Causal Network Influence with Latent Homophily and Measurement Error: An Application to Therapeutic Community

Subhadeep Paul[1]*, Shanjukta Nath[2], Keith Warren[1]

[1]The Ohio State University and [2]Stanford University

Abstract

The Spatial or Network Autoregressive model (SAR, NAM) is popular for modeling the influence network connected neighbors exert on the outcome of individuals. However, many authors have noted that the *causal* network influence or contagion cannot be identified from observational data due to the presence of homophily. We propose a latent homophily-adjusted spatial autoregressive model for networked responses to identify the causal contagion and contextual effects. The latent homophily is estimated from the spectral embedding of the network's adjacency matrix. Separately, we develop maximum likelihood estimators for the parameters of the SAR model correcting for measurement error when covariates are measured with error. We show that the bias corrected MLE are consistent and derive its asymptotic limiting distribution. We propose to estimate network influence using the bias corrected MLE in a SAR model with the estimated latent homophily added as a covariate. Our simulations show that the methods perform well in finite sample. We apply our methodology to a data-set of female criminal offenders in a therapeutic community (TC) for substance abuse and criminal behavior. We provide causal estimates of network influence on graduation from TC and re-incarceration after accounting for latent homophily.

1 Introduction

A common form of data collected in many application problems from diverse domains, including social sciences, public health, political science, psychology, and economics, is a

*SP's research was partially supported by National Science Foundation grant DMS-1830547. We are grateful to Prof. Cosma Shalizi and Prof. Joshua Cape for helpful discussions and feedback.

network on a set of entities and node-level measurements. Such nodal attributes include behavioral outcomes, economic variables, and health outcomes. There is more than half a century of research on understanding how the network and the attributes affect each other [41, 49, 40, 68, 16, 17, 1, 10, 50, 45]. It has also been observed that they are intertwined and evolve together. For example researchers have found evidence of social influence or contagion effect on emotions of individuals [38, 19, 10], patterns of exercising [1], and health and mental outcomes [16, 24, 25]. On the other hand, biological and social networks have been demonstrated to display homophily, or social selection, whereby individuals who are similar in characteristics tend to be linked in a network [52, 41, 22, 68, 77].

A fundamental scientific problem associated with such network-linked data is *identifying the influence* network-connected neighbors exert on individuals' outcomes and detecting *influential entities* within the network. In the literature, such network effects have been called "contagion", "social or network influence", "peer effects," etc. Estimating the *causal* network influence has been an active topic of research for almost three decades [49, 68, 77, 78, 7]. As [49] noted, there are two mechanisms of such effects - endogenous effects or contagion where the response of the peers affect the response of the ego and exogenous or contextual (spillover) effects where the covariates of peers affect the response of the ego. [49] showed that these two effects cannot be identified separately from observational data in linear-in-means models. Moreover, a third effect might be a correlated unobserved variable (latent homophily), which creates omitted variable bias. The contagion effect cannot generally be separated from this latent homophily, using observational social network data [68].

The Spatial autoregression model (SAR) [57], originally developed for spatial data analysis has been adopted for estimating network influence both when a single time point measurement of the network and outcome is available as well as in the longitudinal settings [7, 43, 47, 44, 42, 35, 89, 45]. Recently the network autoregressive model (NAR) [89] and extensions of it [14, 88, 87, 85], have been proposed as tools for modeling and predicting networked time series. Such models are closely related to the longitudinal version of the SAR models where both the outcome and the network are observed over multiple time points.

Separately, in longitudinal settings, a number of high impact papers in public health estimating significant contagion effect on obesity, smoking, happiness [16, 17, 24] led to a series of articles describing issues with modeling and statistical assumptions and invalidity of causal relationships [18, 54, 48, 68]. Since then a number of solutions to circumvent the issues have also been proposed [77, 79, 51, 56].

It is generally possible to separately identify the endogenous contagion effect and the

exogenous contextual effects in a social network using the SAR models. [7, 43, 47, 44, 28]. The key requirements for such identification is that the social network is more general than just a collection of connected peer groups and the error term of the SAR model is independent of the covariates and the network [7, 28]. However the latent homophily still causes non-identification. Recently, several authors have put forth partial solutions to this problem, including using latent communities from a stochastic block model [51], and joint modeling an outcome equation and social network formation model [28, 34]. In the longitudinal case [51] presents a model where the true latent communities are part of the response equation and in practice are replaced with estimated values from a stochastic block model or a latent space model. Taking advantage of the large literature on consistent (and minimax optimal) estimation of community structure in SBM, they show that the least squares estimates of social influence is asymptotically unbiased. [28] suggests joint modeling of an outcome equation and social network formation model both of which contain latent variables and proposes a Bayesian solution. [34] suggests a joint SAR outcome model and network formation model based on the latent space model of [32]. We propose strategies to control bias due to homophily in finite samples in both longitudinal and cross-sectional studies of network influence.

We develop a method to estimate network influence adjusting for homophily by combining a latent position random graph model with measurement error models that performs bias correction in non-asymptotic settings (in finite sample). In longitudinal settings, this approach improves upon the asymptotic bias correction in [51]. Further, while [51] focused only with the case where a time series of outcome is available (and therefore estimation of network influence boils down to linear regression), we also develop methods for the more challenging case of the SAR model when the outcome is available at only one time point and therefore linear regression is not possible. Unlike the Bayesian MCMC procedures in [34, 28] our two-step estimation method involves spectral decomposition and regression, which makes it computationally attractive. Our methodological contributions are two-fold:

Measurement error correction in SAR model: There is extensive literature on correcting for measurement error in covariates in many linear and non linear models [13, 9, 53, 71]. However, there is a lack of methodology for bias correction due to measurement errors in covariates in the SAR models. We develop a methodology for applying bias correction in maximum likelihood estimation of the spatial or network influence parameter in spatial autoregressive model. We show the bias corrected estimator is consistent, and we derive an asymptotic distribution of the estimator. Our simulation studies verify the bias correction property of the resulting estimator.

Homophily adjusted causal network influence estimation: We develop a method that adjusts for latent homophily in network formation while estimating network influence using a cross sectional or longitudinal SAR model. For this purpose, we model the network with a Stochastic Block Model (SBM) and estimate the latent homophily factors from a spectral embedding of the adjacency matrix. The SBM and its extensions are random graph models with a latent community structure which have been extensively studied in the literature [33, 65, 6, 86, 46, 60, 2, 61]. The latent factors estimated through spectral embeddings have been shown to be consistent and minimax optimal for estimating the latent community structure under different setups [46, 26, 75, 15, 63, 62, 76]. We then include these estimated factors in our outcome model.

1.1 Graduation and Recidivism: Therapeutic Community

The methodological developments of this paper are motivated by an application problem involving data on 472 female residents from a therapeutic community in a Midwestern city in the United States. The facility is a low-security correctional facility for felony offenders. The offenses of the residents include possession of drugs, robbery, burglary, domestic violence among others. We track the female residents over three years between 2005 and 2008. A critical characteristic of TCs is that peer affirmations are commonly used for mutual monitoring. We use electronic records of over 61,000 such peer affirmations to form a residents' social network [11, 84, 82, 81]).

Figure 1 shows the affirmations network links received by four nodes (egos) in the facility. We observe that the graduation status of the ego is associated with the density of affirmations received and the graduation status of residents who sent those affirmations. Residents who did not graduate have received fewer affirmations than residents who graduated. Additionally, the fraction of red nodes (did not graduate) is significantly higher for individuals (b), (c), and (d) who themselves did not graduate. Several studies

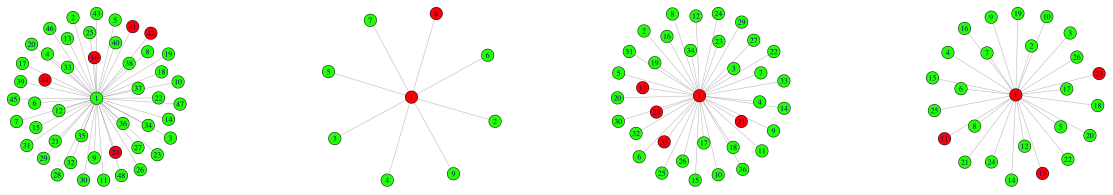


Figure 1: Affirmation Ego Network by Graduation Status for 4 nodes: from top left the central (ego) nodes are (a) graduated and (b), (c) and (d) did not graduate. Green node indicates that resident graduated from TC and red node indicates that resident could not graduate from the TC. An edge indicates an affirmation was sent to the ego node.

in the literature on crime have suggested the presence of peer effects or network influence in criminal behavior. [73] finds contagion in juvenile incarceration and suggests that exposure to aggressive peers increases the likelihood of crime post exit from the correctional facility. Similarly, [4] finds evidence on peer influence in criminal behavior using a sample of juvenile offenders in Florida.

Our contribution to the peer effects literature on crime is twofold. First, we provide estimates on network influence accounting for latent homophily. Our primary goal is to estimate network influence on the propensity to graduate from the TC and recidivism post exit from the facility using a network of written affirmations among the residents. Our results suggest that failure to account for latent homophily can result in biased estimates in our setting. After controlling for homophily and correcting for bias, we find network influence is critical for predicting graduation from the facility but not for recidivism in the future. Additionally, we add to the limited research on peer effects of female offenders and their behavior in correctional facilities. Our work suggests that network interactions for females in TC do not last long and only impact shorter-term outcomes such as graduation from the facility.

2 Spatial autoregressive model with latent homophily

We consider a network encoding relational data among a set of n entities whose adjacency matrix is A. The adjacency matrix A is possibly weighted and directed. The diagonal elements of A are assumed to be 0. We observe an n dimensional vector Y of univariate responses at the vertices of the network. Define the diagonal matrix of degrees D such that $D_i = \sum_j A_{ij}$. The Laplacian matrix or the row normalized adjacency matrix is defined as $L = D^{-1}A$. We further observe an $n \times p$ matrix Z of measurements of p dimensional covariates at each node. Our goal is to model and predict Y using the information contained in the network dependence and the covariate information Z.

We propose the following Spatial Auto-regressive model (SAR) with latent homophily variables as the generative model for the response

$$Y = \rho LY + U\beta + Z\gamma + V, \quad V_i \overset{i.i.d.}{\sim} N(0, \sigma^2) \tag{2.1}$$

where U is a $n \times d$ matrix of latent homophily variables. Each row U_i of the matrix represents a vector of latent variable values for a node and ρ is the network influence parameter of interest. For any node i, the variable $(LY)_i = \sum_j L_{ij}Y_j$, measures a weighted average of the responses of the network connected neighbors of i. Therefore the above model asserts that outcome of i is a function of weighted average of outcomes of its network

connected neighbors, values of the covariates, and a set of latent variables representing unobserved characteristics. A simple rearrangement of the terms in Equation 2.1, taking expectation, and power series expansion shows

$$E[Y] = (I - \rho L)^{-1}(U\beta + Z\gamma) \approx (I + \rho L + \rho^2 L^2 + \rho^3 L^3)(U\beta + Z\gamma)$$

Therefore the model implies a global spillover or contagion process which translates expected direct effects of covariates of some node to effects on outcomes of nodes which are distant in network connections [29]. However, the strength of such spillovers steadily decreases as the network distance increases.

We model the network to be generated following the Random Dot Product Graph (RDPG) model [75, 2, 3, 76].

$$A \sim Bernoulli(P), \quad P = U_{n\times d} U_{d\times n}^T, \quad U_i^T U_j \in [0, 1], \tag{2.2}$$

where U is the same matrix of latent factors as in the outcome equation. We consider the dimension of the latent variable d to be unknown. The RDPG model contains the (positive semideifinite) Stochastic Block Model (SBM), Degree corrected SBM and mixed membership SBM as special cases [2]. The latent factors are therefore assumed to be obtained from a d dimensional continuous latent space that satisfies the specified constraint.

We assume that the nodal covariates Z_is do not directly affect the formation of the network, and is therefore not part of the network generating process in Equation 2.2. We do not assume that the unobserved latent variables U_is and the observed covariates Z_is are uncorrelated. Therefore U_i might be unobserved confounders that affect both the network links and the covariates simultaneously. In determining the effect of the covariates Z we wish to adjust for the latent homophily confounders U. However we do make the assumption that the latent variable U captures all homophily in the network formation.

We will also consider an extension of the model in Equation 2.1, which includes the network lagged covariate effects or the contextual effects:

$$Y = \rho LY + U\beta + Z\gamma + LZ\gamma_1 + V \quad V_i \overset{i.i.d.}{\sim} N(0, \sigma^2) \tag{2.3}$$

Here the vector $(LZ)_i = \sum_j L_{ij} Z_j$ obtains the weighted average of the covariate vectors for all network connected neighbors of i. This model can help us separate the endogenous effect of network influence, the contextual effects of covariates, and the latent homophily.

We estimate the latent factors through a d dimensional spectral embedding of the ad-

jacency matrix. The spectral embedding performs a singular value decomposition (SVD) of the adjacency matrix X. Let $V_{n\times d}$ be the singular vectors corresponding to the d largest singular values and Σ be the diagonal matrix containing those singular values. Then we estimate the latent factor as $\hat{U} = V\Sigma^{1/2}$. In the second step we use these estimated factors as predictors in the SAR model

$$Y = \rho LY + \hat{U}\beta + Z\gamma + V \quad V_i \overset{i.i.d.}{\sim} N(0, \sigma^2). \tag{2.4}$$

Define the parameter vector $\theta = \{\rho, \beta, \gamma, \sigma^2\}$. Note that even when $\hat{U}$ is known, the parameters cannot be estimated through an usual regression model since we have Y on both the left and the right hand side of the equation. Further the term LY is clearly correlated with the error term ϵ. However, we can obtain the parameters through a maximum likelihood estimation. The log-likelihood function given $\hat{U}$ is given by [57, 42]

$$l(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}[((I-\rho L)Y - Z\gamma - \hat{U}\beta)^T((I-\rho L)Y - Z\gamma - \hat{U}\beta)] + \log|I-\rho L|$$

However, similar to the method considered in [51], simply replacing U with estimated $\hat{U}$ will lead to bias in the estimates. For example, in the context of linear regression, when regressing Y on $\hat{U}$, the usual least squares estimator for the vector of parameters assuming $\hat{U}$ is a fixed design matrix is $\hat{\beta} = (\hat{U}^T\hat{U})^{-1}\hat{U}^TY$. However, due to presence of estimation error in $\hat{U}$, this estimator is biased in finite sample and is biased asymptotically unless the estimation error vanishes [31, 5]. We construct a bias corrected estimator where the central idea is to correct for the finite sample bias using the corrected score function methodology from the well developed theory of measurement error models [70, 72, 67, 53, 55].

In the next section we will develop bias corrected estimation methods given the *knowledge* of the covariance matrix of $\hat{U}$ conditional on U. However, in practice it is difficult to find such an estimate of the covariance matrix. We propose to use the result from [3, 76] on estimates of covariance matrix for a *finite* number of nodes. We will further focus our attention only to the context of SBM, where the matrix U has only k unique rows. Let C_q denote the unique row corresponding to the qth community. Then to apply corollary 2.3 of [76], we compute

$$\Delta_F = E[U_1U_1^T] = \sum_k \pi_k C_k C_k^T, \text{ and } \Sigma(C_q) = \Delta_F^{-1}(\sum_k \pi_k C_k C_k^T (C_q^T C_k - (C_q^T C_k)^2))\Delta_F^{-1}.$$

Therefore, for a fixed node i, $(\sqrt{n}(O_n\hat{U}_i - C_q)|C_i = q) \sim N(0, \Sigma(C_q))$, for some orthogonal matrix O_n. We propose to estimate C_q with its natural estimate $\hat{C}_q$ which are the cluster centers, and π_q with $\hat{\pi}_q$ which are the cluster size proportions. Therefore a

plug in estimator for the covariance matrix $\hat{\Delta}_q$ for any $\hat{U}_i$ whose true community is q, is

$$\hat{\Delta}_q = \frac{1}{n}\left(\sum_k \hat{\pi}_k \hat{U}_k \hat{U}_k^T\right)^{-1}\left(\sum_k \hat{\pi}_q(\hat{U}_q^T U_k - (\hat{U}_q^T \hat{U}_k)^2)\hat{U}_k \hat{U}_k^T\right)\left(\sum_k \hat{\pi}_k \hat{U}_k \hat{U}_k^T\right)^{-1}, \quad (2.5)$$

Then the sum of the covariance matrices for the latent vectors can be obtained as the weighted sum $n\sum_q \pi_q \Delta_q$. We note that the results in [76, 3] hold nodewise and therefore does not hold simultaneously for all n nodes. However, we show in the simulations that our measurement bias correction methods with this estimate of the covariance matrix provides effective bias corrections, especially in small samples.

