Essays on Social Networks in Development Economics

by

Arun Gautham Chandrasekhar

B.A. Mathematics, B.A. Economics, Columbia University (2007)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosopy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

Author . . . . . .
Department of Economics
May 15, 2012

Certified by . . . . . .
Abhijit Banerjee
Ford International Professor of Economics
Thesis Supervisor

Certified by . . . . . .
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Thesis Supervisor

Certified by . . . . . .
Victor Chernozhukov
Professor of Economics
Thesis Supervisor

Accepted by . . . . . .
Michael Greenstone
3M Professor of Environmental Economics
Chairman, Departmental Committee on Graduate Studies

Essays on Social Networks in Development Economics

by

Arun Gautham Chandrasekhar

Submitted to the Department of Economics
on May 15, 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosopy

Abstract

This thesis examines the role that social networks play in developing economies. The first two chapters analyze econometric issues that arise when researchers work with sampled network data. The final two chapters study how the embedding of agents in a network affects a group's ability to overcome weak contracting institutions and what models of social learning are important in describing the diffusion of information. These chapters make use of experiments that I conducted in rural Karnataka, India.

The first chapter (co-authored with Randall Lewis) examines the econometric difficulties that applied researchers face when using partially observed network data. In applied work, researchers generally construct networks from data collected from a partial sample of nodes. Treating this sampled network as the true network of interest, the researcher constructs statistics to describe the network or specific nodes and employs these statistics in regression or GMM analysis. This chapter shows that even if nodes are selected randomly, partial sampling leads to non-classical measurement error and therefore bias in estimates of the regression coefficients or GMM parameters. We provide analytical and numerical examples to illustrate the severity of the biases in common applications and discuss possible solutions. Our analysis of the sampling problem as well as the proposed solutions are applied to rich network data of Banerjee et al. (2012) from 43 villages in Karnataka, India.

In the second chapter, I develop an econometric method to cope with sampled network data. I develop a method, graphical reconstruction, by which a researcher can consistently estimate the economic parameters of interest. Graphical reconstruction uses the available (partial) network data to predict the missing links and uses these predictions to mitigate the biases. As each network may be generated by a different network formation model, the asymptotic theory allows for heterogeneity in the network formation process across graphs.

The third chapter (co-authored with Cynthia Kinnan and Horacio Larreguy) analyzes how social networks affect the provision of informal insurance. Social networks are understood to play an important role in smoothing consumption risk, particularly in developing countries where formal contracts are limited and financial development is low. Yet understanding *why* social networks matter is confounded by endogeneity of risk-sharing partners. This chapter, first, examines the causal effect of close social ties between individuals on their ability to informally insure one another. Second, we examine how the interaction of social proximity and access to savings affects consumption smoothing. Theoretically, they could be complements or substitutes. Savings access may crowd out insurance unless social proximity is high, in which case it benefits the highly connected. Or savings may crowd out risk sharing among the highly connected while helping the less connected smooth risk intertemporally. By conducting a framed field experiment in Karnataka, India, we study the relationships between inability to commit to insurance, ability to save, and social proximity. We find that limited commitment reduces risk sharing, but social proximity sub-

stitutes for commitment. On net, savings allows individuals to smooth risk that cannot be shared interpersonally, with the largest benefits for those who are weakly connected in the network.

The final chapter (co-authored with my classmates Horacio Larreguy and Juan Pablo Xandri) attempts to determine which models of social learning on networks best describe empirical behavior. Theory has focused on two leading models of social learning on networks: Bayesian and DeGroot rules of thumb learning. These models can yield greatly divergent behavior; individuals employing rules of thumb often double-count information and may not exhibit convergent behavior in the long run. By conducting a unique lab experiment in rural Karnataka, India, set up to exactly differentiate between these two models, we test which model best describes social learning processes on networks. We study experiments in which seven individuals are placed into a network, each with full knowledge of its structure. The participants attempt to learn the underlying (binary) state of the world. Individuals receive independent, identically distributed signals about the state in the first period only; thereafter, individuals make guesses about the underlying state of the world and these guesses are transmitted to their neighbors at the beginning of the following round. We consider various environments including incomplete information Bayesian models and provide evidence that individuals are best described by DeGroot models wherein they either take simple majority of opinions in their neighborhood.

Thesis Supervisor: Abhijit Banerjee
Title: Ford International Professor of Economics

Thesis Supervisor: Esther Duflo
Title: Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics

Thesis Supervisor: Victor Chernozhukov
Title: Professor of Economics

For Appa, who redefines the role of supportive father. You show an unyielding enthusiasm and interest in all that I do. Your relentless confidence in me, no matter the circumstance, continues to drive me.

Acknowledgments

I am deeply indebted to my advisors Abhijit Banerjee, Esther Duflo, Matt Jackson and Victor Chernozhukov. Abhijit and Esther taught me how to think and construct a viewpoint in economics. They nurtured my curiosity and entertained my often odd-ball perspectives, teaching me when to trust myself and when to lean on established results. Moreover, they showed me that compelling empirical work often requires a careful mix of different skills and taught me the value of collaboration and co-authorship. In addition to having the opportunity to learn from the both of them, I feel blessed to have been a part of the wonderful development economics community they built at MIT and J-PAL.

I had the privilege of learning about social networks from Matt Jackson. Matt is an incredibly warm and supportive mentor, and he took me in as one of his students despite my being from another university. He spent significant time providing advice on my research across varying topics, including empirical networks analyses, experimental methodology, and even econometric theory. Matt has taught me how to keep my work clear and simple by making the core arguments required without adding baroque trappings. I look forward to his friendly spirit and generosity as his colleague in the coming years.

Victor Chernozhukov encouraged me from an early stage to study the intersection of econometrics and network analysis and apply my findings to development. Victor's advice and insights on both theoretical and applied research are extremely helpful, and he has pushed me to be meticulous and rigorous in my own work. His inquisitive nature and thirst for deepening his knowledge of methodological tools is infectious. Victor has repeatedly encouraged me to "learn by doing" and conduct research on new and varied topics whenever possible. I am privileged to have him for an advisor.

