Network Peer Effects without Network Data: Identification and Estimation of Network Structure in the Presence of Latent, Unobserved Factors

Alan Griffith*
University of Washington

Sida Peng
Microsoft Research

October 11, 2021

Abstract

Identification and estimation of network-based peer effects is often limited by lack of data on network connections. In the absence of such data, we provide new identification results for a version of the linear-in-means model in which covariance in outcomes across agents is due to both network effects and latent, unobserved factors. Restrictions on network density and maximum degree are shown to be sufficient to separately identify both network structure and latent factors. When there is a single factor, network density can be as high as $\frac{1}{4}$, while maximum degree may be as large as $\frac{5}{12}$ of all agents. For estimation, we propose a proximal gradient descent algorithm that uses $L_{1,1}$ and $L_{1,2}$ norms. We perform a number of simulations, showing that the algorithm performs well in recovering network structure, with predictably better performance when more time periods are observed. In contrast to existing methods, our results allow for identification of network structure from covariance in outcomes, allow for latent factors, do not require exogeneity assumptions or even observation of time-varying covariates, and allow for identification of the absence of peer effects.

1 Introduction

Recent decades have witnessed a large amount of research investigating network-based peer effects.[1] This broad research agenda is due at least in part to the availability of datasets that contain rich network data, such as AddHealth (Harris, 2009) and the Diffusion of Microfinance dataset (Banerjee et al., 2012). Such datasets allow the researcher to observe *who* one's peers are, although

*Corresponding author: alangrif@uw.edu. We thank Jungyoun Kim for superb research assistance. All errors remain our own.

[1] Analysis of peer effects *through networks* is contrasted with the older literature on group-based peer effects, whereby an individual's "peers" is defined as all of those within some group, such as a classroom. A rich literature, some of it randomized, provides substantial causal evidence of classroom-based peer effects (see Epple and Romano, 2011; Sacerdote, 2011, for review of this large literature)

often noisily, and similar network data is often assumed to be a necessary input into any study of peer effects.[2]

Network data, however, is often unavailable or difficult to collect; in retrospective research, collecting network data may be impossible. Additionally, even when such data is available, it may be imperfect for a number of reasons, and noisily-measured network data may lead to predictable errors in estimation and inference.[3] Given these challenges, the implications of partially- or noisily-observed network data on estimation of model parameters is currently a very active research area (see Boucher and Houndetoungan, 2019; Chandrasekhar and Lewis, 2011; Griffith, 2021; Thirkettle, 2019).

In this paper, we propose a novel method to recover network structure in panel settings where network data is completely unavailable. Our model begins with a standard linear-in-means specification (see Manski, 1993), augmented with unobserved, individual-specific, low-dimensional latent "factors" (see Anderson and Rubin, 1956; Bai, 2009). For purposes of identification, we assume that the researcher observes the positive-definite $N \times N$ covariance matrix of outcomes across time. The main identification challenge lies in separating covariance due to the latent factors from covariance due to the unobserved network structure. Similar to Manresa (2016), our model has no "common parameters": the "parameter" of interest is the structure of the network itself, along with the latent factor structure.

Our main identification result pairs Turán's Theorem—a result in extremal graph theory that relates network density to existence of cliques—with insights from the matrix completion and matrix decomposition literature. We show that assumptions on the sparsity of the network—in terms of both average and maximum degree—are sufficient to recover the entire network structure, even in the presence of latent, unobserved factors. Further, these assumptions are not especially

[2] Typical analysis of network-based peer effects consist of the following steps. First, gather individual-level data on outcomes y_{it} and covariates x_{it}, where i indexes individuals and t indexes time. Pair this with data on social links. From this network data, construct a peer mean of one's peers' outcomes and covariates. The third step is to correlate one's own outcomes y_{it} with (1) (an average of) the outcomes of one's neighbors y_{jt}, where $j \neq i$; (2) one's own characteristics x_{it}, and (3) (an average of) the characteristics of one's neighbors x_{jt}. In the canonical case, this consists of estimating the parameters of a linear-in-means model a la Manski (1993), while noting that network data facilitates identification in many cases where group-based interactions may not (see, e.g. Blume et al., 2015; Bramoullé, Djebbari and Fortin, 2009; DeGiorgi, Pelllizzari and Radaelli, 2010). These analyses may or may not contain group-level fixed effects to control for correlated effects.

[3] As an example, network censoring—whereby individuals may list only a small number of friends—leads to attenuation in the reduced-form peer effects (see Griffith, 2021), as well as inconsistent estimates of other features of networks (Fosdick and Hoff, 2015; Hoff et al., 2013).

biting: when there is a single latent factor, we can allow for average density of approximately $\frac{1}{4}$ and maximum degree of approximately $\frac{5}{12}$ of all possible links. With partial knowledge of the network structure—such as when agents are known to be in isolated groups—these bounds can be further relaxed.

Given these identification results, we propose a procedure to jointly estimate the network and the latent factor structure. We propose an iterative Proximal Gradient Descent algorithm, an "accelerated" algorithm that iterates between gradient descent and shrinkage in estimation. Shrinkage is imposed via mixed $L_{1,1}$ and $L_{1,2}$ norms on the estimated network structure. Penalized estimation via these norms imposes soft-threshold versions of the average and maximum density assumptions from our identification results (see Chiong and Moon, 2018).

Through a series of simulation exercises, we show that this procedure performs well in recovering network structure, using networks collected by Banerjee et al. (2012). As expected, these results show that, for proper values of the penalization parameters, the algorithm performs well in recovering network structure. We desmonstrate estimation using single villages as well as multiple villages, where the latter employs covariance in outcomes across villages to aid in identifying the latent factor structure. We further test the algorithm's performance as we add more noise to the data, modeled by observing outcomes across fewer and fewer time periods. As expected, estimates become much less precise as the number of repeated observations decreases.

Our identification results have four key features that contrast them to methods existing in the literature. First, our results infer network structure solely from covariance in *outcomes*, without the need to observe time-varying covariates. This allows for potentially many more applications, since covariates available in data are often fixed demographic variables, and "panel" structure of data is often due to multiple outcomes observed at the same time.[4] This also dispenses the need to make strong assumptions about exogeneity of covariates, in sharp contrast to leading results in the literature such as de Paula, Rasul and Souza (2020) and Manresa (2016).[5] We note however, that these additional features come at a cost: in contrast to de Paula, Rasul and Souza (2020), we cannot identify the structure of dense networks.

[4] As an example of the latter, in the Tennessee STAR data, multiple academic outcomes for each student are observed in each year in the study. In AddHealth, multiple academic and behavioral outcomes per survey wave are observed but only one set of baseline covariates.

[5] While they do not seek to identify network structure itself, Lewbel, Qu and Tang (2021*b*) require a similar exogeneity assumption.

Second, modeling a latent factor structure allows for richer individual-specific heterogeneity that may not be available in observed data. Third, our method allows for identification of the *absence* of peer effects. This is often treated as a knife-edge case of non-identification (see, e.g. Bramoullé, Djebbari and Fortin, 2009; de Paula, Rasul and Souza, 2020).[6] Our results allow us to identify situations in which there is no covariance in outcomes, and we need not make any assumptions about combinations of parameters.

Fourth, our method generalizes the graphical lasso estimator proposed in Battaglini et al. (2021). While Battaglini et al. (2021) only allows for observed common shocks, we consider the case of a low dimensional unobserved common shock structure. This extension leads to a more challenging identification problem and to the best of our knowledge, it is not a structure that nested in any of the existing literature.

Our results here are most closely related to Manresa (2016), de Paula, Rasul and Souza (2020) and Battaglini et al. (2021), which we have contrasted already.[7] Our method is also related to a set of works that seek to estimate (common) peer effects parameters from either no or limited data. Lewbel, Qu and Tang (2021*b*) achieve identification of common parameters by observing the distribution of correlations of outcomes and covariates. Boucher and Houndetoungan (2019) achieve identification under the assumption that a consistent estimate of the distribution of the network is available. Relatedly, Griffith (2021) and Lewbel, Qu and Tang (2021*a*) analyze consistency/inconsistency of peer effects estimators when network data is partially but not fully observed.[8]

In seeking to infer network structure, our methods are also related to the broader literature that seeks to infer network structure from other, more readily-available data, such as those that use Aggregate Relational Data (see Breza et al., 2020; Alidaee, Auerbach and Leung, 2020). Further, our use of concepts from matrix decomposition and matrix completion share insights from leading computer science results.[9] Our estimation via penalized mixed norms employs an insight from

[6] That is, a stated conditions of Propositions 1-5 of Bramoullé, Djebbari and Fortin (2009) and Assumption (A3) of de Paula, Rasul and Souza (2020) require that the peer effects do not cancel out. Our results, in contrast, allow for identification of the absence of peer effects

[7] Miraldo, Propper and Rose (2021) is another example of using panel structure to identify peer effects, but peer effects in their context have a group rather than network structure, and identification is achieved by observing agents moving across groups over time.

[8] Chandrasekhar and Lewis (2011) and Breza et al. (2020) study related issues in estimating diffusion models rather than peer effects.

[9] Our decomposition lemma (Lemma 1) is related to notions of "rank-sparsity incoherence" as defined in, e.g., Chan-

Chiong and Moon (2018), which advocates use of the $L_{1,2}$ norm as a means of imposing maximum degree in graphical models. Identification via second moments and variance restrictions relates our work to a set of results that infer peer effects via variance restrictions including theoretical (Graham, 2008; Rose, 2017) and applied (Lyle, 2009) works.

This paper is organized as follows. Section 2 introduces the model. Our main identification results are in Section 3, with extensions and a special case in Section 4. We introduce our estimation algorithm and discuss its relation to our identification results in Section 5. We present simlated estimation results in Section 6, with sensitivity discussed in Section 7. Section 8 concludes.

2 Model Set-up

2.1 Base

Our model is set up as Equation (1), is a simple generalization of the standard linear-in-means model (Manski, 1993) that allows for heterogeneity among the peer effects.

$$y_{it} = \sum_{j \neq i} \gamma_{ij} y_{jt} + \delta x_{it} + \sum_{j \neq i} \gamma_{ij} x_{jt} + \epsilon_{it} \tag{1}$$

Individuals are indexed by i and "time" is indexed by t. y_{it} is some outcome for individual i at time t, while x_{it} is some covariate for this same agent. In the taxonomy of Manski (1993), the term $\sum_{j \neq i} \gamma_{ij} y_{jt}$ identifies the *endogenous peer effect*, $\sum_{j \neq i} \gamma_{ij} x_{jt}$ the *exogenous peer effect*, and correlation among ϵ_{it} identifies *correlated effects*.

To simplify notation, assume that $\mathbf{x_t} = 0$ uniformly.[10] Rewrite Equation (1) in matrix form, yielding Equation (2).

$$\mathbf{y_t} = \Gamma \mathbf{y_t} + \epsilon_t \tag{2}$$

In this equation, Γ is an adjacency matrix, or "sociomatrix" in the terminology of Blume et al. (2015).

drasekaran et al. (2011). The matrix completion strategy is related to results in, for example, Candès and Recht (2009), except our result requires deterministic rather than high-probability completability.

[10] This assumption may seem strong. However, if the analysis is performed after differencing out individual-level means in a panel setting, then this is equivalent to assuming that x_{it} is fixed across time for each agent i. In other words, x_{it} being a time-invariant characteristic is sufficient for this assumption.

Assumption 1. The following conditions are imposed:

[1] $\Gamma_{(i,i)} = 0 \; \forall \; i$

[2] $\rho(\Gamma) < 1$, where $\rho()$ is the spectral radius of Γ

[3] $\mathbb{E}[\epsilon_t \epsilon_t']$ is bounded and positive definite

In all versions of the model, we restrict the DGP by the restrictions in Assumption 1. Part [1] is standard in the literature and simply rules out self-reflection. Part [2] is a stability condition, versions of which are standard in the literature. We need not impose any row-sum normalization, in contrast to, for example, Bramoullé, Djebbari and Fortin (2009) and de Paula, Rasul and Souza (2020).[11] Taken together, [2] and [3] imply that $\mathbf{y_t}$ can be rewritten as Equation (3).

$$\mathbf{y_t} = (\mathbf{I} - \Gamma)^{-1} \epsilon_t \tag{3}$$

2.2 Latent Factor Structure

While the linear-in-means model has been used extensively by applied researchers to study peer effects, common shocks remain a concern for those studies. For example, even in the case when individual fixed effects are differenced out in equation (1), a latent variable may still generate correlation across ϵ_{it} (for any given t) and thus introduce bias in the peer effect estimator. In Manski (1993), such dependence is explicitly modeled by assuming the dependence structure is the same as the observed network. In de Paula, Rasul and Souza (2020), common shocks are modeled as a common time-varying scalar that can be differenced out from the local average outcomes of neighbors.

We take a different approach. In order to allow for covariance among unobserved ϵ_{it}, we model a latent factor structure. That is, $\epsilon_{it} = \eta_i f_t + u_{it}$, where η_i is $1 \times R$, f_t is $R \times 1$. In matrix notation,

$$\epsilon_t = \eta f_t + \mathbf{u_t} \tag{4}$$

Formally, the latent factor structure is defined in Assumption 2. These assumptions are standard in the literature going back as far as Anderson and Rubin (1956), amounting to a normalization of f_t

[11] In our model, a row-sum normalization would impose that $\sum_{j \neq i} \gamma_{ij} = k$ for some $k \in (0, 1)$.

(and in turn η_i). We maintain Assumption 2 throughout.[12] Condition [5] is simply a normalization that sets the sign of each η_{ir}.

Assumption 2. For each t, $\epsilon_t = \eta f_t + \mathbf{u_t}$, where

[1] $\eta'\eta$ is diagonal with decreasing entries along the diagonal

[2] $\mathbb{E}[\eta \mathbf{u_t}'] = 0$ for all t

[3] $\mathbb{E}[f_t f_t' | \mathbf{u_t}] = \mathbf{I}_R$

[4] $\mathbb{E}[\mathbf{u_t}\mathbf{u_t}']$ is diagonal

[5] $\eta_{11} \geq 0$

2.3 What is Observed

As stated in Assumption 3, we assume that the covariance of outcomes $\mathbf{y_t}$ across all agents $i = 1, ..., N$, is observed. Further, this covariance matrix $\mathbb{E}[\mathbf{y_t}\mathbf{y_t}']$ is positive definite, and thus can be inverted.

Assumption 3. $\mathbb{E}[\mathbf{y_t}\mathbf{y_t}']$ is observed and is positive definite.

Assumptions 1 - 3 guarantee observation of both sides of Equation (5).

$$\mathbb{E}[\mathbf{y_t}\mathbf{y_t}'] = (\mathbf{I} - \Gamma)^{-1}(\eta\eta' + \mathbf{U})(\mathbf{I} - \Gamma')^{-1} \tag{5}$$

Due to Assumption 3, both sides of Equation (5) are positive definite. Therefore, the precision matrix $\mathbf{\Sigma_y} = (\mathbb{E}[\mathbf{y_t}\mathbf{y_t}'])^{-1}$ exists and is observed/identified. It is written in Equation (6).

$$\mathbf{\Sigma_y} = (\mathbf{I} - \Gamma')(\eta\eta' + \mathbf{U})^{-1}(\mathbf{I} - \Gamma) \tag{6}$$

The identification problem thus lies in separating Γ, η, and $\mathbf{U}$ in the precision matrix. The Woodburery Inverse Formula allows us to restate Equation (6) as the sum of two other matrices as shown in Equation (7).

$$\mathbf{\Sigma_y} = (\mathbf{I} - \Gamma')\mathbf{U}^{-1}(\mathbf{I} - \Gamma) - (\mathbf{I} - \Gamma')\mathbf{U}^{-1}\eta(\mathbf{I}_R + \eta\mathbf{U}^{-1}\eta')^{-1}\eta'\mathbf{U}^{-1}(\mathbf{I} - \Gamma) \tag{7}$$

[12] Special cases that assume away the factor structure effectively add the assumption $\eta = 0$.

2.4 Unovserved Factors: A Simple Example

To highlight the importance of accounting for unobserved factors, suppose that we want to infer network structure through a penalized procedure as is standard (see Manresa, 2016; de Paula, Rasul and Souza, 2020). In the absence of any network effects ($\Gamma = 0$) and assuming $R = 1$ and $\mathbf{U} = \sigma_u \mathbf{I}$, the precision matrix is given in Equation (8).

$$\mathbf{\Sigma_y} = \frac{1}{\sigma_u}\mathbf{I} - \frac{1}{\sigma_u(\sigma_u + \eta'\eta)}\eta\eta' \tag{8}$$

This implies that, for all $i, j \neq i$ (off-diagonal terms), element (i, j) of $\mathbf{\Sigma_y}$ is given by $-\frac{1}{\sigma_u(\sigma_u+\eta'\eta)}\eta_i\eta_j$. If we naively inferred network links from $\mathbf{\Sigma_y}$ without considering the factor structure, we would infer "links" wherever $\eta_i\eta_j$ is large in magnitude. In other words, we would infer links as existing between agents holding the largest latent factors. Additionally, we would inaccurately infer high "degree" for agents who have high η_i in magnitude.

3 Identification

Here, we derive our main identification result. First, we define terms, including what we mean by identification as well as terms from graph theory.

3.1 Definitions

3.1.1 Identification

As stated in Assumption 3, in all cases we assume that the precision/covariance matrix is observed. By "identification," we mean a unique mapping from these moments to parameters of the model. This is the notion of (global) identification as defined in Rothenberg (1971), among others.[13] That is, given a set of observed moments Ψ, a set of parameters θ is identified if there is a unique mapping from Ψ to θ, given the assumed data-generating process.

Definition 1. A parameter θ is identified if, for all possible observed moments $\Psi(\theta)$, there exists a unique "to one" mapping from $\Psi(\theta)$ to θ.

[13] For a fuller discussion of this definitian its relation to other notions of point identification, see Lewbel (2019).

Identification is defined in Definition 1. In our setting, Assumption 3 states that we observe the positive-definite covariance matrix $\mathbb{E}[\mathbf{y_t}\mathbf{y_t'}]$ and therefore its inverse $\mathbf{\Sigma_y}$, the precision matrix as given in Equation (7). Accordingly, we are interested in conditions under which there is a mapping from each possible $\mathbf{\Sigma_y}$ to unique parameters (Γ, η).

We further define a notion of generic identification in Definition 2. Many of our results hold generically but not globally, since in knife-edge cases combinations of parameters may preclude identification.[14]

Definition 2. A statement holds generically for some set $\mathbf{B}$ if the set $\mathbf{B}' \subset \mathbf{B}$ in which it does not hold has Lebesgue measure zero.

3.1.2 Network/Graph Theory Definitions

By assumption, the observed "reduced-form" parameter $\mathbf{\Sigma_y}$ is a symmetric, positive-definite matrix. Accordingly, it encodes (at most) $\frac{N(N+1)}{2}$ restrictions. Therefore, simply counting the number of parameters suggests that we cannot identify an arbitrarily dense Γ, which might include as many as $N(N-1)$ nonzero entries. Further, η contains an additional $N \times R$ parameters that must be identified. Therefore, without additional restrictions, the model contains $N(N+R-1) >> \frac{N(N+1)}{2}$ parameters.

To set up the discussion of identification conditions, we define some notions from graph theory. These definitions are standard in the graph theory literature (see, e.g. Diestel, 2005; Jackson, 2008). First, we note that, in our terminology, a network is completely characterized by its adjacency matrix $\mathbf{G}$, which may be weighted and/or directed. One such $\mathbf{G}$ of interest is the Γ that we seek to identify. By convention, $\mathbf{G}_{(i,i)} = 0$ for all i (there are no self-links).

Definition 3. Given an adjacency matrix $\mathbf{G}$,

[1] $V(\mathbf{G})$ is the vertex set of $\mathbf{G}$

[2] $E(\mathbf{G})$ is the edge set of $\mathbf{G}$. That is, $(i,j) \in E(\mathbf{G})$ if $\mathbf{G}_{(i,j)} \neq 0$, with its complement defined as $\overline{E(\mathbf{G})}$.