Finally we remark that in our model while $U\beta$ is identifiable, the parameter β cannot be identified separately. This is because the latent variable U can be estimated only up to the ambiguity of an orthogonal matrix O. Therefore we will only be interested in identifying ρ, γ, σ^2 parameters correctly.

3 SAR model with measurement error

In this section we develop a bias corrected maximum likelihood estimator for the SAR model when a set of covariates are measured with error. Mimicking the SAR model in the previous section, we assume that the model has two sets of covariates, where one set of covariates U_i are measured with an additive measurement error and we can write

$$\hat{U}_i = U_i + \xi_i, \quad cov(\xi_i) = \Delta_i.$$

The second set of covariate vectors Z_i are assumed to be observed without any error. We derive a maximum likelihood estimator for the SAR model similar to [57, 42], but with a modified score function to correct for the bias introduced by the error in the estimation of U. Using the notation of [53], define E^+ as the expectation with respect to V conditional on $\hat{U}$ and E^* as the expectation with respect to ξ. Let $E = E^+E^*$ denote the unconditional expectation. We wish to find a corrected likelihood function $l^*(\theta, \hat{U}, Y, Z)$ such that $E^*[l^*(\theta, \hat{U}, Y, Z)] = l(\theta, U, Y, Z)$ [53]. It will be convenient to define a few notations.

- Let d_1 and d_2 be the dimensions of U_i and Z_i respectively. Define $\tilde{X} = [\hat{U} \quad Z]$ and $X = [U \quad Z]$. Further define $\eta = [\xi \quad 0_{d_2}]^T$, where 0_{d_2} is the d_2 dimensional 0 vector. Finally, let $\delta = [\beta \quad \gamma]^T$. Then the data generating model can be compactly written as

$$Y = \rho LY + X\delta + V, \quad V_i|X_i \overset{i.i.d}{\sim} N(0, \sigma^2)$$

$$\tilde{X}_i = X_i + \eta_i, \quad cov(\eta_i) = \Omega_i = \begin{pmatrix} \Delta_i & 0 \\ 0 & 0 \end{pmatrix}$$

The fitted SAR model replaces X with $\tilde{X}$ in the equation for Y.

- The parameter vector is $\theta = [\beta, \gamma, \rho, \sigma^2]$.
- $S(\rho) = (I - \rho L), \quad G(\rho) = LS(\rho)^{-1}, \quad \tilde{H}(\rho) = G(\rho)(\hat{U}\beta + Z\gamma) = G(\rho)\tilde{X}\delta$
- $\tilde{V}(\rho) = (I - \rho L)Y - Z\gamma - \hat{U}\beta = S(\rho)Y - \tilde{X}\delta$

Now the conditional expectation of the log likelihood function given U is (i.e., the log likelihood function we would have if we had U available),

$$E^*[l(\theta, \hat{U}, Y, Z)] = l(\theta, U, Y, Z) - \frac{1}{2\sigma^2}\beta^T(\sum_i \Delta_i)\beta.$$

Then the corrected log-likelihood function is given by

$$l^*(\theta, \hat{U}, Y, Z) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}[\sum_i \tilde{V}(\rho)_i\tilde{V}(\rho)_i^T - \beta^T(\sum_i \Delta_i)\beta] + \log|I - \rho L|$$

We define the solutions to the corrected score equations $\nabla_\theta l^*(\theta, \hat{U}, Y, Z) = 0$ as the bias corrected maximum likelihood estimators. The following quantities will be useful for the derivation:

$$M = I - \begin{pmatrix} \hat{U} & Z \end{pmatrix} \begin{pmatrix} \hat{U}^T\hat{U} - \sum_i \Delta_i & \hat{U}^T Z \\ Z^T\hat{U} & Z^T Z \end{pmatrix}^{-1} \begin{pmatrix} \hat{U} \\ Z \end{pmatrix} = I - \tilde{X}(\tilde{X}^T\tilde{X} - \sum_i \Omega_i)^{-1}\tilde{X}^T.$$

$$K = \tilde{X}(\tilde{X}^T\tilde{X} - (\sum_i \Omega_i))^{-1}(\sum_i \Omega_i)(\tilde{X}^T\tilde{X} - (\sum_i \Omega_i))^{-1}\tilde{X}^T$$

The corrected score function with respect to β, γ is

$$\nabla_\beta l^*(\theta) = -\frac{1}{\sigma^2}[-\hat{U}^T((I - \theta L)Y - Z\gamma) + \hat{U}^T\hat{U}\beta - \sum_i \Delta_i\beta],$$

$$\nabla_\gamma l^*(\theta) = -\frac{1}{\sigma^2}[-Z^T((I - \theta L)Y - \hat{U}\beta) + Z^T Z\gamma]$$

Equating this function to 0, yields the solution

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} \hat{U}^T\hat{U} - \sum_i \Delta_i & \hat{U}^T Z \\ Z^T\hat{U} & Z^T Z \end{pmatrix}^{-1} \begin{pmatrix} \hat{U} \\ Z \end{pmatrix} (I - \rho L)Y = (\tilde{X}^T\tilde{X} - \sum_i \Omega_i)^{-1}\tilde{X}^T S(\rho)Y.$$

Using this solution for β, γ, we have the estimate for σ^2 as

$$\hat{\sigma}^2 = \frac{1}{n}\{((I-\rho L)Y - \hat{U}\hat{\beta} - Z\hat{\gamma})^T((I-\rho L)Y - \hat{U}\hat{\beta} - Z\hat{\gamma}) - \hat{\beta}^T(\sum_i \Delta)\hat{\beta}\}$$

The first part can be simplified by noticing,

$$(I-\rho L)Y - \hat{U}\hat{\beta} - Z\gamma = S(\rho)Y - \tilde{X}\hat{\delta} = MS(\rho)Y.$$

Therefore, the first term becomes

$$S(\rho)^T Y^T M^T M Y S(\rho).$$

Note in this case the matrix M is not idempotent. For the second part first notice $\beta^T(\sum_i \Delta_i)\beta = \delta^T(\sum_i \Omega_i)\delta$. Then the second part is

$$\begin{aligned}\hat{\delta}^T(\sum_i \Omega_i)\hat{\delta} &= Y^T S(\rho)^T \tilde{X}(\tilde{X}^T\tilde{X} - (\sum_i \Omega_i))^{-1}(\sum_i \Omega_i)(\tilde{X}^T\tilde{X} - (\sum_i \Omega_i))^{-1}\tilde{X}^T S(\rho)Y \\ &= S(\rho)^T Y^T K Y S(\rho)\end{aligned}$$

However, with a little algebra we note (in Appendix A.1)

$$M^T M - K = I - \tilde{X}(\tilde{X}^T\tilde{X} - (\sum_i \Omega_i))^{-1}\tilde{X}^T = M.$$

Therefore, combining the two terms we have

$$\hat{\sigma}^2 = \frac{1}{n}\{Y^T(I-\rho L)^T M(I-\rho L)Y\} = \frac{1}{n}\{Y^T S(\rho)^T M S(\rho)Y\}.$$

The parameter ρ can be estimated by minimizing the negative of the corrected concentrated log likelihood function obtained by replacing β and σ^2 by their estimates as follows:

$$l^*(\rho) = \frac{n}{2}(-\frac{2}{n}\log|1-\rho L| + \log(\hat{\sigma}^2)).$$

If A is undirected, the minimization can be performed by writing the determinant as product of the real eigenvalues and then using a Newton-Raphson algorithm similar to [57]. The first and the second derivative of $l^*(\rho)$ with respect to ρ is given in the Appendix. For non-symmetric A matrix we directly perform optimization. The entire algorithm including homophily estimation using spectral embedding methods and the measurement error bias correction is summarized in Algorithm 1.

Algorithm 1: Bias corrected network influence estimation

Input: Network Adjacency matrix A, Response Y, Observed Covariates Z, dimension of latent factors d, number of communities k

Result: Model parameters $(\hat{\beta}, \hat{\gamma}, \hat{\sigma}^2, \hat{\rho})$

1: SVD: $A = V\Lambda V^T, \quad \hat{U} = V[1:d](\Lambda[1:d])^{1/2}$

2: $[\hat{C}, centers, \pi] \longleftarrow Kmeans(\hat{U}, k)$

3: Estimate Δ_k using Equation 2.5.

4: $\tilde{X} = \begin{pmatrix} \hat{U} & Z \end{pmatrix}, \quad \Omega_i = \begin{pmatrix} \Delta_i & 0 \\ 0 & 0 \end{pmatrix}$

5: $L = D^{-1}A$ where $D_{ii} = \sum_j A_{ij}$ and $D_{ij} = 0$ for $j \neq i$

6: $M = I - \tilde{X}(\tilde{X}^T\tilde{X} - \sum_i \Omega_i)^{-1}\tilde{X}^T$

7: $\sigma^2(\rho) = \{Y^T S(\rho)^T M S(\rho) Y\}/n.$

8: $\hat{\rho}$ =minimize $f(\rho) = \frac{n}{2}(-\frac{2}{n}\log|1 - \rho L|) + \log(\sigma^2))$.

9: $\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = (\tilde{X}^T\tilde{X} - \sum_i \Omega_i)^{-1}\tilde{X}^T(I - \hat{\rho}L)Y$

10: $\hat{\sigma}^2 = \sigma^2(\hat{\rho})$

11: **return** $[\hat{\beta}, \hat{\gamma}, \hat{\sigma}^2, \hat{\rho}]$

3.1 Asymptotic theory and inference on parameters

We study the asymptotic properties of the bias corrected estimators and derive asymptotic standard errors. We augment the approach in [42] in the context of measurement error in variables to find the consistency and limiting distribution of the bias corrected MLE. For simplicity we make an explicit assumption of normality of the error term in the SAR model. Our results can be extended to non normal error distribution similar to the QMLE approach of [42]. However, despite the assumption of normality, our estimator is still not an MLE due to the presence of measurement error. Therefore we apply the generic theory of M estimators. Let the vector $\theta_0 = [\beta_0, \gamma_0, \rho_0, \sigma_0^2]^T$ be the true parameter vector. We consider an asymptotic setting where $n \to \infty$. Therefore in the following we will add a subscript of n to all quantities to denote sequences of quantities. The bias corrected log likelihood function is $l_n^*(\theta)$. The gradient vector $\nabla_\theta l_n^*(\theta)$ and Hessian matrix $\nabla_\theta^2 l_n^*(\theta)$ of the log likelihood function is given in Appendix A.3.

For a matrix S, we define the notations $\|S\|_\infty = \max_i \sum_j |S_{ij}|$ as the maximum row (absolute) sum norm, $\|S\|_1 = \max_j \sum_i |S_{ij}|$, as the maximum (absolute) column sum norm, and $\|S\|_2$, as the spectral norm of the matrix. We further define $\text{tr}(S)$ as the trace of the matrix. For a matrix S, the term "uniformly bounded" in row sum norm (respectively column sum norm) is used to mean $\|S\|_\infty < c$, (respectively $\|S\|_1 < c$) where c is not dependent on n.

Assumptions: We need most of the assumptions in [42] and a few additional as-

sumptions related to the measurement error.

1. Error terms $(V_n)_i \overset{iid}{\sim} N(0, \sigma^2), i = 1, \ldots, n$, with $\sigma^2 < \infty$, and $E[(V_n)_i^{4+\eta}] < \infty$.

2. The matrix $S_n = S_n(\rho_0)$ is non-singular

3. The elements of the network adjacency matrix A_n are non-negative and uniformly bounded. The degrees $(d_n)_i = O(h_n)$ uniformly for all i and consequently, the elements of the matrix L_n are uniformly bounded by $O(1/h_n)$ for some sequence $h_n \to \infty$. Further $h_n = o(n)$, i.e., the degrees grow with n but at a rate slower than n.

4. Sequence of matrices L_n and S_n^{-1} are uniformly bounded in both row and column sum norms.

5. $S(\rho)^{-1}$ is bounded in row sums uniformly for all $\rho \in R$, where the parameter space R is compact. The true parameter ρ_0 is in the interior of R.

6. For each i, $(\hat{U}_n)_i = (U_n)_i + (\xi_n)_i$, with $E[(\xi_n)_i] = 0$, and $cov((\xi_n)_i) = (\Delta_n)_i < \infty$ for all n and the $(\Delta_n)_i$ are assumed to be known. Further $E[(\xi_n)_i^{4+\eta}] < \infty$. The vectors $(U_n)_i$ and $(Z_n)_i$ are assumed to be non-random. The error terms $(V_n)_i$ and $(\xi_n)_i$ are independent of each other.

7. Let $X_n = [U_n \quad Z_n]$. We assume X_n has full column rank, i.e., $X_n^T X_n$ is non singular. The elements of X_n are bounded constants as a function of n, consequently, $\frac{1}{n} X_n^T X_n = O(1)$.

8. Matrix $\tilde{X}_n(\tilde{X}_n^T \tilde{X}_n - \Omega)^{-1} \tilde{X}_n^T$ are uniformly bounded in row and column sum norms.

Theorem 1. *Under Assumptions A1-A8, the bias corrected maximum likelihood estimator $\hat{\theta}_n$ are consistent estimators of the true population parameters, i.e., $\hat{\theta}_n \xrightarrow{p} \theta_0$.*

Short proof outline: The complete proof of this theorem is given in Appendix B.1. Here we describe a brief outline of the main arguments.