Special thanks to Anna Mikusheva and Rob Townsend, who both put in countless hours helping me hone my ideas. I have also benefited from conversations with Daron Acemoglu, Rachel Glennerster, Rema Hanna, Jerry Hausman, Whitney Newey, Ben Olken and Tavneet Suri. I owe thanks to Leigh Linden, who introduced me to development while I was at Columbia, and Ted Miguel, who encouraged me to pursue a PhD in economics.

I am grateful for having an amazing set of classmates from whom I learned a great deal. I am even more fortunate that my class became a very tight-knit group. In particular, I am lucky to have gotten to know Emily Breza, Horacio Larreguy, Juan Pablo Xandri and Luis Zermeño whose friendship, support and advice in both research and life means the world to me. I am also grateful to have become friends and colleagues with Gabe Carroll, Paul Goldsmith-Pinkham, Ben Golub, Nathan Hendren, Felipe Iachan, Anil Jain, Cynthia Kinnan, Randall Lewis, Anton Kolotilin, Paul Schrimpf, Ashish Shenoy, Ashley Swanson, and Xiao Yu May Wang. You have all impacted my work in numerous ways and, moreover, socializing over ping pong, pub trivia, video games and other forms of general merriment with each of you has made the last five years a wonderful experience.

I must also mention Albert Fang, Isaac Mbiti, Adam Sacarny, JY Ping and Jan Zilinsky who, over the years, have shaped my perspectives in many ways while also spending countless hours killing zombies, playing ping pong, football, and fútbol, or just sitting in coffee shops with me.

My graduate research would not have been possible without the financial (and often logistical) support of the National Science Foundation's Graduate Research Fellowship, the George and Obie Shultz Fund, the Center for Microfinance at the Institute for Financial Management and Research, the Consortium on Financial Systems and Poverty, the Russell Sage Foundation, MIT's Center for International Studies, Innovations for Poverty Action, MIT International Science and Technology Initiatives, and the Sanford S. Parker Prize. Moreover, my field work went smoothly only due to the tireless and amazing work of both Gowri Nagraj, who has been my rock and guide in many ways, and Manaswini Rao, who has been an excellent research manager.

I owe a very large debt of gratitude to my (even larger) family, most of whom reside in Bangalore. Conducting my research in a region in which I have the support of a vast and deep familial network has been a tremendous experience. My grandfather, aunts, uncles, cousins, nephews and nieces have all in their own way contributed to this process and have built a nurturing environment for me in Bangalore. In particular Bali and Kalpana provided me with a home away from home from which I could conduct field work.

I must thank my parents and sister for the support they have given me over the years. Amma and Appa have always encouraged me to learn and develop my intellectual interests in my own, often haphazard way. They have spent countless hours helping me translate research protocols into Kannada and providing logistical help. My younger sister and best friend, Radhika, often plays the role of an older sister in many ways. I am ever grateful for her support.

Finally, this would not have been possible without Archana Prakash, who I will have the privilege of calling my wife come mid-August of this year. Arch has supported me in innumerable ways. She has an uncanny ability to know exactly how I am feeling and, more importantly, how to pick me up no matter what the situation. Arch is the most loving person I know and I am grateful for her unwavering support and faith in me throughout this exhausting process, and as we begin our life together.

Contents

Chapter 1

Econometrics of Sampled Networks

1.1 Introduction

A growing literature examines social networks and their implications for economic outcomes (see e.g., (Jackson 2008c; Jackson 2009a; Jackson 2009b) for an extensive survey of the literature). A network represents a set of connections (edges) among a collection of agents (nodes). For example, in a village network, nodes may represent households and edges may represent risk-sharing ties between households. Applied researchers typically construct a network from data that has been collected from a partial sample of nodes rather than from all nodes in the network. Henceforth, such a network will be called the "sampled network." This sampled network is naively treated as the true network of interest. The researcher uses a collection of sampled networks to estimate how network structure affects economic outcomes. This paper highlights, examines, and addresses econometric problems that arise when a researcher studies these network effects using sampled network data.

Concrete examples of network-based regressions in applied work include (Kremer and Miguel 2007), who study the diffusion of deworming pill take-up, and (Hochberg, Ljungqvist, and Lu 2007), who regress fund performance on measures of network importance of venture capital firms.[1] The applied work typically has low sampling rates (the share of nodes sampled), with a median of 25%, and 2/3 of the papers having a sampling rate below 51% (see Figure 1-1). Despite the prevalence of partial sampling, its implications for the estimation of economic parameters are rarely considered. One exception is (Conley and Udry 2010) who study the diffusion of information among pineapple farmers in Ghana. Aware of the sampling problem, they conduct robustness exercises.

Our goal is to analyze the effect of using sampled network data on the estimation of parameters

[1]There are numerous other examples. (Kinnan and Townsend 2011) study whether whether households that are socially closer to credit sources smooth consumption better. (Leider, Möbius, Rosenblat, and Do 2009) and (Goeree, McConnell, Mitchell, Tromp, and Yariv 2010) study the effect of social proximity between pairs on the offers made in dictator games. (Alatas, Banerjee, Chandrasekhar, Olken, and Hanna 2011) and examines whether networks with better diffusion properties actually induced greater information spreading. (De Giorgi, Pellizzari, and Redaelli 2010) study how network neighbors' major choices affect a student's own major choice.

in network models of economic behavior. Henceforth, we call these the "economic parameters" without meaning to suggest that network formation is not economic.[2] In general, we are interested in parameters in a generalized method of moments (GMM) model, motivated by theory, describing the behavior of nodes in a network. The biases in estimates of economic parameters have not yet been systematically dealt with. While GMM is a general framework, two common classes of models allow us to explicitly characterize biases and are easier to work with due to their linearity: regressions of economic outcomes on network characteristics and regressions of a node's outcomes on its network neighbors' outcomes. After characterizing the biases, we propose two new strategies to correct such biases: a set of analytical corrections for commonly used network statistics and a two-step estimation procedure using graphical reconstruction that can be applied more broadly.

We focus on a running example throughout the paper: the diffusion of microfinance in 43 villages in rural Karnataka, India ((Banerjee, Chandrasekhar, Duflo, and Jackson 2011)). A microfinance institution (MFI) based in Bengaluru expanded into these villages. Upon entering a village, the MFI informed certain households about its intentions and asked them to encourage others to join. The researcher wants to estimate how networks affect the diffusion of microfinance participation through these villages.

