \mathcal{H}$ is bounded below by

$$\mathbb{P}\left((\boldsymbol{Y}, \boldsymbol{Z}) \in \mathcal{C}(\delta_N) \cap \mathcal{H}\right) \geq 1 - \mathbb{P}\left((\boldsymbol{Y}, \boldsymbol{Z}) \notin \mathcal{C}(\delta_N)\right) - \mathbb{P}\left((\boldsymbol{Y}, \boldsymbol{Z}) \notin \mathcal{H}\right) = 1 - o(1).$$

The above inequality stems from a union bound, while the identity follows from the assumption that the probabilities of the events $(\boldsymbol{Y}, \boldsymbol{Z}) \notin \mathcal{C}(\delta_N)$ and $(\boldsymbol{Y}, \boldsymbol{Z}) \notin \mathcal{H}$ satisfy

$$\mathbb{P}\left((\boldsymbol{Y}, \boldsymbol{Z}) \notin \mathcal{C}(\delta_N)\right) = \mathbb{P}\left(\|\boldsymbol{s}(\boldsymbol{\theta}^\star; \boldsymbol{Y}, \boldsymbol{Z}) - \mathbb{E}\,\boldsymbol{s}(\boldsymbol{\theta}^\star; \boldsymbol{Y}, \boldsymbol{Z})\|_\infty \geq \delta_N\right) = o(1)$$

$$\mathbb{P}\left((\boldsymbol{Y}, \boldsymbol{Z}) \notin \mathcal{H}\right) = o(1),$$

where the first result leverages the fact that $\mathbb{E}\,\boldsymbol{s}(\boldsymbol{\theta}^\star;\,\boldsymbol{Y},\boldsymbol{Z}) = \mathbf{0}$ by Lemma 7 of S25.

Conclusion. Combining (C.6) with (C.7) establishes that, for all $N > N_0$, the random set $\widehat{\boldsymbol{\Theta}}(\delta_N)$ is non-empty and, with probability $1 - o(1)$, satisfies

$$\widehat{\boldsymbol{\Theta}}(\delta_N) \ \subseteq \ \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\,\rho_N).$$

D   Corollaries 1 and 2

To state and prove Corollaries 1 and 2, we first introduce notation along with background on conditional independence graphs (Maathuis et al., 2019) and couplings (Lindvall, 2002).

D.1   Notation and Background

We consider the model of Corollary 1, with joint probability mass function

$$\mathbb{P}_{\boldsymbol{\theta}}\left((\boldsymbol{Y},\,\boldsymbol{Z}) = (\boldsymbol{y},\,\boldsymbol{z}) \mid \boldsymbol{X} = \boldsymbol{x}\right) \ \propto \ \exp\left(\boldsymbol{\theta}^\top \boldsymbol{b}(\boldsymbol{x},\,\boldsymbol{y},\,\boldsymbol{z})\right). \tag{D.1}$$

The parameter vector is $\boldsymbol{\theta} := (\alpha_{Z,1},\ldots,\alpha_{Z,N},\gamma_{Z,Z},\,\gamma_{X,Y,Z}) \ \in \ \mathbb{R}^{N+2}$ and the vector of sufficient statistics is $\boldsymbol{b}(\boldsymbol{x},\,\boldsymbol{y},\,\boldsymbol{z}) \in \mathbb{R}^{N+2}$, with coordinates

- $b_i(\boldsymbol{x},\,\boldsymbol{y},\,\boldsymbol{z}) := \sum_{j \in \mathcal{P}_N \setminus \{i\}} z_{i,j}$ $(i = 1,\ldots,N)$,

- $b_{N+1}(\boldsymbol{x},\,\boldsymbol{y},\,\boldsymbol{z}) := \sum_{i=1}^{N} \sum_{j=i+1}^{N} d_{i,j}(\boldsymbol{z})\, z_{i,j}$,

- $b_{N+2}(\boldsymbol{x},\,\boldsymbol{y},\,\boldsymbol{z}) := \sum_{i=1}^{N} \sum_{j=i+1}^{N} c_{i,j}\, (x_i\, y_j + x_j\, y_i)\, z_{i,j}$,

where the terms $c_{i,j}$ and $d_{i,j}(\boldsymbol{z})$ are defined as follows:

$$c_{i,j} := \mathbb{1}(\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset)$$

$$d_{i,j}(\boldsymbol{z}) := \mathbb{1}(\exists\, k \in \mathcal{N}_i \cap \mathcal{N}_j \,:\, z_{i,k} = z_{k,j} = 1). \tag{D.2}$$

In light of $\psi := 1$, we do not distinguish between $\boldsymbol{y}$ and $\boldsymbol{y}^\star$ or y_i and $y_i^\star$. To ease the presentation, we write $Y_i \mid \boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z}$ rather than $Y_i \mid (\boldsymbol{X}, \boldsymbol{Y}_{-i}, \boldsymbol{Z}) = (\boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z})$, and $Z_{i,j} \mid \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}}$ rather than $Z_{i,j} \mid (\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}_{-\{i,j\}}) = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}})$. Expectations, variances, and covariances with respect to the conditional distributions of $Y_i \mid \boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z}$ and $Z_{i,j} \mid \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}}$ are denoted by $\mathbb{E}_{y,i}, \mathbb{V}_{y,i}, \mathbb{C}_{y,i}$ and $\mathbb{E}_{z,i,j}, \mathbb{V}_{z,i,j}, \mathbb{C}_{z,i,j}$, respectively.

Conditional independence graph. Let $M := N + \binom{N}{2}$ be the total number of responses and connections and

$$\boldsymbol{W} := (W_1, \ldots, W_M) := (Y_1, \ldots, Y_N, Z_{1,2}, \ldots, Z_{N-1,N}) \in \mathcal{W} := \{0, 1\}^{N+\binom{N}{2}} \quad (\text{D.3})$$

be the vector consisting of responses and connections. The conditional independence structure of the model can be represented by a conditional independence graph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$ with a set of vertices $\mathcal{V} := \{W_1, \ldots, W_M\}$ and a set of undirected edges $\mathcal{E}$. We refer to elements of $\mathcal{V}$ and $\mathcal{E}$ as vertices and edges of $\mathcal{G}$. There are two distinct subsets of vertices in $\mathcal{G}$:

- the subset $\mathcal{V}_y := \{W_1, \ldots, W_N\}$ corresponding to responses $Y_1, \ldots, Y_N$;

- the subset $\mathcal{V}_z := \{W_{N+1}, \ldots, W_M\}$ corresponding to connections $Z_{1,2}, \ldots, Z_{N-1,N}$.

An undirected edge between two vertices in $\mathcal{G}$ represents dependence of the two corresponding random variables conditional on all other random variables. The vertices in $\mathcal{G}$ are connected to the following subsets of vertices (neighborhoods):

- The neighborhood of Y_i in $\mathcal{G}$ consists of all Y_j and all $Z_{i,j}$ such that $j \in \mathcal{P}_N \setminus \{i\}$ and $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$.

- The neighborhood of $Z_{i,j}$ in $\mathcal{G}$ consists of

  1. Y_i and Y_j;

  2. all $Z_{i,h}$ and $Z_{j,h}$ such that $h \in \mathcal{P}_N \setminus \{i, j\}$ and $h \in \mathcal{N}_i \cap \mathcal{N}_j$;

9

3. all $Z_{i,h}$ and $Z_{j,h}$ such that $h \in \mathcal{P}_N \setminus \{i, j\}$ and $h \notin \mathcal{N}_i \cap \mathcal{N}_j$ provided that either

   $j \in \mathcal{N}_i \cap \mathcal{N}_h$ holds or $i \in \mathcal{N}_j \cap \mathcal{N}_h$ holds.

Let $d_{\mathcal{G}}(i, j)$ be the length of the shortest path from vertex $W_i \in \mathcal{V}$ to vertex $W_j \in \mathcal{V}$ in $\mathcal{G}$

and let $\mathcal{S}_{\mathcal{G},i,k}$ be the set of vertices with distance $k \in \{1, 2, \ldots\}$ to the ith vertex W_i in $\mathcal{G}$:

$$\mathcal{S}_{\mathcal{G},i,k} \ := \ \{W_j \in \mathcal{V} \setminus \{W_i\} : d_{\mathcal{G}}(i, j) = k\}.$$

We define the maximum degree of vertices relating to connections in $\mathcal{G}$ as follows:

$$D_N \ := \ \max_{1 \leq i \leq M} |\mathcal{S}_{\mathcal{G},i,1}|. \tag{D.4}$$

Coupling matrix. Let $\boldsymbol{W}_{a:b} := (W_a, \ldots, W_b) \in \mathcal{W}_{a:b}$ be the subvector consisting of

responses and connections with indices $1 \leq a \leq b \leq M$. The set of random variables

excluding the random variable $W_v \in \mathcal{V}$ with $v \in \{1, \ldots, M\}$ is denoted by $\boldsymbol{w}_{-v} \in \mathcal{W}_{-v}$.

Consider any $\boldsymbol{a} \in \{0, 1\}^{M-i}$ and define

$$\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{w}_{1:(i-1)}, w_i}(\boldsymbol{W}_{(i+1):M} = \boldsymbol{a}) \ := \ \mathbb{P}_{\boldsymbol{\theta}^\star}(\boldsymbol{W}_{(i+1):M} = \boldsymbol{a} \mid (\boldsymbol{W}_{1:(i-1)}, W_i) = (\boldsymbol{w}_{1:(i-1)}, w_i)).$$

We use the total variation distance between the conditional distributions $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{w}_{1:(i-1)}, 0}$ and

$\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{w}_{1:(i-1)}, 1}$ for quantifying the amount of dependence induced by the model, where $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta}$

is the data-generating parameter vector. The total variation distance between $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{w}_{1:(i-1)}, 0}$

and $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{w}_{1:(i-1)}, 1}$ can be bounded from above by using coupling methods (Lindvall, 2002).

A coupling of $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{w}_{1:(i-1)}, 0}$ and $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{w}_{1:(i-1)}, 1}$ is a joint probability distribution $\mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{w}_{1:(i-1)}}$

for a pair of random vectors $(\boldsymbol{W}^\star_{(i+1):M}, \boldsymbol{W}^{\star\star}_{(i+1):M}) \in \{0, 1\}^{M-i} \times \{0, 1\}^{M-i}$ with marginals

$\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{w}_{1:(i-1)}, 0}$ and $\mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{w}_{1:(i-1)}, 1}$. For convenience, we define $(\boldsymbol{W}^\star, \boldsymbol{W}^{\star\star}) \in \{0, 1\}^M \times \{0, 1\}^M$,

where the first i elements are given by $\boldsymbol{W}^\star_{1:i} = (\boldsymbol{w}_{1:(i-1)}, 0)$ and $\boldsymbol{W}^{\star\star}_{1:i} = (\boldsymbol{w}_{1:(i-1)}, 1)$, re-

spectively. The basic coupling inequality (Lindvall, 2002, Theorem 5.2, p. 19) shows that

any coupling satisfies

$$\left\| \mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{w}_{1:(i-1)}, 0} - \mathbb{P}_{\boldsymbol{\theta}^\star, \boldsymbol{w}_{1:(i-1)}, 1} \right\|_{\mathrm{TV}} \ \leq \ \mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{w}_{1:(i-1)}}(\boldsymbol{W}^\star_{(i+1):M} \neq \boldsymbol{W}^{\star\star}_{(i+1):M}), \tag{D.5}$$

where $\|.\|_{\mathrm{TV}}$ denotes the total variance distance between probability measures. If the two sides in Equation (D.5) are equal, the coupling is called optimal. An optimal coupling is guaranteed to exist, but may not be unique (Lindvall, 2002, pp. 99–107). To prove Corollary 1, we need an upper bound on the spectral norm $\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2$ of the coupling matrix $\mathcal{D}(\boldsymbol{\theta}^\star)$, so we construct a coupling that is convenient but may not be optimal.

A coupling $\mathbb{Q}_{\boldsymbol{\theta}^\star,i,\boldsymbol{w}_{1:(i-1)}}$ of $\mathbb{P}_{\boldsymbol{\theta}^\star,\boldsymbol{w}_{1:(i-1)},0}$ and $\mathbb{P}_{\boldsymbol{\theta}^\star,\boldsymbol{w}_{1:(i-1)},1}$ can be constructed as follows:

Step 1: Set $\mathcal{U} = \{1,\dots,i\}$ and $\mathcal{K} = \{1,\dots,M\}$.

Step 2: Set $\mathcal{A} = \{j \in \mathcal{K}\setminus\mathcal{U} : (W_i, W_j) \in \mathcal{E}$ with $i \in \mathcal{U}$ and $j \in \mathcal{K}\setminus\mathcal{U}$ such that $W_j^\star \neq W_j^{\star\star}\}$.

(a) If $\mathcal{A} \neq \emptyset$, pick the smallest element $j \in \mathcal{A}$ and let $(W_j^\star, W_j^{\star\star})$ be distributed according to an optimal coupling of $\mathbb{P}_{\boldsymbol{\theta}^\star}(W_j = \cdot \mid \boldsymbol{W}_\mathcal{U} = \boldsymbol{w}_\mathcal{U}^\star)$ and $\mathbb{P}_{\boldsymbol{\theta}^\star}(W_j = \cdot \mid \boldsymbol{W}_\mathcal{U} = \boldsymbol{w}_\mathcal{U}^{\star\star})$.

(b) If $\mathcal{A} = \emptyset$, pick the smallest element $j \in \mathcal{K}\setminus\mathcal{U}$ and let $(W_j^\star, W_j^{\star\star})$ be distributed according to an optimal coupling of $\mathbb{P}_{\boldsymbol{\theta}^\star}(W_j = \cdot \mid \boldsymbol{W}_\mathcal{U} = \boldsymbol{w}_\mathcal{U}^\star)$ and $\mathbb{P}_{\boldsymbol{\theta}^\star}(W_j = \cdot \mid \boldsymbol{W}_\mathcal{U} = \boldsymbol{w}_\mathcal{U}^{\star\star})$.

Step 3: Replace $\mathcal{U}$ by $\mathcal{U} \cup \{j\}$ and repeat Step 2 until $\mathcal{K}\setminus\mathcal{U} = \emptyset$.

Based on $\mathbb{Q}_{\boldsymbol{\theta}^\star,i,\boldsymbol{w}_{1:(i-1)}}$, we construct a coupling matrix $\mathcal{D}(\boldsymbol{\theta}^\star) \in \mathbb{R}^{M\times M}$ with elements

$$
\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star) := \begin{cases} 0 & \text{if } i < j \\[2mm] 1 & \text{if } i = j \\[2mm] \displaystyle\max_{\boldsymbol{w}_{1:(i-1)} \in \mathcal{W}_{1:i-1}} \mathbb{Q}_{\boldsymbol{\theta}^\star,i,\boldsymbol{w}_{1:(i-1)}}(W_j^\star \neq W_j^{\star\star}) & \text{if } i > j. \end{cases}
$$

Overlapping subpopulations. To obtain convergence rates based on a single observation of dependent random variables $\boldsymbol{W}$, we need to control the dependence of $\boldsymbol{W}$ in the

form of $\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2$. In line with the simulation setting in Section 5, we therefore assume that overlapping subpopulations $\mathcal{A}_1, \mathcal{A}_2, \ldots$ characterize the neighborhoods. The neighborhood $\mathcal{N}_i$ of unit $i \in \mathcal{P}_N$ is then defined as

$$\mathcal{N}_i \;\coloneqq\; \{j \in \mathcal{P}_N : \text{ there exists } k \in \{1, 2, \ldots\} \text{ such that } i \in \mathcal{A}_k \text{ and } j \in \mathcal{A}_k\}. \qquad \text{(D.6)}$$

Let $\mathcal{G}_\mathcal{A}$ be a subpopulation graph with a set of vertices $\mathcal{V}_\mathcal{A} \coloneqq \{\mathcal{A}_1, \mathcal{A}_2, \ldots\}$ and a set of edges connecting distinct subpopulations $\mathcal{A}_k$ and $\mathcal{A}_l$ with $\mathcal{A}_k \cap \mathcal{A}_l \neq \emptyset$. Define

$$\mathcal{S}_{\mathcal{G}_\mathcal{A}, i, k} \;\coloneqq\; \{\mathcal{A}_j \in \mathcal{V}_\mathcal{A} \setminus \{\mathcal{A}_i\} : \; d_{\mathcal{G}_\mathcal{A}}(i, j) = k\}. \qquad \text{(D.7)}$$

Using the background introduced above, we restate Condition 3 more formally.

Condition 4: Dependence. *The population $\mathcal{P}$ consists of intersecting subpopulations $\mathcal{A}_1, \mathcal{A}_2, \ldots$, whose intersections are represented by subpopulation graph $\mathcal{G}_\mathcal{A}$. Let $D_N \in \{2, 3, \ldots\}$ be defined by (D.4) and $\mathcal{S}_{\mathcal{G}_\mathcal{A}, i, k}$ be defined by (D.7), and assume that*

$$\max_{k \in \{1, 2, \ldots\}} |\mathcal{S}_{\mathcal{G}_\mathcal{A}, k, l}| \;\leq\; \omega_1 + \frac{\omega_2}{2\, D_N^3} \log(l + 1), \qquad l = 1, 2, \ldots,$$

where $\omega_1 \geq 0$ and $0 \leq \omega_2 \leq \min\{\omega_1, 1/((\omega_1 + 1)\,|\log(1 - U)|)\}$ with $U \coloneqq (1 + \exp(-A))^{-1} > 0$. The constant $A > 0$ is identical to the constant A in Condition 2. In addition, for each unit $i \in \mathcal{P}_i$, the neighborhood $\mathcal{N}_i$ is defined by (D.6), and there exists a constant $B \in (0, +\infty)$ such that $\max_{1 \leq i \leq N} |\mathcal{N}_i| < B$.

The assumption $\max_{1 \leq i \leq N} |\mathcal{N}_i| < B$ implies that D_N is bounded above by a constant $D \in \{2, 3, \ldots\}$.

D.2   Proof of Corollary 1

To prove Corollary 1, define

$$\mathcal{H} \; := \; \mathcal{H}_1 \cap \mathcal{H}_2$$

$$\mathcal{H}_1 \; := \; \left\{ \boldsymbol{w} \in \mathcal{W} : \sum_{i=1}^{N} \|\boldsymbol{H}_{i,1}(\boldsymbol{w})\|_\infty \geq \frac{N}{2\,(1+\chi(\boldsymbol{\theta}^\star))^2} \right\}$$

$$\mathcal{H}_2 \; := \; \left\{ \boldsymbol{w} \in \mathcal{W} : \sum_{i=1}^{N} \|\boldsymbol{H}_{i,2}(\boldsymbol{w})\|_\infty \geq \frac{c^2\,N}{2\,(1+\chi(\boldsymbol{\theta}^\star))} \right\} \tag{D.8}$$

$$\boldsymbol{H}_{i,1}(\boldsymbol{w}) \; := \; (d_{i,1}(\boldsymbol{z}), \,\ldots,\, d_{i,i-1}(\boldsymbol{z}), \, d_{i,i+1}(\boldsymbol{z}), \,\ldots,\, d_{i,N}(\boldsymbol{z}))$$

$$\boldsymbol{H}_{i,2}(\boldsymbol{w}) \; := \; (c_{i,1}\, x_1^2\, z_{i,1}, \,\ldots,\, c_{i,i-1}\, x_{i-1}^2\, z_{i,i-1}, \, c_{i,i+1}\, x_{i+1}^2\, z_{i,i+1}, \,\ldots,\, c_{i,N}\, x_N^2\, z_{i,N})$$

and

$$\chi(\boldsymbol{\theta}^\star) \; := \; \exp(C\,D^2\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)), \tag{D.9}$$

where the constants $0 < c < C < \infty$ and $D \in \{2, 3, \ldots\}$ are identical to the corresponding constants defined in Condition 1 and Equation (D.4), respectively.