Recall the corrected loglikelihood function maximized to obtain the estimators is

$$l_n^*(\theta) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2}[\tilde{V}_n(\rho)^T \tilde{V}_n(\rho) - \delta^T (\sum_i \Omega_i)\delta] + \log|S_n(\rho)|.$$

Now define

$$Q_n(\rho) = \max_{\beta, \sigma^2} E[l_n^*(\theta)] = \max_{\beta, \sigma^2} E^+ E^*[l_n^*(\theta)],$$

as the concentrated unconditional expectation of the corrected log likelihood function. Solving the optimization problem leads to solutions for δ, σ^2 for a given ρ

$$\begin{aligned}\delta(\rho) &= (X_n^T X_n)^{-1} X_n^T S_n(\rho) E[Y_n] = (X_n^T X_n)^{-1} X_n^T S_n(\rho) S_n^{-1} X_n \delta_0 \\ \sigma^2(\rho) &= \frac{1}{n}\left((\rho_0 - \rho)^2 (G_n X_n \delta_0)^T M_{1n} (G_n X_n \delta_0)\right. \\ &\quad \left. + \frac{\sigma_0^2}{n} tr\left((S_n^T)^{-1} S_n(\rho)^T S_n(\rho) S_n^{-1}\right),\right. \end{aligned} \tag{3.1}$$

where $M_{1n} = (I_n - X_n(X_n^T X_n)^{-1} X_n^T)$, and $G_n = L_n S_n^{-1}$. The arguments in [42] proves uniqueness of ρ_0 as the global maximizer of $Q_n(\rho)$ in the compact parameter space R. However we need to show the uniform convergence of the concentrated corrected log likelihood $l_n^*(\rho)$ to $Q_n(\rho)$ over the compact parameter space R, which requires further analysis. Now we have

$$\frac{1}{n}(l_n^*(\rho) - Q_n(\rho)) = -\frac{1}{2}(\log(\hat{\sigma}_n^2) - \log(\sigma^2(\rho))),$$

where $\sigma^2(\rho)$ is as defined in Equation 3.1 and

$$\hat{\sigma}_n^2(\rho) = \frac{1}{n} Y_n^T S_n^T(\rho) M_{2n} S_n(\rho) Y_n, \tag{3.2}$$

with $M_{2n} = I_n - \tilde{X}_n(\tilde{X}_n^T \tilde{X}_n - \Omega)^{-1} \tilde{X}_n^T$. We can expand $\hat{\sigma}_n^2(\rho)$ to obtain

$$\begin{aligned}\hat{\sigma}_n^2(\rho) &= \frac{1}{n}(S_n^{-1} X_n \delta_0 + S_n^{-1} V_n)^T S_n^T(\rho) M_{2n} S_n(\rho)(S_n^{-1} X_n \delta_0 + S_n^{-1} V_n) \\ &= (X_n \delta_0 + (\rho_0 - \rho) G_n X_n \delta_0 + S_n(\rho) S_n^{-1} V_n)^T M_{2n} (X_n \delta_0 + (\rho_0 - \rho) G_n X_n \delta_0 + S_n(\rho) S_n^{-1} V_n) \\ &= \frac{1}{n}\Bigg((\rho_0 - \rho)^2 (G_n X_n \delta_0)' M_{2n} (G_n X_n \delta_0) + (X_n \delta_0)^T M_{2n} (X_n \delta_0) \\ &\quad + V_n^T (S_n^T)^{-1} S_n^T(\rho) M_{2n} S_n(\rho) S_n^{-1} V_n + 2(\rho_0 - \rho)(G_n X_n \delta_0)^T M_{2n} S_n(\rho) S_n^{-1} V_n \\ &\quad + 2(\rho_0 - \rho)(X_n \delta_0)^T M_{2n} (G_n X_n \delta_0) + 2(X_n \delta_0)^T M_{2n} S_n(\rho) S_n^{-1} V_n \Bigg)\end{aligned}$$

The decomposition of $\hat{\sigma}^2(\rho)$ produces 6 terms. We show that under the assumptions of A1-A8, the first term converges to the first term in Equation 3.1 uniformly in ρ, while the terms 2, 4, 5, 6 uniformly converges to 0. The 3rd term is a quadratic form in the error vector V_n with a matrix $B_n(\rho) = (S_n^T)^{-1} S_n^T(\rho) M_{2n} S_n(\rho) S_n^{-1}$ which itself is stochastic. In contrast the commonly encountered quadratic forms are of the form $y^T A y$, with A being a non random matrix. We show this term uniformly converges to the second term in Equation 3.1 for all $\rho \in R$. Once the convergence of $\hat{\rho}_n$ to ρ_0 has been established, the

convergence of $\hat{\tilde{\delta}}_n$ can be easily proved:

$$\hat{\tilde{\delta}}_n = (\tilde{X}_n^T \tilde{X}_n - \sum_i \Omega_i)^{-1} \tilde{X}_n^T (I + (\rho_0 - \rho) G_n)(X_n \delta_0 + V_n) = \delta_0 + o_p(1).$$

The next results establishes that the bias corrected estimator is asymptotically normal. The result also gives us the asymptotic variance-covariance matrix of the bias corrected estimator which can be used to compute the standard errors of the estimates.

Theorem 2. *If assumptions A1-A8 hold and $\hat{\theta}_n$ is a consistent estimator of θ_0, then*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, I(\theta_0, U, Z)^{-1} \Sigma(\theta_0, U, Z) I(\theta_0, U, Z)^{-1}),$$

where $I(\theta_0, U, Z)$ is the unconditional Fisher information matrix given by

$$I(\theta_0, U, Z) = E[\frac{1}{n} \nabla_\theta^2 l^*(\theta_0, \hat{U}, Z)]$$

and $\Sigma(\theta_0, U, Z)$ is the unconditional variance of $\frac{1}{n} \nabla_\theta l_n^*(\theta_0, \hat{U}, Z)$ evaluated at θ_0:

$$\Sigma(\theta_0, U, Z) = Var[\frac{1}{n} \nabla_\theta l_n^*(\theta_0, \hat{U}, Z].$$

We estimate the matrix $I(\theta_0, U, Z)$ with the corrected Fisher information matrix

$$I^*(\theta_0, \hat{U}, Z) = E^+[\frac{1}{n} \nabla_\theta^2 l^*(\theta_0, \hat{U}, Z)],$$

which can be computed as:

$$\frac{1}{n} \begin{pmatrix} \frac{1}{\sigma_0^2}(\hat{U}_n^T \hat{U}_n - \sum_i \Delta_i) & \frac{1}{\sigma_0^2} \hat{U}_n^T Z_n & \frac{1}{\sigma_0^2} \hat{U}_n^T H_n & \frac{1}{\sigma_0^4} \sum_i \Delta_i \beta_0 \\ \frac{1}{\sigma_0^2} Z_n^T \hat{U}_n & \frac{1}{\sigma_0^2} Z_n^T Z_n & \frac{1}{\sigma_0^2} Z_n^T H_n & 0 \\ \frac{1}{\sigma_0^2} H_n^T \hat{U}_n & \frac{1}{\sigma_0^2} H_n^T Z_n & \frac{1}{\sigma_0^2} H_n^T H_n + tr(G_n G_n) & \frac{1}{\sigma_0^2} tr(G_n) \\ \frac{1}{\sigma_0^4} \beta_0^T \sum_i \Delta_i & 0 & \frac{1}{\sigma_0^2} tr(G_n) & \frac{n}{2\sigma_0^4} - \frac{1}{\sigma_0^6} \beta_0^T \sum_i \Delta_i \beta_0, \end{pmatrix}$$

where as before $S_n = (I - \rho_0 L)$, $G_n = L_n S_n^{-1}$, and $H_n = G_n(\hat{U}_n \beta_0 + Z_n \gamma_0)$

Since the score vector is composed of linear and quadratic forms, the matrix $\Sigma(\theta_0, U, Z)$, which is the unconditional variance of the score vector at θ_0 is estimated using the following estimator

$$\frac{1}{n} \sum_i (\nabla_\theta l_i^*(\theta_0, \hat{U}_i, Z_i))(\nabla_\theta l_i^*(\theta_0, \hat{U}_i, Z_i)^T,$$

where $\nabla_\theta l_i^*(\theta_0, \hat{U}_i, Z_i)$ is an estimate of the score vector for observation i evaluated at θ_0 such that $\nabla_\theta l_n^*(\theta_0, \hat{U}_i, Z_i) = \sum_i \nabla_\theta l_i^*(\theta_0, \hat{U}_i, Z_i)$, described below:

$$\begin{pmatrix} \frac{1}{\sigma_0^2}[\hat{U}_i\tilde{V}_i(\rho_0)^T - \hat{U}_i\hat{U}_i^T\beta_0 + (\Delta_i)\beta_0] \\ \frac{1}{\sigma_0^2}[Z_i\tilde{V}_i(\rho_0)^T - Z_iZ_i^T\gamma_0] \\ \frac{1}{\sigma_0^2}(LY)_i\tilde{V}_i(\rho_0)^T - (G_n)_{ii}, \\ -\frac{1}{2\sigma_0^2} + \frac{1}{2\sigma_0^4}\{\tilde{V}_i(\rho_0)\tilde{V}_i(\rho_0)^T - \beta_0^T(\Delta_i)\beta_0\}, \end{pmatrix}$$

For both of these matrices we replace θ_0 with the estimator $\hat{\theta}_n$ to obtain their plug-in estimators.

Short Proof Outline: The complete proof of this theorem can be found in the Appendix B.2. Here we provide a short outline of the main arguments. Several of these arguments are are similar to [42] (See also the lecture notes by [66]). From Taylor expansion with intermediate value theorem we have

$$\nabla_\theta l_n^*(\hat{\theta}_n, \tilde{X}_n) = \nabla_\theta l_n^*(\theta_0, \tilde{X}_n) + \nabla_\theta^2 l_n^*(\tilde{\theta}_n, \tilde{X}_n)(\hat{\theta}_n - \theta_0),$$

for some $\tilde{\theta}_n$ which is intermediate value between $\hat{\theta}_n$ and θ_0 (therefore $\tilde{\theta}_n \xrightarrow{p} \theta_0$). Since $\nabla_\theta l^*(\hat{\theta}_n, \tilde{X}_n) = 0$, by definition of $\hat{\theta}_n$, this implies

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = (-\frac{1}{n}\nabla_\theta^2 l_n^*(\tilde{\theta}, \tilde{X}_n))^{-1}\frac{1}{\sqrt{n}}\nabla_\theta l_n^*(\theta_0, \tilde{X}_n)$$

We show the following two convergence in probability results by analyzing the convergence of each element of the corresponding matrices:

$$\frac{1}{n}\nabla_\theta^2 l_n^*(\tilde{\theta}_n, \tilde{X}_n) \xrightarrow{p} \frac{1}{n}\nabla_\theta^2 l_n^*(\theta_0, \tilde{X}_n) \qquad \text{and} \qquad \frac{1}{n}\nabla_\theta^2 l^*(\theta_0, \tilde{X}_n) \xrightarrow{p} I(\theta_0, X_n).$$

Then using the central limit theorem for linear and quadratic forms in [36], we obtain,

$$\sqrt{n}(\frac{1}{n}\nabla_\theta l_n^*(\theta_0, \tilde{X}_n) \xrightarrow{D} N(0, \Sigma(\theta_0, X_n)).$$

Combining these two results leads to the claimed theorem.

4 Simulation Studies

To verify the finite sample performance of the proposed methods in terms of providing unbiased estimates of the network influence parameters, we design a simulation study. We generate the networks from a Stochastic Block Model (SBM) with increasing number of

nodes $n = \{50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600\}$, fixed dimensions $d = 2$, and number of communities $k = 4$. Accordingly the matrix of probabilities P is generated as $P = UU^T$, where the matrix $U_{n\times 2}$ is generated such that it has only 4 unique rows. The resulting block matrix of probabilities is

$$\begin{pmatrix} 0.53 & 0.19 & 0.18 & 0.45 \\ 0.19 & 0.37 & 0.14 & 0.35 \\ 0.18 & 0.14 & 0.08 & 0.20 \\ 0.45 & 0.35 & 0.20 & 0.50 \end{pmatrix}$$

The network edges A_{ij} are generated independently from Bernoulli distribution with parameters P_{ij}. The covariate matrix Z is correlated with the matrix U and response Y is generated from a multivariate normal distribution as follows:

$$Y \sim N\left((I - \rho_0 L)^{-1}(U\beta_0 + Z\gamma_0), \sigma_0^2 (I - \rho_0 L)^{-1}(I - \rho_0 L)^{-1}\right).$$

We set $\rho_0 = 0.4$, $\beta_0 = (1, 2)$, $\sigma_0 = 0.8$ and $\gamma_0 = (0.2, -0.3)$. We compare the three methods (a) no latent factor (b) latent factors estimated from SBM but without any bias correction and (c) estimated latent factor with bias correction, in terms of bias of estimating the network influence parameter ρ and the parameters of the covariates γ_1, γ_2. Figure 2 shows a comparison of the estimates of the network influence parameter from the 3 methods. The estimator without homophily control is substantially biased even when the sample size increases. This shows the benefit of homophily correction in SAR model. The figure further shows that the proposed measurement error bias corrected estimator for the network effect parameter ρ has less bias compared to the estimator with homophily control but no bias correction, especially in small samples. The bias in the parameter estimate goes close to 0 more quickly with the measurement error correction. Therefore the proposed bias correction methodology works well.

Both the homophily and measurement error bias correction also have substantial effect on the accuracy of the parameter estimates for the covariates as depicted in Figures 3. Overall the simulation study shows that the methods of homophily correction and bias correction using the finite node asymptotic covariance estimate from the network model performs well in practice.

Finally we verify the accuracy of our estimate of standard error from Theorem 2. For this purpose, in Figure 4(A) we display box plots of the estimated standard errors over the 200 repetitions with increasing number of nodes. We compare the boxplots with observed standard deviation of the estimates of ρ over the 200 repetitions (red dots). We

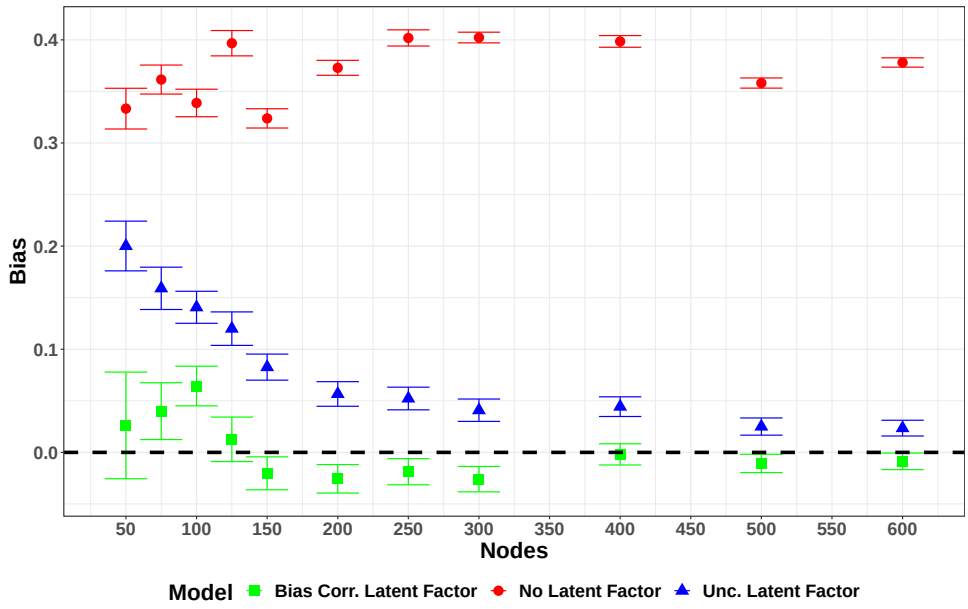


Figure 2: Comparison of estimates of the network influence parameter ρ from the naive model (model with no latent factors), the model with latent factors but no bias correction, and the model with latent factors and bias correction. The points represent mean bias and the error bars represent 1.96*SE for mean bias.

notice that the estimated standard errors closely match the observed sampling standard deviations for different values of n validating the result in Theorem 2.

4.1 Out of sample performance and prediction

As noted in [30], there are several ways to define out of sample predictions in an SAR model. We define the leave one out prediction for a network connected individual as the "trend-signal-noise predictor". An additional complexity in our context on top of what the SAR model presents is that we also have latent homophily variables in the model that needs to be estimated for new test observations. In this aspect, our model is a mixed effects model. However, the latent variables for new observations can be estimated given the observation of the whole network. A yet another difficulty is that due to ambiguity of an unitary diagonal matrix in estimating U, the parameter estimate of β is not identified.