The present paper makes three core contributions. Our first contribution is to highlight and analyze the biases in estimates of economic parameters when using sampled network data. We develop analytical examples for commonly used network statistics, motivated by a number of applied questions concerning diffusion of information, social collateral, and risk-sharing. Next, we derive the corresponding biases that emerge when each of these statistics is used in regression. We show that the standard argument for attenuation due to classical measurement error does not apply; coefficients may expand, attenuate, or switch signs depending on the network statistic of interest. In addition, we consider a model in which a node's outcome depends on its peers' outcomes and a node's peer group is defined by the set of its social connections ((Bramoulle, Djebbari, and Fortin 2009; De Giorgi, Pellizzari, and Redaelli 2010)). We show that the instrumentation technique used in the literature to overcome the reflection problem ((Manski 1993)) in such models is invalid since the measurement error in the instrument will be correlated with the measurement error in the endogenous variable. Similarly, we consider GMM estimation of the (Jackson and Rogers 2007) model of diffusion and show that sampling the network induces expansion bias in the diffusion parameter. We supplement our analysis with numerical evidence for a wide array of examples to illustrate how sensitive econometric estimation is to the sampling of a network. In our numerical experiments, we estimate many models across a number of network statistics. At a sampling rate of $1/3$ we find that the estimates of the economic parameters have a mean absolute bias of 90% with a maximum of 260% for network-level regressions and a mean absolute bias of 63% with a maximum of 91% for node-level regressions.

[2]Parameters which describe the process by which networks are formed certainly are economic. We reserve "economic parameters" in our environment for parameters that describe a process that occurs on fixed networks.

Our second contribution is to develop two strategies to alleviate the biases: analytical corrections that apply to commonly used network statistics and two-step estimation using graphical reconstruction, which uses the observed part of the data to probabilistically reconstruct the missing part and then estimate the economic parameter accordingly. By explicitly characterizing the biases, we derive simple bias corrections when the problem is tractable. We discuss several corrections and explore their reliability in addressing the biases. While computationally simple and easy to implement, these methods are typically limited to network-level regressions and are dependent on the particular network statistic of interest. Thus we discuss the application of a second, more general method that works well in practice – estimation by graphical reconstruction – to consistently estimate the economic parameter. This technique does not limit the researcher to network-level regressions nor to specific and tractable network statistics.[3]

To demonstrate another practical application of our results, in our third contribution we describe how researchers can employ our framework to make better decisions in collecting sampled network data, given their budget constraints. We provide an algorithm to assess the trade-off between the number of networks in a sample and the sampling rate a researcher uses. This exercise is similar in spirit to power calculations frequently used in applied field work. First, the researcher obtains 100%-sampled network data for a small number of randomly chosen villages, using a pilot budget. Second, the researcher performs a numerical experiment by simulating outcome data from a specification that the researcher anticipates studying. In our microfinance example, the researcher simulates outcome data as a function of the path length from the initially informed households by assuming a regression coefficient and an R-squared. Third, the researcher draws, with replacement, a set of villages sampled at a given rate such that her budget is exhausted. By applying graphical reconstruction, the researcher can assess the mean-squared error minimizing choice of sampling rate.

We then apply our analysis of the sampling problem as well as the proposed solutions to sampled network data, collected in part by the authors, from 43 villages in Karnataka, India. (Banerjee, Chandrasekhar, Duflo, and Jackson 2011) study the diffusion of microfinance and, inspired by this analysis, we study natural specifications motivated by diffusion theory. We examine parameter estimates using the raw sampled data and compare them to those obtained by applying graphical reconstruction or analytical corrections. We find that applying our methods at times greatly changes parameter estimates and economic inferences. For instance, the impact of the network importance of initially informed households on the microfinance take-up rate in the village is under-estimated by 33% using the raw sampled network data when compared to using graphical reconstruction. In addition, a regression of a node's take-up decision on its neighbors' decisions shows that endogenous network effects may be severely under-estimated (with a 60% bias relative to the corrected estimate) or even switch signs (with a 166% bias relative to the corrected estimate). Moreover, regression coefficients in several specifications are not significantly different from zero at conventional levels

[3]The theory for this method is developed in Chandrasekhar (2012).

when using the raw sampled data but are significantly different when applying the reconstruction estimator.

Related literatures across a number of fields including economics, epidemiology, statistics, sociology, and computer science have noted problems due to partial network data. The classical literature begins with (Granovetter 1973), (Frank 1980; Frank 1981), and (Snijders 1992) who identify how average degree and clustering are affected by several modes of random sampling. (Rothenberg 1995) provides an excellent overview of the literature. More recently, the literature has focused on two classes of numerical experiments, typically with a single network.[4] The first class documents biases that emerge when estimating parameters in a network formation model with partial data (e.g., in economics, (Santos and Barrett 2008)). Second, the literature numerically describes behavior of certain network statistics under sampling (e.g., in epidemiology, (Ghani, Donnelly, and Garnett 1998) and sociology, (Kossinets 2006)). (Handcock and Gile 2010) offer the straightforward solution to the first problem: by augmenting the likelihood to include the sampling scheme one can, in expectation, recover the correct network formation parameter.[5] Finally, (Golub and Jackson 2010b) explain how selection bias affects the estimation of a diffusion process and solve the puzzle raised by (Liben-Nowell and Kleinberg 2008) as to why chain letter data exhibited very long average paths.

Our work builds on the above literature, with several key differences. The literature typically has not focused on nor developed a methodology to consistently estimate parameters from models of behavior on networks with sampled data. The substantive distinction here must be stressed. We are not interested in recovering the structural properties of the unobserved part of the network *per se*; instead, our goal is to understand the biases in estimation of these economic parameters and develop a method to recover them. Moreover, given that we are interested in a setting with many networks, we are forced to respect the substantial heterogeneity across networks. The technical requirements to control the incidental parameter problem are akin to those made in nonlinear panel with fixed effects literature and are discussed extensively in Chandrasekhar (2012).