PROOF OF COROLLARY 1. We prove Corollary 1 using Theorem 1 in five steps:

Step 1: We bound

$$\mathbb{P}\left(\|\nabla_{\boldsymbol{\theta}}\,\ell(\boldsymbol{\theta}; \boldsymbol{W})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} - \mathbb{E}\,\nabla_{\boldsymbol{\theta}}\,\ell(\boldsymbol{\theta}; \boldsymbol{W})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star}\|_\infty < \frac{\rho_N}{2\,\Lambda_N(\boldsymbol{\theta}^\star)} \right) \; \geq \; 1 - \tau\left(\frac{\rho_N}{2\,\Lambda_N(\boldsymbol{\theta}^\star)} \right),$$

and choose ρ_N so that $1 - \tau(\rho_N/(2\,\Lambda_N(\boldsymbol{\theta}^\star))) \geq 1 - 2\,/\max\{N,\,p\}^2$.

Step 2: We show that $-\nabla_{\boldsymbol{\theta}}^2\,\ell(\boldsymbol{\theta}; \boldsymbol{w})$ is invertible for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$ and all $\boldsymbol{w} \in \mathcal{H}$.

Step 3: We prove that the event $\boldsymbol{W} \in \mathcal{H}$ occurs with probability at least $1 - \upsilon(\rho_N/(2\,\Lambda_N(\boldsymbol{\theta}^\star))) \geq 1 - 4\,/\max\{N,\,p\}^2$.

Step 4: We bound δ_N.

Step 5: We bound ρ_N.

The proof of Corollary 1 leverages auxiliary results supplied by Lemmas 3, 4, and 5, which show that there exists an integer $N_1 \in \{3, 4, \dots\}$ such that, for all $N > N_1$,

$$\Lambda_N(\boldsymbol{\theta}^\star) \quad \leq \quad C_1 \frac{\chi(\boldsymbol{\theta}^\star)^9}{N} \qquad \text{by Lemma 3}$$

$$\sqrt{N/2} \quad \leq \quad \Psi_N \leq C_2 \sqrt{N} \quad \text{by Lemma 4}$$

$$\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2 \quad \leq \quad C_3 \qquad \text{by Lemma 5,}$$

where $C_1 > 0$, $C_2 > 0$, and $C_3 \geq 1$ are constants.

Step 1: Since $\boldsymbol{W} \in \{0, 1\}^{M \times M}$, Lemma 6 of S25 establishes

$$\mathbb{P}\left(\|\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}; \boldsymbol{W})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} - \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}; \boldsymbol{W})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \|_\infty < \frac{\rho_N}{2\, \Lambda_N(\boldsymbol{\theta}^\star)} \right) \quad \geq \quad 1 - \tau\left(\frac{\rho_N}{2\, \Lambda_N(\boldsymbol{\theta}^\star)} \right),$$

where

$$\tau\left(\frac{\rho_N}{2\, \Lambda_N(\boldsymbol{\theta}^\star)} \right) \quad := \quad 2 \exp\left(-\frac{\rho_N^2}{32\, \Lambda_N(\boldsymbol{\theta}^\star)^2\, (1+D)^2\, \|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2^2\, \Psi_N^2} + \log p \right),$$

with $D \in \{2, 3, \dots\}$ defined in (D.4). Choosing

$$\rho_N \quad := \quad \sqrt{96}\, \Lambda_N(\boldsymbol{\theta}^\star)\, (1+D)\, \|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2\, \Psi_N\, \sqrt{\log \max\{N, p\}} \qquad \text{(D.10)}$$

implies that the event

$$\|\nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}; \boldsymbol{W})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} - \mathbb{E}\, \nabla_{\boldsymbol{\theta}}\, \ell(\boldsymbol{\theta}; \boldsymbol{W})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^\star} \|_\infty \quad < \quad \frac{\rho_N}{2\, \Lambda_N(\boldsymbol{\theta}^\star)}$$

occurs with probability at least

$$1 - \tau\left(\frac{\rho_N}{2\, \Lambda_N(\boldsymbol{\theta}^\star)} \right) \quad \geq \quad 1 - \frac{2}{\max\{N, p\}^2}.$$

Step 2: Let $\mathcal{H}$ be defined in (D.8). Lemma 3 establishes that $-\nabla_{\boldsymbol{\theta}}^2\, \ell(\boldsymbol{\theta}; \boldsymbol{w})$ is invertible for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$ and all $\boldsymbol{w} \in \mathcal{H}$.

14

Step 3: Lemma 11 shows that there exists an integer $N_2 \in \{3, 4, \ldots\}$ such that, for all $N > N_2$, the event $\boldsymbol{W} \in \mathcal{H}$ occurs with probability at least

$$1 - \upsilon(\delta_N) \;=\; 1 - \frac{4}{\max\{N, p\}^2}.$$

Step 4: The quantity

$$\delta_N \;:=\; \frac{\rho_N}{2 \, \Lambda_N(\boldsymbol{\theta}^\star)} \;=\; \sqrt{24} \, (1 + D) \, \|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2 \, \Psi_N \, \sqrt{\log \max\{N, p\}}$$

is bounded below by

$$\delta_N \;\geq\; \sqrt{24} \, D \, \sqrt{N/2} \, \sqrt{\log N} \;=\; \sqrt{12} \, D \, \sqrt{N \log N}$$

and is bounded above by

$$\delta_N \;\leq\; \sqrt{24} \, C_2 \, C_3 \, (2 \, D) \, \sqrt{N} \, \sqrt{2 \log N} \;=\; \sqrt{192} \, C_2 \, C_3 \, D \, \sqrt{N \log N},$$

using $D \in \{2, 3, \ldots\}$, $1 \leq \|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2 \leq C_3$, $\sqrt{N/2} \leq \Psi_N \leq C_2 \sqrt{N}$, and $\max\{N, p\} = p = N + 2$. Since $C_2 > 0$, $C_3 \geq 1$, and $D \in \{2, 3, \ldots\}$ defined in (D.4) are constants, there exist constants $0 < L \leq U < \infty$ such that

$$L \, \sqrt{N \log N} \;\leq\; \delta_N \;\leq\; U \, \sqrt{N \log N}.$$

Step 5: Substituting the bounds on $\Lambda_N(\boldsymbol{\theta}^\star)$, Ψ_N, and $\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2$ supplied by Lemmas 3, 4, and 5 into (D.10) reveals that

$$\begin{aligned}
\rho_N \;&:=\; \sqrt{96} \, \Lambda_N(\boldsymbol{\theta}^\star) \, (1 + D) \, \|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2 \, \Psi_N \sqrt{\log \max\{N, p\}} \\
&\leq\; \sqrt{96} \, C_1 \, C_2 \, C_3 \, (2 \, D) \, \frac{\chi(\boldsymbol{\theta}^\star)^9}{N} \, \sqrt{N \log(N + 2)} \quad\quad\quad \text{(D.11)} \\
&\leq\; \sqrt{768} \, C_1 \, C_2 \, C_3 \, D \, \chi(\boldsymbol{\theta}^\star)^9 \, \sqrt{\frac{\log N}{N}},
\end{aligned}$$

using $\max\{N, p\} = p = N + 2$ and $\log(N + 2) \leq \log(2 \, N) \leq 2 \log N$ $(N \geq 2)$. To bound $\chi(\boldsymbol{\theta}^\star)$, we invoke Condition 2:

$$\chi(\boldsymbol{\theta}^\star)^9 \;:=\; \exp(C \, D^2 \, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^9 \;\leq\; \exp(C \, D^2 \, (A + \epsilon^\star))^9 \;=\; \exp(9 \, C \, D^2 \, (A + \epsilon^\star)).$$

Define

$$K := \sqrt{768} \; C_1 \, C_2 \, C_3 \, D \, \exp(9 \, C \, D^2 \, (A + \epsilon^\star)) \; > \; 0.$$

Since A, C, C_1, C_2, C_3, D, and $\epsilon^\star$ are independent of N, so is K. We conclude that

$$\rho_N \; \leq \; K \, \sqrt{\frac{\log N}{N}} \; \to \; 0 \quad \text{as} \quad N \to \infty.$$

Conclusion. Theorem 1 implies that, for all $N > N_0 := \max\{N_1, N_2\}$, the random set $\widehat{\Theta}(\delta_N)$ is non-empty and satisfies

$$\widehat{\Theta}(\delta_N) \; \subseteq \; \mathcal{B}_\infty \left(\boldsymbol{\theta}^\star, \; K \, \sqrt{\frac{\log N}{N}} \right)$$

with probability at least

$$1 - \tau(\delta_N) - \upsilon(\delta_N) \; \geq \; 1 - \frac{6}{\max\{N, \, p\}^2} \; \geq \; 1 - \frac{6}{N^2},$$

using $\max\{N, \, p\}^2 = p^2 = (N+2)^2 \geq N^2$.

D.3   Statement and Proof of Corollary 2

If subpopulations do not overlap, $\|\boldsymbol{\theta}^\star\|_\infty$ can grow as a function N. Condition 5 details how fast $\|\boldsymbol{\theta}^\star\|_\infty$ can grow.

Condition 5. *The parameter space is* $\boldsymbol{\Theta} = \mathbb{R}^{N+2}$ *and the data-generating parameter vector* $\boldsymbol{\theta}^\star \in \mathbb{R}^{N+2}$ *satisfies*

$$\|\boldsymbol{\theta}^\star\|_\infty \; \leq \; \frac{E + \vartheta \, \log N}{C \, D^2} - \epsilon^\star,$$

where $E \geq 0$ *and* $\vartheta \in [0, \, 1/18)$ *are constants,* $C > 0$ *is identical to the constant* C *in Condition 1,* $D \in \{2, 3, \ldots\}$ *is identical to the constant* D *in* (D.4), *and* $\epsilon^\star > 0$ *is identical to the constant* $\epsilon^\star$ *in the definition of* $\Lambda_N(\boldsymbol{\theta}^\star)$ *in Section 4.*

16

Corollary 2 replaces Condition 2 by Condition 5. Resulting from this, the constant U coming up in Condition 4 is redefined as $U := (1 + \exp(-D))^{-1} > 0$.

Corollary 2. *Consider a single observation of dependent responses and connections $(\boldsymbol{Y}, \boldsymbol{Z})$ generated by the model with parameter vector $\boldsymbol{\theta}^\star := (\alpha_{\mathcal{Z},1}^\star, \ldots, \alpha_{\mathcal{Z},N}^\star, \gamma_{\mathcal{Z},\mathcal{Z}}^\star, \gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}}^\star) \in \mathbb{R}^{N+2}$. If Conditions 1, 4, and 5 are satisfied with $\vartheta \in [0, 1/18)$, there exist constants $K \in (0, +\infty)$ and $0 < L \le U < +\infty$ along with an integer $N_0 \in \{3, 4, \ldots\}$ such that, for all $N > N_0$, the quantity δ_N satisfies*

$$L \sqrt{N \log N} \;\le\; \delta_N \;\le\; U \sqrt{N \log N},$$

and the random set $\widehat{\boldsymbol{\Theta}}(\delta_N)$ is non-empty and satisfies

$$\widehat{\boldsymbol{\Theta}}(\delta_N) \;\subseteq\; \mathcal{B}_\infty\left(\boldsymbol{\theta}^\star, \; K \sqrt{\frac{\log N}{N^{1-18\,\vartheta}}}\right)$$

with probability at least $1 - 6/N^2$.

PROOF OF COROLLARY 2. The proof of Corollary 2 resembles the proof of Corollary 1, with Condition 2 replaced by Condition 5. The proof of Corollary 1 shows that

$$\rho_N \;\le\; \sqrt{768} \; C_1 \, C_2 \, C_3 \, D \, \chi(\boldsymbol{\theta}^\star)^9 \sqrt{\frac{\log N}{N}},$$

where the constants $C_1 > 0$, $C_2 > 0$, $C_3 \ge 1$, and $D \in \{2, 3, \ldots\}$ are defined in Lemmas 3, 4, and 5, and Equation (D.4), respectively. Condition 5 implies that

$$\chi(\boldsymbol{\theta}^\star)^9 := \exp(C\,D^2\,(\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star))^9 \;\le\; \exp\left(C\,D^2\left(\frac{E + \vartheta \log N}{C\,D^2}\right)\right)^9 \;=\; \exp(9\,E)\,N^{9\,\vartheta},$$

which in turn implies that

$$\rho_N \;\le\; \sqrt{768} \; C_1 \, C_2 \, C_3 \, D \, \exp(9\,E) \sqrt{\frac{\log N}{N^{1-18\,\vartheta}}} \;=\; K \sqrt{\frac{\log N}{N^{1-18\,\vartheta}}},$$

where $K := \sqrt{768} \; C_1 \, C_2 \, C_3 \, D \, \exp(9\,E) > 0$. The remainder of the proof of Corollary 2 resembles the proof of Corollary 1. We conclude that there exists an integer $N_0 \in \{3, 4, \ldots\}$

such that, for all $N > N_0$, the random set $\widehat{\boldsymbol{\Theta}}(\delta_N)$ is non-empty and satisfies

$$\widehat{\boldsymbol{\Theta}}(\delta_N) \;\subseteq\; \mathcal{B}_\infty\left(\boldsymbol{\theta}^\star,\; K\,\sqrt{\frac{\log N}{N^{1-18\,\vartheta}}}\right)$$

with probability at least $1 - 6\,/N^2$.

D.4 Bounding $\Lambda_N(\boldsymbol{\theta}^\star)$

Lemma 3. *Consider the model of Corollary 1. If Conditions 1 and 4 are satisfied along with either Condition 2 or Condition 5 with $\vartheta \in [0, 1/18)$, there exists a constant $C_1 > 0$ along with an integer $N_0 \in \{3, 4, \dots\}$ such that, for all $N > N_0$,*

- $(-\nabla_{\boldsymbol{\theta}}^2\, \ell(\boldsymbol{\theta};\, \boldsymbol{w}))^{-1}$ *is invertible for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ and all $\boldsymbol{w} \in \mathcal{H}$,*

- *the event $\boldsymbol{W} \in \mathcal{H}$ occurs with probability at least $1 - 4\,/\max\{N,\, p\}^2$,*

- $\Lambda_N(\boldsymbol{\theta}^\star) \;:=\; \sup\limits_{\boldsymbol{w} \in \mathcal{H}}\; \sup\limits_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)}\; \|(-\nabla_{\boldsymbol{\theta}}^2\, \ell(\boldsymbol{\theta};\, \boldsymbol{w}))^{-1}\|_\infty \;\leq\; C_1\, \dfrac{\chi(\boldsymbol{\theta}^\star)^9}{N},$

where $\mathcal{H}$ is defined in (D.8) *and $\chi(\boldsymbol{\theta}^\star)$ is defined in* (D.9).

PROOF OF LEMMA 3. We first partition $-\nabla_{\boldsymbol{\theta}}^2\, \ell(\boldsymbol{\theta};\, \boldsymbol{w})$ in accordance with $\boldsymbol{\theta} := (\boldsymbol{\theta}_1,\, \boldsymbol{\theta}_2)$, given by $\boldsymbol{\theta}_1 := (\alpha_{\mathcal{Z},1}, \dots, \alpha_{\mathcal{Z},N}) \in \mathbb{R}^N$ and $\boldsymbol{\theta}_2 := (\gamma_{\mathcal{Z},\mathcal{Z}}, \gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}}) \in \mathbb{R}^2$:

$$-\nabla_{\boldsymbol{\theta}}^2\, \ell(\boldsymbol{\theta};\, \boldsymbol{w}) \;:=\; \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta},\, \boldsymbol{w}) & \boldsymbol{C}(\boldsymbol{\theta},\, \boldsymbol{w}) \\[2mm] \boldsymbol{C}(\boldsymbol{\theta},\, \boldsymbol{w})^\top & \boldsymbol{B}(\boldsymbol{\theta},\, \boldsymbol{w}) \end{pmatrix}, \tag{D.12}$$

where the matrices $\boldsymbol{A}(\boldsymbol{\theta},\, \boldsymbol{w}) \in \mathbb{R}^{N \times N}$ and $\boldsymbol{B}(\boldsymbol{\theta},\, \boldsymbol{w}) \in \mathbb{R}^{2 \times 2}$ define the covariance matrices of the sufficient statistics corresponding to the parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. Define $\boldsymbol{C}(\boldsymbol{\theta},\, \boldsymbol{w}) := (\boldsymbol{C}_1(\boldsymbol{\theta},\, \boldsymbol{w}), \boldsymbol{C}_2(\boldsymbol{\theta},\, \boldsymbol{w})) \in \mathbb{R}^{N \times 2}$, where $\boldsymbol{C}_1(\boldsymbol{\theta},\, \boldsymbol{w}) \in \mathbb{R}^N$ and $\boldsymbol{C}_2(\boldsymbol{\theta},\, \boldsymbol{w}) \in \mathbb{R}^N$ are the covariances of the degree terms with the transitive connection term with weight $\gamma_{\mathcal{Z},\mathcal{Z}}$ and spillover term with weight $\gamma_{\mathcal{X},\mathcal{Y},\mathcal{Z}}$, respectively.

We wish to bound the infinity norm of $(-\nabla_{\boldsymbol{\theta}}^2 \ell(\boldsymbol{\theta}; \boldsymbol{w}))^{-1}$, given by

$$(-\nabla_{\boldsymbol{\theta}}^2 \ell(\boldsymbol{\theta}, \boldsymbol{w}))^{-1} = \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w}) & \boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w}) \\ \boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w})^\top & \boldsymbol{B}(\boldsymbol{\theta}, \boldsymbol{w}) \end{pmatrix}^{-1}$$

$$= \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1} & \boldsymbol{0}_{N,2} \\ \boldsymbol{0}_{2,N} & \boldsymbol{0}_{2,2} \end{pmatrix}$$

$$+ \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1} \boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w})^\top \\ -\boldsymbol{I}_{2,2} \end{pmatrix} \boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{w})^{-1} \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1} \boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w})^\top \\ -\boldsymbol{I}_{2,2} \end{pmatrix}^\top,$$

where $\boldsymbol{0}_{a,b} := \mathrm{diag}(0,\ldots,0) \in \{0,1\}^{a \times b}$ and $\boldsymbol{I}_{a,b} := \mathrm{diag}(1,\ldots,1) \in \{0,1\}^{a \times b}$ $(a, b \in \{1, 2, \ldots\})$ are diagonal matrices, and

$$\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{w}) := \boldsymbol{B}(\boldsymbol{\theta}, \boldsymbol{w}) - \boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w})^\top \boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1} \boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w})$$

is the Schur complement of $-\nabla_{\boldsymbol{\theta}}^2 \ell(\boldsymbol{\theta}, \boldsymbol{w})$ with respect to the block $\boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})$.