We put forth two related but slightly different strategies for predictions, one for obtaining out of sample predictive evaluation of model, and another for predicting the response of a new observation. For model fit validation, we extract the latent variables U from the whole network once and treat it similarly as Z. Let the set o denote one test observation and the set S denotes remaining observations. Therefore the training data consists of $Y_S, Z_S, \hat{U}_S, L_{SS}$ and the test data consists of $Y_o, \hat{U}_o, Z_o, L_{oS}$. We use the training data to

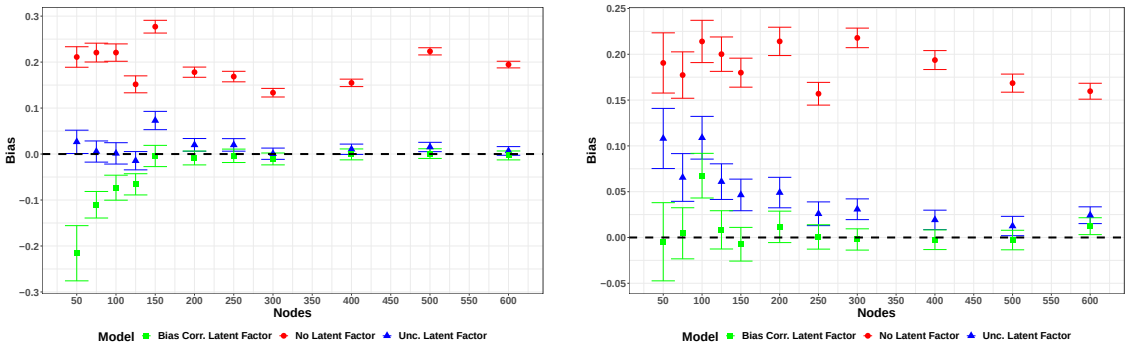


Figure 3: Comparison of estimates of the parameters of the covariates γ_1, γ_2 from the three models. The points represent mean bias and the error bars represent 1.96*SE for mean bias.

estimate the parameters of the model, namely, $\hat{\rho}, \hat{\beta}, \hat{\gamma}$. The leave one out trend signal noise prediction is given by

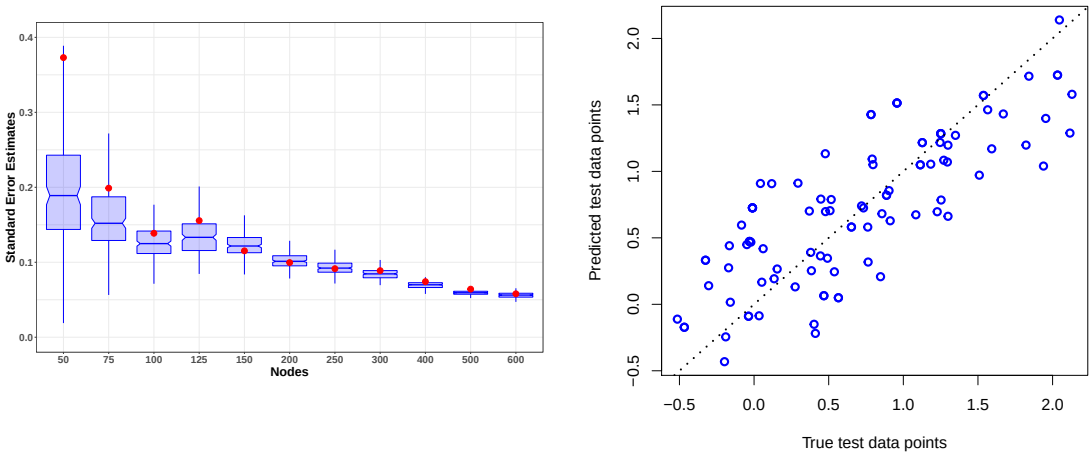
$$\hat{Y}_o = Z_o\hat{\gamma} + \hat{U}_o\hat{\beta} + \hat{\rho}L_{oS}Y_S. \tag{4.1}$$

The predictive R-squared statistic is then computed from these predictions. We illustrate the predictive performance of the bias corrected estimator in a small simulation in Figure 4. The simulation setup is the same as before while we fix the number of nodes as 200. Figure 4 shows that the predicted responses of the out of sample data points are close to their true responses.

For predicting the response of a new node added to the network, we need a slightly different strategy. We assume that we observe the values of the individual characterstics or covariates of this node as well as the network links of this new node with the existing nodes. Therefore we first augment the network including this new node. Then we estimate the latent homophily from this augmented network. We then re-estimate the model parameters with the training dataset $Y_S, Z_S, \hat{U}_S, L_{SS}$ where $\hat{U}_S$ is now the subset of the new latent homophily estimate restricted to the units in the training data. We use the new parameter estimates to predict the response for this new data point using the same equation as Equation 4.1.

4.2 Causal network influence in longitudinal settings

We consider the longitudinal response settings where we have observations on the response variable over t time points along with a network that does not vary over time and the goal is to estimate the network influence or contagion as in [16, 68, 51, 79]. As before define the diagonal matrix of degrees D such that $D_i = \sum_j A_{ij}$, and then the



(a)Accuracy of Standard Error Estimates

(b) Predicted and actual test data

Figure 4: (a) Accuracy of Standard error estimates: Box and Whisker plots (blue) of estimated standard errors with the standard deviation of parameter estimates (red dots) for different sample sizes (b) Out of sample predictions: Predicted and actual test data points

Laplacian matrix as $L = D^{-1}A$. Therefore we have the following data generating model:

$$Y_{it} = U_i\beta + Y_{i,(t-1)} + \rho \sum_j L_{ij} Y_{j(t-1)} + Z_i\gamma + V, \quad V_{it} \overset{i.i.d}{\sim} N(0, \sigma^2). \tag{4.2}$$

with Z being a covariate that is independent of the network. This formulation in [51] avoids the simultaneity problem [48] by regressing Y_i on time lagged outcomes from Y_j. The covariates Z in the above equation are observed and therefore can be controlled for. Note there might be other unobserved environmental covariates who are network irrelavant but affect all outcomes. We can either make an assumption similar to [51], which leads to the conclusion that these covariates only affect the variance and not the bias, or we can try to control for them using another time lagged variable [79]. Our goal is to identify the parameter θ. Since we have used the time lag of the response in the right hand side of the equation, the model can be estimated using least squares for linear response and logistic regression for binary response. It was argued in [51] that when the latent factors in Equation 4.2 are latent communities from the SBM, then replacing estimated communities in place of the true communities, one can obtain an asymptotically unbiased estimate of θ. The authors further proved an upper bound on the rate of decrease of the bias as n increases. The proposed method from the previous section allows us to obtain a

bias corrected estimator of θ, which we show performs better than the estimator in [51].

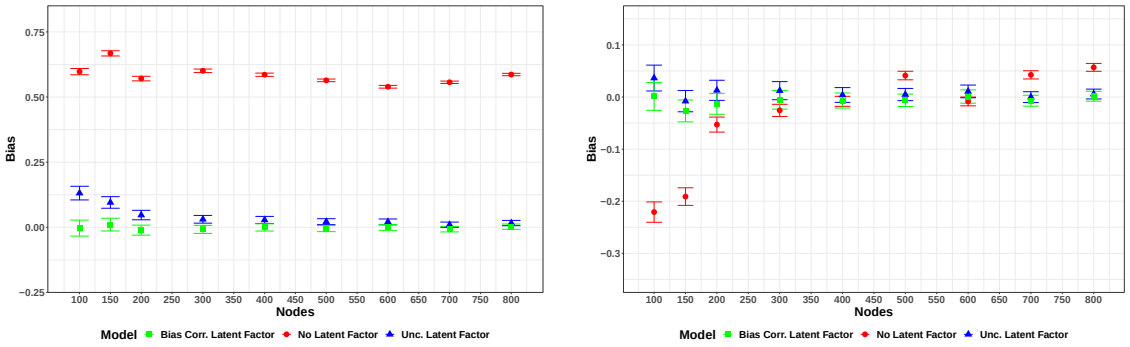
Define $X = [Y_{t-1} \quad LY_{t-1} \quad Z]$ as the matrix collecting all the predictor variables except for $\hat{U}$. Let $\theta = [\beta, \gamma, \rho]$. Define the following quantities.

$$\Omega = \begin{pmatrix} \sum_i \hat{\Delta}_i & 0 \\ 0 & 0 \end{pmatrix}, \quad M_{XU} = [\hat{U} \quad X]^T[\hat{U} \; X], \quad M_Y = [\hat{U} \quad X]^T Y.$$

Then the bias corrected estimator is [53]:

$$\hat{\theta} = (M_{XU} - \Omega)^{-1} M_Y, \quad \hat{\sigma}^2 = \frac{1}{n}(\|Y - [\hat{U} \; X]\hat{\theta})\|_2^2 - \hat{\theta}^T \Omega \hat{\theta}), \quad var(\hat{\theta})_j \; = \hat{\sigma}^2 (M_{XU} - \Omega)^{-1}_{jj}$$

We perform a simulation study to check if the bias corrected estimator using the estimated covariance matrix from the spectral embedding of SBM performs better than the uncorrected estimator. Since OLS is used to estimate the model, we simplify the simulation setup by generating U and Z correlated with each other and generating Y as $Y = U\beta + Z\gamma + \epsilon$. We set the dimensions of U_i and Z_i as 3 and 2 respectively. As before the network A is generated from a SBM with $k = 3$ communities with U being the latent homophily variable (therefore U has 3 unique rows). Our goal is to correctly identify the parameters γ_1, γ_2. Figure 5 shows the performance of the estimators with increasing number of nodes. As noted in [51], the estimator without homophily correction remains biased even with increasing n. We notice that the bias in the bias corrected estimator is very close to 0 for most values of n. In smaller sample sizes, the estimator improves upon the non bias corrected estimator substantially.



(a) Parameter γ_1 (b) Parameter γ_2

Figure 5: Simulation Estimates in Longitudinal Data: Bias and +/- 1.96 se of bias

5 Application to therapeutic community data

5.1 Data

For our empirical analysis, we use therapeutic community (TC) data on female criminal offenders located in a Midwestern city in the United States. This facility consists of one unit for females that can accommodate a maximum of eighty female residents. Our dataset consists of 472 TC inhabitants who entered the facility between August 2005 and August 2008. The residents entered the facility at different points in time in this three year time period (see figure A2 (a)). Figure A2 (b) below shows the distribution of the time these individuals spent in the TC. On an average we find residents have spent around 142 days in this facility. The participants are kept in facility for a maximum of six months.

This therapeutic community maintained records on several critical socio-demographic characteristics, behavioral aspects, as well as the graduation and recidivism status and dates of the residents. Moreover, the officials implemented a system of mutual monitoring among the residents. This monitoring took the form of positive affirmations to appreciate good behavior among the peers ([11, 84, 82, 81]). Our data includes a little over 61,000 instances of affirmations over a three year period. For each of these instances, we observe the ID of the sender, the ID of the receiver and the date of the message. We construct a weighted network for the residents based on these affirmations.

The TC also maintained electronic records on the dates the residents graduated from the facility. About 79.7% of the residents were able to successfully graduate from this facility. Along with the status and date of graduation, we found information on recidivism for these residents. The data records the dates for four instances of recidivism. We use the first date of recidivism for our analysis. The response variables used in our analysis are graduation status and recidivism status. We define these outcome variables below.

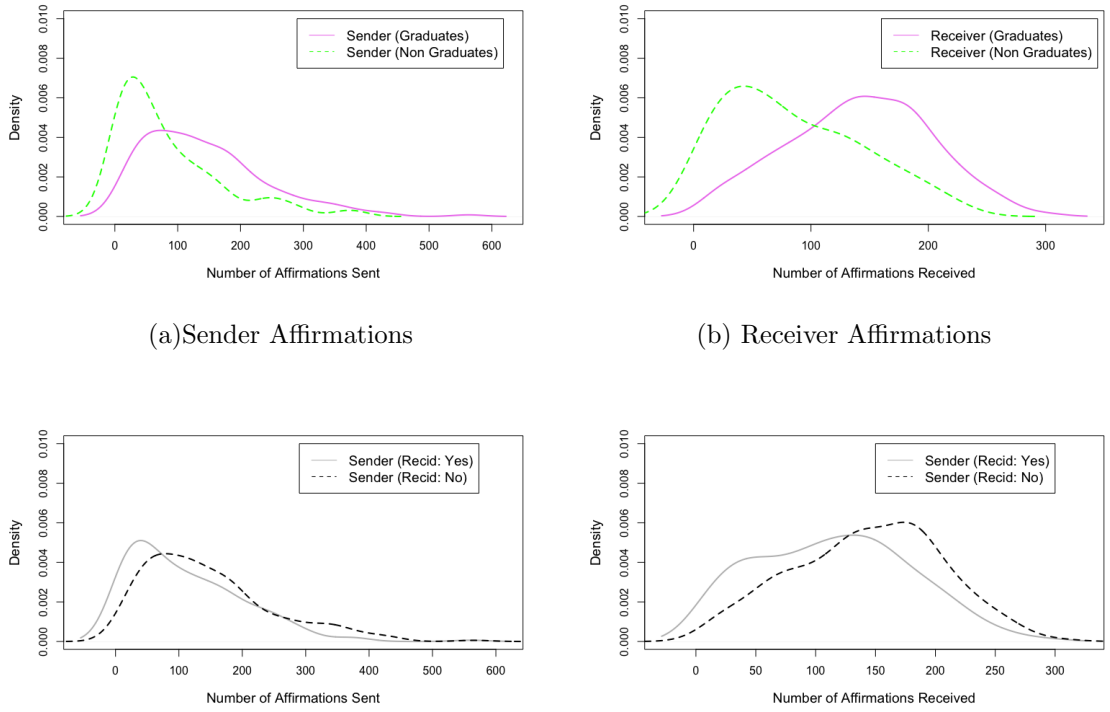
$$\text{Graduation}_i = \begin{cases} 1, & \text{if resident } i \text{ graduated from TC} \\ 0, & \text{otherwise} \end{cases}$$

$$\text{Recid Status}_i = \begin{cases} 1, & \text{if resident } i \text{ was incarcerated post exit from the TC} \\ 0, & \text{otherwise} \end{cases}$$

Understanding the role of network influence on crime has important policy implications. Criminal activities and associations impacts an individual's socio-economic outcomes. In fact, such associations and behavior can have long lasting effects. [80, 8] show that criminal activities adversely impact educational attainment and labor market outcomes. High rates of crime in a society can also result in sub optimal outcomes in terms

of economic mobility and output ([69, 64]). Related work on social networks in crime such as [73, 59, 20, 4, 27] have shown that social networks play a key role in propagating delinquent behavior. However, most of this earlier work relies on accounting for endogeneity of networks using either group fixed effects or average contextual effects to instrument for average outcomes ([27, 4]).

We on the contrary to the earlier literature on peer effects, use the network of affirmations to learn about residents' unobserved homophily. This predicted homophily is then used as a predictor in our outcome equation of interest. Our exploratory analysis shows that the affirmation network of individuals vary between those who succeeded in graduating and those who did not succeed in graduating. Figure 6 (a) and (b) show significant variation in degree distribution for the affirmations network by graduation status. We also see such variation in degree distribution, but to a lesser extent when we distinguish them by recidivism status.



(a)Sender Affirmations

(b) Receiver Affirmations

(c)Sender Affirmations

(d) Receiver Affirmations

Figure 6: Degree distribution of network by graduation status and recidivism status

Table A1 provides further details on the characteristics of these residents. In terms of socio-demographic characteristics, the sample is 80% white. The remaining 19.9% black

with one hispanic resident. The age distribution has a mean of 30 years and a standard deviation of eight years. The facility also recorded a score called Level of Service Inventory (LSI) that is based on multiple factors which is thought to be predictive of recidivism. A higher LSI is indicative of a higher propensity to get re-incarcerated.

The socio-demographic variables such as race, age and LSI are incorporated in our SAR model. In addition, we include contextual effects i.e., the weighted average of the covariate values of the network connected neighbors. This is done to understand the importance of the observed characteristics of network links on a resident's propensity to graduate and likelihood of getting re-incarcerated [49, 90, 12].

5.2 Results

We begin our analysis using a simple specification. We explore the importance of degree distribution of affirmations network on graduation and recidivism. For this we fit the following multiple logistic regression model to the data.

$$Y_i \overset{ind}{\sim} Bernoulli(p_i), \quad \text{logit}(p_i) = Z\gamma + D\eta, \tag{5.1}$$

where D is vector of the sum of the sender and receiver degrees in the affirmation network, Z is the matrix of covariates that includes a race dummy, age, LSI and an intercept term. The results are shown in Table A2. Column (1) and (2) use logistic regression on graduation status and recidivism respectively. For both the outcomes in Table A2, we observe that higher degree in affirmations is associated with pro-social behavior; it improves the propensity to graduate and reduces the propensity to recid. This suggests that affirmations network is likely critical to understand resident's graduation and recidivism behavior. We explore this channel further and provide our main results.

Next we fit the SAR model with endogenous network influence, covariates and contextual effects (weighted average of network connect neighbors) of covariates. As described in section 2, this estimation happens in two steps. First, we estimate the latent factors of the residents using a d dimensional spectral embedding of the adjacency matrix of the affirmations network. Second, we use the estimated latent variable as a covariate in our outcome equation of interest. Since we use estimated latent factors instead of true latent factors, we correct for the finite sample bias using the methodology developed in this paper.