Defintion 3 defines vertex set and edge sets. For any graph/network $\mathbf{G}$, the vertex set is the set

[14] Generic identification is necessary but not sufficient for global identification. This is the same definition of generic identification given in Lewbel (2019) Section 7.2.

of all agents, while the edge set is the set of all links that exist. Since links may be weighted and directed, the edge set is the set of nonzero links or, equivalently, the set of pairs (i, j) for which $\mathbf{G}_{(i,j)} \neq 0$.

Definition 4. For an adjacency matrix $\mathbf{G}$,

[1] $|V(\mathbf{G})|$ is the order of $\mathbf{G}$ (the number of agents)

[2] $\|E(\mathbf{G})\|$ is the size of $\mathbf{G}$ (the number of nonzero elements in $\mathbf{G}$)

[3] $\frac{\|E(\mathbf{G})\|}{|V(\mathbf{G})|(|V(\mathbf{G})|-1)}$ is the density of $\mathbf{G}$

Next, Definition 4 defines order, size, and density. The order of $\mathbf{G}$ is the number of agents, while the size is the number of links, where a "link" is defined to exist whenever $\mathbf{G}_{(i,j)} \neq 0$. The network's density is the number of links (the size) compared to the number of links that could exist, which is $|V(\mathbf{G})|(|V(\mathbf{G})| - 1)$.

Definition 5. For a network $\mathbf{G}$ and $V' \subset V(\mathbf{G})$, $\mathbf{G}'$ is the $|V'| \times |V'|$ induced subnetwork on V'. That is, $\mathbf{G}'$ contains all (and only) links in $\mathbf{G}$ among agents in V'.

Our identification results impose restrictions on density of networks and subnetworks. Accordingly, Definition 5 defines an induced subnetwork, which consists of a subset of agents and the links among those agents. That is, an induced subnetwork is completely characterized by $\mathbf{G}$ restricted to rows and columns corresponding to agents in $V' \subset V(\mathbf{G})$. The order, size, and density of $\mathbf{G}'$ are defined analogously to those of the full network in Definition 4.

Finally, we define degree for each agent in $V(\mathbf{G})$ as well as maximum and minimum degree. For each agent i, degree is simply the number of nonzero elements in row i of the adjacency matrix. $\delta(\mathbf{G})$ and $\Delta(\mathbf{G})$ are the minimum and maximum, respectively, across all agents i.

Definition 6. For a network $\mathbf{G}$,

[1] For any agent $i \in V(\mathbf{G})$, $d_i(\mathbf{G}) = \|\mathbf{G_i}\|$ is the degree of i

[2] $\delta(\mathbf{G}) = \min_i d_i(\mathbf{G})$ is the minimum degree of $\mathbf{G}$

[3] $\Delta(\mathbf{G}) = \max_i d_i(\mathbf{G})$ is the maximum degree of $\mathbf{G}$

3.2 Matrix Decomposition Lemma

To simplify notation and clarify the identification problem, we first define $\mathbf{Z}$, a $N \times R$ matrix, in Equation (9). To start,

$$\begin{aligned} \mathbf{Z}\mathbf{Z}' &= (\mathbf{I} - \Gamma')\mathbf{U}^{-1}\eta(\mathbf{I}_R + \eta'\mathbf{U}^{-1}\eta)^{-1}\eta'\mathbf{U}^{-1}(\mathbf{I} - \Gamma), \\ &\text{where } \mathbf{Z} = (\mathbf{I} - \Gamma')\mathbf{U}^{-1}\eta(\mathbf{I}_R + \eta'\mathbf{U}^{-1}\eta)^{-\frac{1}{2}} \end{aligned} \tag{9}$$

Since $(\mathbf{I}_R + \eta\mathbf{U}^{-1}\eta')^{-1}$ is a symmetric, positive-definite matrix, $(\mathbf{I}_R + \eta\mathbf{U}^{-1}\eta')^{-\frac{1}{2}}$ is guaranteed to exist. Therefore, $\mathbf{Z}$ exists, in terms of Γ, η, and $\mathbf{U}$.[15] With $\mathbf{Z}\mathbf{Z}'$ thus defined, we rewrite Equation (7) as in Equation (10).

$$\mathbf{\Sigma_y} = (\mathbf{I} - \Gamma')\mathbf{U}^{-1}(\mathbf{I} - \Gamma) - \mathbf{Z}\mathbf{Z}' \tag{10}$$

Since $\mathbf{\Sigma_y}$ is assumed to be observed/identified, identification of $\mathbf{Z}\mathbf{Z}'$ requires separating the two terms on the right-hand side of Equation (10). Separating these two matrices is fundamentally a matrix decomposition problem.

Lemma 1. Given Assumptions 1-3, identification of $\mathbf{Z}\mathbf{Z}' \Rightarrow$ identification of Γ, $\mathbf{U}$, and η.

Proof. See Appendix. □

Clearly, identification of $\mathbf{Z}\mathbf{Z}'$ immediately implies identification of $(\mathbf{I} - \Gamma')\mathbf{U}^{-1}(\mathbf{I} - \Gamma)$. Lemma 1 gives a stronger result, however: identification of $\mathbf{Z}\mathbf{Z}'$ is sufficient for identification of *all parameters* of the model, not just the composite parameter $(\mathbf{I} - \Gamma')\mathbf{U}^{-1}(\mathbf{I} - \Gamma)$. The intuition for Lemma 1 is straightforward: since $\mathbf{U}^{-1}$ is diagonal and all diagonal elements of Γ are zero, $\mathbf{U}$ is identified by the diagonal elements of $(\mathbf{I} - \Gamma')\mathbf{U}^{-\frac{1}{2}}$. Once $\mathbf{U}$ is known, we we can deduce Γ from $(\mathbf{I} - \Gamma')\mathbf{U}^{-\frac{1}{2}}$. Once $\mathbf{U}$ and Γ are known, η can be recovered from $\mathbf{Z}$.

Lemma 1 states that identification of $\mathbf{Z}\mathbf{Z}'$ is sufficient to identify all parameters of the model. In turn, our main results give sufficient conditions for the logically prior step of identifying $\mathbf{Z}\mathbf{Z}'$ which, together with Lemma 1 state conditions for identification of all parameters in Theorem 1.

[15] Defining $\mathbf{M} = (\mathbf{I}_R + \eta\mathbf{U}^{-1}\eta')^{-\frac{1}{2}}$, $\mathbf{M}_{(r,r)} = (1 + \sum_{i=1}^{N} \frac{\eta_{ir}^2}{u_i^2})^{-\frac{1}{2}}$ for all $r = 1, ..., R$.

Our strategy for decomposing the precision matrix $\mathbf{\Sigma}_{\mathbf{y_t}}$ into its component parts as shown in Equation (10) is related to the notion of "rank-sparsity incoherence" found in the computer science literature (see Agarwal, Negahban and Wainwright, 2012; Chandrasekaran et al., 2011; Hsu, Kakade and Zhang, 2011). Clearly, a single matrix cannot be uniquely decomposed into two additive matrices without additional conditions. The basic insight is that if one of the component parts is sufficiently sparse while the other is of low rank, then there can be a unique decomposition. In our context, sparsity in the network—of a particular type—implies sparsity of the first term $(\mathbf{I} - \Gamma')\mathbf{U}^{-1}(\mathbf{I} - \Gamma)$, while $\mathbf{ZZ}'$ has rank R at most, where R is the number of factors in the factor structure η.

3.3 Identification as Matrix Completion

3.3.1 Matrix Completion Set-up

Lemma 1 states the identification problem as one of matrix *decomposition*. Here, we show that this matrix decomposition problem is in essence an exercise in matrix *completion*, where the sparsity structure of the network provides for partial observability of $\mathbf{ZZ}'$.

To fix ideas, consider $\mathbf{ZZ}'^{*}$, a partially-observed version of $\mathbf{ZZ}'$. For example, we might observe

$$\mathbf{ZZ}'^{*} = \begin{bmatrix} * & z_1'z_2 & z_1'z_3 & \dots \\ z_1'z_2 & * & * & \dots \\ z_1'z_3 & * & \dots & \\ \dots & & & \end{bmatrix} \tag{11}$$

where $*$ indicates missing entries. In the jargon of, for example, Király and Tomioka (2012), there exists some binary matrix $\mathbf{M}$ that defines a "mask" on $\mathbf{ZZ}'$. In this example, $\mathbf{M}_{(1,2)} = \mathbf{M}_{(1,3)} = 1$ while $\mathbf{M}_{(2,3)} = 0$ as are all diagonal elements of $\mathbf{M}$.

In this context, identification of $\mathbf{ZZ}'$ requires conditions to ensure observation of a sufficient number and placement of unmasked (observed) elements in $\mathbf{ZZ}'^{*}$. Given identification of $\mathbf{ZZ}'_{(i,j)}$ for some (i, j), identification of the rest of the entries of $\mathbf{ZZ}'$ is a *maxtrix completion* problem.

There is a large literature in computer science that deals with the problem of low-rank matrix completion. Early examples are Candès and Recht (2009) and Candès and Tao (2010). However,

these results are not directly applicable since they generally formulate the problem in terms of *random* missing entries, and generally derive results for unique completability "with high probaility." In contrast, here we are concerned with non-random missingness patterns, and identification requires deterministic unique completability. To this end, a few papers have investigated the problem of deterministic matrix completion (see, e.g. Bhojanapalli and Jain, 2014; Bishop and Yu, 2014; Király and Tomioka, 2012; Singer and Cucuringu, 2010). Our results share insights from these works, but none are directly on point for the problem at hand.

3.3.2 Constructive Results

Before proceeding to our main result, we state a set of lemmas. These results allow us to fill in certain missing elements in $\mathbf{ZZ}'^{*}$, which in turn allows for iterative construction of $\mathbf{ZZ}'$ from observed subsets of entries, corresponding to induced subgraphs of the adjacency matrix.

Lemma 2. If $\mathbf{M}$ is a matrix of at most rank R, then for any $R+1 \times R+1$ sub-matrix, identification of all but one elements implies identification of the final element, generically.

Proof. See Appendix. □

First, Lemma 2 says that, for any matrix that has rank less than $R+1$, we can infer the final entry of any all-but-one-element identified matrix of size $R+1$. This result is similar to that given in Proposition 2.12 of Király and Tomioka (2012). When $R=1$, this implies that for any sub-matrix as in Expression (12), identification of any three elements is sufficient to identify the fourth.

$$\begin{bmatrix} \mathbf{ZZ}'_{(i,j)} & \mathbf{ZZ}'_{(i,k)} \\ \mathbf{ZZ}'_{(l,j)} & * \end{bmatrix} \tag{12}$$

Identification of the missing elements follows from the fact that this matrix must be rank-deficient, and thus its determinant must be zero. Therefore, $\mathbf{ZZ}'_{(l,k)} = \frac{\mathbf{ZZ}'_{(i,k)}\mathbf{ZZ}'_{(l,j)}}{\mathbf{ZZ}'_{(i,j)}}$, noting that the genericity condition ensures that $\mathbf{ZZ}'_{(i,j)} \neq 0$ except in a knife-edge case. When $R=2$, Lemma 2 provides identification of single missing elements in any 3×3 sub-matrix, and analogously for larger R.

Lemma 3. If there exist $V_1', V_2' \subset V$ such that

[1] $|V_1'|, |V_2'| \geq R$

[2] $V_1' \cap V_2' = \emptyset$

[3] $\mathbf{ZZ'}_{(i,j)}$ is identifed for all $i, j \in V_1'$, $\mathbf{ZZ'}_{(k,l)}$ is identifed for all $k, l \in V_2'$

then $\mathbf{ZZ'}_{(i,k)}$ is identified for all $i, k \in V_1' \cup V_2'$.

Next, Lemma 3 provides for filling in off-diagonal blocks in $\mathbf{ZZ'}$ when sufficiently large blocks along the diagonal have been identified.That is, suppose that $\mathbf{ZZ'}_{11}$ and $\mathbf{ZZ'}_{22}$ are identified in the following matrix, and each is square and at least $R \times R$.

$$\begin{bmatrix} \mathbf{ZZ'}_{11} & \mathbf{ZZ'}_{12} \\ \mathbf{ZZ''}_{12} & \mathbf{ZZ'}_{22} \end{bmatrix} \tag{13}$$

Lemma 3 implies that $\mathbf{ZZ'}_{12}$ (and also $\mathbf{ZZ''}_{12}$) are identified. When $R = 1$, then this provides that identifcation of elements $\mathbf{ZZ'}_{(i,i)}$ and $\mathbf{ZZ'}_{(j,j)}$ are sufficient to identify $\mathbf{ZZ'}_{(i,j)}$ (and $\mathbf{ZZ'}_{(j,i)}$). More generally, Lemma 3 says that identification of disjoint principal sub-matrices immediately implies identification of the off-diagonal blocks. Accordingly, Lemma 3 states that identification of sufficiently many principal sub-matrices, each of sufficient size, can identify the entire matrix.[16]

3.4 Conditions on Observability

Lemmas 2 and 3 allow for adding to known elements of partially-observed $\mathbf{ZZ'}^*$, but these results require that at least some elements are already observed. Thus, their usefulness requires conditions under which some entries of $\mathbf{ZZ'}$ are observed. Here, we state conditions that relate sparsity in the network $E(\mathbf{G}) \cup E(\mathbf{G'}) \cup E(\mathbf{G'G})$ to observability of entries of $\mathbf{ZZ'}$.

Lemma 4. For any adjacency matrix $\mathbf{G}$ and diagonal $\mathbf{D}$, for any $j \neq i$, $(i,j) \notin E(\mathbf{G}) \cup E(\mathbf{G'}) \cup E(\mathbf{G'G}) \Rightarrow$ element (i,j) of $(\mathbf{I} - \mathbf{G'})\mathbf{D}(\mathbf{I} - \mathbf{G})$ is zero.

Proof. See Appendix. □

First, Lemma 4 relates observability of elements of $\mathbf{ZZ'}$ to the existence or absence of network links. We note that, in the linear-in-means framework of, e.g., Manski (1993), $\mathbf{G}_{(i,j)} \neq 0$ implies

[16] This result is analogous to the strategy in Bishop and Yu (2014).

that agent j influences agent i. Therefore, $E(\mathbf{G}'\mathbf{G})$ defines the set of common influences: $(i,j) \in E(\mathbf{G}'\mathbf{G})$ whenever i and j influence a common third party k.[17] Therefore, Lemma 4 says that $(\mathbf{I}-\mathbf{G}')\mathbf{U}^{-1}(\mathbf{I}-\mathbf{G})$ is zero whenever i and j neither directly influence each other nor commonly influence a third agent.

The importance of Lemma 4 is that, whenever (i,j) is not in the union of the networks in Lemma 4, then element (i,j) of $(\mathbf{I}-\mathbf{G}')\mathbf{U}^{-1}(\mathbf{I}-\mathbf{G})$ is zero. In turn, via Equation (10), element (i,j) of $\mathbf{Z}\mathbf{Z}'$ is observed. That is, Lemma 4 gives a sufficient condition for observation of individual elements of $\mathbf{Z}\mathbf{Z}'$. Together with sparsity conditions, Lemma 4 guarantees a number of elements of $\mathbf{Z}\mathbf{Z}'$ are observed, without identifying which particular elements are known.

Lemma 5. For any graph $\mathbf{G_1}$ and associated vertex set V_1, if $\|E(\mathbf{G_1}) \cup E(\mathbf{G_1}') \cup E(\mathbf{G_1}'\mathbf{G_1})\| < \frac{|V_1|}{2}(\frac{|V_1|}{R}-2)$, then there exists $V_1' \subset V_1$ such that

[1] $|V_1'| \geq 2R+1$

[2] $\mathbf{Z}\mathbf{Z}'_{(i,j)}$ is identified for all $i,j \in V_1'$

[3] Each agent in $\overline{V_1'}$ has at most $R-1$ links to agents in V_1' in the complement of the edge set edge set $E(\mathbf{G_1}) \cup E(\mathbf{G_1}') \cup E(\mathbf{G_1}'\mathbf{G_1})$, (defined as $\overline{E(\mathbf{G_1}) \cup E(\mathbf{G_1}') \cup E(\mathbf{G_1}'\mathbf{G_1})}$)

Proof. See Appendix. □

Second, we use a well-known result from extremal graph theory to give conditions under which *cliques*—defined as maximally-connected induced subgraphs—must exist in a graph. Lemma 5 employs Turán's Theorem, which relates the density of any graph (or subgraph) to the existence of cliques of a certain size (see, e.g., Diestel, 2005, Ch. 7.1), where a clique is defined as a complete induced sub-network. In particular, Lemma 5 uses Turán's Theorem to give sufficient conditions for the existence of cliques of at least size $2R+1$, where R is the number of latent factors in the model.

This result is extremely powerful in our setting. For all agents in the clique of size $2R+1$, it immediately implies identification of the off-diagonal elements of $\mathbf{Z}\mathbf{Z}'$ for all agents in the induced subgraph. In turn, application of Lemma 2 then provides for identification of the missing diagonal elements. Therefore, given sufficient density, we identify all elements in the block.

[17] Note that this is a one-way implication: no common influence implies element (i,j) of $(\mathbf{I}-\mathbf{G}')\mathbf{D}(\mathbf{I}-\mathbf{G})$ is zero, but the reverse may not hold if, for example, multiple common influences exactly cancel each other.

3.5 Main Result

Theorem 1 gives our main identification result, of which Lemmas 1 - 5 are intermediate steps. Density conditions together with Lemmas 4 and 5 provide that some elements of $\mathbf{ZZ}'$ are necessarily identified. From this starting point, Lemmas 2 and 3 allow iterative construction of a unique, low-rank $\mathbf{ZZ}'$. Finally, Lemma 1 states that this unique $\mathbf{ZZ}'$ implies identification of all other model parameters, including both the network structure Γ and the latent factors η.

To simplify notation somewhat, before stating Theorem 1, define $m = 1 - \frac{1}{2R}$, and note that m is strictly increasing in R. Theorem 1 says that conditions on the average degree (network density) and maximum degree of the network $E(\Gamma) \cup E(\Gamma') \cup E(\Gamma'\Gamma)$ are sufficient for identification of all parameters of the model. That is, as long as average density and maximum degree of the union of the direct and common-influence networks are not too large, there exists a unique mapping from the precision matrix $\Sigma_{\mathbf{y_t}}$, which is assumed to be observed, to the underlying graph and factor structure.

Theorem 1. Given Assumptions 1 - 3, if the following conditions hold

[1] $|E(\Gamma) \cup E(\Gamma') \cup E(\Gamma'\Gamma)| < \frac{N(\frac{N}{R}-2)}{4}$ (density)

[2] $\Delta(E(\Gamma) \cup E(\Gamma') \cup E(\Gamma'\Gamma) \leq \frac{N(1-m)+P_1(R)m-R}{2}$ (max degree),

(where $P_1(R) = \frac{1+2mN+2(R-1)+\left((1+2mN+2(R-1))^2-8N(m+1)(R-1)\right)^{\frac{1}{2}}}{2(1+m)}$)

then $(\Gamma, \eta, \mathbf{U})$ is identified.

Proof. See Appendix. □

The full proof of Theorem 1 is technical and in the appendix, but the intuition is as follows. First, assume that there exist $\Gamma^{(1)}, \Gamma^{(2)}$ as well as $\eta^{(1)}, \eta^{(2)}$. Define an edge set H as the union of the direct influence and common-influence links under the two networks. In other words,

$$H = E(\Gamma^{(1)}) \cup E(\Gamma^{(1)'}) \cup E(\Gamma^{(1)'}\Gamma^{(1)}) \cup E(\Gamma^{(2)}) \cup E(\Gamma^{(2)'}) \cup E(\Gamma^{(2)'}\Gamma^{(2)}) \tag{14}$$

The maximum density conditions imply that this network has at most $\frac{N(\frac{N}{R}-2)}{2}$ links. Therefore, by Lemma 5, its complement $\overline{H}$ must contain a clique of size $2R+1$. By Lemma 4, $\mathbf{ZZ}'_{(i,j)}$ is identified

for all (i, j) in this clique. Define the set of agents in this clique as $V' \subset V$. Lemma 3 says that we can add agents who are connected to at least R agents in this clique through Lemma 3.