Finally, we note that partially observed networks constitute a special case of a broader class of network measurement problems.[6] Our framework for analyzing the nature of biases in economic parameter estimates may provide intuition in these environments as well. If the researcher misspecifies the network, biases will emerge. For example, if one is interested in economic behavior of individuals in a social network, but one defines the network based on social media data, biases will be induced as individuals may meaningfully interact with a subset of their social media neighbors. Another example is survey fatigue. Imagine that an individual forgets to name each of her

[4](Santos and Barrett 2008) also provide an extensive discussion of survey methodology and (Thompson 2006) discusses sampling methodology and inferences on the degree distribution and network size.

[5]In simulations (Huisman 2009) shows that for small amounts of missing data, an ad hoc method of imputing edges may reduce the bias in the network characteristics, though this is insufficient for larger amounts of missing data.

[6]Observed network data may miss information for a variety of reasons (see (Kossinets 2006) for a discussion).

connections independently with probability $1-p$. Studying the OR network (where we denote two nodes as connected if either names the other) yields a network with only $1-(1-p)^2$ of the average number of friendship ties, while studying the AND network (where we denote two nodes as connected if both name each other) produces a network with only p^2 of the average number of friendship ties. A straightforward argument extends the results in section 1.3, with minor modification, to characterize the biases present in this example. Another common environment that induces bias is survey top-coding, where the survey limits the number of edges an individual can name.

The rest of the paper is organized as follows. Section 1.2 establishes the framework. The main results are in sections 1.3 and 1.4. Section 1.3 provides analytical examples of bias along with corrections. Section 1.4 discusses graphical reconstruction estimation. Section 1.5 contains numerical experiments which supplement sections 1.3 and 1.4. In section 1.6 we offer an algorithm for researchers to trade off the sampling rate against the number of networks. Section 1.7 concludes. All proofs are in the appendices.

1.2 Framework

In this section we establish the framework. Section 1.2.1 introduces the notation, section 1.2.2 presents the econometric environment, and section 1.2.3 previews the asymptotic frame.

1.2.1 Notation and Setup

A network or a *graph* is a pair $G=(V,E)$ consisting of a set V of *nodes* and a set E of *edges*, with $n:=|V|$. Nodes i and j are either connected or unconnected (the graph is unweighted) and if i is connected to j, then j is connected to i (the graph is undirected). Most of what follows in this paper is applicable to directed and weighted graphs, though following the bulk of the applied research we restrict our attention to the undirected, unweighted case. A graph with n nodes is a member of the set of all undirected, unweighted graphs, denoted by $\mathcal{G}_n$.

A graph is represented by its *adjacency matrix*, $A:=A(G)$. It is a matrix of 0s and 1s that depicts whether two nodes are connected, where $A_{ij}=\mathbf{1}\{ij\in E\}$ with the convention that $A_{ii}=0$. We denote the *neighborhood* of i, the set of nodes it is connected to, by $N_i:=\{j\in V: ij\in E\}$. Researchers are interested in economic models where an economic behavior or outcome is predicted by network statistics. We let $w(G)$ represent a d_w-dimensional vector of these network statistics. Since the data set may contain multiple networks, we use R to denote the number of graphs. The researcher is interested in *economic parameter* β_0.

1.2.1.1 Sampling

Typically researchers obtain one of two types of sampled network data. First, the researcher may survey a set of m nodes and ask each node about the social connections with the other $m-1$ nodes

in that data set. We call this the *induced subgraph*, as it restricts the network among those who are sampled. Second, the researcher may have a list of the nodes in the network (e.g., a household census list in a village). From this list, a sample of m nodes may be surveyed. These nodes can name their social connections, not only to other $m-1$ surveyed nodes, but connections to anyone from the list of n nodes. We call this the *star subgraph*.

Let S be the set of surveyed nodes, randomly chosen from V, with $m = |S|$. Let $m = \lfloor \psi n \rfloor$, where ψ is the *sampling rate*. The researcher obtains a subgraph of the graph in question. There are two potential resulting networks: the induced subgraph $G^{|S} = (S, E^{|S})$, which consists of the sampled nodes and the edges restricted to the set of surveyed nodes ($E^{|S}$), and the star subgraph $G^S = (V, E^S)$, where E^S are edges such that at least one of the nodes is in S.

Figure 1-2 provides an illustration of the problem that this paper intends to address. Figure 1-2(a) displays G, the target network, Figure 1-2(c) shows the induced subgraph and Figure 1-2(e) depicts the star subgraph. We will also write $\boldsymbol{A} = (A^{obs}, A^{mis})$ to denote the observed and missing part of the adjacency matrix, which are random variables under the sampling procedure. Although this framework idealizes the random sampling used in many applied contexts, our setting can easily be extended to other sampling methods such as independent edge sampling or snowball sampling.[7]

1.2.2 Econometric Models

The researcher intends to study economic behavior on R networks, $\{G_r : r = 1, ..., R\}$. For simplicity, we assume every network has n nodes. An economic process has taken place on every network and can be described by an econometric model depending on an *economic parameter* β_0. Returning to the microfinance example, information about microfinance has been introduced to certain households in every village and households decide to participate as the information propagates throughout the villages. Our goal is to estimate an economic parameter. We could easily do so if the networks were fully observed. The general framework for analyzing such models is to presume that a conditional moment restriction is satisfied,

$$\mathrm{E}\left[m(y, w(\boldsymbol{G}); \beta_0) | \boldsymbol{G}\right] = 0. \tag{1.1}$$

where $y \in \mathbb{R}^{d_y}$ is an outcome random variable, $m(\cdot, \cdot; \cdot)$ is a moment function, $w(\cdot)$ is function on $\mathcal{G}_n$, and $\beta \in \mathcal{B}$ is a parameter with true value β_0.

Examples include discrete choice models, stopping time models (e.g., (Iyer and Puri ming)), quantile regression (e.g., Angelucci et al., 2010), and network-based matching models (e.g., (Aral and Walker ming; Banerjee, Chandrasekhar, Duflo, and Jackson 2011)). More generally, our results apply to indexed GMM models with parameter $\beta_0(u)$ where $u \in \mathcal{U}$ (e.g., time in a stopping time model or quantile in quantile regression). Partial sampling will generally generate biases as the

[7] Our graphical reconstruction solutions often apply to missing-at-random samples, where the probability of graph information being missing is independent of the missing data itself ((Rubin 1976)).

moment will be a nonlinear function of the network statistic, so the estimated parameter will be inconsistent.