Table 1 provides the results on graduation status using three versions of the model. Specification model (1) does not account for latent homophily. In model (2) we adjust for latent homophily but do not correct for bias. Finally, in model (3) we incorporate latent homophily as well as adjust for the bias.

Table 1: SAR with latent homophily and bias correction

	Dependent Variable: Graduation Status		
	Model 1	Model 2	Model 3
Network Influence	0.333*	0.416**	0.422***
	(0.141)	(0.143)	(0.099)
Age	-0.000	-0.000	-0.000
	(0.002)	(0.002)	(0.002)
White	0.058	0.054	0.053
	(0.039)	(0.039)	(0.043)
LSI	-0.010***	-0.010***	-0.010***
	(0.002)	(0.002)	(0.002)
Time in TC	0.007***	0.007***	0.007***
	(0.000)	(0.000)	(0.000)
Adjacency Matrix × Age	0.006	0.006	0.006
	(0.007)	(0.008)	(0.008)
Adjacency Matrix × White	0.402*	0.471*	0.476
	(0.190)	(0.193)	(0.254)
Adjacency Matrix × LSI	0.026**	0.030**	0.031**
	(0.009)	(0.010)	(0.010)
Adjacency Matrix × Time in TC	-0.009***	-0.011***	-0.011***
	(0.001)	(0.002)	(0.002)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Standard errors are provided in parenthesis. The sample consists of 472 females in TC. Estimation is done in two steps for model 2 and model 3.

We find several interesting results. First, a failure to account for latent homophily results in under-estimating the network influence on graduation status. The coefficient on network influence increases by about 30% once we account for latent homophily in columns (2) and (3). Our results indicate that a higher graduation rate among the affirmations network-connected neighbors improves a resident's chances of graduation.

Second, age and race do not substantially change the propensity to graduate from TC. However, own and network weighted average of LSI significantly impacts graduation. A higher own LSI reduces the chances of graduation. However, higher LSI among the network-connected neighbors positively affects graduation. Third, we find that own time spent in TC improves the chances of graduation. However, having more network connected neighbors who have remained for a long time in the facility reduces the propensity to graduate. These are likely measuring a willingness to reach out to weaker members of the TC. Those who affirm peers who took less time to graduate probably affirmed some who were eventually terminated without graduation. A plausible explanation for this could be that resident i holds a central position in the network, and being surrounded by weaker

peers puts more pressure to set examples of pro-social behavior.

Table 2: SAR with latent homophily and bias correction

	Dependent Variable: Recidivism Dummy		
	Model 1	Model 2	Model 3
Network Influence	0.067	-0.233	-0.241
	(0.199)	(0.221)	(0.196)
Age	-0.003	-0.002	-0.002
	(0.003)	(0.003)	(0.003)
White	-0.140*	-0.134*	-0.133*
	(0.064)	(0.063)	(0.061)
LSI	0.006*	0.005*	0.005
	(0.003)	(0.003)	(0.003)
Time in TC	-0.005***	-0.005***	-0.004***
	(0.001)	(0.001)	(0.001)
Adjacency Matrix ×Age	0.025*	0.021	0.021
	(0.011)	(0.011)	(0.011)
Adjacency Matrix ×White	0.521	0.304	0.293
	(0.299)	(0.309)	(0.310)
Adjacency Matrix ×LSI	-0.026	-0.038*	-0.038*
	(0.015)	(0.016)	(0.015)
Adjacency Matrix ×Time in TC	0.004*	0.009***	0.009***
	(0.002)	(0.002)	(0.003)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. Standard errors are provided in parenthesis. The sample consists of 472 females in the facility. Estimation is done in two steps for model 2 and model 3.

In Table 2 we estimate network influence on the propensity of the first instance of recidivism post their exit from the TC. Similar to the previous table, in model (1) we omit homophily and do not correct for bias, model (2) includes latent homophily and model (3) incorporates latent factors as well as is corrected for bias.

Unlike the graduation status, we do not find any significant impact of network influence on the propensity to get re-incarcerated. This is possible as the affirmations network is formed inside the TC, and it is likely to impact short-term outcomes like graduation from the facility but not longer-term outcomes such as recidivism. As for the own covariates, we find that race influences recidivism. Being white reduces the likelihood of recidivism in the future. This result is in line with the findings in the literature. For example, [39, 37] suggest that discrimination in the criminal justice system is a common underlying mechanism that results in disparities in crime based on race. In our model 3, we do not find any significant impact of age and own LSI on recidivism.

Own time spent in TC substantially reduces the chances of recidivism. Residents

seemed to have gained from their time spent in the correctional facility, which encouraged them to practice pro-social behavior in the future. Nevertheless, having more network connections who took a long time to graduate increases one's propensity for recidivism in future. This dependence on time spent in TC is similar to what we observed for the outcome of graduation. Therefore, own time spent in TC substantially increases chance of graduation and decreases chance of future recidivism, while having network connected neighbors who took more time to graduate decreases chance of graduation and increases chance of recidivism.

5.3 Assessing Model fit with Out of Sample Prediction

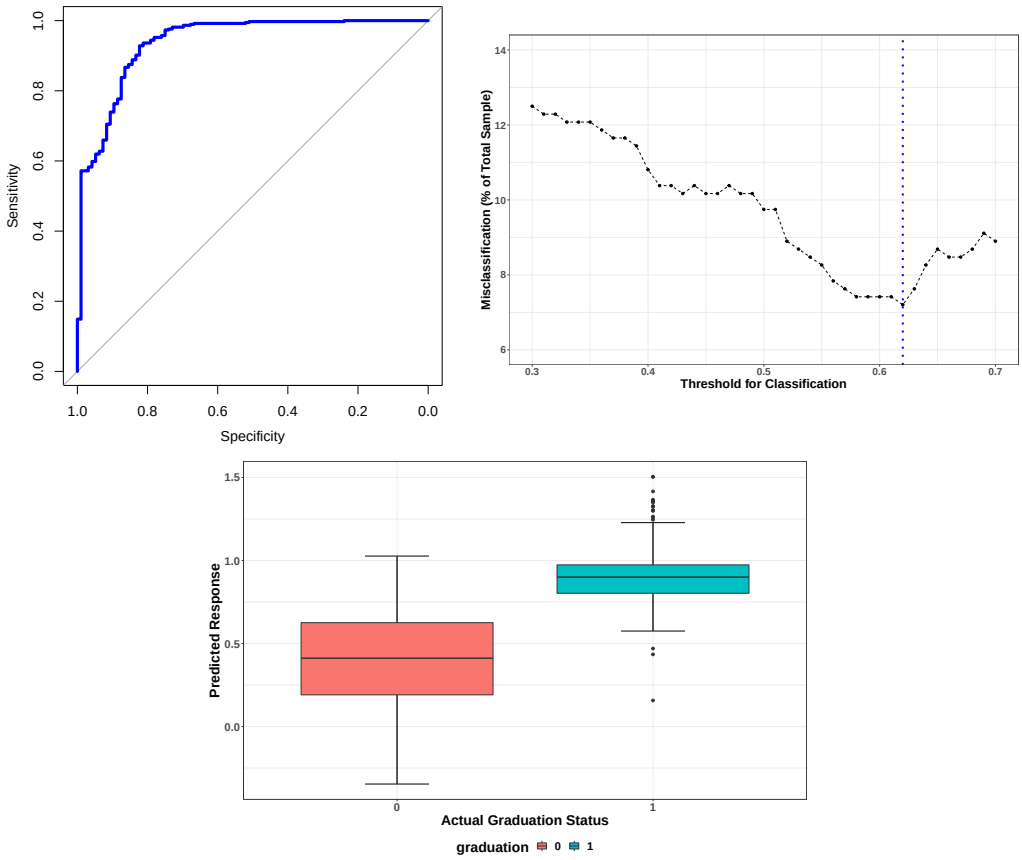


Figure 7: Leave one out prediction for graduation status : (a) ROC curve, AUC value is 0.9376. (b) misclassification error (as percentage of total sample) for different classification threshold (c) Box plot of predicted responses by actual graduation status.

We also access the fit of our final model with out of sample predictive performance. As discussed above, we first estimate the latent variables $\hat{U}$ from the adjacency matrix of the whole network. Then we hold out one test data point and consider the remaining

data points as training data $Y_S, Z_S, \hat{U}, L_{SS}$. We obtain parameter estimates and predict the response for the held out data point Y_o. We repeat this procedure multiple times such that we have out of sample prediction for every datapoint. We then obtain a Receiver Operating Characteristics (ROC) plot (Figure 7(a)), a plot of misclassification error with varying threshold of classification (Figure 7(b)), and a plot comparing predicted responses with actual responses (Figure 7(c)). The Area Under the Curve (AUC) value for the plot is 0.9376. The ROC plot along with the AUC value shows that the model provides a good fit to the data and is able to predict the graduation status accurately. Figure 7(b) further shows that a classification procedure based on the predicted responses with a few thresholds around 0.5 have high classification accuracy for graduation status. Figure 7(c) shows that predicted responses from the model are generally higher when the true response is 1 than when the true response is 0, further giving evidence of good model fit.

We note that there is a limitation of the methodology in terms of being appropriate for analyzing this data. While the methods developed in the paper are appropriate for continuous outcomes, both the outcomes in our application problem are binary. Therefore there might be some concerns with the validity of the results obtained from this model on such a dataset. The model fit results in this section lends some credibility to the fit and utility of the model even through the outcome variable is binary. Developing homophily and bias corrected estimator for binary response SAR models with logit or probit link functions is a future research direction.

5.4 Discussion

These models carry several lessons for TC clinicians. Most obviously, the finding that network connections predict graduation, [21] confirms the importance of the community as a method of treatment in the TC. Previous research has found the existence of homophily among TC graduates [83]. However, this paper shows that network influence, once we control for homophily, substantially impacts the propensity to graduate from TC.

Residents who had more connections with peers who were in the program for longer were less likely to graduate and more likely to be reincarcerated, while residents who had more connections with peers who had higher LSI scores were more likely to graduate and less likely to be reincarcerated. One possible explanation for this finding that would be consistent with TC theory is that a willingness to reach out to all peers in the TC, even those struggling, is a predictor of positive outcomes. This would also be consistent with studies of twelve-step groups, which find that members who help peers are more likely to maintain sobriety [58, 74].

This analysis suggests that more research into those aspects of social networks that predict outcomes in TCs is warranted. The social learning emphasis of TC clinical theory

[84] would imply that social networks play their classic role as conduits of information, which would explain the direct effect of network connections on graduation. But this does not explain why individuals who connect with weaker peers have a better outcome. Previous research suggests that social network roles influence and constrain residents as they go through TC treatment [11, 81]. It is plausible that interaction between stronger and weaker members of the TC allows the former to experience the role of helper to the latter and that such experience is of value in recovery.

6 Conclusions

We develop a new method for providing causal estimates of network influence by separating it from the latent homophily of network nodes. We do the estimation in two steps. A Stochastic Block Model is used to model the network and we estimate the latent homophily from a spectral embedding of the adjacency matrix. Since we use the estimated latent homophily instead of true latent factors, we provide a new methodology to correct this finite sample bias. We show that the bias-corrected estimator is consistent and derive its asymptotic limiting distribution. Our method is computationally fast and can be applied to large network linked data-sets as we use spectral decomposition and regression for estimation.

We apply our methodology to a Therapeutic Community dataset of female criminal offenders. The dataset provides us with electronic records of residents' entry and exit dates, socio-demographic characteristics, graduation status, and recidivism post exit from the TC. Additionally, we also observe a rich network of peer affirmations used as a tool for mutual monitoring in the TC. We provide causal estimates of network influence on graduation and recidivism by controlling for latent homophily using the method developed in this paper. We find that network influence for the female offenders impacts short-term outcomes such as graduation from TC but has no implications for recidivism. Moreover, we find that contextual effects significantly impact both graduation and recidivism. We find that having more peers with a high LSI or shorter time in the facility increases the likelihood of graduation and lowers the propensity of re-incarceration. This result merits research in the direction of network roles. It is plausible that residents constrain their anti-social behavior, especially if they are in an influential position in the network and exercise a role model effect on the weaker peers.

References

[1] S. Aral and C. Nicolaides. Exercise contagion in a global social network. *Nature communications*, 8:14753, 2017.

[2] A. Athreya, D. E. Fishkind, M. Tang, C. E. Priebe, Y. Park, J. T. Vogelstein, K. Levin, V. Lyzinski, and Y. Qin. Statistical inference on random dot product graphs: a survey. *The Journal of Machine Learning Research*, 18(1):8393–8484, 2017.

[3] A. Athreya, C. E. Priebe, M. Tang, V. Lyzinski, D. J. Marchette, and D. L. Sussman. A limit theorem for scaled eigenvectors of random dot product graphs. *Sankhya A*, 78(1):1–18, 2016.

[4] P. Bayer, R. Hjalmarsson, and D. Pozen. Building criminal capital behind bars: Peer effects in juvenile corrections. *The Quarterly Journal of Economics*, 124(1):105–147, 2009.

[5] D. BB and B. Mutton. The effect of errors in the independent variables in linear regression. *Biometrika*, 62(2):383–391, 1975.

[6] P. J. Bickel and A. Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073, 2009.

[7] Y. Bramoullé, H. Djebbari, and B. Fortin. Identification of peer effects through social networks. *Journal of econometrics*, 150(1):41–55, 2009.

[8] R. Brown and A. Velásquez. The effect of violent crime on the human capital accumulation of young adults. *Journal of development economics*, 127:1–12, 2017.

[9] J. P. Buonaccorsi. *Measurement error: models, methods, and applications*. Chapman and Hall/CRC, 2010.

[10] J. T. Cacioppo, J. H. Fowler, and N. A. Christakis. Alone in the crowd: the structure and spread of loneliness in a large social network. *Journal of personality and social psychology*, 97(6):977, 2009.

[11] B. Campbell, K. Warren, M. Weiler, and G. De Leon. Eigenvector centrality defines hierarchy and predicts graduation in therapeutic community units. *Plos one*, 16(12):e0261405, 2021.

[12] S. E. Carrell, R. L. Fullerton, and J. E. West. Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3):439–464, 2009.

[13] R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. *Measurement error in nonlinear models: a modern perspective*. Chapman and Hall/CRC, 2006.

[14] E. Y. Chen, J. Fan, and X. Zhu. Community network auto-regression for high-dimensional time series. *arXiv preprint arXiv:2007.05521*, 2020.

[15] P. Chin, A. Rao, and V. Vu. Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *COLT*, pages 391–423, 2015.

[16] N. A. Christakis and J. H. Fowler. The spread of obesity in a large social network over 32 years. *New England journal of medicine*, 357(4):370–379, 2007.

[17] N. A. Christakis and J. H. Fowler. The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21):2249–2258, 2008.

[18] E. Cohen-Cole and J. M. Fletcher. Is obesity contagious? social networks vs. environmental factors in the obesity epidemic. *Journal of health economics*, 27(5):1382–1387, 2008.

[19] L. Coviello, Y. Sohn, A. D. Kramer, C. Marlow, M. Franceschetti, N. A. Christakis, and J. H. Fowler. Detecting emotional contagion in massive social networks. *PloS one*, 9(3):e90315, 2014.

[20] A. P. Damm and C. Dustmann. Does growing up in a high crime neighborhood affect youth criminal behavior? *American Economic Review*, 104(6):1806–32, 2014.

[21] G. De Leon, H. K. Wexler, and N. Jainchill. The therapeutic community: Success and improvement rates 5 years after treatment. *International Journal of the Addictions*, 17(4):703–747, 1982.

[22] D. O. Dean, D. J. Bauer, and M. J. Prinstein. Friendship dissolution within social networks modeled through multilevel event history analysis. *Multivariate behavioral research*, 52(3):271–289, 2017.

[23] E. Dobriban and S. Wager. High-dimensional asymptotics of prediction: Ridge regression and classification. *The Annals of Statistics*, 46(1):247–279, 2018.

[24] J. H. Fowler and N. A. Christakis. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *Bmj*, 337:a2338, 2008.