The rest of the proof relies upon showing that this set $V' \subset V$ must eventually include all agents in V, or else a contradiction is implied. When the number of links in the set is relatively small (but still at least $2R+1$), then the maximum number of possible links in $\overline{H}$ is smaller than the minimum density implied by the maximum density conditions on H. When the number of agents in V' is relatively large, then agents in $\overline{V'}$ cannot have enough links to meet the minimum degree requirement on average, where this minimum degree requirement is implied by the maximum degree condition on H.

3.6 Discussion Of Identification

Theorem 1 requires sparsity conditions, even in the limit. That is, even under the assumption that the covariance matrix is perfectly observed, separating the network and factor structures still requires restrictions on the network structure. This is in contrast to de Paula, Rasul and Souza (2020), who can identify arbitrarily dense network structures, but they do not allow for a factor structure. We note that when we assume there is no factor structure, we can dispense with the sparsity conditions, as shown below in Proposition 1.

We note that the two conditions are dependent on R, the number of factors that need identifying. The density condition is clearly decreasing with R: when $R = 1$, density can be as high as approximately $\frac{1}{4}$; when $R = 2$, this drops to $\frac{1}{8}$.

The maximum degree condition is needed to ensure that we observe a sufficient number of entries in each row/column. We note that this restriction may not be particularly binding in applications: for $R = 1$, the max degree condition is $\frac{5N-4}{12}$. Accordingly, in this case, we can allow for agents to be connected to approximately $\frac{5}{12}$ of other agents as long as the average degree condition still holds.

4 Extensions and Special Cases

4.1 Diagonal Covariance

Here, we consider the case that is closest to those existing in the literature thus far (see Manresa, 2016; de Paula, Rasul and Souza, 2020), which do not have a latent factor structure. The absence of a latent factor structure implies a restriction on the covariance matrix $\mathbf{V}$. Alternatively, this can be thought of as a limiting case of Theorem 1 when $R = 0$.

Proposition 1. Given Assumptions 1 - 3, if $\mathbf{V}$ is diagonal (equivalently, $\eta = 0$), then Γ, $\mathbf{U}$ are identified.

Proof. This result is a straightforward application of Lemma 1. That is, the assumption that $\eta = 0$ implies that $\mathbf{ZZ}' = 0$. Therefore $\mathbf{ZZ}'$ is identified. Lemma 1 immediately provides identification of Γ and $\mathbf{U}$. □

This result states that we can recover $\mathbf{\Gamma}$—even if it is arbitrarily dense—if there are no unobserved factors. This follows from the fact that, when $\eta = 0$ uniformly, any covariance across $\mathbf{y_t}$ must be due to the network structure. This result is analogous to de Paula, Rasul and Souza (2020), but we note that we do not identify the other features of the model that they are interested in, and we do not require a row-sum normalization.

4.2 Partial Knowledge of Network Structure

If we have a priori knowledge of which network links cannot exist, then we can state weaker assumptions for identification. First, we define a *restricted set* in Definition 7. This formalizes the notion that there may be some pairs of agents that cannot be linked. This may be due to, for example, social or geographic distance, such as cases where agents are known to be in different villages or schools.

Definition 7. A Restricted Set J is a set of ordered pairs such that, for all $(i,j) \notin J$, $(i,j) \notin E(\Gamma) \cup E(\Gamma') \cup E(\Gamma'\Gamma)$.

For purposes of identification, a restricted set constrains the set of allowable links. That is, whenever $(i,j) \notin J$, agents $i, j \neq i$ cannot be connected in $E(\Gamma) \cup E(\Gamma') \cup E(\Gamma'\Gamma)$.[18] Due to

[18] In the case where we impose no restrictions, $(i,j) \in J$ for all $i, j \neq i$.

Lemma 4, for all $(i,j) \notin J$, we always observe $\mathbf{ZZ}'_{(i,j)}$. With a sufficient number and placement of links that are known not to exist, we can uniquely identify $\mathbf{ZZ}'$, regardless of the structure of the network within the restricted set.

With the restricted set thus defined, Theorem 2 states an important identification result. Its proof is in the Appendix, but the intuition is straightforward. The two hypotheses of Theorem 2 imply that each agent belongs to a clique in $\overline{J}$ (the complement of the restricted set), of size at least $2R+1$. From this, we can use similar arguments to those used in Lemma 5 to identify $\mathbf{ZZ}'_{(i,j)}$ for all i,j in that clique, including when $i=j$. Lemma 3 then provides for identification of all elements involving agents in different cliques, and thus $\mathbf{ZZ}'$ is identified. Finally, Lemma 1 provides for identification of Γ and η.

Theorem 2. Suppose that there exists as restricted set J such that, for each $i \in V$, there exists some $V' \subset V$ such that $i \in V'$ and

[1] $|V'| \geq 2R+1$

[2] $(i,j) \in \overline{J} \; \forall \; i,j \neq i \in V'$

Then, (Γ, η) are identified.

Theorem 2 is particularly useful when we want to simultaneously identify networks (Γ) and unobserved factors (η) among agents across many disjoint groups, such as villages or schools. If the set of agents belongs to at least $2R+1$ groups that we know have no connections across groups, then Theorem 2 states that Γ is identified, regardless of network density *witthin* groups. Intuitively, we infer η from the cross-group covariances in outcomes across time, since we know these agents have no network connections. Conditional on η, any remaining covariance among agents must be due to the network structure, which gives identification of Γ.

Two notes bear mentioning. First, with knowledge of which links may not exist, we can identify denser networks. Assuming $R=1$ and we observe three equally-sized groups, then network density can be approximately $\frac{1}{3}$, in contrast to max density of approximately $\frac{1}{4}$ as required by Theorem 1. If $R=2$ and we observe five equally-sized groups, then network density can be as large as approximately $\frac{1}{5}$, in contrast to approximately $\frac{1}{8}$ as required by Theorem 1.

Second, Theorem 2 provides a lower bound on the number of disjoint groups that are sufficient for identification in the absence of density restrictions. Regardless of the within-group network

structure, $2R+1$ disjoint groups are sufficient to identify arbitrarily dense within-group networks. That is, if $R=1$, then three groups are sufficient; if $R=2$, then five are sufficient.

5 Estimation

We note here that, as in de Paula, Rasul and Souza (2020), our identification results do not depend on any particular estimator, and we make no claims that the algorithm we propose is "optimal." To show the algorithm's practicality, we provide simulations in Section 6.

5.1 Ideal Problem

Our identification results provide that, for a given R, given the empirical covariance $\mathbf{S} = \mathbb{E}[\mathbf{y_t}\mathbf{y_t'}]$, there exists a unique solution to the following minimization problem.

$$\begin{aligned} \min_{\Gamma,\mathbf{Z}} \|\mathbf{S}^{-1} - (\mathbf{I}-\Gamma')\mathbf{U}^{-1}(\mathbf{I}-\Gamma) + \mathbf{Z}\mathbf{Z}'\|_F \\ \text{s.t. } \|(\mathbf{I}-\Gamma')(\mathbf{I}-\Gamma)\|_{i0} < \frac{N(\frac{N}{R}-2)}{4} \ \forall\, i \\ \|(\mathbf{I}-\Gamma')(\mathbf{I}-\Gamma)\|_0 \leq \frac{5N-4}{12} \end{aligned} \tag{15}$$

where $\|\|_0$ is the L_0 norm of the entire matrix, and $\|\|_{i0}$ is the L_0 norm for each row.

Theorem 1 provides conditions that ensure that this optimization problem has a unique solution. However, in general this problem is non-convex due to the L_0 norm constraints. Accordingly, we implement a convex relaxation that imposes soft-threshold version of the two constraints. As defined in Chiong and Moon (2018), these constraints map naturally to $L_{1,1}$ and $L_{1,2}$ norms respectively, given by

$$\|\Gamma\|_{1,1} = \sum_i \sum_j |\Gamma_{ij}|, \quad \|\Gamma\|_{1,2} = \left(\sum_i (\sum_j |\Gamma_{ij}|)^2\right)^{\frac{1}{2}} \tag{16}$$

That is, $\|\Gamma\|_{1,1}$ penalizes network density, while $\|\Gamma\|_{1,2}$ penalizes high-degree agents.

In principle, we could estimate a soft-threshold version of the optimization problem in (15)

above. This would take the form of Criterion (17).

$$\min_{\Gamma,\mathbf{Z}} \|\Sigma_{\mathbf{y}} - (\mathbf{I}-\Gamma')\mathbf{U}^{-1}(\mathbf{I}-\Gamma) + \mathbf{Z}\mathbf{Z}'\|_F + \lambda_{1,1}\|\Gamma\|_{1,1} + \lambda_{1,2}\|\Gamma\|_{1,2}^2 \tag{17}$$

where $\lambda_{1,1}, \lambda_{1,2} > 0$. However, the Frobenius norm in the criterion does not admit simple calculation and may be non-convex.

5.2 Graphical Lasso Algorithms

To simplify computation, we make parametric assumptions on unobserved variables. This leads to an estimation criterion derived from a Gaussian likelihood function.

Assumption 4. For all t, $\begin{bmatrix} \mathbf{u_t} \\ f_t \end{bmatrix} \sim N\left(0, \begin{bmatrix} \mathbf{I}_N\sigma^2 & 0 \\ 0 & \mathbf{I}_R \end{bmatrix}\right)$

First, Assumption 4 assumes both homoskedasticity across agents and that all random variables at time t follow a Guassian distribution.[19] Independence (0's off the diagonal) is implied by Assumption 2. Note that we make no assumption on the distribution of η, as it is a parameter of the model. Given Assumption 4, it follows that

$$\mathbf{y_t} \sim N\left(0, (\mathbf{I}-\Gamma)^{-1}(\mathbf{I}\sigma^2 + \eta\eta')(\mathbf{I}-\Gamma')^{-1}\right) \tag{18}$$

We use known results to write the likelihood of the sample covariance $\mathbf{S} = \frac{1}{T}\sum_t \mathbf{y_t}\mathbf{y_t}'$.[20] Up to a constant, the likelihood of the precision matrix is

$$L(\mathbf{S};\Gamma,\sigma^2,\eta) = \log(|(\mathbf{I}-\Gamma')(\mathbf{I}\sigma^2+\eta\eta')^{-1}(\mathbf{I}-\Gamma)|) - tr(\mathbf{S}(\mathbf{I}-\Gamma')(\mathbf{I}\sigma^2+\eta\eta')^{-1}(\mathbf{I}-\Gamma)) \tag{19}$$

where $|.|$ defines the determinant function and $tr()$ defines the matrix trace. To simplify, note that

$$\log(|(\mathbf{I}-\Gamma')(\mathbf{I}\sigma^2+\eta\eta')^{-1}(\mathbf{I}-\Gamma)|) = 2\log(|(\mathbf{I}-\Gamma)|) - \log(|(\mathbf{I}\sigma^2+\eta\eta')|) \tag{20}$$

$$= 2\log(|(\mathbf{I}-\Gamma)|) - \log(|\mathbf{I_R}\sigma^2+\eta'\eta|) - (N-R)\log\sigma^2 \tag{21}$$

[19] Note that this is homoskedasticity on the non-factor part of the error term u_{it}. The factor structure will necessarily imply heteroskedasticity in the composite error term $\epsilon_{it} = \eta_i' f_t + u_{it}$.

[20] Note that this assumes that $\mathbf{y_t}$ has been demeaned by agent first. That is, if $\mathbf{y}_t^*$ is raw data for t, $\mathbf{y_t} = \mathbf{y}_t^* - \sum_{s=1}^T \mathbf{y}_s^*$

where the last step employs the Matrix Determinant Lemma. Next, since the trace is a linear operator, we can apply the Woodbury Inverse Formula, which implies

$$tr(S(\mathbf{I}-\Gamma')(\mathbf{I}\sigma^2+\eta\eta')^{-1}(\mathbf{I}-\Gamma)) = \frac{1}{\sigma^2}\left(tr(S(\mathbf{I}-\Gamma')(\mathbf{I}-\Gamma)) - tr(S(\mathbf{I}-\Gamma')\eta(\mathbf{I_R}\sigma^2+\eta'\eta)^{-1}\eta'(\mathbf{I}-\Gamma))\right) \tag{22}$$

Substituting Equations (21) and (22) into (19) thus yields the following likelihood function.

$$\begin{aligned} L(\mathbf{S};\Gamma,\sigma^2,\eta) = {} & 2\log(|(\mathbf{I}-\Gamma)|) - \log(|\mathbf{I_R}\sigma^2+\eta'\eta|) - (N-R)\log\sigma^2 \\ & + \frac{1}{\sigma^2}tr(S(\mathbf{I}-\Gamma')\eta(\mathbf{I_R}\sigma^2+\eta'\eta)^{-1}\eta'(\mathbf{I}-\Gamma)) - \frac{1}{\sigma^2}tr(S(\mathbf{I}-\Gamma')(\mathbf{I}-\Gamma)) \end{aligned} \tag{23}$$

In principle, we could directly maximize the likelihood as given in Equation (23). However, this would estimate an arbitrarily dense Γ and, without imposing sparsity, we cannot be sure that there would be a unique solution in Γ and η. We impose sparsity by minimizing a penalized version of the negative likelihood function, with penalties in terms of the $L_{1,1}$ and $L_{1,2}$ norms of Γ. That is, we seek to solve the problem given in Criterion (24).

$$\min_{\Gamma,\sigma,\eta}\left(-L(\mathbf{S};\Gamma,\sigma^2,\eta) + \lambda_{1,1}\|\Gamma\|_{1,1} + \lambda_{1,2}\|\Gamma\|_{1,2}^2\right) \tag{24}$$

where $L(\mathbf{S};\Gamma,\sigma^2,\eta)$ is defined in Equation (23).

Algorithm 1. (Latent Graphical LASSO with Mixed Norms)

Set-up

Data: $\mathbf{S} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y_t}\mathbf{y_t'}$, an $N \times N$ matrix

Tuning Parameters: $\lambda_{1,1}, \lambda_{1,2} > 0$, $R \in \{1, 2, ..., \}$,

Initialization: $\hat{\eta}^{(0)} = \mathbf{0}_{N\times r}$, $\hat{\Gamma}^{(0)} = \mathbf{0}_{N\times N}$, $\hat{\sigma}^{(0)} = 1$, $\kappa_\sigma, \kappa_\Gamma$ large

Step 1: Estimate $(\hat{\Gamma}, \tilde{\eta}, \hat{\sigma}^2)$ by minimizing Criterion:

$$(\hat{\Gamma}, \tilde{\eta}, \hat{\sigma}) = \arg\max_{\Gamma,\eta,\sigma}\left(-L(\Gamma,\sigma^2,\eta) + +\lambda_{1,1}\|\Gamma\|_{1,1} + \lambda_{1,2}\|\Gamma\|_{1,2}^2\right)$$

Step 2: Given $\tilde{\eta}$, define $\hat{\eta}$:

(a) $\mathbf{E}_\eta = \tilde{\eta}\tilde{\eta}'$

(b) $\hat{\eta} = \mathbf{V}_R \Lambda_R^{\frac{1}{2}}$, where

$$\mathbf{E}_\eta = \mathbf{V}\Lambda\mathbf{V}', \text{ the diagonal decomposition of } \mathbf{E}$$

$$\mathbf{V}_R = \text{first } R \text{ columns of } \mathbf{V}$$

$$\Lambda_R = \text{first } R \text{ rows/columns of } \Lambda, \text{ where } \Lambda_{(i,i)} \geq \Lambda_{(i+1,i+1)}$$

We propose Algorithm 1 to estimate the model parameters. The algorithm has two steps. First, Step 1 minimizes Criterion (24). Conditional on the minimand at Step 1, Step 2 "rotates" $\tilde{\eta}$ so that the estimated $\hat{\eta}$ meets the assumptions for the latent factor structure in Assumption 2; essentially, Step 2 adjusts estimated $\tilde{\eta}$ so that the factors are mutually orthogonal and in order of decreasing variance. When $R = 1$, then Step 2 is unnecessary, since Assumption 2 is necessarily satisfied.

We propose the optimization in Step 1 be performed using an iterative proximal gradient descent method. That is, we iterate between two steps: (1) a step in the direction of steepest gradient, and (2) shrinkage of parameters via a proximity projection. These two steps, including closed forms for the gradient and proximity function, are described in detail in Appendix B.

Solving non-differentiable convex optimization with proximal gradient descent can be traced back to von Neumann (1951). The fundamental idea of the method is to replace the differential part of the target function with a quadratic approximation so that an analytic solution for the minimization problem can be generated. As a gradient decent algorithm, it shares all the advantages of Newton's methods where backtracking line search can be used to determine the step size and it is parallelizable when computing gradients. This method has received a great amount of attention recently as an "accelerated" method of penalized estimation (see Bien, Taylor and Tibshirani, 2013; Boyd et al., 2010; Cevher, Becker and Schmidt, 2014; Tibshirani, 2014).

Finally, we note that Algorithm 1 is easily adaptable to situations in which the researcher has partial knowledge of the network structure, in the form of a restricted set as discussed for identification in Theorem 2. If, for example, the network is known to be block diagonal, where each "block" represents a school, village, etc., we can run the algorithm under the restriction that

$\Gamma_{(i,j)} = 0$ whenever we know that agents i, j cannot be linked, such as if they are in different villages.[21]

5.3 Choice of Tuning Parameters

Estimation of Criterion (24) requires a choice of tuning parameters $\lambda_{1,1}$ and $\lambda_{1,2}$. $\lambda_{1,1}$ penalizes network density, while $\lambda_{1,2}$ penalizes degree per agent. Clearly, detected network density is decreasing in both parameters: in the limiting cases, the algorithm returns a full network ($\lambda_{1,1} = \lambda_{1,2} = 0$) or an empty one ($\lambda_{1,1}$, $\lambda_{1,2}$ both large).

In our results in Sections 6 and 7, we present results across a grid of values for the tuning parameters. As expected, the detected networks become more sparse as these parameters become larger. Accordingly, the choice of tuning parameters works on the interplay between finding "too many" links (many false positives) or "too few" links (many false negatives).

In empirical applications, care should be taken in choosing these tuning parameters. In principle, k-fold cross validation allows for choosing among a range of tuning parameters, while targeting some out-of-sample moment such as the empirical covariance/precision, likelihood, or even possibly network density.

6 Simulations on Diffusion of Microfinance Networks

6.1 DGP

In order to simulate peer effects in networks that resemble "real world" networks, we employ networks from the widely-used Diffusion of Microfinance dataset (see Banerjee et al., 2012). That is, we take the networks we see there as given. In our simulations, we use a subset of their villages and various network definitions, with network statistics given in Table 1.

The networks in the data define an adjacency matrix $\mathbf{L}$, where $L_{(i,j)} \in \{0, 1\}$. We row-normalize the adjacency matrix $\mathbf{L}$, then multiply by $\beta < 1$, which implies that the DGP satisfies Assumption 1 (and thus $(\mathbf{I} - \Gamma)$ is invertible). This defines the "true" network $\mathbf{\Gamma}$. That is, for each $i, j \neq i$,

[21] In our multiple-village results, we simulate just such a procedure, with results given in Table 5.

TABLE 1
DIFFUSION OF MICROFINANCE NETWORK DATA DESCRIPTIVES

		All Connections		Money Borrow & Lend		Rice Come & Go		Visit Come & Go	
Village	Size	Density	Max Degree	Density	Max Degree	Density	Max Degree	Density	Max Degree
3	292	0.0305	46	0.0062	15	0.0089	11	0.0126	16
8	94	0.0897	35	0.0327	17	0.0380	16	0.0467	19
10	77	0.1141	25	0.0547	14	0.0437	10	0.0581	15
58	178	0.0528	30	0.0195	24	0.0232	13	0.0315	19
60	356	0.0225	39	0.0091	23	0.0106	19	0.0141	23

Notes: Table gives summary statistics for select villages from data collected by Banerjee et al. (2012).