While GMM is a general framework, two common classes of econometric models with network data are easier to analyze due to their linearity. The first class consists of models wherein economic outcomes are regressed on network characteristics. The second class consists of models where a node's outcome depends on its network neighbors' outcomes.

Regression of Economic Outcomes on Network Characteristics

A researcher wants to study how network structure affects the economic outcome of interest, y, in regressions of the form[8]

$$y = \alpha + w(G)\beta_0 + \epsilon. \tag{1.2}$$

The researcher can estimate this regression at various observation levels. At the graph level, the data is $\{(y_r, w(G_r)) : r = 1, ..., R\}$ where $w(G_r)$ is a d_w-vector of network statistics (e.g., average degree, clustering) and the regression contains R observations. In our example, the researcher may regress the microfinance take-up rate in a village on the average network importance of the random set of households which were initially informed about microfinance. We anticipate that the centrality of these initial nodes correlates positively with take-up rates.

At the node level, the data is $\{(y_{ir}, w_{ir}(G_r)) : i = 1, ..., n,\ r = 1, ..., R\}$ where $w_{ir}(G_r)$ is a d_w-vector of statistics (e.g., degree of i, eigenvector centrality of i) and the regression has nR observations.[9] In our example, the researcher regresses a household's decision to join microfinance on its centrality. Theory suggests that central nodes will be more likely to learn new information. Similarly, one may estimate regressions at the edge level. Here $w_{ij}(G_r)$ is a d_w-vector of edge level statistics (e.g., social distance between the nodes) and the regression contains $\binom{n}{2} \cdot R$ observations.

Using sampled networks, the researcher runs regressions of the form

$$y = \alpha + w(\bar{G})\beta + u,$$

where $\bar{G}$ is either $G^{|S}$ or G^S, depending on the sampling scheme. In general, the measurement error will not be classical and may result in attenuation bias, expansion bias, or even sign switching. Sections 1.3.1 contain examples of common and economically meaningful network statistics where such biases exist and section 1.5 provides further numerical evidence on these biases.

Regression of Economic Outcomes on Network Neighbors' Outcomes

In a social equilibrium model, an economic outcome, y_i, depends on exogenous covariates of the individual, x_i, as well as the outcome of i's peer group, $\{y_j : j \in N_i\}$. In our running example, y_i is the microfinance meeting attendance rate of a household and x_i represents whether the researcher

[8] A vector of demographic covariates may be included, though we omit it for simplicity.

[9] With missing data, there will be $O(nR)$ observations. For instance with $G^{|S}$, one has $mR = \psi nR$ observations.

has exogenously informed the household. Estimating such a model is difficult in the usual way ((Manski 1993)), but with network data, assuming exogeneity of x_i as in the above examples, (Bramoulle, Djebbari, and Fortin 2009) and (De Giorgi, Pellizzari, and Redaelli 2010) show that the model may be identified as the peer groups for individuals are overlapping but not identical.

Formally, let $y = (y_1, ..., y_n)'$ be the vector of outcome variables, $x = (x_1, ..., x_n)'$ be the vector of exogenous covariates and $\iota = (1, ..., 1)'$. A researcher is interested in estimating

$$y = \alpha_0 \iota + \rho_0 w(G) y + \gamma_0 x + \delta_0 w(G) x + \epsilon, \tag{1.3}$$

where $w(G)$ is a (possibly weighted) adjacency matrix that describes how much y_i is affected by others in the network. The economic parameter is $\beta_0 = (\rho_0, \gamma_0, \delta_0)$. Due to sampling, the researcher mistakenly estimates the model,

$$y = \alpha \iota + \rho w(\bar{G}) y + \gamma x + \delta w(\bar{G}) x + u, \tag{1.4}$$

where $\bar{w}$ is defined analogously with $\bar{G}$ either $G^{|S}$ or G^S. The neighborhoods will be misspecified and the estimator exhibits bias. We discuss this model in Section 1.3.2.

1.2.3 Random Graphs and Asymptotic Framework

Random Graphs

Until now we have described an economic process, such as diffusion, occurring on a given collection of networks. Consider the example of a regression of y on network covariate $w(G)$. With missing data the researcher does not observe the true network statistic. In section 1.3 we demonstrate the biases induced by using $w(\bar{G})$ where $\bar{G}$ is the star or induced subgraph. Section 1.4 develops graphical reconstruction. We think of the network as the realization of a random network formation process. Consider a simple but commonly used model: the probability that individuals i and j are connected, conditional on covariate z_{ij}, is given by

$$\mathrm{P}(A_{ij} = 1 | z_{ij}, \theta_0) = \Phi(z_{ij}' \theta_0),$$

where Φ is some link function. Thinking of the network as a random graph allows us to compute the conditional expectation of the regressor $w(G)$ given the observed portion of the network A^{obs} and the sampling scheme: $\mathrm{E}[w(G)|A^{obs}; \theta_0]$. If we knew the distribution of G we could compute this expectation. By properties of conditional expectation using $\mathrm{E}[w(G)|A^{obs}; \theta_0]$ as a regressor allows us to consistently estimate β_0.

Formally, each network G_r is a random graph that is independently though not identically distributed over the space $\mathcal{G}_{n_R}$. We model the random networks as a triangular array of independent but not identically distributed random graphs, $G_{1,R}, ..., G_{R,R}$. Each graph $G_{r,R}$ is a random draw from a distribution $\mathrm{P}_{r,R}(G_r; \theta_{0r})$ over $\mathcal{G}_{n_R}$, where $\theta_{0r} \in \Theta_{r,R}$ is a parameter governing the

distribution. In what follows, we omit the R subscript indexing the triangular array.

Asymptotic Frame

Graphical reconstruction requires estimating a conditional expectation for every network. Since the parameter θ_{0r} for each network is unknown we must be able to consistently estimate all of these together. Intuitively, we need conditions such that every network has enough information in it so that its parameter can be precisely estimated. This is similar to panel data with non-linear fixed effects, where both the number of individuals and the number of periods grow.

Formally, we will assume that $n_R \to \infty$ as $R \to \infty$. The rate requirements of n and R are discussed in Section 4. Moreover, $\Theta_{r,R}$ is typically finite dimensional, though we discuss an example where its dimension grows as $R \to \infty$. We assume conditions on n, R, and the random graph models such that every network asymptotically contains enough information to estimate θ_{0r} very well. In turn, we can estimate the conditional expectation very accurately and therefore recover the economic parameter β_0.