[25] L. Fratiglioni, H.-X. Wang, K. Ericsson, M. Maytan, and B. Winblad. Influence of social network on occurrence of dementia: a community-based longitudinal study. *The lancet*, 355(9212):1315–1319, 2000.

[26] C. Gao, Z. Ma, A. Y. Zhang, and H. H. Zhou. Achieving optimal misclassification proportion in stochastic block models. *The Journal of Machine Learning Research*, 18(1):1980–2024, 2017.

[27] A. Gaviria and S. Raphael. School-based peer effects and juvenile behavior. *Review of Economics and Statistics*, 83(2):257–268, 2001.

[28] P. Goldsmith-Pinkham and G. W. Imbens. Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, 31(3):253–264, 2013.

[29] A. B. Golgher and P. R. Voss. How to interpret the coefficients of spatial models: Spillovers, direct and indirect effects. *Spatial Demography*, 4(3):175–205, 2016.

[30] M. Goulard, T. Laurent, and C. Thomas-Agnan. About predictions in spatial autoregressive models: Optimal and almost optimal strategies. *Spatial Economic Analysis*, 12(2-3):304–325, 2017.

[31] S. Hodges and P. Moore. Data uncertainties and least squares regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 21(2):185–195, 1972.

[32] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098, 2002.

[33] P. Holland, K. Laskey, and S. Leinhardt. Stochastic blockmodels: some first steps. *Social Networks*, 5:109–137, 1983.

[34] C.-S. Hsieh and L. F. Lee. A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*, 31(2):301–319, 2016.

[35] H. H. Kelejian and I. R. Prucha. A generalized moments estimator for the autoregressive parameter in a spatial model. *International economic review*, 40(2):509–533, 1999.

[36] H. H. Kelejian and I. R. Prucha. On the asymptotic distribution of the moran i test statistic with applications. *Journal of Econometrics*, 104(2):219–257, 2001.

[37] D. Knox, W. Lowe, and J. Mummolo. Administrative records mask racially biased policing. *American Political Science Review*, 114(3):619–637, 2020.

[38] A. D. Kramer, J. E. Guillory, and J. T. Hancock. Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 111(24):8788–8790, 2014.

[39] K. Lang and A. Kahn-Lang Spitzer. Race discrimination: An economic perspective. *Journal of Economic Perspectives*, 34(2):68–89, 2020.

[40] D. Lazer. The co-evolution of individual and network. *Journal of Mathematical Sociology*, 25(1):69–108, 2001.

[41] D. Lazer, B. Rubineau, C. Chetkovich, N. Katz, and M. Neblo. The coevolution of networks and political attitudes. *Political Communication*, 27(3):248–274, 2010.

[42] L.-F. Lee. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6):1899–1925, 2004.

[43] L.-F. Lee. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140(2):333–374, 2007.

[44] L.-f. Lee, X. Liu, and X. Lin. Specification and estimation of social interaction models with network structures. *The Econometrics Journal*, 13(2):145–176, 2010.

[45] R. T. A. Leenders. Modeling social influence through network autocorrelation: constructing the weight matrix. *Social networks*, 24(1):21–47, 2002.

[46] J. Lei, A. Rinaldo, et al. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.

[47] X. Lin. Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, 28(4):825–860, 2010.

[48] R. Lyons. The spread of evidence-poor medicine via flawed social-network analysis. *Statistics, Politics and Policy*, 2(1), 2011.

[49] C. F. Manski. Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542, 1993.

[50] P. V. Marsden and N. E. Friedkin. Network studies of social influence. *Sociological Methods & Research*, 22(1):127–151, 1993.

[51] E. McFowland III and C. R. Shalizi. Estimating causal peer influence in homophilous social networks by inferring latent locations. *Journal of the American Statistical Association*, pages 1–12, 2021.

[52] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[53] T. Nakamura. Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika*, 77(1):127–137, 1990.

[54] H. Noel and B. Nyhan. The "unfriending" problem: The consequences of homophily in friendship retention for causal estimates of social influence. *Social Networks*, 33(3):211–218, 2011.

[55] S. J. Novick and L. A. Stefanski. Corrected score estimation via complex variable simulation extrapolation. *Journal of the American Statistical Association*, 97(458):472–481, 2002.

[56] E. L. Ogburn and T. J. VanderWeele. Vaccines, contagion, and social networks. *The Annals of Applied Statistics*, 11(2):919–948, 2017.

[57] K. Ord. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349):120–126, 1975.

[58] M. E. Pagano, K. B. Friend, J. S. Tonigan, and R. L. Stout. Helping other alcoholics in alcoholics anonymous and drinking outcomes: findings from project match. *Journal of studies on alcohol*, 65(6):766–773, 2004.

[59] E. Patacchini and Y. Zenou. The strength of weak ties in crime. *European Economic Review*, 52(2):209–236, 2008.

[60] S. Paul and Y. Chen. Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. *Electronic Journal of Statistics*, 10(2):3807–3870, 2016.

[61] S. Paul, Y. Chen, et al. A random effects stochastic block model for joint community detection in multiple networks with applications to neuroimaging. *Annals of Applied Statistics*, 14(2):993–1029, 2020.

[62] S. Paul, Y. Chen, et al. Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *The Annals of Statistics*, 48(1):230–250, 2020.

[63] M. Pensky, T. Zhang, et al. Spectral clustering in the dynamic stochastic block model. *Electronic Journal of Statistics*, 13(1):678–709, 2019.

[64] P. Pinotti. The economic costs of organised crime: Evidence from southern italy. *The Economic Journal*, 125(586):F203–F232, 2015.

[65] K. Rohe, S. Chatterjee, and B. Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist*, 39(4):1878–1915, 2011.

[66] M. Sarrias. Notes on spatial econometrics. 2020.

[67] D. W. Schafer. Covariate measurement error in generalized linear models. *Biometrika*, 74(2):385–391, 1987.

[68] C. R. Shalizi and A. C. Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological methods & research*, 40(2):211–239, 2011.

[69] P. Sharkey and G. Torrats-Espinosa. The effect of violent crime on economic mobility. *Journal of Urban Economics*, 102:22–33, 2017.

[70] L. A. Stefanski. The effects of measurement error on parameter estimation. *Biometrika*, 72(3):583–592, 1985.

[71] L. A. Stefanski and R. J. Carroll. Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, 74(4):703–716, 1987.

[72] L. A. Stefanski, R. J. Carroll, et al. Covariate measurement error in logistic regression. *The Annals of Statistics*, 13(4):1335–1351, 1985.

[73] M. Stevenson. Breaking bad: Mechanisms of social influence and the path to criminality in juvenile jails. *Review of Economics and Statistics*, 99(5):824–838, 2017.

[74] M. S. Subbaraman, L. A. Kaskutas, and S. Zemore. Sponsorship and service as mediators of the effects of making alcoholics anonymous easier (maaez), a 12-step facilitation intervention. *Drug and Alcohol Dependence*, 116(1-3):117–124, 2011.

[75] D. L. Sussman, M. Tang, D. E. Fishkind, and C. E. Priebe. A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association*, 107(499):1119–1128, 2012.

[76] M. Tang, C. E. Priebe, et al. Limit theorems for eigenvectors of the normalized laplacian for random graphs. *The Annals of Statistics*, 46(5):2360–2415, 2018.

[77] T. J. VanderWeele. Sensitivity analysis for contagion effects in social networks. *Sociological Methods & Research*, 40(2):240–255, 2011.

[78] T. J. VanderWeele and W. An. Social networks and causal inference. In *Handbook of causal analysis for social research*, pages 353–374. Springer, 2013.

[79] T. J. VanderWeele, E. L. Ogburn, and E. J. T. Tchetgen. Why and when" flawed" social network analyses still yield valid tests of no contagion. *Statistics, Politics and Policy*, 3(1), 2012.

[80] A. Velásquez. The economic burden of crime evidence from mexico. *Journal of Human Resources*, 55(4):1287–1318, 2020.

[81] K. Warren, B. Campbell, and S. Cranmer. Tightly bound: the relationship of network clustering coefficients and reincarceration at three therapeutic communities. *Journal of Studies on Alcohol and Drugs*, 81(5):673–680, 2020.

[82] K. Warren, N. J. Doogan, and F. Doherty. Difference in response to feedback and gender in three therapeutic community units. *Frontiers in Psychiatry*, 12, 2021.

[83] K. L. Warren. Senior therapeutic community members show greater consistency when affirming peers: evidence of social learning. *Therapeutic Communities: The International Journal of Therapeutic Communities*, 2020.

[84] K. L. Warren, N. Doogan, U. Wernekinck, and F. C. Doherty. Resident interactions when affirming and correcting peers in a therapeutic community for women. *Therapeutic Communities: The International Journal of Therapeutic Communities*, 2021.

[85] H. Xu, G. Fang, and X. Zhu. Network group hawkes process model. *arXiv preprint arXiv:2002.08521*, 2020.

[86] Y. Zhao, E. Levina, and J. Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist*, 40:2266–2292, 2012.

[87] J. Zhou, Y. Tu, Y. Chen, and H. Wang. Estimating spatial autocorrelation with sampled network data. *Journal of Business & Economic Statistics*, 35(1):130–138, 2017.

[88] X. Zhu, X. Chang, R. Li, and H. Wang. Portal nodes screening for large scale social networks. *Journal of econometrics*, 209(2):145–157, 2019.

[89] X. Zhu, R. Pan, G. Li, Y. Liu, and H. Wang. Network vector autoregression. *The Annals of Statistics*, 45(3):1096–1123, 2017.

[90] D. J. Zimmerman. Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and statistics*, 85(1):9–23, 2003.

Appendix A Miscellaneous results

A.1 Proof of $M^TM - K = M$

$$\begin{aligned}
M^TM - K &= I - 2\tilde{X}(\tilde{X}^T\tilde{X} - (\sum_i \Omega_i))^{-1}\tilde{X}^T + \tilde{X}(\tilde{X}^T\tilde{X} - (\sum_i \Omega_i))^{-1}\tilde{X}^T\tilde{X}(\tilde{X}^T\tilde{X} - (\sum_i \Omega_i))^{-1}\tilde{X}^T \\
&\quad - \tilde{X}(\tilde{X}^T\tilde{X} - (\sum_i \Omega_i))^{-1}(\sum_i \Omega_i)(\tilde{X}^T\tilde{X} - (\sum_i \Omega_i))^{-1}\tilde{X}^T \\
&= I - \tilde{X}(\tilde{X}^T\tilde{X} - (\sum_i \Omega_i))^{-1}\tilde{X}^T = M.
\end{aligned}$$

A.2 First and second derivative of $l^*(\rho)$

The first and second derivative of the corrected concentrated log likelihood function $l^*(\rho)$ with respect to ρ is given by:

$$\begin{aligned}
\frac{\partial l^*}{\partial \rho} &= \frac{2}{n}\sum_{i=1}^{n}\frac{\lambda_i}{1-\rho\lambda_i} + 2\frac{\rho L^TY^TMYL - Y^TMYL}{\hat{\sigma}^2} \\
\frac{\partial^2 l^*}{\partial \rho} &= \frac{2}{n} * \sum_{i=1}^{n}\frac{\lambda_i^2}{(1-\rho\lambda_i)^2} + \frac{2L^TY^TMYL}{\hat{\sigma}^2} - \frac{4(\rho L^TY^TMYL - Y^TMYL)^2}{(\hat{\sigma}^2)^2}
\end{aligned}$$

A.3 The gradient and Hessian of log likelihood function

$$\begin{aligned}
\nabla_\delta l_n^*(\theta) &= \frac{1}{\sigma^2}[\tilde{X}_n^T(S_n(\rho)Y_n - \tilde{X}_n\delta) + \sum_i(\Omega_i)\delta], \\
\nabla_\rho l_n^*(\theta) &= \frac{1}{\sigma^2}[(LY)^T(S_n(\rho)Y_n - \tilde{X}_n\delta)] - tr(G_n(\rho)), \\
\nabla_{\sigma^2} l_n^*(\theta) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\{(S_n(\rho)Y_n - \tilde{X}_n\delta)^T(S_n(\rho)Y_n - \tilde{X}_n\delta) - \delta^T(\sum_i \Omega_i)\delta\},
\end{aligned}$$

The first two gradients can be conveniently written together as

$$\nabla_\delta l_n^*(\theta) = \frac{1}{\sigma^2}\{\tilde{X}_n^T(S_n(\rho)Y_n - \tilde{X}_n\delta) + (\sum_i \Omega_i)\delta\}.$$

The elements of the Hessian matrix are given by:

$$\nabla^2_{\delta,\delta} l^*_n(\theta) = -\frac{1}{\sigma^2}[-\tilde{X}_n^T \tilde{X}_n - \sum_i \Omega_i],$$

$$\nabla_{\delta,\rho} l^*_n(\theta) = -\frac{1}{\sigma^2}[\tilde{X}_n^T L_n Y_n]$$

$$\nabla_{\delta,\sigma^2} l^*_n(\theta) = -\frac{1}{\sigma^4}\{\tilde{X}_n^T (S_n(\rho)Y_n - \tilde{X}_n\delta) + (\sum_i \Omega_i)\delta\}$$

$$\nabla_{\sigma^2,\sigma^2} l^*_n(\theta) = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\{(S_n(\rho)Y_n - \tilde{X}_n\delta)^T (S_n(\rho)Y_n - \tilde{X}_n\delta) + \delta^T (\sum_i \Omega_i)\delta\}$$

$$\nabla_{\sigma^2,\rho} l^*_n(\theta) = -\frac{1}{\sigma^4}(S_n(\rho)Y_n - \tilde{X}_n\delta)^T L_n Y_n$$

$$\nabla_{\rho,\rho} l^*_n(\theta) = -\frac{1}{\sigma^2}(L_n Y_n)^T L_n Y_n - tr(G_n(\rho)G_n(\rho)).$$

Appendix B Proof of Theorems

B.1 Proof of Theorem 1

Some results follow from the assumptions.

Proposition 1. *Under the assumptions A1-A8, we have the following results*

1. $\|S_n(\rho)S_n^{-1}\|_\infty < \infty$ *i.e.,* $S_n(\rho)S_n^{-1}$ *is uniformly bounded in row sums.*

2. *The elements of the vectors* $X_n\delta_0$ *and* $G_nX_n\delta_0$ *are uniformly bounded in* n*.*

Proof. For the first result, $\|S_n(\rho)S_n^{-1}\|_\infty \leq \|S_n(\rho)\|_\infty\|S_n^{-1}\|_\infty < \infty$, for all ρ. For the second result, since all elements of X_n are uniformly bounded in n, the rowsums of X_n are uniformly bounded and consequently elements of $X_n\delta_0$ are bounded. For $G_nX_n\delta_0$ we note that, $\|G_nX_n\|_\infty \leq \|L_n\|_\infty\|S_n^{-1}\|_\infty\|X_n\|_\infty < \infty$. □

To show consistency of $\hat{\theta}_n$, we first analyze the consistency of $\hat{\rho}_n$ and then that of $\hat{\delta}_n$ and $\hat{\sigma}_n^2$.