$\Gamma_{(i,j)} = \beta \frac{L_{(i,j)}}{\sum_{k \neq i,j} L_{(i,k)}}$. We construct y_{it} according to

$$y_{it} = \sum_{j \neq i} \Gamma_{(i,j)} y_{jt} + f_t \eta_i + u_{it} \tag{25}$$

which corresponds to the assumed DGP in Equation (2), with factor structure as defined in Equation (4).

In all cases, we draw $f_t, u_{it} \sim_{i.i.d.} N(0,1)$ and we set the number of factors as 1. In the results that we present here, we set $\sigma_\eta^2 = 1$ and $\beta = 0.25$.[22]

6.2 Estimation Procedure and Performance Metrics

We simulate data according to the DGP as described in Subsection 6.1. For each set of data-generation and estimation parameters (including the networks used), we run our estimation algorithm across a range of penalization parameters $\lambda_{1,1}$ and $\lambda_{1,2}$. For some sets of results, we run a version where $R = 0$, which corresponds to estimation without accounting for the factor structure.

To measure performance, we define several performance metrics. For any estimation run, define $\hat{\Gamma}$ as the estimated/detected network. We define the following performance metrics.

[1] Network Density = $\frac{\sum_i \sum_{j \neq i} \mathbb{1}\{\hat{\Gamma}_{(i,j)} \neq 0\}}{N(N-1)}$

[2] Positive Predictive Value (PPV) = $\frac{\sum_i \sum_{j \neq i} \mathbb{1}\{\Gamma_{(i,j)} \neq 0\} \mathbb{1}\{\hat{\Gamma}_{(i,j)} \neq 0\}}{\sum_i \sum_{j \neq i} \mathbb{1}\{\hat{\Gamma}_{(i,j)} \neq 0\}}$

[3] Negative Predictive Value (NPV) = $\frac{\sum_i \sum_{j \neq i} \mathbb{1}\{(1 - \Gamma_{(i,j)} \neq 0\})(1 - \mathbb{1}\{\hat{\Gamma}_{(i,j)} \neq 0\})}{\sum_i \sum_{j \neq i} (1 - \mathbb{1}\{\hat{\Gamma}_{(i,j)} \neq 0\})}$

[22] Ongoing work is testing the sensitivity of results to different DGP parameter values. Note the edge cases here. First, $\sigma_\eta^2 = 0$ corresponds to the case where there is no latent factor structure. Second, $\beta = 0$ corresponds to the case in which there are no network-based peer effects.

Estimated network density is simply the probability of detecting a link, among all possible pairs of agents, in the estimated network. PPV is the probability that a detected link is an actual link. A natural benchmark for PPV is (true) network density, as any PPV higher than the network density implies that the algorithm does better than chance at finding network links. Conversely, NPV is the probability that a pair classified as not linked is in fact not linked. A natural benchmark is (1 - density).

6.3 Main Results with a Single Village

First, we present results from a single village. For purposes of this exercise, we set $T = 1,000,000$, and we interpret these results as assessing the case closest to that assumed for purposes of identification. That is, when T is very large, the covariance matrix is close to perfectly observed.[23]

Our first results are given in Table 2. Across a range of $(\lambda_{1,1}, \lambda_{1,2})$, Panel A shows detected network density, Panel B shows PPV, and Panel C shows NPV. As expected, density is decreasing as the penalization parameters increase, corresponding to moving to the right and downward within each panel. PPV is also increasing with the penalization parameters: as we raise the penalty for each link detected, the likelihood of each detected link being a true link is rising. Conversely, NPV is decreasing with the penalty parameters.

The limiting cases are illustrative here. When $\lambda_{1,1} = \lambda_{1,2} = 0$, then there is no penalization for detected links. As such, due to noise in the data, all pairs are detected as linked some nonzero amount, which implies that detected density is 1 and PPV is simply network density. Conversely, when both penalization parameters are high, the algorithm detects almost no links, and network density is close to zero. In this case, NPV approaches (1 - density).

The sensitivity of results to choice of penalization parameters illustrates the importance of the choice of penalty terms. When penalties are low, the algorithm returns too many “false positives"; when they are high, it returns too many “false negatives." Intermediate values of the penalty terms $\lambda_{1,1}$ and $\lambda_{1,2}$ trade off these two concerns.

We also report results for $R = 0$ in estimation, which corresponds to a “misspecified" model that ignores the latent factor structure. These results are shown in Table 3, which gives analogous

[23] We assess performance for smaller T in Subsection 7.1.

TABLE 2
PERFORMANCE METRICS FOR SINGLE VILLAGE
($R = 1$)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-6}$	0	0.1	1	10	100	1000	10000
Panel A: Network Density							
0.000	1.0000	0.9957	0.9135	0.0659	0.0226	0.0143	0.0073
0.010	0.9994	0.9926	0.8985	0.0642	0.0225	0.0140	0.0073
0.025	0.9948	0.9872	0.8717	0.0620	0.0223	0.0138	0.0074
0.050	0.9863	0.9738	0.8176	0.0598	0.0219	0.0133	0.0073
0.075	0.9687	0.9480	0.7390	0.0581	0.0215	0.0131	0.0073
0.100	0.9326	0.9067	0.6359	0.0563	0.0211	0.0129	0.0073
0.250	0.1500	0.1378	0.0779	0.0472	0.0196	0.0117	0.0070
0.500	0.0487	0.0485	0.0478	0.0386	0.0171	0.0103	0.0069
0.750	0.0405	0.0404	0.0392	0.0319	0.0155	0.0097	0.0066
1.000	0.0312	0.0312	0.0304	0.0258	0.0136	0.0088	0.0064
Panel B: Positive Predictive Value							
0.000	0.0528	0.0530	0.0577	0.7475	0.9215	0.9468	0.9610
0.010	0.0528	0.0532	0.0587	0.7651	0.9251	0.9501	0.9610
0.025	0.0531	0.0535	0.0605	0.7844	0.9260	0.9517	0.9612
0.050	0.0535	0.0542	0.0645	0.8083	0.9289	0.9523	0.9610
0.075	0.0545	0.0556	0.0713	0.8250	0.9349	0.9517	0.9652
0.100	0.0566	0.0582	0.0828	0.8417	0.9429	0.9507	0.9651
0.250	0.3477	0.3778	0.6619	0.9476	0.9676	0.9730	0.9955
0.500	0.9961	0.9961	0.9960	0.9984	1.0000	1.0000	1.0000
0.750	0.9984	0.9984	0.9984	0.9980	1.0000	1.0000	1.0000
1.000	0.9980	0.9980	0.9979	0.9975	1.0000	1.0000	1.0000
Panel C: Negative Predictive Value							
0.000	n/a	1.0000	0.9993	0.9962	0.9673	0.9602	0.9539
0.010	1.0000	1.0000	0.9994	0.9960	0.9672	0.9599	0.9539
0.025	1.0000	1.0000	0.9995	0.9955	0.9671	0.9598	0.9539
0.050	1.0000	0.9988	0.9995	0.9952	0.9668	0.9593	0.9539
0.075	0.9980	0.9988	0.9996	0.9948	0.9665	0.9592	0.9539
0.100	0.9991	0.9993	0.9996	0.9943	0.9664	0.9589	0.9539
0.250	0.9992	0.9991	0.9987	0.9915	0.9655	0.9581	0.9538
0.500	0.9954	0.9952	0.9945	0.9851	0.9637	0.9571	0.9538
0.750	0.9871	0.9870	0.9858	0.9783	0.9621	0.9565	0.9535
1.000	0.9777	0.9777	0.9768	0.9722	0.9603	0.9556	0.9533

Notes: $\beta = 0.25$. Village 58, network of All Relationships. See Table 1 for N and density. $T = 1,000,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

results to Table 2. Note that the classification problem is “easier" in a sense here since the number of parameters is much smaller. The “misspecified" algorithm does detect systematically sparser networks for a given combination of penalty parameters.

6.4 Multiple Villages

Next, we present results pooling multiple villages. This illustrates the utility of using cross-village covariances to identify η, with increases in the precision of detected network links. We perform this exercise under three different estimation procedures:

[1] Impose no prior knowledge of village structure (Unrestricted)

[2] Restrict all cross-village links to be zero (Restricted)

[3] Estimate villages separately (Separate)

Table 4 gives results for the unrestricted algorithm. Note that NPV and PPV are generally smaller than in the single-village case in Table 2, which follows from the fact that this is a “harder" problem as there are more parameters to estimate. In Panel D, we see that with relatively low values of penalization parameters, the algorithm detects a non-trivial number of cross-village links that are clearly erroneous. In contrast, with higher penalty values, the network detects very few cross-villlage links.

In order to prevent the algorithm from detecting cross-village links, we run a restricted version of the algorithm, where we restrict cross-village pairs to be unlinked. These results are shown in Table 5. Note that the algorithm here by design cannot detect cross-village links. Further, comparing Tables 4 to 5 shows that the restricted version generally performs better in the sense of higher PPV and NPV for a given choice of penalty parameters.

Finally, in Table 6 we show results where the algorithm is run separately for each of the two villages. This procedure employs less “data" than the Restricted version (Table 5), since it does not utilize cross-village correlation in estimating the factor structure. The algorithm generally performs worse than the Restricted version, in terms of both NPV and PPV.[24]

[24]Note that the range of $\lambda_{1,1}$ and $\lambda_{1,2}$ is different for the results in Table 6 as compared to Tables 4-5.

TABLE 3
PERFORMANCE METRICS FOR SINGLE VILLAGE
($R = 0$)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-6}$	0	0.1	1	10	100	1000	10000
Panel A: Network Density							
0.000	1.0000	0.8294	0.2224	0.0599	0.0244	0.0157	0.0071
0.010	0.5307	0.4344	0.1593	0.0569	0.0237	0.0154	0.0071
0.025	0.2330	0.2084	0.1100	0.0534	0.0229	0.0150	0.0070
0.050	0.1108	0.1048	0.0778	0.0522	0.0222	0.0147	0.0068
0.075	0.0770	0.0754	0.0660	0.0516	0.0219	0.0146	0.0068
0.100	0.0651	0.0645	0.0606	0.0508	0.0216	0.0143	0.0068
0.250	0.0539	0.0538	0.0536	0.0461	0.0191	0.0131	0.0067
0.500	0.0452	0.0451	0.0439	0.0346	0.0162	0.0114	0.0063
0.750	0.0280	0.0278	0.0272	0.0237	0.0127	0.0097	0.0060
1.000	0.0178	0.0178	0.0175	0.0157	0.0100	0.0083	0.0055
Panel B: Positive Predictive Value							
0.000	0.0528	0.0637	0.2375	0.8606	0.9544	0.9615	0.9686
0.010	0.0995	0.1216	0.3316	0.9041	0.9692	0.9731	0.9686
0.025	0.2267	0.2535	0.4800	0.9619	0.9931	0.9894	0.9818
0.050	0.4768	0.5038	0.6792	0.9763	1.0000	1.0000	0.9953
0.075	0.6862	0.7000	0.8008	0.9809	1.0000	1.0000	1.0000
0.100	0.8109	0.8193	0.8712	0.9875	1.0000	1.0000	1.0000
0.250	0.9794	0.9800	0.9816	0.9959	1.0000	1.0000	1.0000
0.500	0.9965	0.9965	0.9971	0.9982	1.0000	1.0000	1.0000
0.750	0.9977	0.9977	0.9977	0.9973	1.0000	1.0000	1.0000
1.000	0.9964	0.9964	0.9964	1.0000	1.0000	1.0000	1.0000
Panel C: Negative Predictive Value							
0.000	n/a	1.0000	1.0000	0.9986	0.9697	0.9617	0.9537
0.010	1.0000	1.0000	1.0000	0.9986	0.9694	0.9615	0.9537
0.025	1.0000	1.0000	1.0000	0.9984	0.9693	0.9614	0.9537
0.050	1.0000	1.0000	1.0000	0.9981	0.9687	0.9613	0.9537
0.075	1.0000	1.0000	1.0000	0.9977	0.9684	0.9612	0.9537
0.100	1.0000	1.0000	1.0000	0.9972	0.9681	0.9609	0.9537
0.250	0.9999	0.9999	0.9998	0.9928	0.9657	0.9598	0.9535
0.500	0.9919	0.9917	0.9905	0.9810	0.9627	0.9581	0.9532
0.750	0.9744	0.9742	0.9736	0.9701	0.9593	0.9565	0.9529
1.000	0.9643	0.9643	0.9639	0.9623	0.9567	0.9551	0.9524

Notes: $\beta = 0.25$. Village 58, network of All Relationships. See Table 1 for N and density. $T = 1,000,000$, $\sigma_\eta = 1$ in all simulations. $R = 0$ in estimation.

TABLE 4
PERFORMANCE METRICS FOR TWO-VILLAGE RESULTS (UNRESTRICTED)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-6}$	0	0.1	1	10	100	1000	10000
Panel A: Network Density							
0.000	1.0000	0.9961	0.9539	0.1879	0.0301	0.0149	0.0076
0.010	0.9942	0.9872	0.9416	0.1636	0.0295	0.0149	0.0076
0.025	0.9811	0.9739	0.9221	0.1364	0.0288	0.0148	0.0076
0.050	0.9556	0.9458	0.8805	0.1018	0.0283	0.0143	0.0076
0.075	0.9145	0.9030	0.8191	0.0788	0.0276	0.0141	0.0076
0.100	0.8563	0.8433	0.7184	0.0644	0.0269	0.0137	0.0076
0.250	0.1065	0.1028	0.0761	0.0463	0.0251	0.0130	0.0075
0.500	0.0433	0.0433	0.0426	0.0373	0.0228	0.0117	0.0072
0.750	0.0337	0.0337	0.0332	0.0303	0.0204	0.0110	0.0068
1.000	0.0279	0.0279	0.0277	0.0264	0.0178	0.0097	0.0067
Panel B: Positive Predictive Value (within Village)							
0.000	0.0499	0.0501	0.0521	0.2562	0.9233	0.9769	0.9820
0.010	0.0502	0.0506	0.0528	0.2936	0.9336	0.9769	0.9820
0.025	0.0508	0.0512	0.0538	0.3522	0.9403	0.9791	0.9820
0.050	0.0521	0.0526	0.0563	0.4706	0.9477	0.9832	0.9820
0.075	0.0544	0.0551	0.0604	0.6045	0.9577	0.9853	0.9819
0.100	0.0580	0.0589	0.0687	0.7368	0.9706	0.9900	0.9818
0.250	0.4591	0.4752	0.6390	0.9643	0.9973	0.9921	0.9817
0.500	0.9857	0.9857	0.9855	0.9871	1.0000	1.0000	1.0000
0.750	0.9898	0.9898	0.9896	0.9955	1.0000	1.0000	1.0000
1.000	0.9951	0.9951	0.9950	0.9948	1.0000	1.0000	1.0000
Panel C: Negative Predictive Value (within Village)							
0.000	n/a	1.0000	0.9948	0.9978	0.9771	0.9641	0.9572
0.010	1.0000	1.0000	0.9959	0.9977	0.9769	0.9640	0.9572
0.025	0.9964	0.9974	0.9960	0.9978	0.9765	0.9640	0.9572
0.050	0.9953	0.9962	0.9965	0.9977	0.9762	0.9636	0.9572
0.075	0.9976	0.9979	0.9975	0.9975	0.9758	0.9634	0.9572
0.100	0.9981	0.9982	0.9978	0.9974	0.9755	0.9631	0.9572
0.250	0.9988	0.9988	0.9986	0.9944	0.9745	0.9625	0.9571
0.500	0.9924	0.9924	0.9917	0.9863	0.9723	0.9613	0.9569
0.750	0.9828	0.9828	0.9824	0.9796	0.9699	0.9606	0.9566
1.000	0.9772	0.9771	0.9769	0.9757	0.9673	0.9594	0.9565
Panel D: Percent of Across-Village Links Detected (within Village)							
0.000	1.0000	0.9997	0.9988	0.2236	0.0023	0.0003	0.0001
0.010	0.9997	0.9997	0.9974	0.1864	0.0019	0.0003	0.0001
0.025	0.9996	0.9994	0.9934	0.1426	0.0016	0.0003	0.0001
0.050	0.9974	0.9963	0.9809	0.0865	0.0013	0.0003	0.0001
0.075	0.9853	0.9828	0.9572	0.0479	0.0011	0.0003	0.0001
0.100	0.9644	0.9606	0.8906	0.0242	0.0008	0.0001	0.0001
0.250	0.1008	0.0944	0.0465	0.0015	0.0001	0.0001	0.0001
0.500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.750	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: $\beta = 0.25$. Villages 8 and 10, network of All Relationships. See Table 1 for N and density. $T = 1,000,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE 5
PERFORMANCE METRICS FOR TWO-VILLAGE RESULTS (RESTRICTED)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-6}$	0	0.1	1	10	100	1000	10000
Panel A: Network Density							
0.000	1.0000	0.9960	0.9531	0.2282	0.0598	0.0236	0.0141
0.010	0.9861	0.9786	0.9347	0.2017	0.0593	0.0233	0.0141
0.025	0.9634	0.9556	0.9062	0.1723	0.0581	0.0230	0.0140
0.050	0.9144	0.9056	0.8462	0.1382	0.0567	0.0222	0.0138
0.075	0.8478	0.8384	0.7648	0.1185	0.0552	0.0222	0.0138
0.100	0.7603	0.7482	0.6462	0.1083	0.0537	0.0221	0.0138
0.250	0.1130	0.1122	0.1065	0.0907	0.0500	0.0207	0.0126
0.500	0.0863	0.0863	0.0849	0.0742	0.0455	0.0190	0.0123
0.750	0.0671	0.0671	0.0662	0.0604	0.0407	0.0179	0.0123
1.000	0.0556	0.0555	0.0551	0.0526	0.0355	0.0167	0.0122
Panel B: Positive Predictive Value (within Village)							
0.000	0.0995	0.0999	0.1041	0.4242	0.9255	0.9739	0.9757
0.010	0.1009	0.1017	0.1062	0.4798	0.9283	0.9765	0.9757
0.025	0.1033	0.1041	0.1093	0.5581	0.9363	0.9762	0.9756
0.050	0.1087	0.1096	0.1170	0.6921	0.9444	0.9784	0.9802
0.075	0.1169	0.1182	0.1293	0.8052	0.9566	0.9784	0.9802
0.100	0.1303	0.1323	0.1528	0.8760	0.9719	0.9814	0.9801
0.250	0.8623	0.8687	0.9118	0.9804	0.9986	0.9967	0.9891
0.500	0.9857	0.9857	0.9855	0.9871	1.0000	1.0000	1.0000
0.750	0.9898	0.9898	0.9896	0.9955	1.0000	1.0000	1.0000
1.000	0.9951	0.9951	0.9950	0.9948	1.0000	1.0000	1.0000
Panel C: Negative Predictive Value (within Village)							
0.000	n/a	1.0000	0.9942	0.9965	0.9530	0.9217	0.9131
0.010	1.0000	1.0000	0.9958	0.9966	0.9527	0.9214	0.9131
0.025	1.0000	1.0000	0.9956	0.9959	0.9521	0.9212	0.9130
0.050	0.9984	0.9978	0.9969	0.9955	0.9513	0.9205	0.9129
0.075	0.9973	0.9975	0.9974	0.9954	0.9506	0.9205	0.9129
0.100	0.9983	0.9981	0.9979	0.9949	0.9500	0.9205	0.9128
0.250	0.9977	0.9977	0.9973	0.9884	0.9478	0.9195	0.9119
0.500	0.9843	0.9842	0.9827	0.9717	0.9434	0.9179	0.9118
0.750	0.9645	0.9645	0.9636	0.9581	0.9387	0.9169	0.9118
1.000	0.9532	0.9531	0.9527	0.9502	0.9336	0.9158	0.9116