Finally, we employ the following notation throughout the paper. $\mathrm{E}[\cdot]$ denotes expectation, $\mathbb{E}_n[\cdot]$ the empirical expectation,[10] $\|\cdot\| = \|\cdot\|_2$ the ℓ^2-norm, $\|\cdot\|_\infty$ the sup-norm, and $\ell^\infty(U)$ the space of bounded functions on $\mathcal{U}$. Also, $f_n \in \Theta(g_n)$ means $\exists k_1, k_2 > 0, n_0$ such that $\forall n > n_0$ $|g_n| \cdot k_1 \leq |f_n| \leq |g_n| \cdot k_2$, and the falling factorial is given by $(n)_j = n(n-1)...(n-j+1)$.

1.3 Analytical Examples of Bias

In this section we provide several analytical examples which demonstrate the biases due sampled network data. We provide three classes of examples: regression of economic outcomes on network statistics in section 1.3.1, regression of outcomes on network neighbors' outcomes in section 1.3.2, and a nonlinear GMM model of diffusion in section 1.3.3.

1.3.1 Regression of Economic Outcomes on Network Characteristics

To gain an intuition, we relate our problem to general measurement error. If the researcher is interested in a regression

$$y_r = w_r \beta_0 + \epsilon_r$$

but instead uses mismeasured regressors $\bar{w}_r$, the resulting estimator satisfies $\operatorname{plim} \widehat{\beta} = \beta_0 \frac{\operatorname{cov}(\bar{w}, w)}{\operatorname{var}(\bar{w})}$. Expansion, attenuation, and sign-switching bias are all possible. In our environment, $w_r = w(G_r)$, the relevant network statistic, but due to sampling the researcher uses $\bar{w}_r = w(\bar{G}_r)$, where $\bar{G}_r$ is the star or induced subgraph. Consequently, we are primarily interested in the covariance of the network statistic with its true value, under the sampling scheme.

[10] For $a = (a_1, ..., a_n)$, $\mathbb{E}_n[a_i] = \frac{1}{n}\sum_i a_i$. Similarly, $\mathbb{E}_R[a_r] = \frac{1}{R}\sum_r a_r$ and $\mathbb{E}_{n,R}[a_{i,r}] = \frac{1}{n}\sum_r \sum_i a_{i,r}$.

The covariance is typically not tractable to characterize. However, sometimes mismeasurement has a scaling effect in expectation. The scaling effect roughly means that $\mathrm{E}[\bar{w}|w] = \pi w + o(1)$, where $\pi = \pi(\psi)$ is some known deterministic function. Thus,

$$\operatorname{plim} \widehat{\beta} = \beta_0 \cdot \underbrace{\pi^{-1}}_{\text{Scale}} \cdot \underbrace{\frac{\operatorname{var}(w)}{\operatorname{var}(w) + \operatorname{var}(v)\pi^{-2}}}_{\text{Classical attenuation}}$$

where $v = \bar{w} - \mathrm{E}[\bar{w}|w]$. There are two sources of biases: a scale effect which depends purely on $\pi(\psi)$ and a dispersion effect which generates attenuation. Average degree and graph clustering are commonly used network statistics that exhibit scale transformations. However, more general statistics such as path length and eigenvalues are not merely scaled in this manner.

Degree and Clustering

The degree of a node is its number of connections. Degree is a common measure of network importance. (Kremer and Miguel 2007), (Hochberg, Ljungqvist, and Lu 2007), (Angelucci, De Giorgi, Rangel, and Rasul 2009), (Banerjee, Chandrasekhar, Duflo, and Jackson 2011), (Alatas, Banerjee, Chandrasekhar, Olken, and Hanna 2011), among others, use degree as a regressor. In addition to its ubiquitous use in applied work, average degree is a transparent, linear example which easily demonstrates the problem and has an intuitive (though not generalizable) solution.

We define $d(G_r) := n^{-1} \sum_{i,j} A_{ijr}$ as the average degree for graph r and use $d(\bar{G}_r)$ for $\bar{G} \in \{G^S, G^{|S}\}$ to denote the average degree computed with sampled data. As above, we are interested in the conditional expectation under a sampling scheme. One can show that

$$\mathrm{E}\left[d(G_r^S)|G_r\right] = \left(1 - (1-\psi)^2 + o(1)\right) d(G_r) \text{ and } \mathrm{E}\left[d(G_r^{|S})|G_r\right] = (\psi + o(1))\, d(G_r).$$

Intuitively, a node will miss links in proportion to the number of links it has. For instance, in the star subgraph an edge only appears in the data if at least one of the two nodes are sampled, yielding $1-(1-\psi)^2$. Notice that similar computations can be performed for other random sampling schemes the researcher may face. We define the limiting cross-network variance in the degree and the disturbance terms,

$$\sigma^2_{\langle d \rangle} := \operatorname*{plim}_{R\to\infty} a_R^{-2} \mathbb{E}_R \left[d(G_r) - \mathbb{E}_R d(G_r)\right]^2 \text{ and } \sigma^2_{vj} := a_R^{-2} \operatorname*{plim}_{n\to\infty} \mathbb{E}_R v_{rj}^2 \text{ for } j \in \{S,\ |S\}$$

where $v_{rj} := d(\bar{G}_r) - \mathrm{E}[d(\bar{G}_r)|G_r]$ and (a_R) is some sequence of normalizing constants.