The true data generating process is

$$Y_n = S_n^{-1}X_n\delta_0 + S_n^{-1}V_n, \quad V_n \sim N(0, \sigma^2 I_n).$$

Recall the corrected loglikelihood function maximized to obtain the estimators is

$$l_n^*(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}[\tilde{V}_n(\rho)^T\tilde{V}_n(\rho) - \delta^T(\sum_i \Omega_i)\delta] + \log|S_n(\rho)|,$$

where

$$\tilde{V}_n(\rho) = S_n(\rho)Y_n - \tilde{X}_n\delta.$$

Now define

$$Q_n(\rho) = \max_{\beta,\sigma^2} E[l_n^*(\theta)] = \max_{\beta,\sigma^2} E^+E^*[l_n^*(\theta)],$$

as the concentrated unconditional expectation of the corrected log likelihood function. Solving the optimization problem we obtain the solutions for δ, σ^2 for a given ρ as

$$\begin{aligned}
\delta(\rho) &= (X_n^TX_n)^{-1}X_n^TS_n(\rho)E[Y_n] = (X_n^TX_n)^{-1}X_n^TS_n(\rho)S_n^{-1}X_n\delta_0 \\
\sigma^2(\rho) &= \frac{1}{n}\left((\rho_0-\rho)^2(G_nX_n\delta_0)^TM_{1n}(G_nX_n\delta_0\right) + \frac{\sigma_0^2}{n}tr\left((S_n^T)^{-1}S_n(\rho)^TS_n(\rho)S_n^{-1}\right),
\end{aligned}$$

where $M_{1n} = (I_n - X_n(X_n^TX_n)^{-1}X_n^T)$. We note that therefore the function $Q_n(\rho)$ is identical to the function analyzed in [42]. Therefore the same arguments in [42] proves

uniqueness of ρ_0 as the global maximizer of $Q_n(\rho)$ in the compact parameter space R. In particular, for any $\epsilon > 0$, we have

$$\limsup_{n\to\infty} \max_{\rho \in N_\epsilon(\rho_0)} |Q_n(\rho) - Q_n(\rho_0)| < 0.$$

Therefore we only need to show the uniform convergence of the concentrated corrected log likelihood $l_n^*(\rho)$ to $Q_n(\rho)$ over the compact parameter space R. Now we have

$$\frac{1}{n}(l_n^*(\rho) - Q_n(\rho)) = -\frac{1}{2}(\log(\hat{\sigma}_n^2) - \log(\sigma^2(\rho))),$$

where

$$\hat{\sigma}_n^2(\rho) = \frac{1}{n} Y_n^T S_n^T(\rho) M_{2n} S_n(\rho) Y_n, \tag{B.1}$$

with

$$M_{2n} = I_n - \tilde{X}_n (\tilde{X}_n^T \tilde{X}_n - \Omega)^{-1} \tilde{X}_n^T.$$

Since

$$I + (\rho_0 - \rho) G_n = I + \rho_0 L_n S_n^{-1} - \rho L_n S_n^{-1} = I + (I - S_n) S_n^{-1} - \rho L_n S_n^{-1} = S_n^{-1} - \rho L_n S_n^{-1} = S_n(\rho) S_n^{-1},$$

We can expand $\hat{\sigma}_n^2(\rho)$ to obtain

$$\begin{aligned}
\hat{\sigma}_n^2(\rho) &= \frac{1}{n}(S_n^{-1} X_n \delta_0 + S_n^{-1} V_n)^T S_n^T(\rho) M_{2n} S_n(\rho)(S_n^{-1} X_n \delta_0 + S_n^{-1} V_n) \\
&= (X_n \delta_0 + (\rho_0 - \rho) G_n X_n \delta_0 + S_n(\rho) S_n^{-1} V_n)^T M_{2n} (X_n \delta_0 + (\rho_0 - \rho) G_n X_n \delta_0 + S_n(\rho) S_n^{-1} V_n) \\
&= \frac{1}{n} \Bigg((\rho_0 - \rho)^2 (G_n X_n \delta_0)' M_{2n} (G_n X_n \delta_0) + (X_n \delta_0)^T M_{2n} (X_n \delta_0) \\
&\quad + V_n^T (S_n^T)^{-1} S_n^T(\rho) M_{2n} S_n(\rho) S_n^{-1} V_n + 2(\rho_0 - \rho)(G_n X_n \delta_0)^T M_{2n} S_n(\rho) S_n^{-1} V_n \\
&\quad + 2(\rho_0 - \rho)(X_n \delta_0)^T M_{2n} (G_n X_n \delta_0) + 2(X_n \delta_0)^T M_{2n} S_n(\rho) S_n^{-1} V_n \Bigg)
\end{aligned}$$

The decomposition of $\hat{\sigma}^2(\rho)$ produces 6 terms that we analyze in groups of terms that can be analyzes with similar techniques.

Terms 1, 2 and 5: We note the following convergences for terms that do not contain ρ from the weak law of large numbers.

$$\frac{1}{n} \begin{pmatrix} \tilde{X}_n^T X_n \delta_0 \\ \tilde{X}_n^T G_n X_n \delta_0 \\ \tilde{X}_n^T \tilde{X}_n - \Omega \end{pmatrix} \equiv \begin{pmatrix} \frac{1}{n} \sum_i (X_n \delta_0)_i \tilde{X}_{in}^T \\ \frac{1}{n} \sum_i (G_n X_n \delta_0)_i \tilde{X}_{in}^T \\ \frac{1}{n} \sum_i (\tilde{X}_{in} \tilde{X}_{in}^T - \Omega/n) \end{pmatrix} \xrightarrow{p} \frac{1}{n} \begin{pmatrix} X_n^T X_n \delta_0 \\ X_n^T G_n X_n \delta_0 \\ X_n^T X_n \end{pmatrix}.$$

Therefore by the multivariate continuous mapping theorem

$$\begin{aligned}
&(G_nX_n\beta_0)^TM_{2n}(G_nX_n\beta_0)\\
&=(G_nX_n\beta_0)^T(G_nX_n\beta_0)-(G_nX_n\beta_0)^T\tilde{X}_n(\tilde{X}_n^T\tilde{X}_n-\Omega)^{-1}\tilde{X}_n)(G_nX_n\beta_0)\\
&\xrightarrow{p}(G_nX_n\beta_0)^T(G_nX_n\beta_0)-(G_nX_n\beta_0)^TX_n(X_n^TX_n)^{-1}X_n)(G_nX_n\beta_0)\\
&=(G_nX_n\beta_0)^TM_{1n}(G_nX_n\beta_0).
\end{aligned}$$

In addition, $(\rho_0-\rho)^2$ is bounded by a constant since $\rho_0,\rho\in R$, which is a compact set. Therefore the first term converges in probability to first term in Equation 3.1 uniformly for all ρ.

For the second term, the multivariate continuous mapping theorem gives

$$\begin{aligned}
(X_n\beta_0)^TM_{2n}(X_n\beta_0)&=(X_n\beta_0)^T(X_n\beta_0)-(X_n\beta_0)^T\tilde{X}_n(\tilde{X}_n^T\tilde{X}_n-\Omega)^{-1}\tilde{X}_n^TX_n\beta_0\\
&\xrightarrow{p}(X_n\beta_0)^T(X_n\beta_0)-(X_n\beta_0)^TX_n(X_n^TX_n)^{-1}X_n^TX_n\beta_0\\
&=0.
\end{aligned}$$

Finally, for the fifth term a similar argument shows

$$\begin{aligned}
(X_n\delta_0)^TM_{2n}(G_nX_n\delta_0)&=(X_n\delta_0)^T(G_nX_n\delta_0)-\delta_0^TX_n^T\tilde{X}_n(\tilde{X}_n^T\tilde{X}_n-\Omega)^{-1}\tilde{X}_n^TG_nX_n\delta_0\\
&\xrightarrow{p}(X_n\delta_0)^T(G_nX_n\delta_0)-\delta_0^TX_n^TX_n(X_n^TX_n)^{-1}X_n^TG_nX_n\delta_0\\
&=0.
\end{aligned}$$

Terms 4 and 6 : Notice

$$(X_n\delta_0)^TM_{2n}(S_n(\rho)S_n^{-1}V_n=(X_n\delta_0)^T(S_n(\rho)S_n^{-1}V_n-(X_n\delta_0)^T\tilde{X}_n(\tilde{X}_n^T\tilde{X}_n-\Omega)^{-1}\tilde{X}_n^T(S_n(\rho)S_n^{-1}V_n.$$

For the first term in the above expression, we apply the Uniform WLLN to the function $g(V_i,\rho)=\frac{1}{n}(X_n\delta_0)^T(S_n(\rho)S_n^{-1}V_n)=\frac{1}{n}\sum_i(S_n(\rho)S_n^{-1}V_n)_i(X_n\delta_0)_i$. Further from Proposition 1 we have the elements of $X_n\delta_0$ are uniformly bounded and the absolute row sums of $(S_n(\rho)S_n^{-1}$ are uniformly bounded. Therefore we have

$$\begin{aligned}
E|(S_n(\rho)S_n^{-1}V_n)_1(X_n\delta_0)_1|&\le E[|(S_n(\rho)S_n^{-1}V_n)_1||(X_n\delta_0)_1|]\\
&\le|(X_n\delta_0)_1|E\sum_j|(S_n(\rho)S_n^{-1})_{1j}(V_n)_j|\\
&\le|(X_n\delta_0)_1|\sum_j|(S_n(\rho)S_n^{-1})_{1j}|E|(V_n)_j|<\infty,
\end{aligned}$$

uniformly for all $\rho \in R$. We can also compute the expectation as

$$E[(S_n(\rho)S_n^{-1}V_n)_1(X_n\delta_0)_1] = 0.$$

Therefore applying Uniform LLN we conclude,

$$(X_n\delta_0)^T S_n(\rho)S_n^{-1}V_n \xrightarrow{p} 0$$

uniformly in ρ.

Next for the second part of the expression, due to independence of η_n and V_n, and since $E[V_{in}] = 0$, we have

$$E[(S_n(\rho)S_n^{-1}V_n)_i((X_n\delta_0)^T\tilde{X}_n(\tilde{X}_n^T\tilde{X}_n - \Omega)^{-1}\tilde{X}_n^T)_i] = 0.$$

From the assumptions, $\tilde{X}_n(\tilde{X}_n^T\tilde{X}_n - \Omega)^{-1}\tilde{X}_n^T$ is uniformly bounded in absolute row sum. Call $\tilde{X}_n(\tilde{X}_n^T\tilde{X}_n - \Omega)^{-1}\tilde{X}_n^T S_n(\rho)S_n^{-1} = K_n$. Then $\|K_n\|_\infty < \infty$. This implies,

$$(X_n\delta_0)^T\tilde{X}_n(\tilde{X}_n^T\tilde{X}_n - \Omega)^{-1}\tilde{X}_n^T(S_n(\rho)S_n^{-1}V_n \xrightarrow{p} 0,$$

uniformly in ρ.

The convergence of Term 4 follows similarly by noting that the elements of $G_nX_n\delta_0$ are also uniformly bounded.

Term 3: The 3rd term is a quadratic form in the error vector V_n with a matrix $B_n(\rho) = (S_n^T)^{-1}S_n^T(\rho)M_{2n}S_n(\rho)S_n^{-1}$ which itself is stochastic. In contrast the commonly encountered quadratic forms are of the form y^TAy, with A being a non random matrix. We first note that from the Assumptions we have and the relationship among matrix norms, namely, $\|A\|_2 \leq \sqrt{\|A\|_\infty\|A\|_1}$, the spectral norm of the matrix $B_n(\rho)$ is uniformly bounded in ρ:

$$\|B_n(\rho)\|_2 \leq \|(S_n^T)^{-1}\|_2^2\|S_n^T(\rho)\|_2^2\|M_{2n}\|_2^2 \leq C$$

Under this condition, the result in [23] is applicable and we have the following convergence result

$$\frac{1}{n}V_n^T(S_n^T)^{-1}S_n^T(\rho)M_{2n}S_n(\rho)S_n^{-1}V_n \xrightarrow{p} \frac{\sigma_0^2}{n}tr((S_n^T)^{-1}S_n^T(\rho)M_{2n}S_n(\rho)S_n^{-1})$$

uniformly in ρ.

Now note that

$$tr((S_n^T)^{-1}S_n^T(\rho)M_{2n}S_n(\rho)S_n^{-1}) = \sum_i (S_n(\rho)S_n^{-1})_i^T M_{2n} S_n(\rho)S_n^{-1})_i$$

where $(S_n(\rho)S_n^{-1})_i$ denotes the ith column of the matrix $(S_n(\rho)S_n^{-1})$.

Now applying Uniform WLLN we obtain

$$\frac{1}{n}\sum_j (S_n(\rho)S_n^{-1})_{ij}(\tilde{X}_n)_j \xrightarrow{p} \frac{1}{n}\sum_j (S_n(\rho)S_n^{-1})_{ij}(X_n)_j,$$

uniformly for all ρ, and the middle term which is not dependent on ρ converges as $\frac{1}{n}(\tilde{X}^T\tilde{X} - \Omega) \xrightarrow{p} \frac{1}{n}X^TX$. Therefore for a given i, we have

$$\frac{1}{n}(S_n(\rho)S_n^{-1})_i^T M_{2n} S_n(\rho)S_n^{-1})_i \xrightarrow{p} (S_n(\rho)S_n^{-1})_i^T M_{1n} S_n(\rho)S_n^{-1})_i$$

Taking summation over i, we have the desired result on the convergence of $\hat{\rho}_n$.

Once convergence of $\hat{\rho}_n$ to ρ_0 is established, the convergence of $\hat{\delta}_n$ can be established. We note

$$\hat{\delta}_n = (\tilde{X}_n^T\tilde{X}_n - \sum_i \Omega_i)^{-1}\tilde{X}_n^T(I + (\rho_0 - \rho)G_n)(X_n\delta_0 + V_n)$$

Now

$$\frac{1}{n}(\tilde{X}_n^T\tilde{X}_n - \sum_i \Omega_i) = \frac{1}{n}X_n^TX_n + o_p(1)$$

$$\frac{1}{n}(\tilde{X}_n^TX_n\delta_0 + (\rho_0 - \rho)\tilde{X}_n^TG_nX_n\delta_0) = \frac{1}{n}X_n^TX_n\delta_0 + o_p(1)$$

$$\frac{1}{n}(\tilde{X}_n^TV_n + (\rho_0 - \rho)\tilde{X}_n^TG_nV_n) = o_p(1).$$

Therefore

$$\hat{\delta}_n = \delta_0 + o_p(1).$$

B.2 Proof of Theorem 2

We start by noting the following two results:

$$\tilde{V}_n(\rho_0) = V_n - \eta_n\delta_0 \quad \text{and} \quad L_nY_n = L_n(S_n^{-1}X_n\delta_0 + S_n^{-1}V_n) = G_nX_n\delta_0 + G_nV_n.$$

The derivations here follow similar arguments as in [42] (See also the lecture notes by

[66]).

Convergence of $\frac{1}{n}\nabla^2_\theta l^*_n(\tilde{\theta}_n)$ to $\frac{1}{n}\nabla^2_\theta l^*_n(\theta_0)$.

Since $\tilde{\theta}_n$ is an intermediate value between $\hat{\theta}_n$ and θ_0, we have $\tilde{\theta}_n - \theta_0 = o_p(1)$. Now we show that as a consequence, each element of the matrix $\frac{1}{n}\nabla^2_\theta l^*_n(\tilde{\theta}_n) - \frac{1}{n}\nabla^2_\theta l^*_n(\theta_0)$ is $o_p(1)$.

We start with the following result:

$$\begin{aligned}\frac{1}{n}\nabla^2_{\delta,\delta} l^*_n(\tilde{\theta}_n) - \frac{1}{n}\nabla^2_{\delta,\delta} l^*_n(\theta_0) = (\frac{1}{\sigma_0^2} - \frac{1}{\tilde{\sigma}_n^2})[-\frac{\tilde{X}_n^T\tilde{X}_n}{n}] = o_p(1)(\frac{X_n^T X_n}{n} + o_p(1)) \\ = o_p(1)O(1) = o_p(1), \qquad \text{(B.2)}\end{aligned}$$

since $\frac{1}{\sigma_0^2} - \frac{1}{\tilde{\sigma}_n^2} = o_p(1)$ by continuous mapping theorem, and $\frac{1}{n}\tilde{X}_n^T\tilde{X}_n$ is bounded by assumption that elements of the matrix $\tilde{X}_n$ are uniformly bounded.