The ℓ_∞-induced norm is submultiplicative, so

$$\|(-\nabla_{\boldsymbol{\theta}}^2 \ell(\boldsymbol{\theta}, \boldsymbol{w}))^{-1}\|_\infty$$

$$\leq \|\boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty$$

$$+ \left\| \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1} \boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w}) \\ -\boldsymbol{I}_{p,p} \end{pmatrix} \right\|_\infty \|\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty \left\| \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1} \boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w}) \\ -\boldsymbol{I}_{p,p} \end{pmatrix}^\top \right\|_\infty$$

$$\leq \|\boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty + \max\{1, \|\boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1} \boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w})\|_\infty\}$$

$$\times \|\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty (\|\boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w})^\top \boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty + 1)$$

$$\leq \|\boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty + \max\{1, \|\boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty \|\boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w})\|_\infty\}$$

$$\times \|\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty (\|\boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w})^\top\|_\infty \|\boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty + 1).$$

$$(\text{D}.13)$$

We bound the terms $\|\boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty$, $\|\boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w})^\top\|_\infty$, and $\|\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty$ one by one.

Bounding $\|A(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty$. The proof of Lemma 9 in S25 shows that

$$\|A(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty \;\leq\; \frac{18\, \chi(\boldsymbol{\theta}^\star)^2}{N} \tag{D.14}$$

for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$, where $\chi(\boldsymbol{\theta}^\star)$ is an upper bound on the inverse standard deviation of connections $Z_{i,j}$ of pairs of units $\{i,j\} \subset \mathcal{P}_N$ with $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ conditional on $\boldsymbol{X}$, $\boldsymbol{Y}$, $\boldsymbol{Z}_{-\{i,j\}}$. Under the model considered here, the conditional distribution of $Z_{i,j}$ is Bernoulli, as shown in Section 2.2.2. Therefore, $\mathbb{V}_{z,i,j}(Z_{i,j})$ is given by

$$\mathbb{V}_{z,i,j}(Z_{i,j}) \;=\; \mathbb{P}(Z_{i,j}=1 \mid \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}}) \times (1 - \mathbb{P}(Z_{i,j}=1 \mid \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}})).$$

Applying the bounds on $\mathbb{P}(Z_{i,j}=1 \mid \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}})$ supplied by Lemma 7 gives

$$\mathbb{V}_{z,i,j}(Z_{i,j}) \;\geq\; \frac{1}{(\exp(C\, D^2\, \|\boldsymbol{\theta}\|_\infty))^2} \;\geq\; \frac{1}{(\exp(C\, D^2\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)))^2}, \tag{D.15}$$

provided $D \in \{2, 3, \ldots\}$, where D corresponds to the constant D defined in (D.4) and C corresponds to the constant C in Condition 1. For the second inequality of (D.15), we use the fact that $\|\boldsymbol{\theta}\|_\infty \leq \|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star$ for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$. With

$$\chi(\boldsymbol{\theta}^\star) \;:=\; \exp(C\, D^2\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)),$$

we therefore deduce that $\chi(\boldsymbol{\theta}^\star)$ is an bound on the inverse standard deviation of connections $Z_{i,j}$:

$$\chi(\boldsymbol{\theta}^\star) \;\geq\; \frac{1}{\sqrt{\mathbb{V}_{z,i,j}(Z_{i,j})}}.$$

Bounding $\|C(\boldsymbol{\theta}, \boldsymbol{w})^\top\|_\infty$. Define $C(\boldsymbol{\theta}, \boldsymbol{w}) := (C_1(\boldsymbol{\theta}, \boldsymbol{w}), C_2(\boldsymbol{\theta}, \boldsymbol{w}))$, where $C_1(\boldsymbol{\theta}, \boldsymbol{w}) \in \mathbb{R}^N$ and $C_2(\boldsymbol{\theta}, \boldsymbol{w}) \in \mathbb{R}^N$ are the covariance terms of the degree terms with the sufficient statistics pertaining to the transitive connection term weighted by $\gamma_{z,z}$ and the spillover term weighted by $\gamma_{x,y,z}$, respectively. Then

$$\|C(\boldsymbol{\theta}, \boldsymbol{w})^\top\|_\infty \;\leq\; \|C_1(\boldsymbol{\theta}, \boldsymbol{w})\|_\infty + \|C_2(\boldsymbol{\theta}, \boldsymbol{w})\|_\infty.$$

We bound the terms $\|\boldsymbol{C}_1(\boldsymbol{\theta},\,\boldsymbol{w})\|_\infty$ and $\|\boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w})\|_\infty$ one by one.

By Lemma 13 of S25, $\|\boldsymbol{C}_1(\boldsymbol{\theta},\,\boldsymbol{w})\|_\infty \leq 3\,D^3$. The term $\boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w}) \coloneqq (C_{2,1}(\boldsymbol{\theta},\,\boldsymbol{w}),\ldots,C_{2,N}(\boldsymbol{\theta},\,\boldsymbol{w})) \in \mathbb{R}^N$ refers to the covariances between the degrees $b_i(\boldsymbol{x},\,\boldsymbol{y},\,\boldsymbol{z})$ of units $i \in \{1,\ldots,N\}$ and $b_{N+2}(\boldsymbol{x},\,\boldsymbol{y},\,\boldsymbol{z})$. An upper bound on t-th element of $\boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w})$ can be obtained by

$$
\begin{aligned}
|C_{2,t}(\boldsymbol{\theta},\,\boldsymbol{w})| &= \left| \sum_{i=1}^N \sum_{j=i+1}^N \mathbb{C}_{\mathrm{z},i,j}\left(b_t(\boldsymbol{x},\,\boldsymbol{y},\,\boldsymbol{Z}),\, b_{N+2}(\boldsymbol{x},\,\boldsymbol{y},\,\boldsymbol{Z})\right) \right| \\
&= \left| \sum_{i=1}^N \sum_{j=i+1}^N \mathbb{C}_{\mathrm{z},i,j}\left(\sum_{h\neq t} Z_{h,t},\ \sum_{h=1}^N \sum_{k=h+1}^N c_{h,k}\left(x_h\,y_k + x_k\,y_h\right) Z_{h,k}\right) \right| \\
&= \left| \sum_{i\neq t:\ \mathcal{N}_i \cap \mathcal{N}_t \neq \emptyset} \mathbb{C}_{\mathrm{z},i,t}\left(Z_{i,t},\ (x_i\,y_t + y_i\,x_t)\,Z_{i,t}\right) \right| \\
&= \left| \sum_{i\neq t:\ \mathcal{N}_i \cap \mathcal{N}_t \neq \emptyset} (x_i\,y_t + y_i\,x_t)\,\mathbb{V}_{\mathrm{z},i,t}\left(Z_{i,t}\right) \right| \\
&\leq \left| \frac{C}{2} \sum_{i\neq t:\ \mathcal{N}_i \cap \mathcal{N}_t \neq \emptyset}^N 1 \right| \ \leq\ C\,D^2,
\end{aligned}
\tag{D.16}
$$

where C corresponds to the constant from Condition 1 and D is defined in (D.4). On the third line, note that $b_t(\boldsymbol{x},\,\boldsymbol{y},\,\boldsymbol{z})$ only depends on connection $Z_{i,j}$ if $t \in \{i,j\}$. Therefore, the covariance of $b_t(\boldsymbol{x},\,\boldsymbol{y},\,\boldsymbol{z})$ with respect to any other connection is 0. The first inequality follows from the observation that $x_i\,y_j + x_j\,y_i \leq 2\,C$ and $\mathbb{V}_{\mathrm{z},i,j}\left(Z_{i,j}\right) \leq 1/4$, which follows from $0 \leq x_i \leq C < \infty$ by Condition 1 and $Y_i \in \{0,1\}$. The second inequality follows from Lemma 15 in S25 bounding the pairs of units i and t such that $\mathcal{N}_i \cap \mathcal{N}_t \neq \emptyset$ from above by D^2. Since the bound from (D.16) holds for all $t \in \{1,\ldots,N\}$, we obtain $\|\boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w})\|_\infty \leq C\,D^2$. Taken together,

$$
\left\|\boldsymbol{C}(\boldsymbol{\theta},\,\boldsymbol{w})^\top\right\|_\infty \ \leq\ 3\,D^3 + C\,D^2 \ \leq\ \max\{3,\,C\}\,D^3.
\tag{D.17}
$$

21

Bounding $\|V(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty$. Write

$$\boldsymbol{B}(\boldsymbol{\theta}, \boldsymbol{w}) \;\; \coloneqq \;\; \begin{pmatrix} B_{1,1}(\boldsymbol{\theta}, \boldsymbol{w}) & B_{1,2}(\boldsymbol{\theta}, \boldsymbol{w}) \\ B_{1,2}(\boldsymbol{\theta}, \boldsymbol{w}) & B_{2,2}(\boldsymbol{\theta}, \boldsymbol{w}) \end{pmatrix}$$

$$\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{w}) \;\; \coloneqq \;\; \begin{pmatrix} V_{1,1}(\boldsymbol{\theta}, \boldsymbol{w}) & V_{1,2}(\boldsymbol{\theta}, \boldsymbol{w}) \\ V_{1,2}(\boldsymbol{\theta}, \boldsymbol{w}) & V_{2,2}(\boldsymbol{\theta}, \boldsymbol{w}) \end{pmatrix}.$$

The elements of $\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{w})$ are then given by

$$V_{i,j}(\boldsymbol{\theta}, \boldsymbol{w}) \;\; = \;\; B_{i,j}(\boldsymbol{\theta}, \boldsymbol{w}) - \boldsymbol{C}_i(\boldsymbol{\theta}, \boldsymbol{w})^\top \boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1} \boldsymbol{C}_j(\boldsymbol{\theta}, \boldsymbol{w}).$$

The inverse of $\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{w})$ is

$$\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{w})^{-1} \;\; = \;\; \frac{1}{V_{1,1}(\boldsymbol{\theta}, \boldsymbol{w})\, V_{2,2}(\boldsymbol{\theta}, \boldsymbol{w}) - V_{1,2}(\boldsymbol{\theta}, \boldsymbol{w})^2} \begin{pmatrix} V_{2,2}(\boldsymbol{\theta}, \boldsymbol{w}) & -V_{1,2}(\boldsymbol{\theta}, \boldsymbol{w}) \\ -V_{1,2}(\boldsymbol{\theta}, \boldsymbol{w}) & V_{1,1}(\boldsymbol{\theta}, \boldsymbol{w}) \end{pmatrix},$$

implying that

$$\|(\boldsymbol{V}(\boldsymbol{\theta}, \boldsymbol{w}))^{-1}\|_\infty \;\; \leq \;\; \frac{\max\left\{V_{1,1}(\boldsymbol{\theta}, \boldsymbol{w}), V_{2,2}(\boldsymbol{\theta}, \boldsymbol{w})\right\} + |V_{1,2}(\boldsymbol{\theta}, \boldsymbol{w})|}{|V_{1,1}(\boldsymbol{\theta}, \boldsymbol{w})\, V_{2,2}(\boldsymbol{\theta}, \boldsymbol{w}) - V_{1,2}(\boldsymbol{\theta}, \boldsymbol{w})^2|}. \tag{D.18}$$

Invoking the inequalities from (D.14) and (D.17), we obtain for $i, j \in \{1, 2\}$

$$|\boldsymbol{C}_i(\boldsymbol{\theta}, \boldsymbol{w})^\top \boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1} \boldsymbol{C}_j(\boldsymbol{\theta}, \boldsymbol{w})|$$

$$\leq \;\; N \, \|\boldsymbol{C}_i(\boldsymbol{\theta}, \boldsymbol{w})\|_\infty \|\boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty \|\boldsymbol{C}_i(\boldsymbol{\theta}, \boldsymbol{w})\|_\infty$$

$$\leq \;\; N \, \|\boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w})\|_\infty \|\boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty \|\boldsymbol{C}(\boldsymbol{\theta}, \boldsymbol{w})\|_\infty \tag{D.19}$$

$$\leq \;\; 18 \max\{9, C^2\}\, D^6 \, \chi(\boldsymbol{\theta}^\star)^2,$$

where D corresponds to the constant D defined in (D.4) and C corresponds to the constant C from Condition 1.

By applying Lemma 9 along with (D.19), we get for $i, j \in \{1, 2\}$

$$|V_{i,j}(\boldsymbol{\theta}, \boldsymbol{w})| \;\; = \;\; |B_{i,j}(\boldsymbol{\theta}, \boldsymbol{w})| + |\boldsymbol{C}_i(\boldsymbol{\theta}, \boldsymbol{w})^\top \boldsymbol{A}(\boldsymbol{\theta}, \boldsymbol{w})^{-1} \boldsymbol{C}_j(\boldsymbol{\theta}, \boldsymbol{w})|$$

$$\leq \;\; \max\{1, C^2\} \frac{N\, D^5}{4} + 18 \max\{9, C^2\}\, D^6 \, \chi(\boldsymbol{\theta}^\star)^2$$

$$\leq \;\; \max\{9, C^2\}\, D^5 \left(\frac{N}{4} + 18\, D\, \chi(\boldsymbol{\theta}^\star)^2 \right)$$

22

Thus, the numerator of (D.18) is bounded above by

$$\max\left\{V_{1,1}(\boldsymbol{\theta},\,\boldsymbol{w}),\,V_{2,2}(\boldsymbol{\theta},\,\boldsymbol{w})\right\} + |V_{1,2}(\boldsymbol{\theta},\,\boldsymbol{w})|$$
$$\leq\ \max\{9,\,C^2\}\,D^5\left(\frac{N}{2} + 36\,D\,\chi(\boldsymbol{\theta}^\star)^2\right). \tag{D.20}$$

The denominator of (D.18), which is the determinant of $\boldsymbol{V}(\boldsymbol{\theta},\,\boldsymbol{w})$, is

$$V_{1,1}(\boldsymbol{\theta},\,\boldsymbol{w})\,V_{2,2}(\boldsymbol{\theta},\,\boldsymbol{w}) - V_{1,2}(\boldsymbol{\theta},\,\boldsymbol{w})^2$$
$$=\ (B_{1,1}(\boldsymbol{\theta},\,\boldsymbol{w}) - \boldsymbol{C}_1(\boldsymbol{\theta},\,\boldsymbol{w})^\top\boldsymbol{A}(\boldsymbol{\theta},\,\boldsymbol{w})^{-1}\boldsymbol{C}_1(\boldsymbol{\theta},\,\boldsymbol{w}))$$
$$\times\ (B_{2,2}(\boldsymbol{\theta},\,\boldsymbol{w}) - \boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w})^\top\boldsymbol{A}(\boldsymbol{\theta},\,\boldsymbol{w})^{-1}\boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w}))$$
$$-\ (B_{1,2}(\boldsymbol{\theta},\,\boldsymbol{w}) - \boldsymbol{C}_1(\boldsymbol{\theta},\,\boldsymbol{w})^\top\boldsymbol{A}(\boldsymbol{\theta},\,\boldsymbol{w})^{-1}\boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w}))^2$$
$$=\ B_{1,1}(\boldsymbol{\theta},\,\boldsymbol{w})\,B_{2,2}(\boldsymbol{\theta},\,\boldsymbol{w}) - B_{1,1}(\boldsymbol{\theta},\,\boldsymbol{w})\,\boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w})^\top\boldsymbol{A}(\boldsymbol{\theta},\,\boldsymbol{w})^{-1}\boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w})$$
$$-\ B_{2,2}(\boldsymbol{\theta},\,\boldsymbol{w})\,\boldsymbol{C}_1(\boldsymbol{\theta},\,\boldsymbol{w})^\top\boldsymbol{A}(\boldsymbol{\theta},\,\boldsymbol{w})^{-1}\boldsymbol{C}_1(\boldsymbol{\theta},\,\boldsymbol{w})$$
$$+\ (\boldsymbol{C}_1(\boldsymbol{\theta},\,\boldsymbol{w})^\top\boldsymbol{A}(\boldsymbol{\theta},\,\boldsymbol{w})^{-1}\boldsymbol{C}_1(\boldsymbol{\theta},\,\boldsymbol{w}))\,(\boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w})^\top\boldsymbol{A}(\boldsymbol{\theta},\,\boldsymbol{w})^{-1}\boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w})) - B_{1,2}(\boldsymbol{\theta},\,\boldsymbol{w})^2$$
$$+\ 2\,B_{1,2}(\boldsymbol{\theta},\,\boldsymbol{w})\,(\boldsymbol{C}_1(\boldsymbol{\theta},\,\boldsymbol{w})^\top\boldsymbol{A}(\boldsymbol{\theta},\,\boldsymbol{w})^{-1}\boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w})) - (\boldsymbol{C}_1(\boldsymbol{\theta},\,\boldsymbol{w})^\top\boldsymbol{A}(\boldsymbol{\theta},\,\boldsymbol{w})^{-1}\boldsymbol{C}_2(\boldsymbol{\theta},\,\boldsymbol{w}))^2.$$

Applying the property of positive semidefinite matrices $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ that $(\boldsymbol{a}^\top\boldsymbol{P}\,\boldsymbol{a})\,(\boldsymbol{b}^\top\boldsymbol{P}\,\boldsymbol{b}) \geq (\boldsymbol{a}^\top\boldsymbol{P}\,\boldsymbol{b})^2$ is true for all vectors $\boldsymbol{a} \in \mathbb{R}^n$ and $\boldsymbol{b} \in \mathbb{R}^n$ $(n \geq 1)$, we obtain

$$V_{1,1}(\boldsymbol{\theta},\,\boldsymbol{w})\,V_{2,2}(\boldsymbol{\theta},\,\boldsymbol{w}) - V_{1,2}(\boldsymbol{\theta},\,\boldsymbol{w})^2$$
$$\geq\ B_{1,1}(\boldsymbol{\theta},\,\boldsymbol{w})\,B_{2,2}(\boldsymbol{\theta},\,\boldsymbol{w}) - B_{1,2}(\boldsymbol{\theta},\,\boldsymbol{w})^2$$
$$-\ 4\max_{i,j}|B_{i,j}(\boldsymbol{\theta},\,\boldsymbol{w})|\,\max_{i,j}|\boldsymbol{C}_i(\boldsymbol{\theta},\,\boldsymbol{w})^\top\boldsymbol{A}(\boldsymbol{\theta},\,\boldsymbol{w})^{-1}\boldsymbol{C}_j(\boldsymbol{\theta},\,\boldsymbol{w})|$$
$$\geq\ B_{1,1}(\boldsymbol{\theta},\,\boldsymbol{w})\,B_{2,2}(\boldsymbol{\theta},\,\boldsymbol{w}) - B_{1,2}(\boldsymbol{\theta},\,\boldsymbol{w})^2 - 18\,\max\{81,\,C^4\}\,N\,D^{11}\,\chi(\boldsymbol{\theta}^\star)^2$$
$$=\ U(\boldsymbol{\theta},\,\boldsymbol{w}) - 18\,\max\{81,\,C^4\}\,N\,D^{11}\,\chi(\boldsymbol{\theta}^\star)^2,$$

where

$$U(\boldsymbol{\theta},\,\boldsymbol{w})\ \coloneqq\ B_{1,1}(\boldsymbol{\theta},\,\boldsymbol{w})\,B_{2,2}(\boldsymbol{\theta},\,\boldsymbol{w}) - B_{1,2}(\boldsymbol{\theta},\,\boldsymbol{w})^2. \tag{D.21}$$

The final inequality follows from invoking (D.19) along with Lemma 9.