The clustering coefficient of a node is the fraction of its neighbors that are themselves connected to each other. (Bloch, Genicot, and Ray 2008) study risk-sharing on social networks and find that networks which have theoretical properties that lend themselves to higher levels of risk-sharing

tend to have higher clustering, which motivates clustering as a regressor with risk-sharing data.[11] Moreover, under certain assumptions, (Möbius and Szeidl 2006) and (Karlan, Mobius, Rosenblat, and Szeidl 2009) use a model of trust and social collateral to microfound clustering as a measure of social capital. The clustering of the graph (denoted $c(G)$) is the ratio of the number of triangles (three nodes where every node is connected to every other node, denoted $\rho(G)$) to the number of connected triples (three nodes that have at least two social connections between them, denoted $\tau(G)$). That is, the *graph clustering* is $c(G) := \frac{\rho(G)}{\tau(G)}$, where

$$\rho(G) := 3\sum_i \sum_{k>i} \sum_{j\neq i,k} A_{ij}A_{jk}A_{ki} \text{ and } \tau(G) := \sum_i \sum_{k>i} \sum_{j\neq i,k} A_{ij}A_{jk}.$$

The analysis of graph clustering is similar to degree. One can show that

$$\mathrm{E}\left[\rho(G_r^S)|G_r\right] = (3\psi^2(1-\psi)+\psi^3+o(1))\cdot\rho(G_r) \text{ and } \mathrm{E}\left[\tau(G_r^S)|G_r\right] = (\psi(1-\psi)^2+3\psi^2(1-\psi)+\psi^3+o(1))\cdot\tau(G_r).$$

This comes from the fact that to obtain a triangle one needs to sample at least two of the nodes, while to obtain a connected triple, one can additionally sample the middle node. Meanwhile, for the induced subgraph we have $\mathrm{E}[\rho(G_r^{|S})|G_r] = (\psi^3+o(1))\cdot\rho(G_r)$ and $\mathrm{E}[\tau(G_r^{|S})|G_r] = (\psi^3+o(1))\cdot\tau(G_r)$ since all three nodes must be sampled.

We assume a sequence of graphs which have non-zero clustering such that the sequence also has non-zero clustering in the limit. We define

$$v_{rj} := c(\bar{G}_r) - \mathrm{E}[\rho(\bar{G}_r)|G_r]/\mathrm{E}[\tau(\bar{G}_r)|G_r] \text{ for } j \text{ and } \bar{G} \text{ indexing the sampling scheme.}$$

In turn, $\sigma^2_{\langle c\rangle}$, σ^2_{vS}, and $\sigma^2_{v|S}$ are defined analogously as in the case of degree.

PROPOSITION 1.3.1. *Consider a regression of $y_r = \alpha + w(G_r)\beta_0$ with w being either the average degree or graph clustering. Assume the data $(y_{rR}, d(G_{rR}))$ and $(y_{rR}, c(G_{rR}))$ is a triangular array satisfying the regularity conditions of Assumption 1.B.2.*[12]

1. *Average degree:*[13]
$$\widehat{\beta}(G^{|S}) \xrightarrow{\mathrm{P}} \frac{\beta_0}{\psi}\cdot\frac{\sigma^2_{\langle d\rangle}}{\sigma^2_{\langle d\rangle}+\psi^{-2}\sigma^2_{v|S}} \text{ and } \widehat{\beta}(G^S) \xrightarrow{\mathrm{P}} \frac{\beta_0}{1-(1-\psi)^2}\cdot\frac{\sigma^2_{\langle d\rangle}}{\sigma^2_{\langle d\rangle}+(\psi(2-\psi))^{-2}\sigma^2_{vS}}.$$
2. *Clustering:*
$$\widehat{\beta}(G^{|S}) \xrightarrow{\mathrm{P}} \beta_0\cdot\frac{\sigma^2_{\langle c\rangle}}{\sigma^2_{\langle c\rangle}+\sigma^2_{v|S}} \text{ and } \widehat{\beta}(G^S) \xrightarrow{\mathrm{P}} \left(\frac{\psi(3-2\psi)}{1+\psi(1-\psi)}\right)^{-1}\beta_0\cdot\frac{\sigma^2_{\langle c\rangle}}{\sigma^2_{\langle c\rangle}+\left(\frac{\psi(3-2\psi)}{1+\psi(1-\psi)}\right)^{-2}\sigma^2_{vS}}.$$

In expectation, the average degree will be scaled down as a function of the sampling rate, since only a share of social connections are observed. Because the regressors are scaled down the coefficient

[11] One may also be interested in the *support* of a graph defined in (Jackson, Rodriguez-Barraquer, and Tan ming). The analysis of the bias in using support as a regressor will follow in a similar manner

[12] In the case of clustering we require $\sup_{R\geq 1}\sup_{r\leq R} a_R\rho_r \vee a_R\tau_r \leq \bar{\alpha}$, $\inf_{R\geq 1}\inf_{r\leq R} a_R\rho_r \vee a_R\tau_r \geq \underline{\alpha}$.

[13] In Appendix 1.C we demonstrate how this can be done without assuming $\sigma^2_{v|S} \to 0$ by instead estimating $\widehat{\sigma}^2_{v|S}$.

expands, while dispersion around this expectation induces attenuation. Similarly, as random sampling yields on average the same share of connected edges between each possible triangle, $c(G^{|S})$ consistently estimates $c(G)$. Dispersion about the mean yields the attenuation bias in regression. For G^S, there is an expansion bias owing to the fact that a triangle appears as an intransitive triad (where exactly two of the three nodes are connected) when an edge of the triangle is missing due to sampling. The usual attenuation term also applies.

The simple structure of average degree and graph clustering allows for straightforward solutions. First consider average degree. In the case of the induced graph, one can simply rescale the estimated coefficient by ψ and, under regularity conditions, the resulting estimator is consistent.[14] Meanwhile, for the star subgraph, presumably the researcher would use the degree of only surveyed nodes to begin with, which would deliver consistent estimates. Second, let us consider clustering. The induced subgraph yields an unbiased estimate of the clustering but due to the dispersion term, generates attenuated coefficients. The star subgraph requires a rescaling as discussed above.

Path Length and Eigenvalues

The path length between two nodes i and j is given by the minimum number of steps taken on the graph to get from i to j, denoted $\gamma(i,j) := \min_{l \in \mathbb{N} \cup \infty} [A^k]_{ij} > 0$. If there is no such finite path, we put $\gamma(i,j) = \infty$. The average path length of a graph is the mean taken over all finite paths,

$$\gamma(G) := \sum_{i,j:\gamma(i,j)<\infty} \gamma(i,j) / \left| \{(i,j) \in V^2 : \gamma(i,j) < \infty\} \right|.$$

Models of diffusion of information, flows of finance, risk-sharing, nepotism, and other phenomena, build on the principle that the farther apart agents are, the less is transmitted between them. For example, (Kinnan and Townsend 2011) study how the network distance to a bank affects consumption smoothing. Other papers that use path length or average path length include Golub and Jackson (2010) who simulate diffusion processes; (Leider, Möbius, Rosenblat, and Do 2009) and (Goeree, McConnell, Mitchell, Tromp, and Yariv 2010) who study dictator games between members of a school; (Alatas, Banerjee, Chandrasekhar, Olken, and Hanna 2011) who look at the diffusion of information about poverty; and (Banerjee, Chandrasekhar, Duflo, and Jackson 2011) who study the diffusion of microfinance.