Next since G_n is uniformly bounded in row sum , we have $E[|(G_nV_n)_1(\tilde{X}_n)_1^T|] < \infty$. Then noting that $E[(G_nV_n)_1(\tilde{X}_n)_1^T] = 0$, leads to

$$\frac{1}{n}\tilde{X}_n^T G_n V_n \xrightarrow{p} 0.$$

Further we have seen before from WLLN that $\frac{1}{n}\tilde{X}_n^T G_n X_n \delta_0 \xrightarrow{p} \frac{1}{n}X_n^T G_n X_n \delta_0$. Therefore,

$$\frac{1}{n}\tilde{X}_n^T L_n Y_n = \frac{1}{n}\tilde{X}_n^T G_n X_n \delta_0 + \frac{1}{n}\tilde{X}_n^T G_n V_n = \frac{1}{n}X_n^T G_n X_n \delta_0 + o_p(1) + o_p(1) = O_p(1).$$

Consequently,

$$\frac{1}{n}\nabla^2_{\delta,\rho} l^*_n(\tilde{\theta}_n) - \frac{1}{n}\nabla^2_{\delta,\rho} l^*_n(\theta_0) = (\frac{1}{\sigma_0^2} - \frac{1}{\tilde{\sigma}_n^2})\frac{1}{n}(\tilde{X}_n^T L_n Y_n) = o_p(1)O_p(1) = o_p(1). \qquad \text{(B.3)}$$

Next we analyze $\frac{1}{n}\nabla^2_{\delta,\sigma^2} l^*_n(\tilde{\theta}_n) - \frac{1}{n}\nabla^2_{\delta,\sigma^2} l^*_n(\theta_0)$. We need the following intermediate results:

$$\begin{aligned} V_n(\tilde{\theta}_n) &= V_n(\theta_0) - (\tilde{\rho}_n - \rho_0)L_n Y_n - \tilde{X}_n(\tilde{\delta}_n - \delta_0) \\ \frac{1}{n}\tilde{X}_n^T V_n(\theta_0) &= \frac{1}{n}\tilde{X}_n^T V_n + \frac{1}{n}X_n^T\eta_n\delta_0 - \frac{1}{n}\eta_n^T\eta_n\delta_0 = o_p(1) + o_p(1) + (\Omega\delta_0 + o_p(1)) = O_p(1) \\ \frac{1}{n}\tilde{X}_n^T\tilde{X}_n &= \frac{1}{n}X_n^T X_n + \frac{1}{n}\sum_i \Omega_i + o_p(1) = O_p(1). \end{aligned}$$

Then

$$\frac{1}{n}\tilde{X}_n^T(\frac{V_n(\tilde{\theta}_n)}{\tilde{\sigma}^4} - \frac{V_n(\theta_0)}{\sigma_0^4}) = (\frac{1}{\tilde{\sigma}_n^4} - \frac{1}{\sigma_0^4})\frac{1}{n}\tilde{X}_n^T V_n(\theta_0) + \frac{1}{\tilde{\sigma}_n^2}\frac{1}{n}\tilde{X}_n^T\tilde{X}_n((\tilde{\delta}_n - \delta_0) + \frac{1}{\tilde{\sigma}_n^2}\frac{1}{n}\tilde{X}_n^T L_n Y_n(\tilde{\rho} - \rho_0)$$

$$= o_p(1)O_p(1) + O_p(1)O_p(1)o_p(1) + O_p(1)O_p(1)o_p(1) = o_p(1)$$

$$\frac{1}{n}\nabla^2_{\delta,\sigma^2} l^*_n(\tilde{\theta}_n) - \frac{1}{n}\nabla^2_{\delta,\sigma^2} l^*_n(\theta_0) = \frac{1}{n}(\sum_i \Omega_i)(\frac{\tilde{\delta}_n}{\tilde{\sigma}^4_n} - \frac{\delta_0}{\sigma^4_0}) - \frac{1}{n}\{\tilde{X}^T_n(\frac{V_n(\tilde{\theta}_n)}{\tilde{\sigma}^4} - \frac{V_n(\theta_0)}{\sigma^4_0})\}$$
$$= O_p(1)o_p(1) + o_p(1) = o_p(1). \tag{B.4}$$

For $\frac{1}{n}\nabla^2_{\sigma^2,\sigma^2} l^*_n(\tilde{\theta}_n) - \frac{1}{n}\nabla^2_{\sigma^2,\sigma^2} l^*_n(\theta_0)$ we note the following results:

$$\begin{aligned}\frac{1}{n}V_n(\tilde{\theta}_n)^T V_n(\tilde{\theta}_n) &= \frac{1}{n}V_n(\theta_0)^T V_n(\theta_0) + (\tilde{\delta}_n - \delta_0)^T\frac{\tilde{X}^T_n\tilde{X}_n}{n} + (\tilde{\rho} - \rho_0)^2\frac{(L_nY_n)^T(L_nY_n)}{n} \\ &\quad + 2(\tilde{\rho} - \rho_0)(\tilde{\delta}_n - \delta_0)^T\frac{\tilde{X}^T_n L_nY_n}{n} + 2(\tilde{\delta}_n - \delta_0)^T\frac{\tilde{X}^T_n V_n(\theta_0)}{n} + 2(\tilde{\rho} - \rho_0)\frac{(L_nY_n)^T V_n(\theta_0)}{n} \\ &= \frac{1}{n}V_n(\theta_0)^T V_n(\theta_0) + o_p(1).\end{aligned}$$

Note

$$\frac{1}{n}V_n(\theta_0)^T V_n(\theta_0) = \sigma^2_0 + \frac{1}{n}\sum_i \mathrm{tr}(\Omega_i)\delta^T_0\delta_0 + o_p(1) = O_p(1).$$

Therefore,

$$\begin{aligned}&\frac{1}{n}\nabla^2_{\sigma^2,\sigma^2} l^*_n(\tilde{\theta}_n) - \frac{1}{n}\nabla^2_{\sigma^2,\sigma^2} l^*_n(\theta_0) \\ &= \frac{1}{2}(\frac{1}{\tilde{\sigma}^4_n} - \frac{1}{\sigma^4_0}) - \frac{1}{2}(\frac{1}{\tilde{\sigma}^4_n} - \frac{1}{\sigma^4_0})\frac{1}{n}V_n(\theta_0)^T V_n(\theta_0) + (\tilde{\delta}^T_n\sum_i(\Omega_i)\tilde{\delta}_n - \delta^T_0\sum_i(\Omega_i)\delta_0) \\ &= o_p(1) + o_p(1)O_p(1) + o_p(1) = o_p(1).\end{aligned} \tag{B.5}$$

Next

$$\begin{aligned}\frac{1}{n}\nabla^2_{\sigma^2,\rho} l^*_n(\tilde{\theta}_n) - \frac{1}{n}\nabla^2_{\sigma^2,\rho} l^*_n(\theta_0) &= -\frac{1}{\tilde{\sigma}^4}\frac{1}{n}(Y^T_n L^T_n\tilde{X}_n(\tilde{\delta}_n - \delta_0) + \frac{1}{\tilde{\sigma}^4}\frac{1}{n}(Y^T_n L^T_n L_nY_n(\tilde{\rho}_n - \rho_0) \\ &\quad - (\frac{1}{\tilde{\sigma}^4} - \frac{1}{\sigma^4_0})\frac{1}{n}Y^T_n L^T_n V_n(\theta_0) \\ &= O_p(1)O_p(1)o_p(1) + O_p(1)O_p(1)o_p(1) + o_p(1)O_p(1) = o_p(1).\end{aligned} \tag{B.6}$$

For the final result from intermediate value theorem we have

$$\frac{1}{n}(tr(G_n(\tilde{\rho})^2) - tr(G_n(\rho_0)^2) = 2\frac{1}{n}tr(G_n(\bar{\rho})^3)(\tilde{\rho} - \rho_0) = \frac{1}{n}O(n/h_n)o_p(1) = o_p(1),$$

and $\frac{1}{n}(Y_nL_n)^T(Y_nL_n) = \frac{1}{n}O_p(n/h_n) = o_p(1)$. Then

$$\frac{1}{n}\nabla^2_{\rho,\rho}l^*_n(\tilde{\theta}_n) - \frac{1}{n}\nabla^2_{\rho,\rho}l^*_n(\theta_0) = o_p(1). \tag{B.7}$$

The **convergences of $\nabla^2_\theta l^*_n(\theta_0)$ to $E[\nabla^2_\theta l^*_n(\theta_0)]$** hold since all the elements of the unconditional expectation matrix are finite.

Limiting distribution of the scaled score vector The score vector, i.e., the gradient of the corrected log-likelihood function $\nabla_\theta l^*_n(\theta)$ evaluated at the true parameter vector θ_0 is as follows:

$$\begin{aligned}
\nabla_\delta l^*_n(\theta_0) &= \frac{1}{\sigma_0^2}[X_n^TV_n + \eta_n^TV_n - X_n^T\eta_n\delta_0 - \eta_n^T\eta_n\delta_0 + \Omega\delta_0],\\
\nabla_\rho l^*(\theta_0) &= \frac{1}{\sigma_0^2}[(G_nX_n\delta_0)^TV_n + V_n^TG_nV_n - (G_nX_n\delta_0)^T\eta_n\delta_0 - V_n^TG_n\eta_n\delta_0] - \mathrm{tr}(G_n)\\
\nabla_{\sigma^2} l^*(\theta_0) &= -\frac{n}{2\sigma_0^2} + \frac{1}{2\sigma_0^4}\{V_n^TV_n - V_n^T\eta_n\delta_0 + \delta_0^T\eta_n^T\eta_n\delta_0 - n\delta_0^T\Omega\delta_0\}.
\end{aligned}$$

We first verify that the unconditional expectation of the gradient vector is 0:

$$\begin{aligned}
E[\nabla_\delta l^*_n(\theta_0)] &= \frac{1}{\sigma_0^2}[-E[\eta_n^T\eta_n\delta_0 + \Omega\delta_0] = 0,\\
E[\nabla_\rho l^*(\theta_0)] &= \frac{1}{\sigma_0^2}[(G_nX_n\delta_0)^TE[V_n] + \sigma_0^2 tr(G_n) - (G_nX_n\delta_0)^TE[\eta_n]\delta_0 - E[V_n^TG_n\eta_n]\delta_0] - tr(G_n) = 0\\
E[\nabla_{\sigma^2} l^*(\theta_0)] &= -\frac{n}{2\sigma_0^2} + \frac{1}{2\sigma_0^4}\{n\sigma_0^2 - E[V_n^T\eta_n]\delta_0 + \delta_0^TE[\eta_n^T\eta_n]\delta_0 - \delta_0^T(\sum_i \Omega_i)\delta_0\} = 0,
\end{aligned}$$

The unconditonal variance is given by:

$$var[\frac{1}{\sqrt{n}}\nabla_\theta l^*_n(\theta_0)] = E[(\frac{1}{\sqrt{n}}\nabla_\theta l^*_n(\theta_0))(\frac{1}{\sqrt{n}}\nabla_\theta l^*_n(\theta_0))^T]$$

We note that the elements of the score vector consist of linear and quadratic forms on V_n and η_n. The random variables $(\eta_n)_i, (V_n)_i, (\eta_n)_i(V_n)_i$ are independent and their means are 0. The matrix for quadratic form G_n is uniformly bounded in row sums, the elements of the vector $G_nX_n\delta_0$ are uniformly bounded. Finally $\sup E|(\eta_n)_i|^{4+\epsilon} < \infty$, $\sup E|(V_n)_i|^{4+\epsilon} < \infty$ by assumptions. This also implies $\sup E|(\eta_n)_i(V_n)_i|^{4+\epsilon} < \infty$ Therefore the central limit theorem for linear and quadratic forms in [35] and [42] can be applied and we conclude

$$\sqrt{n}(\frac{1}{n}\nabla_\theta l^*_n(\theta_0)) \xrightarrow{p} N(0, \Sigma(\theta_0, X))$$

Appendix C Additional Simulations

The figure A1 shows the density plots of the estimates of the network influence parameter from the simulation study. As already noted in the simulation study section, the bias and homophily corrected estimator outperforms the other estimators in terms of bias.

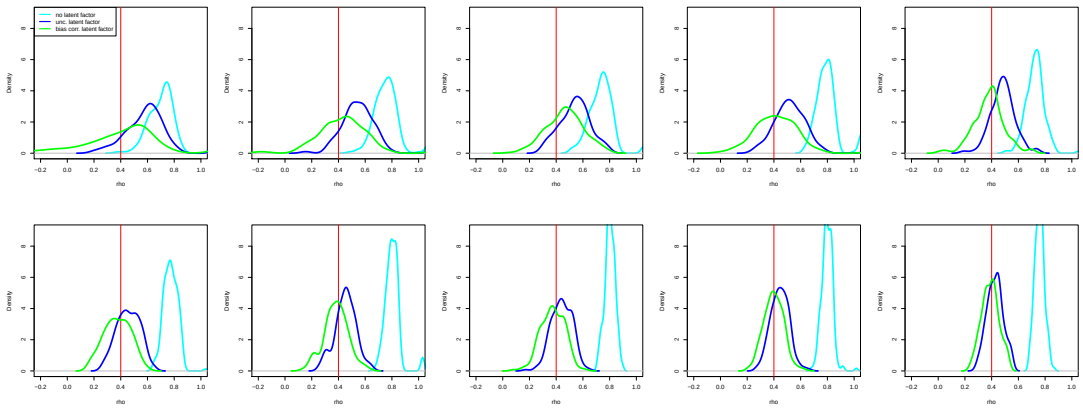
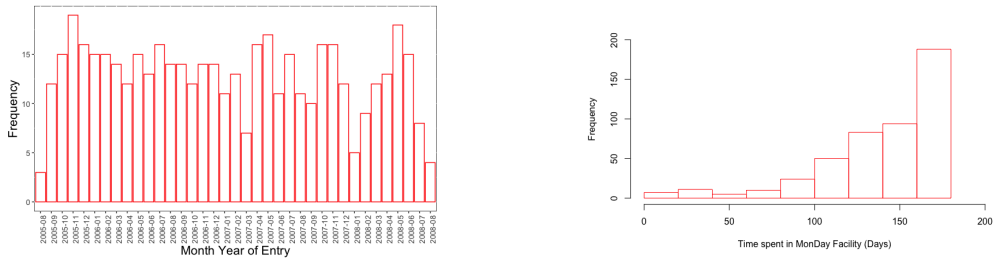


Figure A1: Comparison of estimates of the network influence parameter from the naive model, the model with latent factors but no bias correction, and the model with latent factors and bias correction.

Appendix D Additional results from data application

In this section, we provide some additional figures and tables on our data application. Figure A2 shows the the variation in entry dates and the amount of time spent in TC by the female residents. Table A1 provides a detailed summary of the outcome variables and the covariates used in our main results in section 5.2. Finally, table A2 suggests that that network degree significantly impacts graduation status and the likelihood of re-incarceration in the future.



(a)Month and Year of Entry (b) Time Spent in TC

Figure A2: Variation in entry dates and time spent at Facility

Table A1: Summary statistics: TC

Variable	Mean	Stdev	Min	Median	Max
Recidivism Dummy	0.462	0.499	0	0.0	1
Graduation Dummy	0.797	0.403	0	1.0	1
LSI	25.862	8.378	0	27.0	57
Age	30.358	8.203	18	28.5	60
White	0.799	0.401	0	1.0	1
Black	0.199	0.400	0	0.0	1
Hispanic	0.002	0.046	0	0.0	1
Time spent in TC (days)	142.305	37.450	4	150.0	179

Notes: The sample here consists of 472 residents who lived at the facility between 2005 and 2008 in the female unit. The selected sample consists of those residents in the facility for whom we could observe non-missing values for all the response and predictor variables. Also every individual in the selected sample is part of the receiver and sender set of affirmations.

Table A2: Regression of Affirmation Network Degree on Graduation and Recidivism

	Graduation Status	Recidivism Status
	(1)	(2)
Network Degree	0.006***	-0.003***
	(0.001)	(0.001)
White Dummy	0.609*	-0.336
	(0.327)	(0.244)
Age	-0.009	0.00004
	(0.018)	(0.012)
LSI	-0.223***	0.048***
	(0.030)	(0.013)
Intercept	6.237***	-0.419
	(1.104)	(0.591)
Observations	472	472
Log Likelihood	-159.709	-305.323
Akaike Inf. Crit.	329.418	620.645

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$. The standard errors are provided in the parenthesis. For the degree vector, we add the affirmations sent and received during their stay in the facility.