For (D.21), we obtain

$$
\begin{aligned}
U(\boldsymbol{\theta}, \boldsymbol{w}) \;=\;& B_{1,1}(\boldsymbol{\theta}, \boldsymbol{w})\, B_{2,2}(\boldsymbol{\theta}, \boldsymbol{w}) - B_{1,2}(\boldsymbol{\theta}, \boldsymbol{w})^2 \\[2mm]
=\;& \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j}\left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \right) \\[2mm]
& \times\; \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{y,i}\left(b_{N+2}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{z})\right) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j}\left(b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \right) \\[2mm]
& -\; \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{C}_{z,i,j}\left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}),\, b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \right)^{2} \\[2mm]
=\;& \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j}\left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \right) \left(\sum_{i=1}^{N} \mathbb{V}_{y,i}\left(b_{N+2}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{z})\right) \right) \\[2mm]
& +\; \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j}\left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \right) \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j}\left(b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \right) \\[2mm]
& -\; \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{C}_{z,i,j}\left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}),\, b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \right)^{2}.
\end{aligned}
$$

Next, we show that the third term

$$
\left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{C}_{z,i,j}\left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}), b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \right)^{2}
$$

is smaller than the second term

$$
\left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j}\left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \right) \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j}\left(b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \right).
$$

Define

$$
u_{1,i,j} := \sqrt{\mathbb{V}_{z,i,j}\left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right)} \quad \text{and} \quad u_{2,i,j} := \sqrt{\mathbb{V}_{z,i,j}\left(b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right)}, \quad i = 1, \ldots, N.
$$

Then the second term can be restated as follows:

$$
\begin{aligned}
& \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j}\left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \right) \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j}\left(b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \right) \\[2mm]
&= \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} u_{1,i,j}^{2} \right) \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} u_{2,i,j}^{2} \right),
\end{aligned}
$$

while the third term is

$$\left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{C}_{z,i,j} \left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}), b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}) \right) \right)^2$$

$$\leq \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \left| \mathbb{C}_{z,i,j} \left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}), b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}) \right) \right| \right)^2$$

$$\leq \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \sqrt{\mathbb{V}_{z,i,j} \left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}) \right) \mathbb{V}_{z,i,j} \left(b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}) \right)} \right)^2$$

$$= \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} u_{1,i,j} \, u_{2,i,j} \right)^2$$

$$\leq \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} u_{1,i,j}^2 \right) \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} u_{2,i,j}^2 \right),$$

where the Cauchy-Schwarz inequality is invoked on the third and last line. This translates to the following lower bound on $U(\boldsymbol{\theta}, \boldsymbol{w})$:

$$U(\boldsymbol{\theta}, \boldsymbol{w}) = B_{1,1}(\boldsymbol{\theta}, \boldsymbol{w}) \, B_{2,2}(\boldsymbol{\theta}, \boldsymbol{w}) - B_{1,2}(\boldsymbol{\theta}, \boldsymbol{w})^2$$

$$\geq \left(\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j} \left(b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}) \right) \right) \left(\sum_{i=1}^{N} \mathbb{V}_{y,i} \left(b_{N+2}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{z}) \right) \right)$$

$$\geq \left(\sum_{i=1}^{N} \frac{\|\boldsymbol{H}_{i,1}(\boldsymbol{w})\|_{\infty}}{(1 + \chi(\boldsymbol{\theta}^{\star}))^2} \right) \sum_{i=1}^{N} \left(\sum_{j \neq i}^{N} c_{i,j} \, x_j \, z_{i,j} \right)^2 \mathbb{V}_{y,i} \left(Y_i \right),$$

where $\boldsymbol{H}_{i,1}(\boldsymbol{w})$ is defined in (D.8) and the function $c_{i,j}$ is defined in (D.2). For the second inequality, we use the result from the proof of Lemma 13 in S25, which implies that

$$\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j} \left(b_{N+1}(\boldsymbol{y}, \boldsymbol{Z}) \right) \geq \sum_{i=1}^{N} \sum_{j=i+1}^{N} c_{i,j} \, d_{i,j}(\boldsymbol{z}) \, \mathbb{V}_{z,i,j} \left(Z_{i,j} \right) \tag{D.22}$$

where the function $d_{i,j}(\boldsymbol{Z})$ is defined in (D.2). By Lemma 7, we get

$$\mathbb{V}_{z,i,j}(Z_{i,j}) = \mathbb{P}(Z_{i,j} \mid \boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z}) \times \left(1 - \mathbb{P}(Z_{i,j} \mid \boldsymbol{x}, \boldsymbol{y}_{-i}, \boldsymbol{z}) \right) \geq \frac{1}{(1 + \chi(\boldsymbol{\theta}^{\star}))^2},$$

where $\chi(\boldsymbol{\theta}^\star)$ is defined in (D.9). When combined with (D.22), this results in

$$\sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j}\left(b_{N+1}(\boldsymbol{y}, \boldsymbol{Z})\right) \geq \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} c_{i,j}\, d_{i,j}(\boldsymbol{z})}{(1 + \chi(\boldsymbol{\theta}^\star))^2}$$

$$\geq \sum_{i=1}^{N} \frac{\|\boldsymbol{H}_{i,1}(\boldsymbol{w})\|_\infty}{2\,(1 + \chi(\boldsymbol{\theta}^\star))^2},$$

where the second inequality is again from the proof of Lemma 13 in S25.

By applying Lemma 8 and expanding the quadratic term, we obtain

$$\begin{aligned}
U(\boldsymbol{\theta}, \boldsymbol{w}) \;\geq\; & \left(\sum_{i=1}^{N} \frac{\|\boldsymbol{H}_{i,1}(\boldsymbol{w})\|_\infty}{2\,(1 + \chi(\boldsymbol{\theta}^\star))^2}\right) \sum_{i=1}^{N} \mathbb{V}_{y,i}\left(Y_i\right) \\
& \times \left(\sum_{j=i+1}^{N} x_j^2\, c_{i,j}\, z_{i,j} + \sum_{h=1}^{N} \sum_{k \neq h}^{N} c_{i,h}\, c_{i,k}\, x_h\, x_k\, z_{i,h}\, z_{i,k}\right) \\
\geq\; & \left(\sum_{i=1}^{N} \frac{\|\boldsymbol{H}_{i,1}(\boldsymbol{w})\|_\infty}{2\,(1 + \chi(\boldsymbol{\theta}^\star))^2}\right) \left(\sum_{i=1}^{N} \mathbb{V}_{y,i}\left(Y_i\right) \sum_{j=i+1}^{N} x_j^2\, c_{i,j}\, z_{i,j}\right) \\
\geq\; & \frac{\left(\sum_{i=1}^{N} \|\boldsymbol{H}_{i,1}(\boldsymbol{w})\|_\infty\right) \left(\sum_{i=1}^{N} \|\boldsymbol{H}_{i,2}(\boldsymbol{w})\|_\infty\right)}{2\,(1 + \chi(\boldsymbol{\theta}^\star))^4},
\end{aligned}$$

where $\boldsymbol{H}_{i,2}(\boldsymbol{w})$ is defined in (D.8) and C corresponds to the constant from Condition 1.

The second inequality follows from the assumption $x_i \in [0,\, C]$ by Condition 1 along with $c_{i,j} \in \{0, 1\}$ and $z_{i,j} \in \{0, 1\}$. Lemma 11 shows that

$$\mathbb{P}(\boldsymbol{W} \in \mathcal{H}) \;\geq\; 1 - \frac{4}{\max\{N,\, p\}^2},$$

where $\mathcal{H}$ is defined in (D.8). For all $\boldsymbol{w} \in \mathcal{H}$, we obtain by definition

$$U(\boldsymbol{\theta}, \boldsymbol{w}) \;\geq\; \frac{\left(\sum_{i=1}^{N} \|\boldsymbol{H}_{i,1}(\boldsymbol{w})\|_\infty\right) \left(\sum_{i=1}^{N} \|\boldsymbol{H}_{i,2}(\boldsymbol{w})\|_\infty\right)}{2\,(1 + \chi(\boldsymbol{\theta}^\star))^4} \;\geq\; \frac{c^2\, N^2}{8\,(1 + \chi(\boldsymbol{\theta}^\star))^7},$$

which results in the following bound for the denominator of (D.18):

$$\begin{aligned}
V_{1,1}(\boldsymbol{\theta}, \boldsymbol{w})\, V_{2,2}(\boldsymbol{\theta}, \boldsymbol{w}) - V_{1,2}(\boldsymbol{\theta}, \boldsymbol{w})^2 \;\geq\; & U(\boldsymbol{\theta}, \boldsymbol{w}) - 18\, \max\{81,\, C^4\}\, N\, D^{11}\, \chi(\boldsymbol{\theta}^\star)^2 \\
\geq\; & \frac{c^2\, N^2}{8\,(1 + \chi(\boldsymbol{\theta}^\star))^7} - 18\, \max\{81,\, C^4\}\, N\, D^{11}\, \chi(\boldsymbol{\theta}^\star)^2 \\
>\; & \frac{c^2\, N^2}{8\,(1 + \chi(\boldsymbol{\theta}^\star))^7} \left(1 - \frac{18432\, \max\{81,\, C^4\}\, D^{11}\, \chi(\boldsymbol{\theta}^\star)^9}{c^2\, N}\right),
\end{aligned}$$

26

using the fact that $C > 0$, $D \geq 2$, and $\epsilon^\star > 0$, which implies that

$$\chi(\boldsymbol{\theta}^\star) \;\coloneqq\; \exp(C\, D^2\, (\|\boldsymbol{\theta}^\star\|_\infty + \epsilon^\star)) \;>\; 1.$$

Under Conditions 2 and 5 with $\vartheta \in [0,\, 1/18)$, we have, for all $\boldsymbol{w} \in \mathcal{H}$,

$$\frac{18432\, \max\{81,\, C^4\}\, D^{11}\, \chi(\boldsymbol{\theta}^\star)^9}{c^2\, N} \;\to\; 0 \quad \text{as} \quad N \to \infty.$$

Thus, there exists a real number $\epsilon > 0$ along with an integer $N_3 \in \{3, 4, \ldots\}$ such that

$$\frac{18432\, \max\{81,\, C^4\}\, D^{11}\, \chi(\boldsymbol{\theta}^\star)^9}{c^2\, N} \;\leq\; \epsilon$$

for all $N > N_3$, which implies that

$$V_{1,1}(\boldsymbol{\theta},\, \boldsymbol{w})\, V_{2,2}(\boldsymbol{\theta},\, \boldsymbol{w}) - V_{1,2}(\boldsymbol{\theta},\, \boldsymbol{w})^2 \;\geq\; \frac{c^2\, N^2}{8\,(1 + \chi(\boldsymbol{\theta}^\star))^7}\,(1 - \epsilon). \tag{D.23}$$

Observe that (D.23) provides a positive lower bound on the determinant of $\boldsymbol{V}(\boldsymbol{\theta},\, \boldsymbol{w})$ for $\boldsymbol{w} \in \mathcal{H}$, demonstrating that

$$|V_{1,1}(\boldsymbol{\theta},\, \boldsymbol{w})\, V_{2,2}(\boldsymbol{\theta},\, \boldsymbol{w}) - V_{1,2}(\boldsymbol{\theta},\, \boldsymbol{w})^2| \;=\; V_{1,1}(\boldsymbol{\theta},\, \boldsymbol{w})\, V_{2,2}(\boldsymbol{\theta},\, \boldsymbol{w}) - V_{1,2}(\boldsymbol{\theta},\, \boldsymbol{w})^2. \tag{D.24}$$

Combining (D.20), (D.23), and (D.24) shows that, for all $\boldsymbol{w} \in \mathcal{H}$,

$$
\begin{aligned}
\|\boldsymbol{V}(\boldsymbol{\theta},\, \boldsymbol{w})^{-1}\|_\infty \;&\leq\; \frac{\max\{V_{2,2}(\boldsymbol{\theta},\, \boldsymbol{w}),\, V_{1,1}(\boldsymbol{\theta},\, \boldsymbol{w})\} + V_{1,2}(\boldsymbol{\theta},\, \boldsymbol{w})}{V_{1,1}(\boldsymbol{\theta},\, \boldsymbol{w})\, V_{2,2}(\boldsymbol{\theta},\, \boldsymbol{w}) - V_{1,2}(\boldsymbol{\theta},\, \boldsymbol{w})^2} \\[2mm]
&\leq\; \max\{9,\, C^2\}\, D^5 \left(\frac{N}{2} + 32\, D\, \chi(\boldsymbol{\theta}^\star)^2\right) \frac{8\,(1 + \chi(\boldsymbol{\theta}^\star))^7}{c^2\, N^2\,(1 - \epsilon)} \qquad \text{(D.25)} \\[2mm]
&\leq\; K_1\, \frac{D^5\, \chi(\boldsymbol{\theta}^\star)^7}{N}\, \max\left\{1,\, \frac{D\, \chi(\boldsymbol{\theta}^\star)^2}{N}\right\},
\end{aligned}
$$

where $K_1 > 0$ is a constant.

Conclusion. We show in two steps that $-\nabla_{\boldsymbol{\theta}}^2\, \ell(\boldsymbol{\theta},\, \boldsymbol{w})$ is invertible for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ and all $\boldsymbol{w} \in \mathcal{H}$. First, by Lemma 9 in S25, the matrix $\boldsymbol{A}(\boldsymbol{\theta},\, \boldsymbol{w})$ is invertible for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star,\, \epsilon^\star)$ and all $\boldsymbol{w} \in \mathcal{H}$. Second, (D.24) demonstrates that the determinant of

27

$V(\boldsymbol{\theta}, \boldsymbol{w})$ is bounded away from 0 for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$ and all $\boldsymbol{w} \in \mathcal{H}$. Thus, $V(\boldsymbol{\theta}, \boldsymbol{w})$ is nonsingular for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$ and all $\boldsymbol{w} \in \mathcal{H}$, and so is $-\nabla^2_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \boldsymbol{w})$ by Theorem 8.5.11 of Harville (1997, p. 99). Combining (D.13), (D.14), (D.17), and (D.25) shows that, for all $\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)$ and all $\boldsymbol{w} \in \mathcal{H}$,

$$
\begin{aligned}
\|(-\nabla^2_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \boldsymbol{w}))^{-1}\|_\infty \;\leq\;\; & \|A(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty \\[4pt]
+\;\; & \max\{1, \, \|A(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty \|C(\boldsymbol{\theta}, \boldsymbol{w})\|_\infty\} \, \|V(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty \\[4pt]
\times\;\; & (N\, \|C(\boldsymbol{\theta}, \boldsymbol{w})\|_\infty \|A(\boldsymbol{\theta}, \boldsymbol{w})^{-1}\|_\infty + 1) \\[6pt]
\leq\;\; & \frac{18\, \chi(\boldsymbol{\theta}^\star)^2}{N} \\[6pt]
+\;\; & \max\left\{1, \, \max\{3, C\}\, D^3\, \frac{18\, \chi(\boldsymbol{\theta}^\star)^2}{N}\right\} K_1\, \frac{D^5\, \chi(\boldsymbol{\theta}^\star)^7}{N} \\[6pt]
\times\;\; & \max\left\{1, \, \frac{D^2\, \chi(\boldsymbol{\theta}^\star)^2}{N}\right\} (\max\{3, C\}\, D^3\, 18\, \chi(\boldsymbol{\theta}^\star)^2 + 1).
\end{aligned}
$$

Conditions 2 and 5 with $\vartheta \in [0, 1/18)$ imply that

$$
\frac{\chi(\boldsymbol{\theta}^\star)^2}{N} \;<\; \frac{\chi(\boldsymbol{\theta}^\star)^9}{N} \;\to\; 0 \text{ as } N \to \infty.
$$

Thus, there exists an integer $N_0 \in \{3, 4, \dots\}$ such that the two maxima in the upper bound on $\|(-\nabla^2_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, \boldsymbol{w}))^{-1}\|_\infty$ are equal to 1 for all $N > N_0$, so that

$$
\begin{aligned}
\|(-\nabla^2_{\boldsymbol{\theta}} & \ell(\boldsymbol{\theta}, \boldsymbol{w}))^{-1}\|_\infty \\[4pt]
\leq\;\; & \frac{18\, \chi(\boldsymbol{\theta}^\star)^2}{N} + K_1\, \frac{D^5\, \chi(\boldsymbol{\theta}^\star)^7}{N} (\max\{3, C\}\, D^3\, 18\, \chi(\boldsymbol{\theta}^\star)^2 + 1) \\[6pt]
\leq\;\; & \frac{18\, \chi(\boldsymbol{\theta}^\star)^2}{N} + K_2\, \frac{D^8\, \chi(\boldsymbol{\theta}^\star)^9}{N} \\[6pt]
\leq\;\; & C_1\, \frac{\chi(\boldsymbol{\theta}^\star)^9}{N},
\end{aligned}
\tag{D.26}
$$

where $K_2 > 0$ and $C_1 > 0$ are constants. Substituting (D.26) into the definition of $\Lambda_N(\boldsymbol{\theta}^\star)$ concludes the proof of Lemma 3:

$$
\Lambda_N(\boldsymbol{\theta}^\star) \;\coloneqq\; \sup_{\boldsymbol{w} \in \mathcal{H}} \; \sup_{\boldsymbol{\theta} \in \mathcal{B}_\infty(\boldsymbol{\theta}^\star, \epsilon^\star)} \|(-\nabla^2_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}; \boldsymbol{w}))^{-1}\|_\infty \;\leq\; C_1\, \frac{\chi(\boldsymbol{\theta}^\star)^9}{N}.
$$

D.5   Bounding Ψ_N

Lemma 4. *Consider the model of Corollary 1. Then $\sqrt{N/2} \leq \Psi_N \leq C_2 \sqrt{N}$, where $C_2 > 0$ is a constant.*