Notes: $\beta = 0.25$. Villages 8 and 10, network of All Relationships. See Table 1 for N and density. $T = 1,000,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE 6
PERFORMANCE METRICS FOR TWO-VILLAGE RESULTS (SEPARATE)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-4}$	0	0.001	0.01	0.1	1	10	100
Panel A: Network Density							
0.000	1.0000	1.0000	0.9966	0.8821	0.1244	0.0271	0.0152
0.010	0.4804	0.4796	0.4707	0.4191	0.1171	0.0274	0.0152
0.025	0.4229	0.4228	0.4204	0.3869	0.1095	0.0257	0.0149
0.050	0.4062	0.4059	0.4019	0.3465	0.1005	0.0222	0.0139
0.075	0.3934	0.3928	0.3900	0.2842	0.0955	0.0199	0.0119
0.100	0.3872	0.3871	0.3803	0.2367	0.0909	0.0186	0.0108
0.250	0.1836	0.1862	0.1896	0.1327	0.0677	0.0156	0.0088
0.500	0.0510	0.0509	0.0502	0.0465	0.0256	0.0127	0.0070
0.750	0.0189	0.0189	0.0188	0.0182	0.0163	0.0101	0.0067
1.000	0.0149	0.0149	0.0149	0.0147	0.0128	0.0086	0.0062
Panel B: Positive Predictive Value (within Village)							
0.000	0.0995	0.0995	0.0965	0.0802	0.4306	0.9116	0.9369
0.010	0.1753	0.1753	0.1732	0.1503	0.4336	0.9125	0.9369
0.025	0.1601	0.1598	0.1551	0.1300	0.3867	0.9120	0.9450
0.050	0.1255	0.1248	0.1163	0.1082	0.3306	0.9012	0.9409
0.075	0.0972	0.0960	0.0917	0.1066	0.3034	0.8900	0.9306
0.100	0.0842	0.0846	0.0829	0.1126	0.2969	0.8860	0.9367
0.250	0.1295	0.1310	0.1316	0.1606	0.3239	0.9031	0.9302
0.500	0.4274	0.4280	0.4358	0.4654	0.7620	0.9676	0.9804
0.750	0.9746	0.9746	0.9781	0.9849	1.0000	1.0000	1.0000
1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Panel C: Negative Predictive Value (within Village)							
0.000	n/a	n/a	0.0000	0.7565	0.9476	0.9232	0.9134
0.010	0.9706	0.9704	0.9661	0.9371	0.9448	0.9234	0.9134
0.025	0.9449	0.9447	0.9409	0.9197	0.9358	0.9219	0.9133
0.050	0.9183	0.9178	0.9118	0.9051	0.9263	0.9187	0.9124
0.075	0.8990	0.8982	0.8955	0.9033	0.9220	0.9166	0.9105
0.100	0.8909	0.8911	0.8903	0.9046	0.9203	0.9154	0.9097
0.250	0.9073	0.9077	0.9080	0.9099	0.9168	0.9132	0.9079
0.500	0.9181	0.9181	0.9183	0.9184	0.9179	0.9117	0.9067
0.750	0.9174	0.9174	0.9173	0.9169	0.9154	0.9097	0.9066
1.000	0.9141	0.9141	0.9141	0.9139	0.9122	0.9083	0.9062

Notes: $\beta = 0.25$. Villages 8 and 10, network of All Relationships. See Table 1 for N and density. $T = 1,000,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

7 Sensitivity

Here, we extend our simulations in a number of directions. To conserve space, all tables referenced here are in Appendix C.

7.1 Sensitivity to T

Our results in Section 6 demonstrate that the algorithm performs quite well when $T = 1,000,000$, corresponding to the case where the covariance structure is close to perfectly observed. In real applications, we are unlikely to observe this many repeated observations. Accordingly, we assess the algorithm's performance for lower values of T. First, Table A.1 gives results for $T = 1000$. While PPV is never near 1 as in the case with higher T (Table 2), we see that PPV is much higher than network density for many penalty values. Results get much noiser when we set $T = 100$ and $T = 10$, as shown in Tables A.2 and A.3.

We also reproduce our two-village results with $T = 1000$. Results are generally consistent with those shown in Tables 4, 5, and 6, although all results are noisier. That is, the Restricted model (Table A.5) outperforms the Unrestricted and Separate models (Tables A.4 and A.6, respectively).

7.2 Network Density and Number of Agents

Our main results in Table 2 use Village 58, which has density of 0.0528 among All Connections, approximately average of the 75 villages in the Diffusion of Microfinance dataset. To show that our results are not sensitive to this choice, we also estimate on the most and least dense networks that we find in the data. For the most dense, we use the network of All Connections in Village 10, with a density of 0.1141. For the least dense, we use the network of Give Advice / Help Decision for Village 3, with a density of 0.0062. All other parameters of the DGP are the same as those used for Table 2 above. These results are shown in Tables A.7 and A.8, respectively, with results qualitatively very similar to those in Table 2.

We also perform simulations with the largest village in the Diffusion of Microfinance dataset. This is Village 60, which has 356 households, as compared to 178 in Village 58 used in Table 2. Again, the algorithm performs quite similarly for this village, as shown by comparing results in Tables A.9 and 2.

7.3 Number of Villages

Finally, in order to show that estimation is feasible with many villages, we perform our multiple-village analysis with three villages instead of two. As in the two village case, the Restricted model performs better (in terms of PPV and NPV) than either the Unrestricted or Separate ones. Results for three villages are shown in Tables A.10, A.11, and A.12.

8 Discussion

The study of network-based peer effects is often limited by the unavailability of high-quality data on network connections. In the absence of such data, in this paper we provide new results on identification of the network structure itself. Our results bring together insights from matrix completion, matrix decomposition, and graph theory to give conditions under which the network and latent factor structures are separately identified. We also provide an iterative proximal gradient descent algorithm and demonstrate its performance in estimation.

Our results make four major contributions vis-à-vis the existing literature. First, by relying solely on covariance among outcomes, our results require no exogeneity assumptions or even observation of time-varying covariates. Second, the latent factor structure allows for rich individual-level heterogeneity. Third, our method allows for identification of the absence of peer effects. Finally, we extend the estimator of Battaglini et al. (2021) to allow for unobserved, low-dimensional common shocks.

The identification and estimation methods developed here have potentially wide-ranging applications in studying peer effects in settings in which network data is unavailable. A necessary ingredient in any such application is panel data on outcomes, from which can be constructed an empirical covariance matrix of outcomes across time.

We highlight areas of ongoing and future research. First, ongoing efforts are analyzing methods to choose the penalization parameters. Second, future work will more thoroughly analyze the model's performance in "shorter" panel settings, in an effort to open up more applications. Finally, we leave the issue of incorporating time-varying covariates into our identification and estimation methods to future work.

References

Agarwal, Alekh, Sahand Negahban, and Martin J Wainwright. 2012. "Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions." *The Annals of Statistics*, 1171–1197.

Alidaee, Hossein, Eric Auerbach, and Michael P. Leung. 2020. "Recovering Network Structure from Aggregated Relational Data using Penalized Regression." Unpublished Working Paper.

Anderson, Theodore W, and Herman Rubin. 1956. "Statistical inference in factor analysis." Vol. 5, 111–150.

Bai, Jushan. 2009. "Panel data models with interactive fixed effects." *Econometrica*, 77(4): 1229–1279.

Banerjee, Abhijit, Arun G. Chandrasekhar, Ester Duflo, and Matthew O. Jackson. 2012. "The Diffusion of Microfinance." NBER Working Paper No. 17743.

Battaglini, Marco, Forrest W. Crawford, Eleonora Patacchini, and Sida Peng. 2021. "A Graphical Lasso Approach to Estimating Network Connections: The Case of U.S. Lawmakers." Unpublished Working Paper.

Bhojanapalli, Srinadh, and Prateek Jain. 2014. "Universal matrix completion." 1881–1889.

Bien, Jacob, Jonathan Taylor, and Robert Tibshirani. 2013. "A lasso for hierarchical interactions." *Annals of Statistics*, 41(3): 1111–1141.

Bishop, William E, and Byron M Yu. 2014. "Deterministic symmetric positive semidefinite matrix completion." *Advances in Neural Information Processing Systems*, 27: 2762–2770.

Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Rajshri Jayaraman. 2015. "Linear Social Interactions Models." *Journal of Political Economy*, 123(2): 444–496.

Boucher, Vincent, and Aristide Houndetoungan. 2019. "Eestimating Peer Effects Using Partial Network Data." Unpublished Working Paper.

Boyd, Stephen, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. 2010. "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers." *Foundations and Trends in Machine Learning*, 3(1): 1–122.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of Peer Effects through Social Networks." *Journal of Econometrics*, 150(1): 41–55.

Breza, Emily, Arun G. Chandrasekhar, Tyler H. McCormick, and Mengjie Pan. 2020. "Using aggregated relational data to feasibly identify network structure without network data." *American Economic Review*, 110(8): 2454–84.

Candès, Emmanuel J, and Benjamin Recht. 2009. "Exact matrix completion via convex optimization." *Foundations of Computational mathematics*, 9(6): 717–772.

Candès, Emmanuel J, and Terence Tao. 2010. "The power of convex relaxation: Near-optimal matrix completion." *IEEE Transactions on Information Theory*, 56(5): 2053–2080.

Cevher, Volkan, Stephen Becker, and Mark Schmidt. 2014. "Convex Optimization for Big Data: Scalable, randomized, and parallel algorithms for big data analytics." *IEEE Signal Processing Magazine*, 31(5): 32 – 43.

Chandrasekaran, Venkat, Sujay Sanghavi, Pablo A. Parrilo, and Alan S. Willsky. 2011. "rank-sparsity incoherence for matrix decomposition." *SIAM Journal on Optimization*, 21(2): 572–596.

Chandrasekhar, Arun, and Randall Lewis. 2011. "Econometrics of Sampled Networks." Unpublished Working Paper.

Chiong, Khai Xiang, and Hyungsik Roger Moon. 2018. "Estimation of graphical models using the L 1, 2 norm." *The Econometrics Journal*, 21(3): 247–263.

DeGiorgi, Giacomo, Michele Pelllizzari, and Silvia Radaelli. 2010. "Identification of Social Interactions through Partially Overlapping Peer Groups." *American Economic Journal: Applied Economics*, 2: 241–255.

de Paula, Á, Imran Rasul, and Pedro Souza. 2020. "Identifying Network Ties from Panel Data: Theory and an Application to Tax Competition." Unpublished Working Paper.

Diestel, Reinhard. 2005. "Graph theory." *Graduate texts in mathematics*, 173.

Epple, Dennis, and Richard Romano. 2011. "Peer Effects in Education: A Survey of the Theory and Evidence." In *Handbook of Social Economics*. Vol. 1, 1053–1165.

Fosdick, Bailey K., and Peter D. Hoff. 2015. "Testing and Modeling Dependencies between a Network and Nodal Attributes." *Journal of the American Statistical Association*, 110(511): 1047–1056.

Graham, Bryan S. 2008. "Identifying social interactions through conditional variance restrictions." *Econometrica*, 76(3): 643–660.

Gramfort, Alexandre, Matthieu Kowalski, and Matti Hämäläinen. 2012. "Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods." *Physics in Medicine & Biology*, 57(7): 1937.

Griffith, Alan. 2021. "Name Your Friends but Only Five? The Importance of Censoring in Peer Effects Estimates using Social Network Data." *Journal of Labor Economics*. Forthcoming.

Harris, Kathleen Mullan. 2009. *The National Longitudinal Study of Adolescent to Adult Health (Add Health), Waves I and II, 1994-1996.* Carolina Population Center, University of North Carolina at Chapel Hill.

Hoff, Peter, Bailey Fosdick, Alex Volfovsky, and Katherine Stovel. 2013. "Likelihoods for Fixed Rank Nomination Networks." *Network Science*, 1(3): 253–277.

Hsu, Daniel, Sham M Kakade, and Tong Zhang. 2011. "Robust matrix decomposition with sparse corruptions." *IEEE Transactions on Information Theory*, 57(11): 7221–7234.

Jackson, Matthew O. 2008. *Social and Economic Networks.* Princeton, New Jersey:Princeton Press.

Király, Franz, and Ryota Tomioka. 2012. "A combinatorial algebraic approach for the identifiability of low-rank matrix completion." *arXiv preprint arXiv:1206.6470*.

Lewbel, Arthur. 2019. "The identification zoo: Meanings of identification in econometrics." *Journal of Economic Literature*, 57(4): 835–903.

Lewbel, Arthur, Xi Qu, and Xun Tang. 2021*a*. "Social Networks with Mismeasured Links." Unpublished Working Paper.

Lewbel, Arthur, Xi Qu, and Xun Tang. 2021*b*. "Social Networks with Unobserved Links." Unpublished Working Paper.

Lyle, David S. 2009. "The Effects of Peer Group Heterogeneity on the Production of Human Capital at West Point." *American Economic Jouranl: Applied Economics*, 1(4): 69–84.

Manresa, Elena. 2016. "Estimating the Structure of Social Interactions Using Panel Data." Unpublished Working Paper.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60(3): 531–542.

Miraldo, Marisa, Carol Propper, and Christiern Rose. 2021. "Identification of Peer Effects using Panel Data." Unpublished Working Paper.

Rose, Christiern D. 2017. "Identification of peer effects through social networks using variance restrictions." *The Econometrics Journal*, 20(3): S47–S60.

Rothenberg, Thomas J. 1971. "Identification in parametric models." *Econometrica*, 577–591.

Sacerdote, Bruce. 2011. "Peer Effects in Education: How Might They Work, How Big Are They, and How Much Do We Know Thus Far?" In *Hanbook of the Economics of Education*. Vol. 3, 249–277.

Singer, Amit, and Mihai Cucuringu. 2010. "Uniqueness of low-rank matrix completion by rigidity theory." *SIAM Journal on Matrix Analysis and Applications*, 31(4): 1621–1641.

Thirkettle, Matthew. 2019. "Identification and Estimation of Network Statistics with Missing Link Data." Unpublished Working Paper.

Tibshirani, Ryan J. 2014. "Adaptive piecewise polynomial estimation via trend filtering." *Annals of Statistics*, 42(1): 285–323.

von Neumann, John. 1951. *Functional Operators: The Geometry of Orthogonal Spaces.* Princeton Univ. Press.

A Proofs

Lemma 1

Proof. Rewrite Equation (7) as Equation (A.1).

$$\mathbf{\Sigma_y} - \mathbf{ZZ}' = (\mathbf{I} - \Gamma')\mathbf{U}^{-1}(\mathbf{I} - \Gamma) \tag{A.1}$$

Since we assume that $\mathbf{\Sigma_y}$ is always observed, given identification of $\mathbf{ZZ}'$, the LHS of Equation (A.1) is identified. Therefore, the RHS, $(\mathbf{I} - \Gamma')\mathbf{U}^{-1}(\mathbf{I} - \Gamma)$, is also identified. This expression is positive definite and symmetric. Therefore, it has a unique positive-definite square root: $\mathbf{M} = (\mathbf{I} - \Gamma')\mathbf{U}^{-\frac{1}{2}}$ is identified.

For each i, $\mathbf{M}_{(i,i)} = \frac{1}{\sqrt{u_i}}$, which identifies u_i for all i and thus $\mathbf{U}$. It immediately follows that we can identify $(\mathbf{I} - \Gamma') = \mathbf{U}^{\frac{1}{2}}\mathbf{M}$. Therefore, Γ is identified.

Next, from the definition of $\mathbf{Z}$ (Equation (9)),

$$\mathbf{Z} = (\mathbf{I} - \Gamma')\mathbf{U}^{-1}\eta(\mathbf{I}_R + \eta\mathbf{U}^{-1}\eta')^{-\frac{1}{2}} \tag{A.2}$$

Since $(\mathbf{I} - \Gamma')$, $\mathbf{U}^{-1}$ and $(\mathbf{I}_R + \eta\mathbf{U}^{-1}\eta')^{-\frac{1}{2}}$ are positive definite, we can rewrite (A.2) as (A.3).

$$\eta = \mathbf{U}(\mathbf{I} - \Gamma')^{-1}\mathbf{Z}(\mathbf{I}_R + \eta\mathbf{U}^{-1}\eta')^{\frac{1}{2}} \tag{A.3}$$

Since everything on the RHS of Equation (A.3) is identified, this implies that η is also identified. □

Lemma 2

Proof. We use a similar strategy to Király and Tomioka (2012) to identify sub-matrices using a zero-determinant condition. For any matrix $\mathbf{M}$, define $|\mathbf{M}|$ as its determinant. Since $\mathbf{ZZ}'$ is of rank R, each square sub-matrix of $\mathbf{ZZ}'$ must be of rank at most R. Therefore, any $R+1 \times R+1$ sub-matrix $\mathbf{P}$ must have $|\mathbf{P}| = 0$.

Suppose that we observe all but entry (k, l) in a $R+1 \times R+1$ square sub-matrix $\mathbf{P}$. Define a matrix $\mathbf{C}(\mathbf{P})$ as the matrix of cofactors of $\mathbf{P}$. That is, entry (i, j) of $\mathbf{C}(\mathbf{P})$ is the determinant of the

sub-matrix formed by deleting row i and column j. So,

$$|\mathbf{P}| = 0 = \mathbf{P}_{(k,l)}\mathbf{C}(\mathbf{P})_{k,l} + \sum_{m \neq l} \mathbf{P}_{(k,m)}\mathbf{C}(\mathbf{P})_{k,m} \tag{A.4}$$

Generically, $\mathbf{C}(\mathbf{P})_{k,l}$ is not zero. Therefore,

$$\mathbf{P}_{(k,l)} = -\frac{\sum_{m \neq l} \mathbf{P}_{(k,m)}\mathbf{C}(\mathbf{P})_{k,m}}{\mathbf{C}(\mathbf{P})_{k,l}} \tag{A.5}$$

Everything on the RHS of Equation (A.5) is observed by assumption. Therefore, $\mathbf{P}_{(k,l)}$ is identified (generically). □

Lemma 3

Proof. Let $i \in V_1'$, $k \in V_2'$. W.l.o.g., let $|V_1'| = |V_2'| = R$. Define $\mathbf{ZZ'}_{11}$ as the matrix of rows/columns of $\mathbf{ZZ'}$ corresponding to agents in V_1', and $\mathbf{ZZ'}_{22}$ as the matrix of rows/columns of $\mathbf{ZZ'}$ corresponding to agents in V_2'. $\mathbf{ZZ'}_{11}$ and $\mathbf{ZZ'}_{22}$ are identified by assumption [3] in the lemma statement.

Let $\mathbf{ZZ'}_{12}$ be the $R \times R$ matrix of rows/columns where rows correspond to agents in V_1' and columns to agents in V_2'. Define a new block matrix $\mathbf{W}$ as follows:

$$\mathbf{W} = \begin{bmatrix} \mathbf{ZZ'}_{11} & \mathbf{ZZ'}_{12} \\ \mathbf{ZZ''}_{12} & \mathbf{ZZ'}_{22} \end{bmatrix} \tag{A.6}$$

Clearly, $\mathbf{W}$ is a real, symmetric matrix and therefore diagonalizable. Define its diagonal decomposition as $\mathbf{W} = \mathbf{L\Lambda L'}$, where $\mathbf{\Lambda}$ has entries in decreasing order of magnitude along the diagonal.