PROOF OF LEMMA 4. The term Ψ_N is defined in

$$\Psi_N \ := \ \max_{1 \leq a \leq N+2} \ \|\boldsymbol{\Xi}_a\|_2 \,,$$

where

$$\boldsymbol{\Xi}_a \ := \ (\Xi_{\{1\},a}, \ldots, \Xi_{\{N\},a}, \Xi_{\{1,2\},a}, \ldots, \Xi_{\{N,N-1\},a}) = (\boldsymbol{\Xi}_{\mathcal{Y},a}, \boldsymbol{\Xi}_{\mathcal{Z},a}), \ a \in \{1, \ldots, N+2\}.$$

The sensitivity of the sufficient statistic vector $b_a(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ with respect to changes of responses is quantified by the vector $\boldsymbol{\Xi}_{\mathcal{Y},a} \in \mathbb{R}^N$:

$$\boldsymbol{\Xi}_{\mathcal{Y},a} \ := \ (\Xi_{\{1\},a}, \ldots, \Xi_{\{N\},a}),$$

where

$$\Xi_{\{i\},a} \ := \ \max_{(\boldsymbol{w},\boldsymbol{w}') \in \mathcal{W} \times \mathcal{W}: \, y_k = y'_k \text{ for all } k \neq i, \, \boldsymbol{z} = \boldsymbol{z}'} |b_a(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) - b_a(\boldsymbol{x}, \boldsymbol{y}', \boldsymbol{z}')|.$$

The sensitivity of the sufficient statistic vector $b_a(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ with respect to changes of connections is quantified by the vector $\boldsymbol{\Xi}_{\mathcal{Z},a} \in \mathbb{R}^N$:

$$\boldsymbol{\Xi}_{\mathcal{Z},a} \ := \ (\Xi_{\{1,2\},a}, \ldots, \Xi_{\{N,N-1\},a}),$$

where

$$\Xi_{\{i,j\},a} \ := \ \max_{(\boldsymbol{w},\boldsymbol{w}') \in \mathcal{W} \times \mathcal{W}: \, \boldsymbol{y} = \boldsymbol{y}', \, z_{k,l} = z'_{k,l} \text{ for all } \{k,l\} \neq \{i,j\}} |b_a(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) - b_a(\boldsymbol{x}, \boldsymbol{y}', \boldsymbol{z}')|.$$

Define

$$\Psi_N \ = \ \max_{1 \leq a \leq N+2} \sqrt{\sum_{i=1}^{N} |\Xi_{\{i\},a}|^2 + \sum_{i=1}^{N} \sum_{j=i+1}^{N} |\Xi_{\{i,j\},a}|^2}. \tag{D.27}$$

29

- For $a = 1, \ldots, N$, the statistic $b_a(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ refers to the degree effects of unit a:

$$b_a(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \sum_{j=1; j \neq a}^{N} z_{a,j}.$$

The term $\Xi_{\{i,j\},a}$ is 1 if $a \in \{i, j\}$ and is 0 otherwise. Since the statistic is unaffected by the response, $\Xi_{\{i\},a} = 0$ for all $i = 1, \ldots, N$. For the sum in (D.27) over all $i < j$, where $\mathbb{I}(a \in \{i, j\}) = 1$ holds N times, yielding $\|\Xi_a\|_2 \leq \sqrt{N}$ for all $a = 1, \ldots, N$.

- The statistic $b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ refers to the transitive connections effect given by

$$b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} d_{i,j}(\boldsymbol{z}) \, z_{i,j},$$

where the function $d_{i,j}(\boldsymbol{Z})$ is defined in (D.2). Since this statistic is not affected by $\boldsymbol{y}$, $\Xi_{\{i\},a} = 0$ for all $i = 1, \ldots, N$. Following Lemma 18 in S25,

$$\|\Xi_{N+1}\|_2 \leq \sqrt{N D^2 (1 + D)^2} \leq \sqrt{4 N D^4} = 2 D^2 \sqrt{N},$$

where D corresponds to the constant defined in D.4.

- The statistic $b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ refers to the spillover effect given by

$$b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} c_{i,j} \, (x_i \, y_j + y_i \, x_j) \, z_{i,j},$$

where the function $c_{i,j}$ is defined in (D.2). For $\{i, j\} \subset \mathcal{P}_N$, the terms $\Xi_{\{i,j\},N+2}$ are $(y_i \, x_j + y_j \, x_i) \leq 2 C$ if $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and 0 otherwise. For all $i \in \mathcal{P}_N$,

$$\Xi_{\{i\},N+2} = \sum_{j : \mathcal{N}_i \cap \mathcal{N}_j} x_j \, z_{i,j} \leq \sum_{j : \mathcal{N}_i \cap \mathcal{N}_j} C \leq C D^2,$$

because according to Lemma 15 in S25 there are at most D^2 units such that $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and $x_i \leq C$ for $i = 1, \ldots, N$ according to Condition 1. Combining $\Xi_{\{i,j\},N+2}$ and $\Xi_{\{i\},N+2}$ gives

$$\|\Xi_{N+2}\|_2 \leq \sqrt{2 N C D^2 + N C D^2} \leq D \sqrt{3 N C} \leq 2 D \sqrt{N C},$$

where C corresponds to the constant C in Condition 1.

Combining the results for $\|\Xi_a\|_2$ for $a = 1, \ldots, N + 2$ gives

$$\sqrt{N/2} \;\leq\; \Psi_N \;\leq\; 2\,D\,\sqrt{N\,C}$$

$$\sqrt{N/2} \;\leq\; \Psi_N \;\leq\; C_2\,\sqrt{N},$$

where $C_2 := 2\,D\,\sqrt{C} > 0$ is a constant.

D.6   Bounding $\|\mathcal{D}_N(\boldsymbol{\theta}^\star)\|_2$

Lemma 5. *Consider the model of Corollary 1. If Conditions 1, 2, and 4 are satisfied with $\vartheta \in [0, 1/18)$, there exists a constant $C_3 \geq 1$ such that $\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2 \leq C_3$ for all $N \geq 2$. If the population $\mathcal{P}_N$ consists of non-overlapping subpopulations with dependence restricted to subpopulations, the same result holds when Condition 2 is replaced by Condition 5.*

PROOF OF LEMMA 5. To bound $\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2$ from above, we use the Hölder's inequality

$$\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2 \;\leq\; \sqrt{\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_1 \, \|\mathcal{D}(\boldsymbol{\theta}^\star)\|_\infty}, \tag{D.28}$$

where

$$\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_1 \;:=\; \max_{1 \leq j \leq M} \sum_{i=1}^{M} |\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)|$$

$$\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_\infty \;:=\; \max_{1 \leq i \leq M} \sum_{j=1}^{M} |\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star)|.$$

We can therefore bound $\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2$ by bounding the elements of the upper triangular coupling matrix $\mathcal{D}(\boldsymbol{\theta}^\star) \in \mathbb{R}^{M \times M}$ which are

$$\mathcal{D}_{i,j}(\boldsymbol{\theta}^\star) \;:=\; \begin{cases} 0 & \text{if } i < j \\[2ex] 1 & \text{if } i = j \\[2ex] \displaystyle\max_{\boldsymbol{w}_{1:(i-1)} \in \mathcal{W}_{1:i-1}} \mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{w}_{1:(i-1)}}(W_j^\star \neq W_j^{\star\star}) & \text{if } i > j. \end{cases}$$

31

Next, we define a symmetrized version of the coupling matrix denoted by $\mathcal{T}(\boldsymbol{\theta}^\star) \in \mathbb{R}^{M \times M}$ with elements

$$\mathcal{T}_{i,j}(\boldsymbol{\theta}^\star) \;:=\; \begin{cases} \mathcal{D}_{j,i}(\boldsymbol{\theta}^\star) & \text{if } i < j \\[2ex] \mathcal{D}_{i,i}(\boldsymbol{\theta}^\star) & \text{if } i = j \\[2ex] \mathcal{D}_{i,j}(\boldsymbol{\theta}^\star) & \text{if } i > j. \end{cases}$$

The symmetry of $\mathcal{T}(\boldsymbol{\theta}^\star)$ yields the following upper bound for (D.28):

$$\begin{aligned} \|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2 \;&\leq\; \sqrt{\|\mathcal{T}(\boldsymbol{\theta}^\star)\|_1 \, \|\mathcal{T}(\boldsymbol{\theta}^\star)\|_\infty} \;=\; \|\mathcal{T}(\boldsymbol{\theta}^\star)\|_\infty \\ &=\; 1 + \max_{1 \leq i \leq M} \sum_{j=1:\, j \neq i}^{M} \mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{w}_{1:(i-1)}}(W_j^\star \neq W_j^{\star\star}) \end{aligned} \tag{D.29}$$

where the constant 1 in the second line stems from the diagonal elements of $\mathcal{T}(\boldsymbol{\theta}^\star)$.

Consider any $(i,j) \in \{1, \ldots, M\} \times \{1, \ldots, M\}$ such that $i \neq j$ and define the event $W_i \leftrightarrow\!\!\!\!/ \, W_j$ as the event that there exists a path of disagreement between vertices W_i and W_j in $\mathcal{G}$. A path of disagreement between vertices W_i and W_j in $\mathcal{G}$ is a path from W_i to W_j in $\mathcal{G}$ such that the coupling $(W_{(i+1):M}^\star, W_{(i+1):M}^{\star\star})$ with joint probability mass function $\mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{w}_{1:(i-1)}}$ disagrees at each vertex on the path, in the sense that $W^\star \neq W^{\star\star}$ holds for all vertices W on the path. Theorem 1 of van den Berg and Maes (1994, p. 753) shows that

$$\mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{w}_{1:(i-1)}}(W_j^\star \neq W_j^{\star\star}) \;\leq\; \mathbb{B}_{\boldsymbol{\pi}}(W_i \leftrightarrow\!\!\!\!/ \, W_j \text{ in } \mathcal{G}), \tag{D.30}$$

where $\mathbb{B}_{\boldsymbol{\pi}}$ is a Bernoulli product measure based on M independent Bernoulli experiments with success probabilities $\boldsymbol{\pi} := (\pi_1, \ldots, \pi_M) \in [0,1]^M$. With $v \in \{1, \ldots, M\}$, the success probabilities π_v are

$$\pi_v \;:=\; \begin{cases} 0 & \text{if } v \in \{1, \ldots, i-1\} \\[2ex] 1 & \text{if } v = i \\[2ex] \displaystyle\max_{(\boldsymbol{w}_{-v}, \boldsymbol{w}'_{-v}) \in \mathcal{W}_{-v} \times \mathcal{W}_{-v}} \pi_{v, \boldsymbol{w}_{-v}, \boldsymbol{w}'_{-v}} & \text{if } v \in \{i+1, \ldots, M\}, \end{cases}$$

32

where

$$\pi_{v, \boldsymbol{w}_{-v}, \boldsymbol{w}'_{-v}} \quad := \quad \left\| \mathbb{P}_{\boldsymbol{\theta}}(\,\cdot\mid \boldsymbol{w}_{-v}) - \mathbb{P}_{\boldsymbol{\theta}}(\,\cdot\mid \boldsymbol{w}'_{-v}) \right\|_{\mathrm{TV}}. \tag{D.31}$$

Lemma 10 provides the following upper bound:

$$\pi_{v, \boldsymbol{w}_{-v}, \boldsymbol{w}'_{-v}} \quad \leq \quad \frac{1}{1 + \exp(-C\,D^2\,\|\boldsymbol{\theta}^\star\|_\infty)}, \tag{D.32}$$

where C corresponds to the positive constant from Condition 1 and D is defined in (D.4).

Combining (D.32) with Condition 5 shows that

$$\frac{1}{1 + \exp(-C\,D^2\,\|\boldsymbol{\theta}^\star\|_\infty)} \quad \leq \quad \frac{1}{1 + \exp(-E - \vartheta \log N)} \quad =: \quad U_N. \tag{D.33}$$

The constant $U_N = U$ coincides with the constant U considered in Condition 4.

With (D.33), we define the vector $\boldsymbol{\xi} \in [0,1]^M$ with elements

$$\xi_v \quad := \quad \begin{cases} 0 & \text{if } v \in \{1, \ldots, i-1\} \\[2mm] 1 & \text{if } v = i \\[2mm] U_N & \text{if } v \in \{i+1, \ldots, M\}, \end{cases}$$

and obtain

$$\mathbb{B}_{\boldsymbol{\pi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}) \quad \leq \quad \mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}), \tag{D.34}$$

because $\pi_v \leq \xi_v$ for all $v = 1, \ldots, M$.

Next, we construct the set

$$\mathcal{M}_{a,b} \quad := \quad \{\{c,d\}: \ c \in \mathcal{N}_a \cup \mathcal{N}_b, \ d \in \mathcal{N}_a \cup \mathcal{N}_b \setminus \{c\}\} \ \cup \ \{\{c\}: \ c \in \mathcal{N}_a \cup \mathcal{N}_b\}$$

and two additional graphs with the same set of vertices as $\mathcal{G}$:

1. $\mathcal{G}_1 := (\mathcal{V}, \mathcal{E}_1)$:

- Vertex $W \in \mathcal{V}_{\mathcal{Z}}$ relating to connection $Z_{i,j}$ has edges to vertices that relate to all connections $Z_{h,k}$ and responses Y_h with $\{h,k\}, \{h\} \in \mathcal{M}_{i,j}$.

33

- Vertex $W \in \mathcal{V}_y$ relating to attribute Y_i has edges to vertices that relate to all connections $Z_{h,k}$ and responses Y_h with $\{h, k\}, \{h\} \in \mathcal{M}_{i,N+1}$ for a fictional unit $N + 1$ with $\mathcal{N}_{N+1} = \emptyset$.

2. $\mathcal{G}_2 := (\mathcal{V}, \mathcal{E}_1 \cup \mathcal{E}_2)$: The set $\mathcal{E}_2$ includes edges of all vertices $W_i \in \mathcal{V}$ with $i \in \{1, \dots, M\}$ to vertices in $\mathcal{S}_{\mathcal{G}_1, i, 2}$.

The graph $\mathcal{G}_2$ is a covering of $\mathcal{G}$, so

$$\mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}) \leq \mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}_2). \tag{D.35}$$

Combining the previous results gives

$$
\begin{aligned}
\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2 &\leq 1 + \max_{1 \leq i \leq M} \sum_{j=1: j \neq i}^{M} \mathbb{Q}_{\boldsymbol{\theta}^\star, i, \boldsymbol{w}_{1:(i-1)}}(W_j^\star \neq W_j^{\star\star}) \\
&\leq 1 + \max_{1 \leq i \leq M} \sum_{j=1: j \neq i}^{M} \mathbb{B}_{\boldsymbol{\pi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}) \\
&\leq 1 + \max_{1 \leq i \leq M} \sum_{j=1: j \neq i}^{M} \mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}) \\
&\leq 1 + \max_{1 \leq i \leq M} \sum_{j=1: j \neq i}^{M} \mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}_2),
\end{aligned}
\tag{D.36}
$$

using (D.29), (D.30), (D.34), (D.35). Sorting the vertices without W_i by the geodesic distance to W_i (i.e., by the length of the shortest path to W_i), we obtain

$$
\begin{aligned}
\sum_{j=1: j \neq i}^{M} \mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}_2) &\leq |\mathcal{S}_{\mathcal{G}_2, i, 1}| \left(\max_{W_j \in \mathcal{S}_{\mathcal{G}_2, i, 1}} \mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}_2) \right) \\
&+ \sum_{k=2}^{\infty} |\mathcal{S}_{\mathcal{G}_2, i, k}| \left(\max_{W_j \in \mathcal{S}_{\mathcal{G}_2, i, k}} \mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}_2) \right) \\
&\leq |\mathcal{S}_{\mathcal{G}_2, i, 1}| \\
&+ \sum_{k=2}^{\infty} |\mathcal{S}_{\mathcal{G}_2, i, k}| \max_{W_j \in \mathcal{S}_{\mathcal{G}_2, i, k}} \mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}_2).
\end{aligned}
\tag{D.37}
$$

For the event $W_i \not\leftrightarrow W_j$ in $\mathcal{G}_2$ with $W_j \in \mathcal{S}_{\mathcal{G}_2, i, k}$ and $k \geq 2$ to occur, there must exist at least one vertex in each set $\mathcal{S}_{\mathcal{G}_2, i, 1}, \dots, \mathcal{S}_{\mathcal{G}_2, i, k-1}$ at which the coupling disagrees. Therefore,

we next derive bounds on $|\mathcal{S}_{\mathcal{G}_2,i,k}|$ to obtain an upper bound on $\mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}_2)$.

Following Lemma 6, Condition 4 implies that for $i \in \{1, \ldots, M\}$ and $k \in \{2, 3, \ldots\}$

$$|\mathcal{S}_{\mathcal{G}_2,i,k}| \leq K_1 + K_2 \log k \tag{D.38}$$

and $|\mathcal{S}_{\mathcal{G}_2,i,1}| \leq K_3$, with constants $K_1 \geq 0$, $K_2 \geq 0$, and $K_3 > 0$ being functions of the constants $\omega_1 \geq 0$ and $\omega_2 \geq 0$ defined in Condition 4 and the constant $D \in \{2, 3, \ldots\}$ defined in (D.4). The probability of event $W_i \not\leftrightarrow W_j$ in $\mathcal{G}_2$ can then be bounded as follows:

$$\begin{aligned}
\mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}_2) &\leq U_N \left(1 - (1 - U_N)^{K_3}\right) \prod_{l=2}^{k-1} \left[1 - (1 - U_N)^{K_1 + K_2 \log l}\right] \\
&\leq \prod_{l=2}^{k-1} \left[1 - (1 - U_N)^{K_1 + K_2 \log l}\right] \\
&\leq \left[1 - (1 - U_N)^{K_1 + K_2 \log(k-1)}\right]^{k-2},
\end{aligned}$$

The first inequality follows from

$$U_N \left(1 - (1 - U_N)^{K_3}\right) \leq 1,$$

because $U_N \in [0, 1]$ and $K_3 > 0$. Defining $K_N := \exp(-K_1 \, |\log(1 - U_N)|)$, we obtain for $W_j \in \mathcal{S}_{\mathcal{G},i,k}$

$$\begin{aligned}
\mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}_2) &\leq \left[1 - (1 - U_N)^{K_1 + K_2 \log(k-1)}\right]^{k-2} \\
&\leq \exp\left(-K_N \, (k-1)^{1 - K_2 \, |\log(1 - U_N)|}\right)
\end{aligned} \tag{D.39}$$

with the inequality $1 - a \leq \exp(-a)$ for all $a \in (0, 1)$.

Plugging (D.38) and (D.39) in (D.37), we obtain:

$$\begin{aligned}
\sum_{j=1: j \neq i}^{M} \mathbb{B}_{\boldsymbol{\xi}} \quad & (W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}_2) \\
\leq \quad & K_3 + \sum_{k=2}^{\infty} (K_1 + K_2 \log k) \\
\times \quad & \exp\left(-K_N \, (k-1)^{1 - K_2 \, |\log(1 - U_N)|}\right) \\
= \quad & K_3 + K_1 \sum_{k=2}^{\infty} \exp\left(-K_N \, (k-1)^{1 - K_2 \, |\log(1 - U_N)|}\right) \\
+ \quad & K_2 \sum_{k=2}^{\infty} \log k \, \exp\left(-K_N \, (k-1)^{1 - K_2 \, |\log(1 - U_N)|}\right),
\end{aligned} \tag{D.40}$$

resulting in two series that we bound one by one. With $\lceil \cdot \rceil : [0, \infty) \mapsto \{1, 2, \ldots\}$ being the function giving the upper ceiling of a positive real number and $u_N := \lceil 2/(1 - K_2 \,|\log(1 - U_N)|) \rceil$, the first series can be bounded as follows:

$$\sum_{k=2}^{\infty} \exp\left(-K_N \,(k-1)^{1-K_2\,|\log(1-U_N)|}\right)$$

$$= \sum_{k=1}^{\infty} \exp\left(-K_N \,k^{1-K_2\,|\log(1-U_N)|}\right)$$

$$\leq \frac{u_N!}{(K_N)^{u_N}} \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{u_N!\,\pi^2}{(K_N)^{u_N}\,6}.$$

The above bound is based on a Taylor expansion of $\exp(z)$, which establishes the inequality $\exp(z) > z^u / u!$ implying for any $z > 0$ and any $u \in \{1, 2, \ldots\}$. This, in turn, implies the inequality $\exp(-z) < u! / z^u$ for any $z > 0$ and any $u \in \{1, 2, \ldots\}$. With $v_N := \lceil 3/(1 - K_2\,|\log(1 - U_N)|) \rceil$, we apply the same inequality to the second series:

$$\sum_{k=2}^{\infty} \log(k) \, \exp\left(-K_N \,(k-1)^{1-K_2\,|\log(1-U_N)|}\right)$$

$$= \sum_{k=1}^{\infty} \log(k+1) \, \exp\left(-K_N \,k^{1-K_2\,|\log(1-U_N)|}\right)$$

$$\leq \frac{v_N!}{(K_N)^{v_N}} \sum_{k=1}^{\infty} \frac{\log(k+1)}{k^3}$$

$$\leq \frac{v_N!}{(K_N)^{v_N}} \sum_{k=1}^{\infty} \frac{k}{k^3} = \frac{v_N!}{(K_N)^{v_N}} \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{v_N!\,\pi^2}{(K_N)^{v_N}\,6}.$$

Plugging these results into (D.40) gives

$$\sum_{j=1:\,j\neq i}^{M} \mathbb{B}_{\boldsymbol{\xi}}(W_i \not\leftrightarrow W_j \text{ in } \mathcal{G}_2) \;\leq\; K_3 + \frac{\pi^2}{6}\left(K_1 \frac{u_N!}{(K_N)^{u_N}} + K_2 \frac{v_N!}{(K_N)^{v_N}}\right). \tag{D.41}$$

Last but not least, combining (D.41) with (D.36) yields

$$\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2 \;\leq\; 1 + K_3 + \frac{\pi^2}{6}\left(K_1 \frac{u_N!}{(K_N)^{u_N}} + K_2 \frac{v_N!}{(K_N)^{v_N}}\right).$$

Under Condition 2, $\vartheta = 0$ holds, hence $U_{\vartheta,N}, K_N, u_{\vartheta,N}$, and $v_{\vartheta,N}$ in (D.33) reduce to

$$U_N \quad = \quad \frac{1}{1 + \exp(-E)} \qquad =: \quad U$$

$$K_N \quad = \quad \exp(-K_1 \,|\log(1-U)|) \quad =: \quad K_4$$

$$u_N \quad = \quad \left\lceil \frac{2}{1 - K_2 \,|\log U|} \right\rceil \qquad =: \quad u$$

$$v_N \quad = \quad \left\lceil \frac{3}{1 - K_2 \,|\log U|} \right\rceil \qquad =: \quad v,$$

which are constants independent of ϑ and N. The constant U corresponds to the constant U from Condition 4. This translates to

$$\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2 \quad \leq \quad C_3,$$

with $C_3 := 1 + K_3 + (\pi^2/6)\,(K_1 \,u!/K_4^u + K_2 \,v!/K_4^v) \geq 1$. For non-overlapping subpopulations, we have $K_1 = K_2 = 0$ and

$$\|\mathcal{D}(\boldsymbol{\theta}^\star)\|_2 \quad \leq \quad 1 + K_3 \quad = \quad C_3.$$

D.7   Auxiliary Results

Lemma 6. *Consider the model of Corollary 1. Condition 4 implies that there exist constants $K_1 \geq 0$, $K_2 \geq 0$, $K_3 > 0$ such that, for all $k \in \{2, 3, \ldots\}$ and all $i \in \{1, \ldots, M\}$,*

$$|\mathcal{S}_{\mathcal{G}_2,i,k}| \quad \leq \quad K_1 + K_2 \log k$$

$$|\mathcal{S}_{\mathcal{G}_2,i,1}| \quad \leq \quad K_3.$$

PROOF OF LEMMA 6. With the set

$$\mathcal{M}_{a,b} \quad := \quad \{\{c,d\} : c \in \mathcal{N}_a \cup \mathcal{N}_b, \ d \in \mathcal{N}_a \cup \mathcal{N}_b \setminus \{c\}\} \cup \{\{c\} : c \in \mathcal{N}_a \cup \mathcal{N}_b\},$$

we constructed from $\mathcal{G}$ two additional graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ as follows:

1. $\mathcal{G}_1 := (\mathcal{V}, \mathcal{E}_1)$:

   - Vertex $W \in \mathcal{V}_{\mathcal{Z}}$ relating to connection $Z_{i,j}$ has edges to vertices that relate to all connections $Z_{h,k}$ and responses Y_h with $\{h, k\}, \{h\} \in \mathcal{M}_{i,j}$.

   - Vertex $W \in \mathcal{V}_{\mathcal{Y}}$ relating to attribute Y_i has edges to vertices that relate to all connections $Z_{h,k}$ and responses Y_h with $\{h, k\}, \{h\} \in \mathcal{M}_{i,N+1}$ for a fictional unit $N + 1$ with $\mathcal{N}_{N+1} = \emptyset$.

2. $\mathcal{G}_2 := (\mathcal{V}, \mathcal{E}_1 \cup \mathcal{E}_2)$: The set $\mathcal{E}_2$ includes edges of all vertices $W_i \in \mathcal{V}$ with $i \in \{1, \ldots, M\}$ to vertices in $\mathcal{S}_{\mathcal{G}_1, i, 2}$.

The graph $\mathcal{G}_1$ is equivalent to the graph cover $\mathcal{G}^\star$ defined in Lemma 16 of S25. Therefore, we are able to use results from the proof of Lemma 16 in S25 demonstrating that Condition 4 implies the following bound for $\mathcal{S}_{\mathcal{G}_1, i, k}$:

$$|\mathcal{S}_{\mathcal{G}_1, i, k}| \leq (\omega_1 + 1)(2\, D^3\, \omega_1 + \omega_2 \log(k - 1)), \quad k \in \{2, 3, \ldots\},$$

where D corresponds to the constant defined in (D.4) and the constants $\omega_1 \geq 0$ and $0 \leq \omega_2 \leq \min\{\omega_1, 1/((\omega_1 + 1)\,|\log(1 - U)|)\}$ with $U := (1 + \exp(-A))^{-1} > 0$ correspond to the constant from Condition 4. Defining $K_5 := 2\,\omega_1\,(\omega_1 + 1)\,D^3 \geq 0$ and $K_6 := \omega_2\,(\omega_1 + 1) \geq 0$, this bound is:

$$|\mathcal{S}_{\mathcal{G}_1, i, k}| \leq K_5 + K_6 \log(k - 1), \quad k \in \{2, 3, \ldots\}.$$

The bound for $\mathcal{S}_{\mathcal{G}_1, i, 1} \leq 4\, D^2 + D$ differs to the result from S25 since for our definition of $\mathcal{M}_{i,j}$ there are additional $|\mathcal{N}_i \cup \mathcal{N}_j| \leq D$ responses in $\mathcal{M}_{i,j}$.

Adding edges $\mathcal{E}_2$, defined as the edges from vertices to other vertices with a geodesic distance of two in $\mathcal{G}_1$, to $\mathcal{G}_2$ reduces the geodesic distance between all vertices from $k \in \{1, 2, \ldots\}$ in $\mathcal{G}_1$ to $\lceil k/2 \rceil$ in $\mathcal{G}_2$. Therefore, $|\mathcal{S}_{\mathcal{G}_2, i, k}| = |\mathcal{S}_{\mathcal{G}_1, i, 2k}| + |\mathcal{S}_{\mathcal{G}_1, i, 2k-1}|$ holds for $k \in$

$\{1, 2, \ldots\}$ and $i \in \{1, \ldots, M\}$. This allows us to relate the bounds for $|\mathcal{S}_{\mathcal{G}_1, i, k}|$ to bounds for $|\mathcal{S}_{\mathcal{G}_2, i, k}|$ with $k = 2, 3, \ldots$ and $i \in \{1, \ldots, M\}$:

$$
\begin{aligned}
|\mathcal{S}_{\mathcal{G}_2, i, k}| &= |\mathcal{S}_{\mathcal{G}_1, i, 2\, k}| + |\mathcal{S}_{\mathcal{G}_1, i, 2\, k-1}| \\
&\leq 2\, K_5 + K_6 \left(\log(2\, k) + \log(2\, k - 1) \right) \\
&\leq 2\, K_5 + 2\, K_6 \log(2\, k) \\
&= K_1 + K_2 \log k
\end{aligned}
$$

and

$$
|\mathcal{S}_{\mathcal{G}_2, i, 1}| \leq 4\, D^2 + D + K_1 =: K_3,
$$

with $K_1 := 2\, K_5 + 2\, K_6 \log 2$ and $K_2 := 2\, K_6$. This proves the statement with $K_1 \geq 0, K_2 \geq 0$, and $K_3 > 0$.

Lemma 7. *Consider the model of Corollary 1. Then, for any pair of units $\{i, j\} \subset \mathcal{P}_N$ such that $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$,*

$$
\frac{1}{1 + \exp\left(C\, D^2\, \|\boldsymbol{\theta}\|_\infty \right)} \leq \mathbb{P}_{\boldsymbol{\theta}}(Z_{i,j} = 1 \mid \boldsymbol{x},\, \boldsymbol{y},\, \boldsymbol{z}_{-\{i,j\}}) \leq \frac{1}{1 + \exp\left(-C\, D^2\, \|\boldsymbol{\theta}\|_\infty \right)}.
$$

PROOF OF LEMMA 7. For all $\{i, j\} \subset \mathcal{P}_N$ such that $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$, the conditional probability of $Z_{i,j}$ given $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}_{-\{i,j\}}) = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}})$ is

$$
\begin{aligned}
& \mathbb{P}_{\boldsymbol{\theta}}(Z_{i,j} = z_{i,j} \mid \boldsymbol{x},\, \boldsymbol{y},\, \boldsymbol{z}_{-\{i,j\}}) \\
&= \frac{\exp\left(\boldsymbol{\theta}^\top \boldsymbol{b}(\boldsymbol{x},\, \boldsymbol{y},\, \boldsymbol{z}_{-\{i,j\}}, z_{i,j}) \right)}{\exp\left(\boldsymbol{\theta}^\top \boldsymbol{b}(\boldsymbol{x},\, \boldsymbol{y},\, \boldsymbol{z}_{-\{i,j\}}, 1) \right) + \exp\left(\boldsymbol{\theta}^\top \boldsymbol{b}(\boldsymbol{x},\, \boldsymbol{y},\, \boldsymbol{z}_{-\{i,j\}}, 0) \right)} \\
&= \frac{1}{1 + g(1 - z_{i,j};\, \boldsymbol{z}_{-\{i,j\}}, z_{i,j}, \boldsymbol{\theta})},
\end{aligned}
$$

with

$$
g(z;\, \boldsymbol{z}_{-\{i,j\}}, z_{i,j}, \boldsymbol{\theta}) = \exp\left(\boldsymbol{\theta}^\top \left(\boldsymbol{b}(\boldsymbol{x},\, \boldsymbol{y},\, \boldsymbol{z}_{-\{i,j\}}, z) - \boldsymbol{b}(\boldsymbol{x},\, \boldsymbol{y},\, \boldsymbol{z}_{-\{i,j\}}, z_{i,j}) \right) \right).
$$

Note that

$$\max_{z_{-\{i,j\}} \in \mathcal{Z}_{-\{i,j\}}} |b_a(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}}, 0) - b_a(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}}, 1)|$$

$$\leq \begin{cases} 0 & \text{if } a \in \{1, \dots, N\} \setminus \{i, j\} \\ 1 & \text{if } a \in \{i, j\} \\ 1 + D & \text{if } a = N + 1 \\ 2\,C & \text{if } a = N + 2 \end{cases},$$

where $\mathcal{Z}_{-\{i,j\}} := \bigtimes_{(k,h) \neq (i,j)}^{N} \mathcal{Z}_{k,h}$ is the domain of $\boldsymbol{Z}$ excluding $Z_{i,j}$, C corresponds to the constant from Condition 1, and D matches the constant defined in (D.4). . The bounds for $a = 1, \dots, N$ follow from the observation, that the degree statistic of unit a can, first, only affected by connections $z_{i,j}$ with $a \in \{i, j\}$ and, second, be at most 1 if this is the case. For $a = N+1$, the bound follows from Lemma 18 of S25. For $a = N+2$, the sufficient statistic counts the number of connections with overlapping neighborhoods and either $Y_i\, x_j > 0$ or $Y_j\, x_i > 0$. For $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$, the maximal change in the statistic is $2\,C$ since $y_i \in \{0, 1\}$ and $x_i \leq C$ for $i \in \mathcal{P}_N$, otherwise the maximal change is 0.

Upon applying the triangle inequality,

$$|\boldsymbol{\theta}^\top \boldsymbol{b}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}}, z) - \boldsymbol{\theta}^\top \boldsymbol{b}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}}, z_{i,j})| \leq (2 + 2\,C + D)\,\|\boldsymbol{\theta}\|_\infty,$$

we obtain for $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$

$$\exp\left(-(2 + 2\,C + D)\,\|\boldsymbol{\theta}\|_\infty\right) \leq g(1 - z_{i,j}; \boldsymbol{z}_{-\{i,j\}}, z_{i,j}, \boldsymbol{\theta}) \leq \exp\left((2 + 2\,C + D)\,\|\boldsymbol{\theta}\|_\infty\right).$$

Upon collecting terms, we obtain the final result:

$$\frac{1}{1 + \exp\left(C_6\,\|\boldsymbol{\theta}\|_\infty\right)} \leq \mathbb{P}_{\boldsymbol{\theta}}(Z_{i,j} = 1 \mid \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}}) \leq \frac{1}{1 + \exp\left(-C_6\,\|\boldsymbol{\theta}\|_\infty\right)}$$

$$\frac{1}{1 + \exp\left(C\,D^2\,\|\boldsymbol{\theta}\|_\infty\right)} \leq \mathbb{P}_{\boldsymbol{\theta}}(Z_{i,j} = 1 \mid \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}}) \leq \frac{1}{1 + \exp\left(-C\,D^2\,\|\boldsymbol{\theta}\|_\infty\right)}$$

where $D \in \{2, 3, \dots\}$ and $C_6 := 2 + 2\,C + D > 0$ are constants.

Lemma 8. *Consider the model of Corollary 1. Then, for any $i \in \{1, \dots, M\}$*

$$\frac{1}{1 + \exp\left(C\, D^2\, \|\boldsymbol{\theta}\|_\infty\right)} \;\leq\; \mathbb{P}_{\boldsymbol{\theta}}(Y_i = 1 \mid \boldsymbol{x},\, \boldsymbol{y}_{-i},\, \boldsymbol{z}) \;\leq\; \frac{1}{1 + \exp\left(-C\, D^2\, \|\boldsymbol{\theta}\|_\infty\right)}.$$

PROOF OF LEMMA 8. The conditional probability of Y_i given $(\boldsymbol{X}, \boldsymbol{Y}_{-i}, \boldsymbol{Z}) = (\boldsymbol{y}_{-i}, \boldsymbol{z})$ is

$$\mathbb{P}_{\boldsymbol{\theta}}(Y_i = y_i \mid \boldsymbol{x},\, \boldsymbol{y}_{-i},\, \boldsymbol{z}) \;=\; \frac{\exp\left(\boldsymbol{\theta}^\top \boldsymbol{b}(\boldsymbol{x},\, \boldsymbol{y}_{-i},\, y_i,\, \boldsymbol{z})\right)}{\exp\left(\boldsymbol{\theta}^\top \boldsymbol{b}(\boldsymbol{x},\, \boldsymbol{y}_{-i}, 0,\, \boldsymbol{z})\right) + \exp\left(\boldsymbol{\theta}^\top \boldsymbol{b}(\boldsymbol{x},\, \boldsymbol{y}_{-i}, 1,\, \boldsymbol{z})\right)}$$

$$= \frac{1}{g(0; \boldsymbol{y}_{-i},\, y_i,\, \boldsymbol{\theta}) + g(1; \boldsymbol{y}_{-i},\, y_i,\, \boldsymbol{\theta})},$$

where

$$g(y;\, \boldsymbol{y}_{-i},\, y_i,\, \boldsymbol{\theta}) \;=\; \exp\left(\boldsymbol{\theta}^\top \left(\boldsymbol{b}(\boldsymbol{x},\, \boldsymbol{y}_{-i},\, y,\, \boldsymbol{z}) - \boldsymbol{b}(\boldsymbol{x},\, \boldsymbol{y}_{-i},\, y_i,\, \boldsymbol{z})\right)\right).$$

Note that

$$\max_{\boldsymbol{y}_{-i} \in \mathcal{Y}_{-i}} |b_a(\boldsymbol{x},\, \boldsymbol{y}_{-i}, 0,\, \boldsymbol{z}) - b_a(\boldsymbol{x},\, \boldsymbol{y}_{-i}, 1,\, \boldsymbol{z})| \;\leq\; \begin{cases} 0 & \text{if } a \in \{1, \dots, N+1\} \\[2mm] C\, D^2 & \text{if } a = N+2, \end{cases}$$

where $\mathcal{Y}_{-i} := \bigtimes_{j \neq i}^N \mathcal{Y}_j$ is the domain of $\boldsymbol{Y}$ without Y_i, C corresponds to the constant from Condition 1, and D matches the constant defined in (D.4). The bounds for $a = 1, \dots, N+1$ are 0 as the corresponding statistics are not affected by changes in $\boldsymbol{y}$. For $a = N+2$, the maximal change is bounded by the number of units j such that $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$, which is D^2, times the maximal value C of the predictors. The remainder of the proof of Lemma 8 resembles the proof of Lemma 7.