$\mathbf{W}$ is a sub-matrix of $\mathbf{ZZ'}$ and therefore has rank at most R. Accordingly, it has at most R nonzero eigenvalues. Define $\mathbf{\Lambda_{11}}$ as the diagonal $R \times R$ matrix corresponding to the first R rows/columns of $\mathbf{\Lambda}$. Rewrite $\mathbf{W}$ as follows, where all sub-matrices are $R \times R$:

$$\mathbf{W} = \begin{bmatrix} \mathbf{L_{11}} & \mathbf{L_{12}} \\ \mathbf{L_{12}}' & \mathbf{L_{22}} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda_{11}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{L_{11}}' & \mathbf{L_{12}} \\ \mathbf{L_{12}}' & \mathbf{L_{22}}' \end{bmatrix} \tag{A.7}$$

$$= \begin{bmatrix} \mathbf{L_{11}\Lambda_{11}L_{11}}' & \mathbf{L_{11}\Lambda_{11}L_{12}} \\ \mathbf{L_{12}}'\mathbf{\Lambda_{11}L_{11}}' & \mathbf{L_{12}}'\mathbf{\Lambda_{11}L_{12}} \end{bmatrix} \tag{A.8}$$

Combining Equations (A.6) and (A.8) yields

$$\begin{bmatrix} \mathbf{ZZ'}_{11} & \mathbf{ZZ'}_{12} \\ \mathbf{ZZ''}_{12} & \mathbf{ZZ'}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{L_{11}\Lambda_{11}L_{11}}' & \mathbf{L_{11}\Lambda_{11}L_{12}} \\ \mathbf{L_{12}}'\mathbf{\Lambda_{11}L_{11}}' & \mathbf{L_{12}}'\mathbf{\Lambda_{11}L_{12}} \end{bmatrix} \tag{A.9}$$

Since $\mathbf{ZZ'}_{11}$ is a symmetric positive-semidefinite real matrix, it has a unique square root, $\mathbf{ZZ'}_{11}^{\frac{1}{2}}$, which is identified. Similarly, $\mathbf{ZZ'}_{22}$ has a unique square root $\mathbf{ZZ'}_{22}^{\frac{1}{2}}$, which is identified. Therefore, from Equation (A.9),

$$\begin{aligned} \mathbf{ZZ'}_{12} = \mathbf{L_{11}\Lambda_{11}L_{12}} &= \left(\mathbf{L_{11}\Lambda_{11}}^{\frac{1}{2}}\right)\left(\mathbf{L_{12}}'\mathbf{\Lambda_{11}}^{\frac{1}{2}}\right)' \\ &= \left(\mathbf{ZZ'}_{11}^{\frac{1}{2}}\right)\left(\mathbf{ZZ'}_{22}^{\frac{1}{2}}\right)' \end{aligned} \tag{A.10}$$

where $\mathbf{L_{11}\Lambda_{11}}^{\frac{1}{2}} = \mathbf{ZZ'}_{11}^{\frac{1}{2}}$ and $\mathbf{L_{12}}'\mathbf{\Lambda_{11}}^{\frac{1}{2}} = \mathbf{ZZ'}_{22}^{\frac{1}{2}}$ due to Equation (A.9). Since both terms on the RHS of Equation (A.10) are identified, $\mathbf{ZZ'}_{12}$ is identified. Since $\mathbf{ZZ'}_{(i,k)}$ corresponds to an element of $\mathbf{ZZ'}_{12}$, $\mathbf{ZZ'}_{(i,k)}$ is identified. □

Lemma 4

Proof. $\mathbf{D}$ is diagonal. Therefore, for all $j \neq i$, element (i,j) of $(\mathbf{I}-\mathbf{G'})\mathbf{D}(\mathbf{I}-\mathbf{G})$ is given in Equation (A.11).

$$\begin{aligned} (\mathbf{I}-\mathbf{G'})\mathbf{D}(\mathbf{I}-\mathbf{G})_{(i,j)} &= -\Gamma'_{(i,j)}\mathbf{D}_{(j,j)} - \mathbf{D}_{(i,i)}\Gamma_{(i,j)} + \sum_k \Gamma'_{(i,k)}\mathbf{D}_{(k,k)}\Gamma_{(k,j)} \\ &= -\mathbf{D}_{(j,j)}\Gamma_{(j,i)} - \mathbf{D}_{(i,i)}\Gamma_{(i,j)} + \sum_k \mathbf{D}_{(k,k)}\Gamma_{(k,i)}\Gamma_{(k,j)} \end{aligned} \tag{A.11}$$

Clearly, $(i,j) \notin E(\mathbf{G}), E(\mathbf{G'})$ implies that $\Gamma_{(i,j)} = \Gamma_{(j,i)} = 0$. Further, $(i,j) \notin E(\mathbf{G'G})$ implies that $\Gamma_{(k,i)}\Gamma_{(k,j)} = 0$ for all k. Therefore, the RHS of Equation (A.11) must be zero. □

Lemma 5

Proof. By Turán's Theorem, the density condition is sufficient for the existence of a *clique* of size $2R+1$ in the complement of discretized graph defined on the edge set $E(\mathbf{G_1}) \cup E(\mathbf{G_1}') \cup E(\mathbf{G_1}'\mathbf{G_1})$. This clique is a fully-connected induced subgraph of size $2R+1$. Define the set of agents in this

clique as V_1''. Lemma 4 says that $\mathbf{ZZ'}_{(i,j)}$ is identified for all $i, j \neq i \in V_1''$. Therefore, there is a sub-matrix of $\mathbf{ZZ'}$ that is of size $2R+1 \times 2R+1$ such that all but the diagonal elements are identified.

Since all but the diagonal elements of this submatrix are identified, for each $i \in V_1''$, there is a square sub-matrix of size $R+1$ of this all-but-diagonal identified sub-matrix of $\mathbf{ZZ'}$ that contains $\mathbf{ZZ'}_{(i,i)}$ and all elements but $\mathbf{ZZ'}_{(i,i)}$ are identified. Therefore, Lemma 2 guarantees that $\mathbf{ZZ'}_{(i,i)}$ is identified for all $i \in V_1''$.

Next, suppose that an agent $k \in \overline{V_1''}$ is connected to at least R agents in V_1'' in the edge set defined by $\overline{E(\mathbf{G_1}) \cup E(\mathbf{G_1}') \cup E(\mathbf{G_1}'\mathbf{G_1})}$. Therefore, for at least R unique $i \in V_1''$, $\mathbf{ZZ'}_{(k,i)}$ is identified. In turn, for every $j \in V_1''$, there exists a $R+1 \times R+1$ submatrix of $\mathbf{ZZ'}$ such that either (1) all elements are identified, or (2) all elements but $\mathbf{ZZ'}_{(k,j)}$ is identified. If (2), then Lemma 2 implies that $\mathbf{ZZ'}_{(k,j)}$ is identified for all $j \in V_1''$. Therefore, $\mathbf{ZZ'}_{(k,j)}$ is identified for all $j \in V_1''$.

This prior step implies that, given a set of agents V_1'' such that $\mathbf{ZZ'}_{(i,j)}$ is identified for all $i, j \in V_1$, we can iteratively add any agent $k \in V_1''$ who has at least R links to agents in V_1'', and that $\mathbf{ZZ'}_{(i,k)}$ is identified for all $i, k \in V_1'' \cup \{k\}$. Define V_1' as the set of agents that result when this process stops: i.e., when no more agents in $\overline{V_1''}$ are connected to at least R agents in V_1''. The three statements in Lemma 5 are thus true for the set of agents V_1'. □

Theorem 1

Proof. The strategy of this proof is to assume that there exist $(\Gamma^{(1)}, \eta^{(1)}, \mathbf{U}^{(1)})$ and $(\Gamma^{(2)}, \eta^{(2)}, \mathbf{U}^{(2)})$ and to show that these unobserved variables must be equal. First, define the edge set H as follows:

$$H = E(\Gamma^{(1)}) \cup E(\Gamma^{(1)'}) \cup E(\Gamma^{(1)'}\Gamma^{(1)}) \cup E(\Gamma^{(2)}) \cup E(\Gamma^{(2)'}) \cup E(\Gamma^{(2)'}\Gamma^{(2)}) \tag{A.12}$$

with $\overline{H}$ giving its complement. Clearly, for any $\Gamma^{(1)}, \Gamma^{(2)}$, we observe $\mathbf{ZZ'}_{(i,j)}$ whenever $(i,j) \in \overline{H}$.

Step 1: Bounds on $\|E(H)\|$ and $\delta(\overline{H})$

The density condition on Γ implies maximums on the size of H as follows:

$$\begin{aligned}\|E(H)\| &= \|E(\Gamma^{(1)}) \cup E(\Gamma^{(1)'}) \cup E(\Gamma^{(1)'}\Gamma^{(1)}) \cup E(\Gamma^{(2)}) \cup E(\Gamma^{(2)'}) \cup E(\Gamma^{(2)'}\Gamma^{(2)})\| \\ &< \|E(\Gamma^{(1)}) \cup E(\Gamma^{(1)'}) \cup E(\Gamma^{(1)'}\Gamma^{(1)})\| + \|E(\Gamma^{(2)}) \cup E(\Gamma^{(2)'}) \cup E(\Gamma^{(2)'}\Gamma^{(2)})\|\end{aligned}$$

$$< \frac{N(\frac{N}{R}-2)}{2} \tag{A.13}$$

with corresponding lower bound on $\|E(\overline{H})\|$. Similarly, minimum degree for any agent i in $\overline{H}$ is given by

$$\delta(\overline{H}) \geq (N-1) - 2\left(\frac{N(1-m)+P_1(R)m-R}{2}\right) \geq (N-P_1(R))m + (R-1) \tag{A.14}$$

Step 2: Existence of $V_1 \subset V$ where $\mathbf{ZZ'}_{(i,j)}$ identified for all $i, j \in V_1$

The condition $\|E(H)\| < \frac{N(\frac{N}{R}-2)}{2}$ (A.13) implies that for the entire set of agents V, we can apply Lemma 5. Therefore, there must exist a set $V_1 \subset V$, where $|V_1| \geq 2R+1$ and $\mathbf{ZZ'}_{(i,j)}$ is identified for all $i, j \in V_1$. Further, each agent in V_1 may have at most $(R-1)$ links to agents in $\overline{V_1}$.

Step 3: Bounds on Total Degree for agents in V_1 and $\overline{V_1}$

Suppose that $|V_1| = P < N$. By Lemma 5, each agent $i \in V_1$ may have at most $(R-1)$ links to agents in $\overline{V_1}$. Since there may be at most $P(P-1)$ links *among* agents in V_1 (where the upper bound corresponds to the graph retricted to V_1 being a clique of size $|V_1|$), this implies that

$$\sum_{i \in V_1} d_i \leq P(P-1) + (R-1)(N-P) \tag{A.15}$$

Similarly, the maximum sum of degrees for agents in $\overline{V_1}$ is

$$\sum_{i \in \overline{V_1}} d_i \leq (N-P)(N-P-1) + (R-1)(N-P) \tag{A.16}$$

In turn, the total number of links in $\overline{H}$, defined as $\|E(\overline{H})\|$ is bounded as follows:

$$\|E(\overline{H})\| = \sum_i d_i \leq P(P-1) + (N-P)(N-P-1) + (R-1)(N-P) \tag{A.17}$$

Step 4: If $2R+1 \leq P \leq P_1(R) \Rightarrow$ there exists a clique of size $2R+1$ in $\overline{V_1}$

Clearly $P = |V_1| \geq 2R+1$. Suppose that $P \leq P_1(R)$ and there exists no clique of size $2R+1$ in $\overline{V_1}$. So, by Turán's Theorem, the maximum number of links among agents in $\overline{V_1}$ is $m(N-P)^2$.

Therefore,

$$\|E(\overline{H})\| = \sum_i d_i \leq P(P-1) + (N-P)m^2 + 2(R-1)(N-P) \tag{A.18}$$

If $2R+1 \leq P \leq P_1(R)$, then the RHS of Inequality (A.18) is less than mN^2, which contradicts Inequality (A.13) in Step 1: that is, there are not enough total links in $\overline{H}$ to satisfy the minimum condition for $\|E(\overline{H})\|$. Therefore, the supposition that there does not exist a clique of size $2R+1$ must be false whenever $2R+1 \leq P \leq P_1(R)$.

Step 5: If there exists a clique of size $2R+1$ in $\overline{V_1}$, iterate

If there exists a clique of size $2R+1$ in $\overline{V_1}$, then define $V_2 \subset \overline{V_1} \subset V$ as containing all agents in this set. By analogous reasoning to Lemma 5, it must be the case that $\mathbf{ZZ}'_{(i,j)}$ is identified for all $i, j \in V_2$. By Lemma 3, it is also the case that $\mathbf{ZZ}'_{(i,j)}$ is identified for all $i, j \in V_1 \cup V_2$. Define $V_1' = V_1 \cup V_2$.

If there still exists a clique of size $2R+1$ in $\overline{V_1'}$, then define $V_1 = V_1'$ and iterate this process until no longer true. That is, iteratively add agents to V_1' until there is no longer a clique of size $2R+1$ in $\overline{V_1'}$

Step 6: $P_1(R) < P' < N \Rightarrow$ average degree of agents in $\overline{V_1'}$ is too small

Steps 4 and 5 imply that it must be the case that $P' = |V_1'| > P_1(R)$. Now, suppose that $P' < N$ (and thus $\overline{V_1'}$ is non-empty). So,

$$\sum_{i \in \overline{V_1'}} d_i \leq (N-P')^2 m + (R-1)(N-P') \tag{A.19}$$

and therefore, the *average* degree of agents in $\overline{V_1'}$ is

$$\frac{1}{N-P'} \sum_{i \in \overline{V_1'}} d_i \leq (N-P')m + (R-1) < (N - P_1(R))m + (R-1) \tag{A.20}$$

and thus the average degree of agents in $\overline{V_1'}$ is below the minimum degree in $\overline{H}$ as shown in Line (A.14), which implies a contradiction.

Step 7: Bring it all together

Steps 4-5 together imply that we can construct a set V_1' such that $|V_1'| = P' > P_1(R)$ and $\mathbf{ZZ'}_{(i,j)}$ is identified for all $i, j \in V_1'$. Step 6 says that $P_1(R) < P' < N$ contradicts the minimum degree condition for $\overline{H}$. So, $P' = N$. Therefore, $V_1' = V$ and $\mathbf{ZZ'}_{(i,j)}$ is identified for all $i, j \in V$.

Finally, since $\mathbf{ZZ'}$ is identified, Lemma 1 immediately gives identification of $(\Gamma, \eta, \mathbf{U})$.

□

Theorem 2

Proof. Let $i \in V'$. W.l.o.g., suppose that $|V'| = 2R + 1$. First, note that for all $i, j \neq i$, $\mathbf{ZZ'}_{(i,j)}$ is identified by the restriction: since $(i, j) \notin E(\Gamma) \cup E(\Gamma') \cup E(\Gamma'\Gamma)$, Lemma 4 says that we observe $\mathbf{ZZ'}_{(i,j)}$.

Since $\mathbf{ZZ'}_{(i,j)}$ is observed for all $i, j \neq i$, then there exists a $2R + 1$ sub-matrix of $\mathbf{ZZ'}$ along the main diagonal such that all but the diagonal elements are identified. Thus, for each $i \in V'$, there exists a sub-matrix of this sub-matrix of size $R + 1 \times R + 1$ such that all entries except $\mathbf{ZZ'}_{(i,i)}$ are identified. Therefore, by Lemma 2, $\mathbf{ZZ'}_{(i,i)}$ is identified for all $i \in V'$. Accordingly, $\mathbf{ZZ'}_{(i,j)}$ is identified for all $i, j \in V'$.

The prior argument applies to all agents $k, l \neq k$. Suppose that agents k, l are in different sets V_1' and V_2', each of which satisfies the Theorem criteria. Lemma 3 immediately provides that $\mathbf{ZZ'}_{(k,l)}$ is identified. Together with the prior paragraph, this implies that $\mathbf{ZZ'}_{(i,j)}$ is identified for all $i, j \in V$.

Finally, since $\mathbf{ZZ'}$ is identified, Lemma 1 immediately gives identification of $(\Gamma, \eta, \mathbf{U})$.

□

B Estimation Details

Algorithm 1 in the text nests a proximal gradient descent algorithm, which we outline here. This consists of two steps. First, the Gradient Descent step calculates the gradient and takes a step in the direction of steepest descent. When $R = 1$, the gradient is given in Equation (A.21).

$$\nabla L(\Gamma, \eta, \sigma_2) = \begin{bmatrix} \text{vec}(\frac{\partial L}{\partial \Gamma'}) \\ \frac{\partial L}{\partial \eta} \\ \frac{\partial L}{\partial \sigma^2} \end{bmatrix} \tag{A.21}$$

where vec() is the "vector" operator that reshapes a $N \times N$ matrix into a $N^2 \times 1$ vector.

$$\frac{\partial L}{\partial \Gamma} = 2(I - \Gamma')^{-1} + \frac{2}{\sigma^2(\sigma^2 + \eta'\eta)}\eta\eta'(I - \Gamma)\mathbf{S} - \frac{2}{\sigma^2}(I - \Gamma)\mathbf{S} \tag{A.22}$$

$$\begin{aligned} \frac{\partial L}{\partial \eta} &= \frac{2}{\sigma^2 + \eta'\eta}\eta - \frac{2}{\sigma^2(\sigma^2 + \eta'\eta)}(I - \Gamma)\mathbf{S}(I - \Gamma')\eta \\ &\quad + \frac{2\sigma^2}{\sigma^2(\sigma^2 + \eta'\eta)} tr((I - \Gamma)\mathbf{S}(I - \Gamma')\eta\eta')\eta \end{aligned} \tag{A.23}$$

$$\begin{aligned} \frac{\partial L}{\partial \sigma^2} &= \frac{1}{\sigma^2 + \eta'\eta} + \frac{N - 1}{\sigma^2} + \frac{2\sigma^2 + \eta'\eta}{\sigma^2(\sigma^2 + \eta'\eta)} tr((I - \Gamma)\mathbf{S}(I - \Gamma')\eta\eta') \\ &\quad - \frac{1}{(\sigma^2)^2} tr((I - \Gamma)\mathbf{S}(I - \Gamma')) \end{aligned} \tag{A.24}$$

For a given step size, this generates a new estimate $(\tilde{\Gamma}^{(z)}, \hat{\eta}^{(z)}, \hat{\sigma}^{(z)})$ defined as follows

$$\begin{bmatrix} \tilde{\Gamma}^{(z)} \\ \hat{\eta}^{(z)} \\ \hat{\sigma}^{(z)} \end{bmatrix} = \begin{bmatrix} \tilde{\Gamma}^{(z)} \\ \hat{\eta}^{(z)} \\ \hat{\sigma}^{(z)} \end{bmatrix} + s^{(z)} \nabla L(\Gamma^{(z)}, \eta^{(z)}, \sigma^{(z)}) \tag{A.25}$$

where $s^{(z)}$ is the step size at step (z). We use a constant step size with backtracking line-search.

Second, the "prox" step performs shrinkage on $\tilde{\Gamma}^{(z)}$ to arrive at $\hat{\Gamma}^{(z)}$. The prox function is defined as follows:

$$\hat{\Gamma}^{(z)} = \text{prox}(\tilde{\Gamma}^{(z)}) = \arg\min_{\Gamma} \|\Gamma - \tilde{\Gamma}^{(z)}\|_F + \lambda_{1,1}\|\Gamma\|_{1,1} + \lambda_{1,2}\|\Gamma\|_{1,2} \tag{A.26}$$

The solution to this is similar to that given in Gramfort, Kowalski and Hämäläinen (2012), who

solve it for a model using only the $L_{1,2}$ norm. That is, for each $i, j \neq i$, define $r_{ij} = |\Gamma_{(i,j)}| - \lambda_{1,1}$. Reorder these, defining an index (k) such that $r_{i(k)} \geq r_{i(k+1)}$ for all k. Next, define

$$R_{i(l)} = |\Gamma_{(i,(l))}| - \lambda_{1,1} - \frac{\lambda_{1,2}}{1+\lambda_{1,2}} \sum_{(k)=1}^{(l)} (|\Gamma_{(i,(k))}| - \lambda_{1,1}) \tag{A.27}$$

Clearly $R_{i(l)}$ is decreasing with (l). Define$(l^*) = \max_{(l)} |R_{i(l)} > 0$. This implies that

$$\text{prox}(\tilde{\Gamma}^{(z)}_{(i,j)}) = \mathbb{1}\{(k) \leq (l^*)\} \, \text{sign}(\tilde{\Gamma}^{(z)}_{(i,j)}) \left(|\Gamma_{(i,(l))}| - \lambda_{1,1} - \frac{\lambda_{1,2}}{1+\lambda_{1,2}} \sum_{(k)=1}^{(l^*)} (|\Gamma_{(i,(k))}| - \lambda_{1,1}) \right) \tag{A.28}$$

which defines $\hat{\Gamma}^{(i)}$. That is, $\hat{\Gamma}^{(z)}_{(i,j)} = 0$ for all entries other than the largest (l^*), where (l^*) is defined as above.