Lemma 9. *Consider the model of Corollary 1. If Conditions 1, 4, and 5 are satisfied with $\vartheta \in [0, 1/18)$, we obtain the following bounds for all elements of $\boldsymbol{B}(\boldsymbol{\theta}, \boldsymbol{w})$, being the covariance matrix of the sufficient statistics $b_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ and $b_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ defined in*

Section D.1, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and all $\boldsymbol{w} \in \mathcal{W}$:

$$B_{1,1}(\boldsymbol{\theta}, \boldsymbol{w}) \quad \leq \quad \frac{N D^5}{4}$$

$$|B_{1,2}(\boldsymbol{\theta}, \boldsymbol{w})| \quad \leq \quad \frac{N C^2 D^5}{4}$$

$$B_{2,2}(\boldsymbol{\theta}, \boldsymbol{w}) \quad \leq \quad \frac{N C^2 D^5}{4}.$$

PROOF OF LEMMA 9. We first bound $B_{1,1}(\boldsymbol{\theta}, \boldsymbol{w})$ from above as follows:

$$
\begin{aligned}
B_{1,1}(\boldsymbol{\theta}, \boldsymbol{w}) &= \sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j} \left(s_{N+1}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z}) \right) \\
&= \sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j} \left(\sum_{a=1}^{N} \sum_{b=a+1}^{N} d_{a,b}(\boldsymbol{Z}) \, Z_{a,b} \right) \\
&= \sum_{i=1}^{N} \sum_{j=i+1}^{N} c_{i,j} \, \mathbb{V}_{z,i,j} \left(\sum_{a=1}^{N} \sum_{b=a+1}^{N} Z_{a,b} \, d_{a,b}(\boldsymbol{Z}) \right) \\
&\leq \sum_{i=1}^{N} \sum_{j=i+1}^{N} c_{i,j} \, D^2 \left(\sum_{a=1}^{N} \sum_{b=a+1}^{N} \mathbb{V}_{z,i,j} \left(Z_{a,b} \, d_{a,b}(\boldsymbol{Z}) \right) \right)
\end{aligned}
\tag{D.42}
$$

where D matches the constant defined in (D.4) and the function $d_{a,b}(\boldsymbol{Z})$ is defined in (D.2). On the second line of (D.42), we use that the fact that $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$ implies that $d_{i,j}(\boldsymbol{Z}) \, Z_{i,j} = 0$ and $d_{a,b}(\boldsymbol{Z}) \, Z_{a,b}$ does not depend on $Z_{i,j}$ for any $\{a,b\} \neq \{i,j\}$. For the inequality in the last line of (D.42), we use the fact that the number of pairs (a,b) for which $d_{a,b}(\boldsymbol{Z}) \, Z_{a,b}$ is a function of $Z_{i,j}$ is bounded above by D (see proof of Lemma 19 in S25). Invoking Lemma 15 of S25 together with applying

$$
\sum_{a=1}^{N} \sum_{b=a+1}^{N} \mathbb{V}_{z,i,j} \left(d_{a,b}(\boldsymbol{Z}) \, Z_{a,b} \right) \quad \leq \quad \frac{D}{4}
$$

gives:

$$
B_{1,1}(\boldsymbol{\theta}, \boldsymbol{w}) \quad \leq \quad \sum_{i=1}^{N} \sum_{j=i+1}^{N} c_{i,j} \, D^2 \left(\sum_{a=1}^{N} \sum_{b=a+1}^{N} \mathbb{V}_{z,i,j} \left(Z_{a,b} \, d_{a,b}(\boldsymbol{Z}) \right) \right) \quad \leq \quad \frac{N D^5}{4}
$$

We proceed with bounding $B_{2,2}(\boldsymbol{\theta}, \boldsymbol{w})$:

$$
\begin{aligned}
B_{2,2}(\boldsymbol{\theta}, \boldsymbol{w}) &= \sum_{i=1}^{N} \mathbb{V}_{y,i}\left(s_{N+2}(\boldsymbol{x}, \boldsymbol{Y}, \boldsymbol{z})\right) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j}\left(s_{N+2}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{Z})\right) \\
&= \sum_{i=1}^{N} \mathbb{V}_{y,i}\left(\left(\sum_{j=1}^{N} c_{i,j}\, x_j\, z_{i,j}\right) Y_i\right) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \mathbb{V}_{z,i,j}\left(c_{i,j}\left(x_i\, y_j + x_j\, y_i\right) Z_{i,j}\right) \\
&= \sum_{i=1}^{N} \left(\sum_{j=1}^{N} c_{i,j}\, x_j\, z_{i,j}\right)^2 \mathbb{V}_{y,i}\left(Y_i\right) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} c_{i,j}\left(x_i\, y_j + x_j\, y_i\right)^2 \mathbb{V}_{z,i,j}\left(Z_{i,j}\right) \\
&\leq \frac{5\, N\, C^2\, D^4}{4} \leq \frac{N\, C^2\, D^5}{4},
\end{aligned}
$$

because $|x_j| \leq C$ according to Condition 1. For the first inequality, we also use that

$$
\sum_{j=1}^{N} c_{i,j} \leq D^2
$$

by Lemma 15 in S25. We obtain

$$
\max\{B_{1,1}(\boldsymbol{\theta}, \boldsymbol{w}),\, B_{2,2}(\boldsymbol{\theta}, \boldsymbol{w})\} \leq \frac{N\, C^2\, D^5}{4},
$$

which provides an upper bound on $|B_{1,2}(\boldsymbol{\theta}, \boldsymbol{w})|$ by the Cauchy-Schwarz inequality:

$$
|B_{1,2}(\boldsymbol{\theta}, \boldsymbol{w})| \leq \sqrt{B_{1,1}(\boldsymbol{\theta}, \boldsymbol{w})}\, \sqrt{B_{2,2}(\boldsymbol{\theta}, \boldsymbol{w})} \leq \frac{N\, C^2\, D^5}{4}.
$$

Lemma 10. *Consider the model of Corollary 1. Define*

$$
\pi_{v, \boldsymbol{w}_{-v}, \boldsymbol{w}'_{-v}} := \left\|\mathbb{P}_{\boldsymbol{\theta}}(\,\cdot\mid \boldsymbol{w}_{-v}) - \mathbb{P}_{\boldsymbol{\theta}}(\,\cdot\mid \boldsymbol{w}'_{-v})\right\|_{TV}
$$

$$
\pi^{\star} := \max_{1 \leq v \leq M}\ \max_{(\boldsymbol{w}_{-v}, \boldsymbol{w}'_{-v}) \in \mathcal{W}_{-v} \times \mathcal{W}_{-v}}\ \pi_{v, \boldsymbol{w}_{-v}, \boldsymbol{w}'_{-v}}.
$$

Let $D \in \{2, 3, \ldots\}$ be the maximum degree of vertices $Z_{i,j}$ in $\mathcal{G}$. Then

$$
\pi^{\star} \leq \frac{1}{1 + \exp(-C\, D^2\, \|\boldsymbol{\theta}^{\star}\|_{\infty})}.
$$

Proof of Lemma 10. The proof of Lemma 10 resembles the proof of Lemma 21 in

S25, adapted to the bounds on the conditional probabilities derived in Lemmas 7 and 8.

We distinguish four cases, where W_v with $v \in \{1, \ldots, M\}$ relates to:

1. Connection $Z_{i,j}$ of a pair of nodes $\{i, j\} \subset \mathcal{P}_N$ with $\mathcal{N}_i \cap \mathcal{N}_j = \emptyset$.

2. Attribute Y_i with $i \in \mathcal{P}_N$ and $\{j \in \mathcal{P}_N : \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset\} = \emptyset$.

3. Connection $Z_{i,j}$ of a pair of nodes $\{i, j\} \subset \mathcal{P}_N$ with $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$.

4. Attribute Y_i with $i \in \mathcal{P}_N$ and $\{j \in \mathcal{P}_N : \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset\} \neq \emptyset$.

In cases 1 and 2, W_v is independent of $\boldsymbol{W}_{-v}$, so that $\pi_{v,\boldsymbol{w}_{-v},\boldsymbol{w}'_{-v}} = 0$; note that case 2 cannot occur, because Condition 1 ensures that there are no units $i \in \mathcal{P}_N$ with $\{j \in \mathcal{P}_N : \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset\} = \emptyset$. In cases 3 and 4, W_v depends on a non-empty subset of other vertices in $\mathcal{G}$. Consider any $v \in \{1, \ldots, M\}$ such that $\pi_{v,\boldsymbol{w}_{-v},\boldsymbol{w}'_{-v}} > 0$ for some $(\boldsymbol{w}_{-v}, \boldsymbol{w}'_{-v}) \in \mathcal{W}_{-v} \times \mathcal{W}_{-v}$ and define

$$a_{0,v} := \mathbb{P}_{\boldsymbol{\theta}}(W_v = 0 \mid \boldsymbol{W}_{-v} = \boldsymbol{w}_{-v}) \text{ and } a_{1,v} := \mathbb{P}_{\boldsymbol{\theta}}(W_v = 1 \mid \boldsymbol{W}_{-v} = \boldsymbol{w}_{-v})$$

$$b_{0,v} := \mathbb{P}_{\boldsymbol{\theta}}(W_v = 0 \mid \boldsymbol{W}_{-v} = \boldsymbol{w}'_{-v}) \text{ and } b_{1,v} := \mathbb{P}_{\boldsymbol{\theta}}(W_v = 1 \mid \boldsymbol{W}_{-v} = \boldsymbol{w}'_{-v}).$$

Lemma 21 in S25 shows that

$$\pi_{v,\boldsymbol{w}_{-v},\boldsymbol{w}'_{-v}} \leq \min\{\max\{a_{0,v}, b_{0,v}\}, \max\{a_{1,v}, b_{1,v}\}\}.$$

Plugging in the bounds on the conditional probabilities in Lemmas 7 and 8, we obtain

$$\pi_{v,\boldsymbol{w}_{-v},\boldsymbol{w}'_{-v}} \leq \frac{1}{1 + \exp\left(-C\, D^2\, \|\boldsymbol{\theta}^\star\|_\infty\right)}, \quad v \in \mathcal{V}_Z$$

and

$$\pi_{v,\boldsymbol{w}_{-v},\boldsymbol{w}'_{-v}} \leq \frac{1}{1 + \exp\left(-C\, D^2\, \|\boldsymbol{\theta}^\star\|_\infty\right)}, \quad v \in \mathcal{V}_Y.$$

Since $D \in \{2, 3, \ldots\}$, we obtain

$$\pi^\star := \max_{1 \leq v \leq M} \max_{(\boldsymbol{w}_{-v}, \boldsymbol{w}'_{-v}) \in \mathcal{W}_{-v} \times \mathcal{W}_{-v}} \pi_{v,\boldsymbol{w}_{-v},\boldsymbol{w}'_{-v}} \leq \frac{1}{1 + \exp(-C\, D^2\, \|\boldsymbol{\theta}^\star\|_\infty)}.$$

Lemma 11. *Consider the model of Corollary 1. If Conditions 1 and 4 are satisfied along with either Condition 2 or Condition 5 with $\vartheta \in [0, 1/18)$, there exists an integer*

44

$N_0 \in \{3, 4, \ldots\}$ *such that, for all* $N > N_0$,

$$\mathbb{P}(\boldsymbol{W} \notin \mathcal{H}) \;\leq\; \frac{4}{\max\{N,\,p\}^2},$$

where $\mathcal{H}$ *is defined in* (D.8).

PROOF OF LEMMA 11. We prove Lemma 11 by showing that

$$
\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{N} \|\boldsymbol{H}_{i,1}(\boldsymbol{W})\|_{\infty} \;<\; \frac{N}{2\,(1+\chi(\boldsymbol{\theta}^\star))^2}\right) &\;\leq\; \frac{2}{\max\{N,\,p\}^2} \\
\mathbb{P}\left(\sum_{i=1}^{N} \|\boldsymbol{H}_{i,2}(\boldsymbol{W})\|_{\infty} \;<\; \frac{c^2\,N}{2\,(1+\chi(\boldsymbol{\theta}^\star))}\right) &\;\leq\; \frac{2}{\max\{N,\,p\}^2}.
\end{aligned}
\tag{D.43}
$$

To prove the first line of (D.43), we first bound $(1/2)\sum_{i=1}^{N}\|\boldsymbol{H}_{i,1}(\boldsymbol{W})\|_{\infty}$ from below. We then use Theorem 1 of Chazottes et al. (2007, p. 207) to concentrate $\sum_{i=1}^{N}\|\boldsymbol{H}_{i,1}(\boldsymbol{W})\|_{\infty}$. Last, but not least, we show that there exists an integer $N_0 \in \{3, 4, \ldots\}$ such that the obtained lower bound for $(1/2)\sum_{i=1}^{N}\|\boldsymbol{H}_{i,1}(\boldsymbol{W})\|_{\infty}$ is, with high probability, greater than the deviation of $\sum_{i=1}^{N}\|\boldsymbol{H}_{i,1}(\boldsymbol{W})\|_{\infty}$ from its mean. The first line of (D.43) follows from combining these steps. The second line of (D.43) can be established along the same lines. A union bound then establishes the desired result:

$$
\begin{aligned}
\mathbb{P}(\boldsymbol{W} \notin \mathcal{H}) \;\leq\;\; & \mathbb{P}\left(\sum_{i=1}^{N} \|\boldsymbol{H}_{i,1}(\boldsymbol{W})\|_{\infty} \;<\; \frac{N}{2\,(1+\chi(\boldsymbol{\theta}^\star))^2}\right) \\
& +\; \mathbb{P}\left(\sum_{i=1}^{N} \|\boldsymbol{H}_{i,2}(\boldsymbol{W})\|_{\infty} \;<\; \frac{c^2\,N}{2\,(1+\chi(\boldsymbol{\theta}^\star))}\right) \\
\;\leq\;\; & \frac{4}{\max\{N,\,p\}^2}.
\end{aligned}
$$

Step 1: Condition 1 implies that, for each unit $i \in \mathcal{P}_N$, there exists a unit $j \in \mathcal{P}_N \setminus \{i\}$ such that $\mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$ and $x_j \in [c, C]$. Thus, by Lemma 7, Lemma 17 of S25, and Conditions 1 and 2, we obtain

$$\frac{1}{2}\sum_{i=1}^{N}\|\boldsymbol{H}_{i,1}(\boldsymbol{w})\|_{\infty} \;\geq\; \frac{1}{2}\sum_{i=1}^{N}\mathbb{E}\,d_{i,j}(\boldsymbol{Z}) \;\geq\; \frac{N}{2\,(1+\chi(\boldsymbol{\theta}^\star))^2} \;\geq\; \frac{N}{4\,\chi(\boldsymbol{\theta}^\star)^2} \;\geq\; \frac{N^{1-2\,\vartheta}}{4\,\exp(2\,E)}.$$

Theorem 1 of Chazottes et al. (2007, p. 207) implies

$$\mathbb{P}\left(\left|\sum_{i=1}^{N}\|\boldsymbol{H}_{i,1}(\boldsymbol{w})\|_{\infty} - \mathbb{E}\sum_{i=1}^{N}\|\boldsymbol{H}_{i,1}(\boldsymbol{W})\|_{\infty}\right| < t\right) \geq 1 - 2\exp\left(-\frac{2\,t^2}{\Psi_N^2\,\|\!\|\mathcal{D}_N(\boldsymbol{\theta}^{\star})\|\!\|_2^2}\right).$$

Choosing

$$t \ := \ \sqrt{\log\max\{N,\,p\}}\,\Psi_N\,\|\!\|\mathcal{D}_N(\boldsymbol{\theta}^{\star})\|\!\|_2$$

gives

$$\mathbb{P}\left(\left|\sum_{i=1}^{N}\|\boldsymbol{H}_{i,1}(\boldsymbol{W})\|_{\infty} - \mathbb{E}\sum_{i=1}^{N}\|\boldsymbol{H}_{i,1}(\boldsymbol{W})\|_{\infty}\right| < \sqrt{\log\max\{N,\,p\}}\,\Psi_N\,\|\!\|\mathcal{D}_N(\boldsymbol{\theta}^{\star})\|\!\|_2\right)$$
$$\geq \ 1 - \frac{2}{\max\{N,\,p\}^2}.$$

Next, we demonstrate that there exists an integer $N_1 \in \{3, 4, \dots\}$ such that, for all $N > N_1$,

$$\sqrt{\log\max\{N,\,p\}}\,\Psi_N\,\|\!\|\mathcal{D}_N(\boldsymbol{\theta}^{\star})\|\!\|_2 \ \leq \ \frac{N^{1-2\,\vartheta}}{4\,\exp(2\,E)}.$$

To do so, we bound the three terms one by one. Using $\max\{N,\,p\} = N + 2$, the first term, $\sqrt{\log\max\{N,\,p\}}$, is bounded above by $\sqrt{\log\max\{N,\,p\}} \leq 2\sqrt{\log N}$ provided $N \geq 2$. The second term is bounded above by $\Psi_N \ \leq \ D\sqrt{N}$ as shown in the proof of Lemma 14 in S25. The third term is bounded above by $\|\!\|\mathcal{D}_N(\boldsymbol{\theta}^{\star})\|\!\|_2 < C_3$ by Lemma 5, where $C_3 > 0$ is a constant.

Combining these results gives

$$2\sqrt{N\,\log N}\,C_3,\, D \ \leq \ \frac{N^{1-\vartheta}}{4\,\exp(E)}$$
$$8\,C_3\,D\,\exp(E) \ \leq \ \sqrt{\frac{N^{1-2\,\vartheta}}{\log N}}.$$

Similar to the proof of Lemma 14 in S25, this implies

$$\mathbb{P}\left(\sum_{i=1}^{N}\|\boldsymbol{H}_{i,1}(\boldsymbol{W})\|_{\infty} \geq \frac{N}{2\,(1 + \chi(\boldsymbol{\theta}^{\star}))^2}\right) \ \geq \ 1 - \frac{2}{\max\{N,\,p\}^2}.$$

Step 2: Conditions 1 and 2 along with Lemma 7 establish

$$\frac{1}{2} \sum_{i=1}^{N} \|\boldsymbol{H}_{i,2}(\boldsymbol{w})\|_{\infty} \; \geq \; \frac{c^2}{2} \sum_{i=1}^{N} \mathbb{E}\, Z_{i,j} \; \geq \; \frac{c^2\, N}{2\,(1 + \chi(\boldsymbol{\theta}^{\star}))} \; \geq \; \frac{c^2\, N}{4\,\chi(\boldsymbol{\theta}^{\star})} \; \geq \; \frac{c^2\, N^{1-\vartheta}}{4\,\exp(E)}.$$

Once more, we invoke Theorem 1 of Chazottes et al. (2007, p. 207) to obtain

$$\mathbb{P}\left(\left| \sum_{i=1}^{N} \|\boldsymbol{H}_{i,2}(\boldsymbol{W})\|_{\infty} - \mathbb{E} \sum_{i=1}^{N} \|\boldsymbol{H}_{i,2}(\boldsymbol{W})\|_{\infty} \right| < \sqrt{\log \max\{N,\, p\}}\; \Psi_N \, \|\!|\mathcal{D}_N(\boldsymbol{\theta}^{\star})|\!\|_2 \right)$$

$$\geq \; 1 - \frac{2}{\max\{N,\, p\}^2}.$$

We proceed by showing that there exists an integer $N_2 \in \{3, 4, \dots\}$ such that, for all $N > N_2$,

$$\sqrt{\log \max\{N,\, p\}}\; \Psi_N \, \|\!|\mathcal{D}_N(\boldsymbol{\theta}^{\star})|\!\|_2 \; \leq \; \frac{c^2\, N^{1-\vartheta}}{4\,\exp(E)}. \tag{D.44}$$

We bound the three terms on the left-hand side of (D.44) one by one. The bounds on the first term, $\sqrt{\log \max\{N,\, p\}}$, and third term, $\|\!|\mathcal{D}_N(\boldsymbol{\theta}^{\star})|\!\|_2$, are the same as in the first step. With regard to the second term, we obtain $\Psi_N \leq C_2 \sqrt{N}$ by the proof of Lemma 4 with $C_2 > 0$.