C Supplemental Tables and Figures

TABLE A.1
PERFORMANCE METRICS FOR SINGLE VILLAGE
($T = 1000$)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-6}$	0	0.1	1	10	100	1000	10000
Panel A: Network Density							
0.000	1.0000	0.9721	0.7814	0.2974	0.0554	0.0186	0.0070
0.010	0.9412	0.9142	0.7396	0.2861	0.0540	0.0182	0.0071
0.025	0.8743	0.8487	0.6951	0.2745	0.0527	0.0176	0.0071
0.050	0.7943	0.7744	0.6414	0.2612	0.0508	0.0170	0.0071
0.075	0.7342	0.7165	0.5998	0.2503	0.0494	0.0168	0.0072
0.100	0.6885	0.6731	0.5654	0.2418	0.0483	0.0165	0.0072
0.250	0.5209	0.5109	0.4421	0.2091	0.0445	0.0153	0.0074
0.500	0.3878	0.3813	0.3425	0.1774	0.0408	0.0129	0.0075
0.750	0.3170	0.3131	0.2823	0.1561	0.0382	0.0117	0.0075
1.000	0.2644	0.2604	0.2391	0.1408	0.0358	0.0109	0.0072
Panel B: Positive Predictive Value							
0.000	0.0528	0.0529	0.0571	0.0915	0.2417	0.3884	0.5385
0.010	0.0532	0.0535	0.0583	0.0927	0.2432	0.3874	0.5313
0.025	0.0546	0.0554	0.0600	0.0957	0.2470	0.3957	0.5313
0.050	0.0568	0.0573	0.0621	0.0984	0.2513	0.4097	0.5289
0.075	0.0583	0.0593	0.0649	0.1016	0.2540	0.4121	0.5265
0.100	0.0600	0.0606	0.0669	0.1032	0.2564	0.4181	0.5307
0.250	0.0694	0.0693	0.0732	0.1129	0.2673	0.4200	0.5342
0.500	0.0783	0.0793	0.0846	0.1260	0.2749	0.4619	0.5339
0.750	0.0883	0.0889	0.0943	0.1365	0.2849	0.4676	0.5362
1.000	0.0977	0.0982	0.1042	0.1454	0.2905	0.5000	0.5310
Panel C: Negative Predictive Value							
0.000	n/a	0.9488	0.9625	0.9635	0.9583	0.9536	0.9506
0.010	0.9535	0.9549	0.9627	0.9632	0.9581	0.9534	0.9506
0.025	0.9596	0.9618	0.9637	0.9634	0.9580	0.9533	0.9506
0.050	0.9627	0.9624	0.9637	0.9633	0.9578	0.9534	0.9506
0.075	0.9623	0.9636	0.9653	0.9635	0.9576	0.9533	0.9506
0.100	0.9631	0.9633	0.9655	0.9632	0.9575	0.9533	0.9507
0.250	0.9652	0.9644	0.9633	0.9631	0.9572	0.9529	0.9508
0.500	0.9633	0.9635	0.9637	0.9630	0.9566	0.9525	0.9508
0.750	0.9637	0.9636	0.9635	0.9627	0.9564	0.9521	0.9508
1.000	0.9633	0.9632	0.9633	0.9624	0.9560	0.9521	0.9506

Notes: $\beta = 0.25$. Village 58, network of All Relationships. See Table 1 for N and density. $T = 1,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE A.2
PERFORMANCE METRICS FOR SINGLE VILLAGE
($T = 100$)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-6}$	0	0.1	1	10	100	1000	10000
Panel A: Network Density							
0.000	1.0000	0.9999	0.9997	0.9993	0.9892	0.9047	0.4668
0.010	0.6707	0.6707	0.6706	0.6697	0.6648	0.5821	0.2381
0.025	0.5432	0.5329	0.4633	0.4347	0.4280	0.3638	0.1358
0.050	0.4488	0.4426	0.3899	0.2649	0.2545	0.2090	0.0594
0.075	0.3946	0.3896	0.3483	0.1882	0.1696	0.1339	0.0246
0.100	0.3584	0.3538	0.3198	0.1743	0.1192	0.0869	0.0112
0.250	0.2551	0.2522	0.2325	0.1404	0.0451	0.0191	0.0070
0.500	0.1876	0.1866	0.1764	0.1153	0.0350	0.0138	0.0069
0.750	0.1568	0.1560	0.1480	0.1013	0.0327	0.0125	0.0069
1.000	0.1342	0.1336	0.1275	0.0905	0.0307	0.0121	0.0069
Panel B: Positive Predictive Value							
0.000	0.0528	0.0528	0.0528	0.0529	0.0530	0.0533	0.0570
0.010	0.0546	0.0546	0.0546	0.0546	0.0547	0.0557	0.0621
0.025	0.0557	0.0561	0.0566	0.0582	0.0582	0.0594	0.0687
0.050	0.0576	0.0578	0.0590	0.0615	0.0612	0.0639	0.0807
0.075	0.0584	0.0583	0.0591	0.0654	0.0676	0.0690	0.0966
0.100	0.0593	0.0593	0.0598	0.0666	0.0692	0.0741	0.1190
0.250	0.0623	0.0623	0.0629	0.0694	0.0809	0.0965	0.1416
0.500	0.0655	0.0651	0.0667	0.0697	0.0789	0.1080	0.1422
0.750	0.0690	0.0692	0.0701	0.0712	0.0835	0.1142	0.1422
1.000	0.0717	0.0715	0.0709	0.0729	0.0878	0.1155	0.1422
Panel C: Negative Predictive Value							
0.000	n/a	1.0000	1.0000	1.0000	0.9618	0.9520	0.9509
0.010	0.9508	0.9508	0.9509	0.9508	0.9509	0.9512	0.9501
0.025	0.9507	0.9509	0.9504	0.9513	0.9512	0.9510	0.9497
0.050	0.9511	0.9511	0.9511	0.9503	0.9501	0.9501	0.9489
0.075	0.9508	0.9507	0.9506	0.9501	0.9502	0.9497	0.9483
0.100	0.9508	0.9507	0.9505	0.9501	0.9494	0.9492	0.9479
0.250	0.9504	0.9504	0.9502	0.9499	0.9485	0.9480	0.9478
0.500	0.9501	0.9500	0.9502	0.9494	0.9481	0.9480	0.9478
0.750	0.9502	0.9502	0.9502	0.9493	0.9482	0.9480	0.9478
1.000	0.9501	0.9501	0.9498	0.9492	0.9483	0.9480	0.9478

Notes: $\beta = 0.25$. Village 58, network of All Relationships. See Table 1 for N and density. $T = 100$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE A.3
PERFORMANCE METRICS FOR SINGLE VILLAGE
($T = 10$)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-6}$	0	0.1	1	10	100	1000	10000
Panel A: Network Density							
0.000	1.0000	0.9999	0.9999	0.9987	0.9885	0.9077	0.5826
0.010	0.2675	0.2675	0.2674	0.2670	0.2665	0.2705	0.2503
0.025	0.2096	0.2096	0.2095	0.2089	0.2097	0.2132	0.2066
0.050	0.1718	0.1718	0.1718	0.1714	0.1723	0.1780	0.1559
0.075	0.1558	0.1558	0.1558	0.1556	0.1562	0.1598	0.1175
0.100	0.1421	0.1413	0.1377	0.1374	0.1371	0.1354	0.0898
0.250	0.1118	0.1112	0.1061	0.0740	0.0571	0.0551	0.0292
0.500	0.0926	0.0919	0.0875	0.0634	0.0245	0.0217	0.0113
0.750	0.0838	0.0833	0.0791	0.0578	0.0234	0.0136	0.0068
1.000	0.0777	0.0772	0.0732	0.0539	0.0226	0.0111	0.0065
Panel B: Positive Predictive Value							
0.000	0.0528	0.0528	0.0528	0.0529	0.0529	0.0534	0.0549
0.010	0.0572	0.0572	0.0572	0.0572	0.0572	0.0567	0.0567
0.025	0.0557	0.0557	0.0557	0.0559	0.0557	0.0563	0.0567
0.050	0.0558	0.0558	0.0558	0.0559	0.0556	0.0562	0.0564
0.075	0.0562	0.0562	0.0562	0.0563	0.0565	0.0568	0.0564
0.100	0.0581	0.0577	0.0576	0.0578	0.0577	0.0565	0.0569
0.250	0.0579	0.0576	0.0568	0.0557	0.0528	0.0542	0.0686
0.500	0.0569	0.0573	0.0581	0.0561	0.0647	0.0556	0.0560
0.750	0.0568	0.0571	0.0574	0.0532	0.0624	0.0678	0.0421
1.000	0.0564	0.0567	0.0555	0.0536	0.0619	0.0630	0.0437
Panel C: Negative Predictive Value							
0.000	0.5000	1.0000	1.0000	0.9756	0.9529	0.9526	0.9500
0.010	0.9488	0.9488	0.9488	0.9488	0.9488	0.9486	0.9485
0.025	0.9480	0.9480	0.9480	0.9480	0.9479	0.9481	0.9482
0.050	0.9478	0.9478	0.9478	0.9478	0.9478	0.9479	0.9478
0.075	0.9478	0.9478	0.9478	0.9478	0.9479	0.9479	0.9477
0.100	0.9481	0.9480	0.9480	0.9480	0.9480	0.9478	0.9476
0.250	0.9478	0.9478	0.9477	0.9474	0.9472	0.9473	0.9477
0.500	0.9476	0.9476	0.9477	0.9474	0.9475	0.9472	0.9472
0.750	0.9476	0.9476	0.9476	0.9472	0.9474	0.9474	0.9471
1.000	0.9475	0.9475	0.9474	0.9472	0.9474	0.9473	0.9471

Notes: $\beta = 0.25$. Village 58, network of All Relationships. See Table 1 for N and density. $T = 10$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE A.4
PERFORMANCE METRICS FOR TWO-VILLAGE RESULTS
($T = 1000$, UNRESTRICTED)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-6}$	0	0.1	1	10	100	1000	10000
Panel A: Network Density							
0.000	1.0000	0.9454	0.6554	0.1922	0.0545	0.0169	0.0069
0.010	0.9258	0.8770	0.6193	0.1845	0.0472	0.0165	0.0069
0.025	0.8433	0.8027	0.5782	0.1762	0.0396	0.0160	0.0070
0.050	0.7536	0.7178	0.5270	0.1661	0.0329	0.0158	0.0071
0.075	0.6889	0.6583	0.4898	0.1589	0.0298	0.0156	0.0071
0.100	0.6363	0.6098	0.4599	0.1524	0.0291	0.0155	0.0071
0.250	0.4513	0.4361	0.3502	0.1301	0.0265	0.0149	0.0072
0.500	0.3084	0.3002	0.2526	0.1072	0.0241	0.0134	0.0072
0.750	0.2246	0.2201	0.1911	0.0919	0.0223	0.0120	0.0072
1.000	0.1712	0.1681	0.1491	0.0797	0.0208	0.0112	0.0070
Panel B: Positive Predictive Value (within Village)							
0.000	0.0499	0.0518	0.0632	0.1357	0.2721	0.4745	0.6584
0.010	0.0525	0.0542	0.0646	0.1391	0.2928	0.4739	0.6584
0.025	0.0555	0.0570	0.0674	0.1425	0.3249	0.4807	0.6601
0.050	0.0589	0.0604	0.0714	0.1475	0.3490	0.4858	0.6585
0.075	0.0616	0.0630	0.0749	0.1514	0.3622	0.4912	0.6585
0.100	0.0642	0.0656	0.0771	0.1550	0.3660	0.4922	0.6522
0.250	0.0781	0.0795	0.0912	0.1710	0.3839	0.5035	0.6555
0.500	0.0985	0.1007	0.1140	0.1909	0.4094	0.5359	0.6524
0.750	0.1231	0.1249	0.1347	0.2073	0.4268	0.5686	0.6538
1.000	0.1459	0.1484	0.1585	0.2179	0.4470	0.5902	0.6650
Panel C: Negative Predictive Value (within Village)							
0.000	n/a	0.9830	0.9752	0.9704	0.9629	0.9573	0.9543
0.010	0.9814	0.9804	0.9739	0.9702	0.9621	0.9572	0.9543
0.025	0.9798	0.9786	0.9740	0.9699	0.9614	0.9571	0.9543
0.050	0.9775	0.9766	0.9740	0.9695	0.9602	0.9570	0.9544
0.075	0.9758	0.9752	0.9740	0.9692	0.9596	0.9571	0.9544
0.100	0.9749	0.9744	0.9732	0.9690	0.9595	0.9570	0.9544
0.250	0.9732	0.9729	0.9723	0.9682	0.9592	0.9569	0.9544
0.500	0.9717	0.9718	0.9717	0.9670	0.9589	0.9567	0.9544
0.750	0.9713	0.9712	0.9701	0.9660	0.9587	0.9564	0.9544
1.000	0.9699	0.9699	0.9691	0.9646	0.9585	0.9562	0.9544
Panel D: Percent of Across-Village Links Detected (within Village)							
0.000	1.0000	0.9456	0.6591	0.1808	0.0447	0.0095	0.0025
0.010	0.9285	0.8792	0.6221	0.1732	0.0376	0.0090	0.0025
0.025	0.8457	0.8033	0.5794	0.1647	0.0300	0.0086	0.0025
0.050	0.7537	0.7213	0.5242	0.1540	0.0243	0.0082	0.0025
0.075	0.6916	0.6613	0.4851	0.1468	0.0213	0.0080	0.0025
0.100	0.6388	0.6114	0.4541	0.1402	0.0207	0.0080	0.0026
0.250	0.4460	0.4307	0.3450	0.1179	0.0182	0.0074	0.0026
0.500	0.3002	0.2912	0.2426	0.0948	0.0153	0.0061	0.0027
0.750	0.2166	0.2121	0.1811	0.0797	0.0140	0.0055	0.0026
1.000	0.1603	0.1565	0.1368	0.0684	0.0126	0.0047	0.0025

Notes: $\beta = 0.25$. Villages 8 and 10, network of All Relationships. See Table 1 for N and density. $T = 1,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE A.5
PERFORMANCE METRICS FOR TWO-VILLAGE RESULTS
($T = 1000$, RESTRICTED TO ONLY WITHIN-VILLAGE LINKS ONLY)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-7}$	0	1	10	100	1000	10000	100000
Panel A: Network Density							
0.000	1.0000	0.9715	0.7785	0.2949	0.0547	0.0236	0.0133
0.010	0.9614	0.9295	0.7514	0.2888	0.0541	0.0238	0.0133
0.025	0.9072	0.8800	0.7162	0.2801	0.0532	0.0238	0.0133
0.050	0.8381	0.8147	0.6687	0.2675	0.0519	0.0236	0.0134
0.075	0.7811	0.7612	0.6302	0.2568	0.0508	0.0236	0.0134
0.100	0.7387	0.7158	0.5985	0.2478	0.0498	0.0234	0.0134
0.250	0.5510	0.5392	0.4664	0.2108	0.0464	0.0204	0.0134
0.500	0.3870	0.3803	0.3373	0.1716	0.0426	0.0164	0.0123
0.750	0.2901	0.2861	0.2602	0.1449	0.0393	0.0145	0.0122
1.000	0.2207	0.2179	0.2011	0.1216	0.0363	0.0143	0.0121
Panel B: Positive Predictive Value (within Village)							
0.000	0.0995	0.1014	0.1151	0.2012	0.4887	0.6464	0.7371
0.010	0.1020	0.1041	0.1175	0.2036	0.4880	0.6427	0.7371
0.025	0.1057	0.1080	0.1206	0.2067	0.4942	0.6427	0.7371
0.050	0.1109	0.1124	0.1247	0.2129	0.4947	0.6464	0.7333
0.075	0.1147	0.1169	0.1288	0.2199	0.4987	0.6483	0.7333
0.100	0.1185	0.1209	0.1320	0.2253	0.5062	0.6540	0.7333
0.250	0.1393	0.1411	0.1518	0.2499	0.5229	0.6879	0.7333
0.500	0.1698	0.1714	0.1861	0.2838	0.5434	0.7250	0.7598
0.750	0.2043	0.2071	0.2214	0.3092	0.5592	0.7406	0.7640
1.000	0.2465	0.2478	0.2559	0.3403	0.5755	0.7368	0.7627
Panel C: Negative Predictive Value (within Village)							
0.000	n/a	0.9639	0.9554	0.9431	0.9230	0.9137	0.9091
0.010	0.9627	0.9611	0.9551	0.9428	0.9227	0.9137	0.9091
0.025	0.9609	0.9629	0.9536	0.9422	0.9227	0.9137	0.9091
0.050	0.9594	0.9571	0.9514	0.9419	0.9222	0.9137	0.9091
0.075	0.9549	0.9561	0.9505	0.9421	0.9219	0.9138	0.9091
0.100	0.9541	0.9544	0.9490	0.9420	0.9218	0.9138	0.9091
0.250	0.9493	0.9491	0.9462	0.9407	0.9211	0.9128	0.9091
0.500	0.9449	0.9446	0.9446	0.9387	0.9203	0.9110	0.9087
0.750	0.9433	0.9437	0.9434	0.9361	0.9193	0.9100	0.9087
1.000	0.9421	0.9418	0.9399	0.9338	0.9184	0.9098	0.9086

Notes: $\beta = 0.25$. Villages 8 and 10, network of All Relationships. See Table 1 for N and density. $T = 1,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE A.6
PERFORMANCE METRICS FOR TWO-VILLAGE RESULTS
($T = 1000$, ESTIMATED SEPARATELY)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-4}$	0	0.001	0.01	0.1	1	10	100
Panel A: Network Density							
0.000	1.0000	0.9962	0.8554	0.4019	0.0850	0.0214	0.0137
0.010	0.8812	0.8685	0.7548	0.3951	0.0902	0.0223	0.0138
0.025	0.7569	0.7474	0.6720	0.3863	0.0982	0.0234	0.0140
0.050	0.6257	0.6203	0.5746	0.3682	0.1053	0.0250	0.0146
0.075	0.5365	0.5328	0.5049	0.3482	0.1100	0.0260	0.0149
0.100	0.4705	0.4684	0.4473	0.3284	0.1122	0.0269	0.0151
0.250	0.2801	0.2799	0.2746	0.2359	0.1118	0.0299	0.0160
0.500	0.1590	0.1588	0.1581	0.1474	0.0937	0.0316	0.0162
0.750	0.1096	0.1095	0.1091	0.1048	0.0753	0.0312	0.0162
1.000	0.0771	0.0771	0.0767	0.0757	0.0621	0.0298	0.0158
Panel B: Positive Predictive Value (within Village)							
0.000	0.0995	0.0994	0.1025	0.1350	0.2643	0.4249	0.5150
0.010	0.1025	0.1029	0.1057	0.1354	0.2604	0.4215	0.5149
0.025	0.1037	0.1042	0.1084	0.1345	0.2477	0.4106	0.5147
0.050	0.1101	0.1105	0.1155	0.1362	0.2407	0.4082	0.5023
0.075	0.1167	0.1173	0.1197	0.1411	0.2393	0.4000	0.4931
0.100	0.1232	0.1236	0.1262	0.1448	0.2357	0.3929	0.4909
0.250	0.1531	0.1532	0.1542	0.1664	0.2377	0.3799	0.4850
0.500	0.1897	0.1898	0.1903	0.1990	0.2392	0.3731	0.4768
0.750	0.2145	0.2146	0.2148	0.2217	0.2493	0.3714	0.4746
1.000	0.2400	0.2400	0.2402	0.2462	0.2657	0.3770	0.4826
Panel C: Negative Predictive Value (within Village)							
0.000	n/a	0.8750	0.9185	0.9244	0.9158	0.9076	0.9063
0.010	0.9227	0.9229	0.9195	0.9240	0.9165	0.9078	0.9063
0.025	0.9138	0.9145	0.9187	0.9225	0.9166	0.9079	0.9064
0.050	0.9182	0.9184	0.9222	0.9219	0.9171	0.9084	0.9065
0.075	0.9205	0.9208	0.9211	0.9227	0.9178	0.9085	0.9064
0.100	0.9216	0.9218	0.9221	0.9227	0.9177	0.9086	0.9065
0.250	0.9214	0.9214	0.9212	0.9212	0.9179	0.9092	0.9068
0.500	0.9175	0.9176	0.9176	0.9177	0.9149	0.9094	0.9067
0.750	0.9147	0.9147	0.9146	0.9148	0.9127	0.9093	0.9067
1.000	0.9122	0.9122	0.9122	0.9125	0.9115	0.9090	0.9066

Notes: $\beta = 0.25$. Villages 8 and 10, network of All Relationships. See Table 1 for N and density. $T = 1,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE A.7
PERFORMANCE METRICS FOR SINGLE VILLAGE
($R = 1$, MOST DENSE VILLAGE)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-9}$	0	0.01	0.1	1	10	100	1000
Panel A: Network Density							
0	1.0000	1.0000	0.9981	0.9305	0.2182	0.0367	0.0193
10	1.0000	0.9998	0.9945	0.9243	0.2235	0.0374	0.0197
25	0.9935	0.9932	0.9874	0.9055	0.2259	0.0381	0.0198
50	0.9839	0.9831	0.9735	0.8354	0.2262	0.0379	0.0198
75	0.9663	0.9646	0.9578	0.6941	0.2244	0.0386	0.0200
100	0.9554	0.9552	0.9381	0.5800	0.2172	0.0384	0.0198
250	0.4557	0.4622	0.4708	0.3288	0.1668	0.0367	0.0200
500	0.1271	0.1270	0.1251	0.1160	0.0639	0.0316	0.0174
750	0.0472	0.0472	0.0468	0.0453	0.0407	0.0253	0.0167
1000	0.0371	0.0371	0.0371	0.0366	0.0320	0.0215	0.0156
Panel B: Positive Predictive Value							
0	0.1141	0.1141	0.1125	0.0725	0.2670	0.8465	0.8938
10	0.1141	0.1140	0.1093	0.0727	0.2622	0.8493	0.8957
25	0.1084	0.1081	0.1028	0.0728	0.2587	0.8520	0.8966
50	0.0997	0.0989	0.0900	0.0775	0.2583	0.8559	0.8966
75	0.0833	0.0820	0.0776	0.0876	0.2605	0.8584	0.8974
100	0.0744	0.0748	0.0729	0.0969	0.2659	0.8622	0.9138
250	0.1256	0.1272	0.1278	0.1554	0.3156	0.8977	0.9231
500	0.4274	0.4280	0.4358	0.4654	0.7620	0.9676	0.9804
750	0.9746	0.9746	0.9781	0.9849	1.0000	1.0000	1.0000
1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Panel C: Negative Predictive Value							
0	0.5000	0.5000	0.0000	0.3292	0.9285	0.9138	0.9012
10	0.5000	0.0000	0.0000	0.3792	0.9285	0.9144	0.9015
25	0.0000	0.0000	0.0000	0.4901	0.9280	0.9151	0.9017
50	0.0000	0.0000	0.0000	0.6999	0.9280	0.9151	0.9017
75	0.0000	0.0097	0.0567	0.8257	0.9282	0.9157	0.9018
100	0.0345	0.0458	0.2597	0.8621	0.9280	0.9158	0.9020
250	0.8954	0.8970	0.8980	0.9061	0.9262	0.9157	0.9024
500	0.9315	0.9315	0.9318	0.9320	0.9301	0.9137	0.9012
750	0.9284	0.9284	0.9283	0.9272	0.9234	0.9088	0.9009
1000	0.9200	0.9200	0.9200	0.9195	0.9151	0.9053	0.8998

Notes: $\beta = 0.25$. Village 10, network of All Relationships. See Table 1 for N and density. $T = 1,000,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE A.8
PERFORMANCE METRICS FOR SINGLE VILLAGE
($R = 1$, LEAST DENSE VILLAGE 1)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-9}$	0	0.01	0.1	1	10	100	1000
Panel A: Network Density							
0	1.0000	1.0000	0.9999	0.9996	0.9962	0.9634	0.5119
10	0.9419	0.9419	0.9418	0.9411	0.9336	0.8451	0.2811
25	0.7951	0.7950	0.7948	0.7922	0.7702	0.5523	0.0768
50	0.3331	0.3330	0.3328	0.3306	0.3070	0.1523	0.0206
75	0.0554	0.0554	0.0554	0.0550	0.0510	0.0288	0.0141
100	0.0186	0.0186	0.0186	0.0186	0.0182	0.0155	0.0119
250	0.0086	0.0086	0.0086	0.0086	0.0086	0.0085	0.0078
500	0.0066	0.0066	0.0066	0.0066	0.0065	0.0065	0.0062
750	0.0059	0.0059	0.0059	0.0059	0.0059	0.0059	0.0055
1000	0.0050	0.0050	0.0050	0.0050	0.0050	0.0050	0.0047
Panel B: Positive Predictive Value							
0	0.0062	0.0062	0.0062	0.0062	0.0063	0.0065	0.0122
10	0.0066	0.0066	0.0066	0.0066	0.0067	0.0074	0.0222
25	0.0078	0.0078	0.0078	0.0079	0.0081	0.0113	0.0812
50	0.0187	0.0187	0.0187	0.0189	0.0203	0.0410	0.3027
75	0.1125	0.1126	0.1126	0.1134	0.1223	0.2163	0.4416
100	0.3346	0.3346	0.3346	0.3361	0.3435	0.4033	0.5217
250	0.7260	0.7260	0.7260	0.7260	0.7257	0.7323	0.7831
500	0.9140	0.9140	0.9140	0.9140	0.9137	0.9170	0.9209
750	0.9484	0.9484	0.9484	0.9484	0.9484	0.9520	0.9618
1000	0.9858	0.9858	0.9858	0.9858	0.9858	0.9857	0.9851
Panel C: Negative Predictive Value							
0	n/a	n/a	1.0000	1.0000	1.0000	1.0000	1.0000
10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
25	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
50	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
75	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
250	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
500	0.9998	0.9998	0.9998	0.9998	0.9997	0.9997	0.9995
750	0.9994	0.9994	0.9994	0.9994	0.9994	0.9994	0.9991
1000	0.9987	0.9987	0.9987	0.9987	0.9987	0.9986	0.9984

Notes: $\beta = 0.25$. Village 3, network of Give Advice / Help Decision. See Table 1 for N and density. $T = 1,000,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE A.9
PERFORMANCE METRICS FOR SINGLE VILLAGE
($R = 1$, LARGEST VILLAGE)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-8}$	0	10	100	1000	10000	100000	1000000
Panel A: Network Density							
0.000	1.0000	1.0000	1.0000	0.9993	0.9950	0.9568	0.4159
0.010	0.9917	0.9917	0.9917	0.9911	0.9868	0.9400	0.3416
0.025	0.9794	0.9794	0.9794	0.9789	0.9732	0.9060	0.2469
0.050	0.9521	0.9521	0.9520	0.9511	0.9423	0.8151	0.1376
0.075	0.9118	0.9117	0.9115	0.9098	0.8926	0.6805	0.0805
0.100	0.8436	0.8435	0.8432	0.8440	0.8145	0.5218	0.0533
0.250	0.1112	0.1112	0.1111	0.1100	0.1001	0.0560	0.0256
0.500	0.0247	0.0247	0.0247	0.0247	0.0247	0.0244	0.0222
0.750	0.0225	0.0225	0.0225	0.0225	0.0225	0.0222	0.0197
1.000	0.0206	0.0206	0.0206	0.0206	0.0206	0.0203	0.0174
Panel B: Positive Predictive Value							
0.000	0.0225	0.0225	0.0225	0.0225	0.0226	0.0235	0.0538
0.010	0.0227	0.0227	0.0227	0.0227	0.0228	0.0239	0.0654
0.025	0.0229	0.0229	0.0229	0.0230	0.0231	0.0248	0.0905
0.050	0.0236	0.0236	0.0236	0.0236	0.0238	0.0276	0.1621
0.075	0.0246	0.0246	0.0247	0.0247	0.0252	0.0330	0.2769
0.100	0.0266	0.0266	0.0267	0.0266	0.0276	0.0430	0.4173
0.250	0.2016	0.2016	0.2018	0.2038	0.2239	0.3999	0.8571
0.500	0.8999	0.8999	0.8999	0.8999	0.8999	0.9047	0.9400
0.750	0.9507	0.9507	0.9507	0.9507	0.9507	0.9523	0.9643
1.000	0.9693	0.9693	0.9693	0.9693	0.9692	0.9689	0.9726
Panel C: Negative Predictive Value							
0.000	n/a	n/a	1.0000	1.0000	1.0000	1.0000	0.9998
0.010	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
0.025	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
0.050	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
0.075	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
0.100	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
0.250	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9994
0.500	0.9997	0.9997	0.9997	0.9997	0.9997	0.9996	0.9983
0.750	0.9989	0.9989	0.9989	0.9989	0.9989	0.9987	0.9965
1.000	0.9974	0.9974	0.9974	0.9974	0.9974	0.9972	0.9943

Notes: $\beta = 0.25$. Village 60, network of All Relationships . See Table 1 for N and density. $T = 1,000,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE A.10
PERFORMANCE METRICS FOR THREE-VILLAGE RESULTS (UNRESTRICTED)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-7}$	0	1	10	100	1000	10000	100000
Panel A: Network Density							
0.000	1.0000	0.9953	0.9701	0.8280	0.0883	0.0351	0.0136
0.010	0.9893	0.9836	0.9577	0.7341	0.0582	0.0328	0.0127
0.025	0.9722	0.9670	0.9377	0.5945	0.0476	0.0295	0.0119
0.050	0.9409	0.9344	0.8919	0.3673	0.0420	0.0252	0.0108
0.075	0.8952	0.8869	0.8188	0.2125	0.0392	0.0230	0.0102
0.100	0.8233	0.8100	0.6999	0.1167	0.0370	0.0218	0.0093
0.250	0.0682	0.0656	0.0496	0.0372	0.0307	0.0178	0.0074
0.500	0.0354	0.0354	0.0352	0.0336	0.0262	0.0148	0.0060
0.750	0.0316	0.0316	0.0315	0.0299	0.0223	0.0130	0.0055
1.000	0.0263	0.0263	0.0261	0.0245	0.0190	0.0116	0.0052
Panel B: Positive Predictive Value (within Village)							
0.000	0.0352	0.0353	0.0362	0.0425	0.3949	0.8568	0.9495
0.010	0.0355	0.0358	0.0367	0.0479	0.5970	0.8749	0.9598
0.025	0.0362	0.0364	0.0375	0.0591	0.7279	0.8961	0.9664
0.050	0.0374	0.0376	0.0394	0.0956	0.8147	0.9264	0.9693
0.075	0.0393	0.0396	0.0429	0.1651	0.8608	0.9456	0.9757
0.100	0.0427	0.0434	0.0502	0.3003	0.8958	0.9571	0.9807
0.250	0.5153	0.5356	0.7071	0.9352	0.9865	0.9977	0.9981
0.500	0.9728	0.9731	0.9734	0.9861	0.9905	1.0000	1.0000
0.750	0.9887	0.9887	0.9886	0.9890	0.9951	1.0000	1.0000
1.000	0.9906	0.9906	0.9905	0.9899	0.9971	1.0000	1.0000
Panel C: Negative Predictive Value (within Village)							
0.000	n/a	1.0000	1.0000	1.0000	0.9997	0.9947	0.9775
0.010	1.0000	1.0000	1.0000	1.0000	0.9996	0.9933	0.9767
0.025	1.0000	1.0000	1.0000	1.0000	0.9994	0.9910	0.9760
0.050	1.0000	1.0000	1.0000	0.9999	0.9990	0.9879	0.9750
0.075	1.0000	1.0000	1.0000	0.9999	0.9985	0.9863	0.9745
0.100	1.0000	1.0000	1.0000	0.9999	0.9979	0.9854	0.9737
0.250	1.0000	1.0000	0.9999	0.9996	0.9950	0.9823	0.9720
0.500	0.9992	0.9992	0.9991	0.9979	0.9905	0.9794	0.9707
0.750	0.9960	0.9960	0.9958	0.9943	0.9868	0.9775	0.9702
1.000	0.9906	0.9906	0.9905	0.9888	0.9835	0.9761	0.9698
Panel D: Percent of Across-Village Links Detected (within Village)							
0.000	1.0000	1.0000	0.9991	0.9443	0.0651	0.0045	0.0005
0.010	0.9997	0.9994	0.9980	0.8620	0.0243	0.0036	0.0004
0.025	0.9986	0.9984	0.9942	0.7117	0.0114	0.0028	0.0003
0.050	0.9924	0.9914	0.9776	0.4358	0.0059	0.0017	0.0003
0.075	0.9739	0.9710	0.9329	0.2356	0.0039	0.0010	0.0002
0.100	0.9305	0.9205	0.8245	0.1072	0.0026	0.0007	0.0002
0.250	0.0434	0.0397	0.0170	0.0008	0.0000	0.0000	0.0000
0.500	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.750	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Notes: $\beta = 0.25$. Villages 8, 10, and 54, network of All Relationships. See Table 1 for N and density. $T = 1,000,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE A.11
PERFORMANCE METRICS FOR THREE-VILLAGE RESULTS (RESTRICTED TO ONLY WITHIN-VILLAGE LINKS ONLY)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-7}$	0	1	10	100	1000	10000	100000
Panel A: Network Density							
0.000	1.0000	0.9880	0.8814	0.1354	0.0625	0.0270	0.0131
0.010	0.9716	0.9590	0.8533	0.1327	0.0622	0.0270	0.0131
0.025	0.9260	0.9136	0.8084	0.1291	0.0613	0.0269	0.0131
0.050	0.8500	0.8368	0.7119	0.1239	0.0600	0.0263	0.0130
0.075	0.7539	0.7397	0.5844	0.1196	0.0590	0.0261	0.0130
0.100	0.6195	0.6006	0.4295	0.1149	0.0576	0.0260	0.0129
0.250	0.1173	0.1169	0.1136	0.1038	0.0523	0.0241	0.0124
0.500	0.1057	0.1057	0.1047	0.0935	0.0482	0.0227	0.0123
0.750	0.0945	0.0945	0.0933	0.0811	0.0444	0.0214	0.0120
1.000	0.0783	0.0782	0.0768	0.0663	0.0408	0.0203	0.0120
Panel B: Positive Predictive Value (within Village)							
0.000	0.1051	0.1064	0.1193	0.7671	0.9322	0.9604	0.9718
0.010	0.1082	0.1096	0.1232	0.7814	0.9364	0.9603	0.9718
0.025	0.1135	0.1151	0.1300	0.8024	0.9422	0.9602	0.9717
0.050	0.1237	0.1256	0.1477	0.8353	0.9526	0.9672	0.9748
0.075	0.1394	0.1421	0.1799	0.8630	0.9567	0.9732	0.9778
0.100	0.1697	0.1750	0.2447	0.8968	0.9664	0.9746	0.9776
0.250	0.8954	0.8989	0.9228	0.9691	1.0000	1.0000	1.0000
0.500	0.9735	0.9735	0.9752	0.9886	1.0000	1.0000	1.0000
0.750	0.9887	0.9887	0.9885	0.9909	1.0000	1.0000	1.0000
1.000	0.9905	0.9905	0.9904	0.9938	1.0000	1.0000	1.0000
Panel C: Negative Predictive Value (within Village)							
0.000	n/a	1.0000	1.0000	0.9985	0.9500	0.9187	0.9064
0.010	1.0000	1.0000	1.0000	0.9984	0.9500	0.9186	0.9064
0.025	1.0000	1.0000	1.0000	0.9983	0.9495	0.9185	0.9064
0.050	1.0000	1.0000	1.0000	0.9982	0.9489	0.9182	0.9064
0.075	1.0000	1.0000	1.0000	0.9978	0.9483	0.9182	0.9063
0.100	1.0000	1.0000	1.0000	0.9976	0.9475	0.9181	0.9063
0.250	0.9999	0.9999	0.9996	0.9950	0.9442	0.9170	0.9061
0.500	0.9975	0.9975	0.9967	0.9860	0.9402	0.9156	0.9060
0.750	0.9871	0.9870	0.9858	0.9730	0.9364	0.9144	0.9058
1.000	0.9701	0.9700	0.9685	0.9580	0.9329	0.9135	0.9057

Notes: $\beta = 0.25$. Villages 8, 10, and 54, network of All Relationships. See Table 1 for N and density. $T = 1,000,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.

TABLE A.12
PERFORMANCE METRICS FOR THREE-VILLAGE RESULTS (VILLAGES ESTIMATED SEPARATELY)

$\lambda_{1,1}$	$\lambda_{1,2}$						
$\times 10^{-4}$	0	0.001	0.01	0.1	1	10	100
Panel A: Network Density							
0	1.0000	1.0000	0.9953	0.8912	0.1297	0.0272	0.0148
10	0.9523	0.9514	0.9200	0.6771	0.1296	0.0289	0.0154
25	0.6911	0.6905	0.6792	0.4895	0.1041	0.0283	0.0151
50	0.5407	0.5378	0.5120	0.3321	0.0877	0.0262	0.0149
75	0.3800	0.3783	0.3621	0.2347	0.0824	0.0245	0.0143
100	0.2872	0.2867	0.2785	0.1781	0.0772	0.0230	0.0137
250	0.1273	0.1289	0.1309	0.0963	0.0542	0.0176	0.0112
500	0.0386	0.0386	0.0381	0.0359	0.0223	0.0128	0.0087
750	0.0152	0.0152	0.0151	0.0147	0.0134	0.0090	0.0066
1000	0.0116	0.0116	0.0116	0.0115	0.0102	0.0074	0.0057
Panel B: Positive Predictive Value (within Village)							
0	0.1051	0.1051	0.1009	0.0806	0.4279	0.9092	0.9415
10	0.0897	0.0895	0.0893	0.0981	0.4260	0.9161	0.9439
25	0.1151	0.1150	0.1138	0.1361	0.5014	0.9374	0.9538
50	0.1241	0.1244	0.1257	0.1731	0.5230	0.9465	0.9612
75	0.1488	0.1485	0.1515	0.2130	0.5082	0.9430	0.9597
100	0.1726	0.1730	0.1745	0.2500	0.5011	0.9411	0.9639
250	0.2460	0.2459	0.2446	0.3056	0.4932	0.9485	0.9670
500	0.5463	0.5469	0.5540	0.5837	0.8358	0.9806	0.9905
750	0.9810	0.9810	0.9837	0.9888	1.0000	1.0000	1.0000
1000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Panel C: Negative Predictive Value (within Village)							
0	n/a	n/a	0.0000	0.6944	0.9430	0.9174	0.9074
10	0.5872	0.5898	0.7124	0.8802	0.9426	0.9190	0.9080
25	0.9171	0.9169	0.9133	0.9246	0.9409	0.9191	0.9079
50	0.9173	0.9173	0.9165	0.9287	0.9350	0.9175	0.9078
75	0.9217	0.9213	0.9212	0.9280	0.9311	0.9159	0.9073
100	0.9221	0.9222	0.9217	0.9263	0.9280	0.9146	0.9068
250	0.9154	0.9157	0.9159	0.9162	0.9171	0.9100	0.9047
500	0.9126	0.9126	0.9127	0.9127	0.9116	0.9062	0.9026
750	0.9084	0.9084	0.9084	0.9081	0.9070	0.9030	0.9008
1000	0.9054	0.9054	0.9054	0.9053	0.9041	0.9015	0.9000

Notes: $\beta = 0.25$. Villages 8, 10, and 54, network of All Relationships. See Table 1 for N and density. $T = 1,000,000$, $\sigma_\eta = 1$ in all simulations. $R = 1$ in estimation.