Combining these bounds gives

$$2\sqrt{N \log N}\, C_3 \; \leq \; \frac{c^2\, N^{1-\vartheta}}{4\,\exp(E)}$$

$$\frac{8}{c^2}\, C_3 \, \exp(E) \; \leq \; \sqrt{\frac{N^{1-2\vartheta}}{\log N}},$$

which vanishes as $N \to \infty$ under Conditions 2 and 5 with $\vartheta \in [0,\, 1/18)$. Thus, for all $N > N_2$,

$$\mathbb{P}\left(\sum_{i=1}^{N} \|\boldsymbol{H}_{i,2}(\boldsymbol{W})\|_{\infty} \; \geq \; \frac{c^2\, N}{2\,(1 + \chi(\boldsymbol{\theta}^{\star}))} \right) \; \geq \; 1 - \frac{2}{\max\{N,\, p\}^2}.$$

E   Quasi-Newton Acceleration

The two-step algorithm described in Section 3.2 iterates two steps:

Step 1: Update $\boldsymbol{\theta}_1^{(t)}$ given $\boldsymbol{\theta}_2^{(t-1)}$ using a MM algorithm with a linear convergence rate (Böhning and Lindsay, 1988, Theorem 4.1).

Step 2: Update $\boldsymbol{\theta}_2^{(t)}$ given $\boldsymbol{\theta}_1^{(t)}$ using a Newton-Raphson update with a quadratic convergence rate.

To accelerate Step 1, we use quasi-Newton methods (Lange et al., 2000): We approximate the difference between $(\boldsymbol{A}^\star)^{-1}$ and $[\boldsymbol{A}(\boldsymbol{\theta}^{(t)})]^{-1}$, defined in Lemma 2 and Equation (9), respectively, by rank-one updates.

A first-order Taylor approximation of $\nabla_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(t)})$ around $\boldsymbol{\theta}_1^{(t)}$ shows that

$$-\boldsymbol{A}(\boldsymbol{\theta}^{(t)}) \, \boldsymbol{k}(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_1; \boldsymbol{\theta}_2^{(t)}) \; \approx \; \boldsymbol{\theta}_1^{(t)} - \boldsymbol{\theta}_1, \tag{E.1}$$

where

$$\boldsymbol{k}(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_1; \boldsymbol{\theta}_2^{(t)}) \; := \; \nabla_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(t)})\Big|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(t)}} - \nabla_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(t)}). \tag{E.2}$$

Since a standard Newton-Raphson algorithm corresponds to (E.1) with $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(t-1)}$, the change in consecutive estimates carries information on $[\boldsymbol{A}(\boldsymbol{\theta}^{(t)})]^{-1}$, which we want to approximate. More specifically, we approximate the difference between $(\boldsymbol{A}^\star)^{-1}$ and $[\boldsymbol{A}(\boldsymbol{\theta}^{(t)})]^{-1}$. Thus we write $(\boldsymbol{A}^\star)^{-1} - [\boldsymbol{A}(\boldsymbol{\theta}^{(t)})]^{-1} =: \boldsymbol{M}^{(t)}$ and set $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(t-1)}$, so that (E.1) becomes

$$\boldsymbol{M}^{(t)} \, \boldsymbol{k}(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_1^{(t-1)}; \boldsymbol{\theta}_2^{(t)}) \; = \; (\boldsymbol{\theta}_1^{(t)} - \boldsymbol{\theta}_1^{(t-1)}) + (\boldsymbol{A}^\star)^{-1} \, \boldsymbol{k}(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_1^{(t-1)}; \boldsymbol{\theta}_2^{(t)}) =: \boldsymbol{r}^{(t)}, \tag{E.3}$$

which is called the inverse secant condition for updating $\boldsymbol{M}^{(t)}$. Given that (E.3) relates $[\boldsymbol{A}(\boldsymbol{\theta}^{(t)})]^{-1}$ to the score functions through the definition of $\boldsymbol{k}(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_1^{(t-1)}; \boldsymbol{\theta}_2^{(t)})$ in (E.2) and estimates $\boldsymbol{\theta}_1^{(t)}$ and $\boldsymbol{\theta}_1^{(t-1)}$, the updates of $\boldsymbol{M}^{(t+1)}$ will be based on (E.3). We employ the parsimonious symmetric, rank-one update of Davidon (1991) to satisfy (E.3) by updating $\boldsymbol{M}^{(t)}$ as follows:

$$\boldsymbol{M}^{(t)} \; = \; \boldsymbol{M}^{(t-1)} + \boldsymbol{q}^{(t)} \left(\boldsymbol{q}^{(t)}\right)^\top \left[c^{(t)}\right]^{-1}, \tag{E.4}$$

with $\boldsymbol{q}^{(t)} := \boldsymbol{r}^{(t)} - \boldsymbol{M}^{(t-1)} \boldsymbol{k}(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t-1)}; \boldsymbol{\theta}_2^{(t)})$ and $c^{(t)} := (\boldsymbol{q}^{(t)})^\top \boldsymbol{k}(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t-1)}; \boldsymbol{\theta}_2^{(t)})$. We seed the algorithm with the MM update described in Section 3.2 by setting $\boldsymbol{M}^{(0)} = \boldsymbol{0}$, the $N \times N$ null matrix.

In summary, the quasi-Newton acceleration of the MM algorithm updates $\boldsymbol{\theta}_1^{(t)}$ given $\boldsymbol{\theta}_2^{(t)}$ as follows:

Step 1: Calculate $\boldsymbol{k}(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_1^{(t-1)}; \boldsymbol{\theta}_2^{(t)})$ defined in (E.2).

Step 2: Update $\boldsymbol{M}^{(t)}$ according to (E.4).

Step 3: Update $\boldsymbol{\theta}_1^{(t+1)}$ from $\boldsymbol{\theta}_1^{(t)}$:

$$\boldsymbol{\theta}_1^{(t+1)} = \boldsymbol{\theta}_1^{(t)} + \left[(\boldsymbol{A}^\star)^{-1} - \boldsymbol{M}^{(t)} \right] \left[\nabla_{\boldsymbol{\theta}_1} \ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(t)}) \Big|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(t)}} \right]. \tag{E.5}$$

Unlike the unaccelerated MM algorithm, the described quasi-Newton algorithm does not guarantee that $\ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t+1)}) \geq \ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t)})$. Therefore, $\boldsymbol{\theta}_1^{(t+1)}$ is updated by either the quasi-Newton update (E.5) or the MM update (10), whichever gives rise to the higher pseudo-likelihood. The resulting updates slightly increase the computing time per iteration while potentially dramatically decreasing the total number of iterations.

F   MM Algorithm: Directed Connections

If connections are directed, $Z_{i,j}$ may differ from $Z_{j,i}$. In such cases, the pseudo-loglikelihood can be written as

$$\ell(\boldsymbol{\theta}) \quad := \quad \sum_{i=1}^N \ell_i(\boldsymbol{\theta}) + \sum_{i=1}^{N-1} \sum_{j=1, \, j \neq i}^N \ell_{i,j}(\boldsymbol{\theta}),$$

where ℓ_i and $\ell_{i,j}$ are defined by

$$\ell_i(\boldsymbol{\theta}) \quad := \quad \log p_{\boldsymbol{\theta}}(y_i \mid \boldsymbol{y}_{-i}, \boldsymbol{z}) \quad \text{and} \quad \ell_{i,j}(\boldsymbol{\theta}) \quad := \quad \log p_{\boldsymbol{\theta}}(z_{i,j} \mid \boldsymbol{y}, \boldsymbol{z}_{-\{i,j\}}).$$

We partition the parameter vector $\boldsymbol{\theta} := (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \mathbb{R}^{2N+12}$ into

- the nuisance parameter vector: $\boldsymbol{\theta}_1 := (\alpha_{Z,O,1}, \ldots, \alpha_{Z,O,N}, \alpha_{Z,I,1}, \ldots, \alpha_{Z,I,N-1}) \in \mathbb{R}^{2N-1}$;

- the parameter vector of primary interest: $\boldsymbol{\theta}_2 := (\alpha_Y, \beta_{X,Y,1}, \beta_{X,Y,2}, \beta_{X,Y,3}, \lambda, \gamma_{Z,Z,1},$ $\gamma_{Z,Z,2}, \gamma_{X,Z,1}, \gamma_{X,Z,2}, \gamma_{X,Z,3}, \gamma_{X,Z,4}, \gamma_{Y,Z}, \gamma_{X,Y,Z}) \in \mathbb{R}^{13}$.

As explained in Section 6.1, $\alpha_{Z,N,I}$ is set to 0 in order to address identifiability issues. The negative Hessian is partitioned in accordance:

$$-\nabla_{\boldsymbol{\theta}}^2 \ell(\boldsymbol{\theta}) := \begin{pmatrix} \boldsymbol{A}(\boldsymbol{\theta}) & \boldsymbol{B}(\boldsymbol{\theta}) \\ \boldsymbol{B}(\boldsymbol{\theta})^\top & \boldsymbol{C}(\boldsymbol{\theta}) \end{pmatrix},$$

where $\boldsymbol{A}(\boldsymbol{\theta}) \in \mathbb{R}^{(2N-1)\times(2N-1)}$, $\boldsymbol{B}(\boldsymbol{\theta}) \in \mathbb{R}^{(2N-1)\times13}$, and $\boldsymbol{C}(\boldsymbol{\theta}) \in \mathbb{R}^{13\times13}$. Writing $\ell(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ in place of $\ell(\boldsymbol{\theta})$, we compute at iteration $t+1$:

Step 1: Given $\boldsymbol{\theta}_2^{(t)}$, find $\boldsymbol{\theta}_1^{(t+1)}$ satisfying $\ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t)}) \geq \ell(\boldsymbol{\theta}_1^{(t)}, \boldsymbol{\theta}_2^{(t)})$.

Step 2: Given $\boldsymbol{\theta}_1^{(t+1)}$, find $\boldsymbol{\theta}_2^{(t+1)}$ satisfying $\ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t+1)}) \geq \ell(\boldsymbol{\theta}_1^{(t+1)}, \boldsymbol{\theta}_2^{(t)})$.

In Step 1, it is inconvenient to invert the high-dimensional $(2N-1) \times (2N-1)$ matrix

$$\boldsymbol{A}(\boldsymbol{\theta}^{(t)}) := -\sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \nabla_{\boldsymbol{\theta}_1}^2 \ell_{i,j}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(t)}) \Big|_{\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^{(t)}} = \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \pi_{i,j}^{(t)}(1 - \pi_{i,j}^{(t)}) \boldsymbol{e}_{i,j} \boldsymbol{e}_{i,j}^\top.$$

Note that the definition of vector $\boldsymbol{e}_{i,j} \in \mathbb{R}^{2N-1}$ differs from the undirected case described in Section 3.2. For $j \neq N$, let $\boldsymbol{e}_{i,j}$ be the $(2N-1)$-vector whose ith and $(j+N)$th coordinates are 1 and whose other coordinates are 0. For $j = N$, let $\boldsymbol{e}_{i,j}$ be the $(2N-1)$-vector whose ith coordinate is 1 and whose other coordinates are 0. Along the lines of the MM algorithm for undirected connections described in Section 3.2, we increase ℓ by maximizing a minorizing function of ℓ, replacing $\boldsymbol{A}(\boldsymbol{\theta}^{(t)})$ by a constant matrix $\boldsymbol{A}^\star$ that is more convenient to invert. The constant matrix $\boldsymbol{A}^\star$ is defined as

$$\boldsymbol{A}^\star := \begin{pmatrix} \boldsymbol{A}_{1,1}^\star & \boldsymbol{A}_{1,2}^\star \\ (\boldsymbol{A}_{1,2}^\star)^\top & \boldsymbol{A}_{2,2}^\star \end{pmatrix}$$

where

- $\boldsymbol{A}_{1,1}^{\star} \in \mathbb{R}^{N \times N}$ and $\boldsymbol{A}_{2,2}^{\star} \in \mathbb{R}^{(N-1) \times (N-1)}$ are diagonal matrices with elements $(N-1)/4$ on the main diagonal;

- $\boldsymbol{A}_{1,2}^{\star} \in \mathbb{R}^{N \times (N-1)}$ is a matrix with vanishing elements on its main diagonal and off-diagonal elements $1/4$.

Applying Theorem 8.5.11 in Harville (1997) to $\boldsymbol{A}_{1,2}^{\star}$ and $\boldsymbol{A}^{\star}$ shows that matrix can be inverted in $O(N)$ operations. With the above change in the constant matrix $\boldsymbol{A}^{\star}$, we estimate $\boldsymbol{\theta}$ along the lines of Section 3.2.

G Hate Speech on X: Additional Information

G.1 Data

For the application, we use posts of $N = 2{,}191$ U.S. state legislators on the social media platform X collected by Kim et al. (2022) in the six months leading up to and including the insurrection at the United States Capitol on January 6, 2021. We restrict attention to active legislators, that is, legislators who posted during the aforementioned period and mentioned or reposted content from other active legislators. Since reposts do not necessarily reflect politicians' opinions, we exclude all reposts and non-unique posts that are direct copies of other users' messages to gather information on responses. Employing large language models of Camacho-Collados et al. (2022) pre-trained on these posts enables categorizing the 109,974 posts into those containing hate speech statements versus those that do not. Accordingly, the binary attribute Y_i equals 1 if the corresponding legislator sent at least one post classified as hate speech and 0 otherwise. The algorithm of Camacho-Collados et al. (2022) provides for each Tweet a continuous value between 0 and 1. We classify

the respective Tweet as using hate speech if its value is larger than 0.5. The attribute $x_{i,1} \in \{0,1\}$ is 1 if legislator i is a Republican and 0 otherwise. In addition, we incorporate information on each legislator's gender ($x_{i,2} = 1$ if legislator i is female and 0 otherwise), race ($x_{i,3} = 1$ if legislator is white and 0 otherwise), and state ($x_{i,4}$). On the social media platform X, users have the ability to either mention or repost other users' posts. The resulting network, denoted as $\boldsymbol{Z}$, is based on the mentions and reposts exchanged between January 6, 2020 and January 6, 2021: $Z_{i,j} = 1$ if legislator i mentioned or reposted legislator j in a post.

G.2 Plots

In addition to the goodness-of-fit checks reported in Section 6, we assess whether the model preserves salient characteristics of connections $\boldsymbol{Z}$. Figure 5 suggests that the proposed model captures the shared partner distribution, i.e., the numbers of connected pairs of legislators $\{i, j\} \subset \mathcal{P}_N$ with 1, 2, ... shared partners.
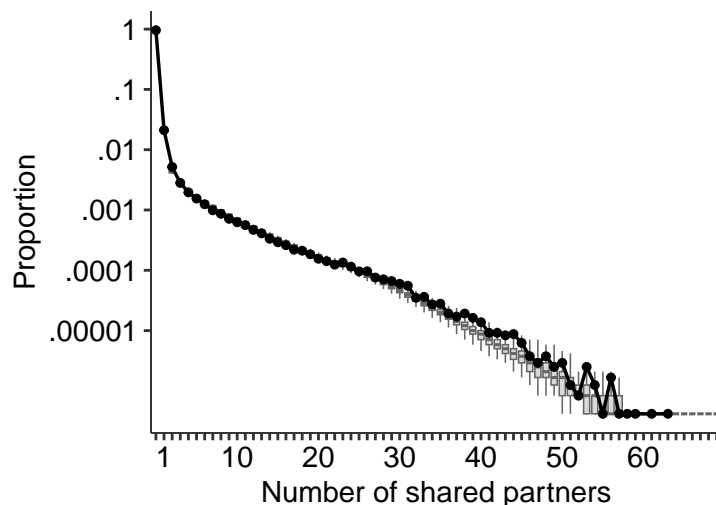
Figure 5: Hate speech on X: The red line indicates the observed shared partners distribution of the network of repost and mention interactions of U.S. legislators, while the boxplots represent the shared partners distributions of simulated networks from the estimated model.

Supplementary Material References

Böhning, D. and B. G. Lindsay (1988). Monotonicity of quadratic-approximation algorithms. *Annals of the Institute of Statistical Mathematics 40*, 641–663.

Camacho-Collados, J., K. Rezaee, T. Riahi, A. Ushio, D. Loureiro, D. Antypas, J. Boisson, L. Espinosa-Anke, F. Liu, E. Martínez-Cámara, G. Medina, T. Buhrmann, L. Neves, and F. Barbieri (2022). TweetNLP: Cutting-Edge Natural Language Processing for Social Media. In *Proceedings of the 2022 conference on empirical methods in natural language processing: System demonstrations.*

Chazottes, J. R., P. Collet, C. Külske, and F. Redig (2007). Concentration inequalities for random fields via coupling. *Probability Theory and Related Fields 137*, 201–225.

Davidon, W. C. (1991). Variable metric method for minimization. *SIAM Journal on Optimization 1*, 1–17.

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. New York, NY: Springer.

Kim, T., N. Nakka, I. Gopal, B. A. Desmarais, A. Mancinelli, J. J. Harden, H. Ko, and F. J. Boehmke (2022). Attention to the COVID-19 pandemic on Twitter: Partisan differences among U.S. state legislators. *Legislative Studies Quarterly 47*, 1023–1041.

Lange, K., D. R. Hunter, and I. Yang (2000). Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics 9*, 1–20.

Lindvall, T. (2002). *Lectures On The Coupling Method*. Courier Corporation.

Maathuis, M., M. Drton, S. Lauritzen, and M. Wainwright (2019). *Handbook of Graphical Models*. Boca Raton, Florida: CRC Press.

Magnus, J. R. and H. Neudecker (2019). *Matrix Differential Calculus with Applications in Statistics and Econometrics* (Third Edition ed.). Oxford, UK: John Wiley & Sons Ltd.

Rue, H. and L. Held (2005). *Gaussian Markov random fields: theory and applications*. CRC press.

Stewart, J. R. and M. Schweinberger (2025). Pseudo-likelihood-based M-estimators for random graphs with dependent edges and parameter vectors of increasing dimension. *The Annals of Statistics*. To appear.

van den Berg, J. and C. Maes (1994). Disagreement percolation in the study of Markov fields. *The Annals of Probability 22*, 749–763.