The average path length of a network is known to be a very difficult object to study analytically.[15] Both the economics and statistical physics literatures study an object we term the *graph span*, which mimics average path length. (Jackson 2008a) shows that for a general family of random graph models the ratio of the graph span to average path length asymptotically almost surely is one. Moreover, the statistical physics literature uses such an approximation as

[14] In Appendix 1.C we demonstrate how this can be done without assuming $\sigma^2_{v|S} \to 0$ by instead estimating $\widehat{\sigma}^2_{v|S}$.

[15] Bollobas (2001) approaches path length from an exact analytical perspective but only for a very specific random graph family. This approach is not suitable for gaining intuition for broader classes of graphs.

well (see, for instance, (Newman, Strogatz, and Watts 2001; Watts and Strogatz 1998), (Watts and Strogatz 1998)). These literatures motivate the study of the graph span as a regressor. Let $d_2(G) := n^{-1}\sum_{i=1}^{n}\sum_{j>i}^{n}\sum_{k\neq i,j} A_{ij}A_{jk}$ be the average number of second neighbors.[16] The *graph span* is

$$\ell(G) := \frac{\log n - \log d(G)}{\log d_2(G) - \log d(G)} + 1.$$

Larger networks have higher spans. Networks that are more expansive in the sense that the number of second neighbors far exceeds the number of neighbors have lower spans; it takes fewer steps to walk across the network. We first describe some useful properties of average degree and average number second neighbors for randomly sampled graphs and then study regression bias.

LEMMA 1.3.1. *Put* $k(\psi) = \psi + \psi^2 - \psi^3$. *For any sequence of random graphs* $(G_n)_{n\in\mathbb{N}}$ *satisfying, as* $n \to \infty$, $d(G)/a_{1n} \xrightarrow{\mathrm{P}} c_1$, $d_2(G)/a_{2n} \xrightarrow{\mathrm{P}} c_2$, $a_{1n}, a_{2n} \in o(n)$, *and* $c_1, c_2 > 0$,

1. $\left|d(G^{|S}) - \psi d(G)\right| = o_{\mathrm{P}}(1)$, $\left|d_2(G^{|S}) - \psi^2 d_2(G)\right| = o_{\mathrm{P}}(1)$,
2. $\left|d(G^S) - (1-(1-\psi)^2)d(G)\right| = o_{\mathrm{P}}(1)$, $\left|d_2(G^S) - k(\psi)d_2(G)\right| = o_{\mathrm{P}}(1)$.

This observation is general in the sense that it only requires that degree and the number of second neighbors to grow sufficiently slowly, which is reasonable for realistic applications. We now study the behavior of regressions with $\ell(G^{|S})$ and $\ell(G^S)$. If a sufficiently high fraction of nodes are sampled, the estimator exhibits attenuation. Meanwhile, if a sufficiently low fraction of nodes are sampled, the estimator may exhibit sign-switching.

PROPOSITION 1.3.2. *Assume the data sequence satisfies the regularity conditions of Assumption 1.B.3 and put* $\xi_{rR} := d(G_{rR})/d_2(G_{rR})$, $0 < c_0 < c_1 < 1$, $c_0 := \inf_{R\geq 1}\inf_{r\leq R}\xi_{rR}$, *and* $c_1 := \sup_{R\geq 1}\sup_{r\leq R}\xi_{rR}$.

1. $\widehat{\beta}$ *is sign-consistent with attenuation if* $\psi \in (c_1, 1)$ *or* $k(\psi)/(1-(1-\psi)^2) \in (c_1, 1)$:

$$\operatorname{plim}|\widehat{\beta}(G^{|S})| < |\beta_0| \text{ and } \operatorname{plim}|\widehat{\beta}(G^S)| < |\beta_0|.$$

2. $\widehat{\beta}$ *may be sign-inconsistent otherwise.*

Put $\widetilde{d}(\bar{G})$ as the sampling-corrected estimate of degree from the preceding subsection, $\widetilde{d}_2(G^S) = d_2(G^S)/k(\psi)$, and $\widetilde{d}_2(G^{|S}) = d_2(G^{|S})/\psi^2$. Define

$$\widetilde{\ell}(G^S) := \frac{\log n - \log \widetilde{d}(G^S)}{\log \widetilde{d}_2(G^S) - \log \widetilde{d}(G^S)} + 1 \text{ and } \widetilde{\ell}(G^{|S}) := \frac{\log(\psi^{-1}m) - \log \widetilde{d}(G^{|S})}{\log \widetilde{d}_2(G^{|S}) - \log \widetilde{d}(G^{|S})} + 1,$$

which we use in the adjusted least squares estimators.

[16]Notice this defines second neighbor in the sense of taking a random node and then counting the number of neighbors of each of the node's neighbors. The definition is different from counting the number of distinct nodes at path length two from a given node, which would be $\frac{1}{n}\sum_i\sum_{k>i}\sum_{j\neq i,k} A_{ij}A_{jk}(1-A_{ik})$.

PROPOSITION 1.3.3. *Analytical corrections are consistent,* $\operatorname{plim}\widetilde{\beta}(G^{|S}) = \beta_0$ *and* $\operatorname{plim}\widetilde{\beta}(G^S) = \beta_0$. Sampling a network thins out the set of edges, resulting in a higher graph span. As the graph span approximates behavior of average path length, it captures the intuition that due to sampling, paths on graphs seem longer than they truly are. The expansion of the graph span has a slope effect on $\widehat{\beta}$, and as $\log\psi < 0$ and $\log(k(\psi)/(1-(1-\psi)^2)) < 0$, the effect is either attenuation unless the sampling rate is too low, in which case sign-switching becomes a possibility. One must proceed with caution when discussing cases where the sampling probability is too low. In this case the network can shatter, yielding "islands" of disconnected sets of nodes which have short average path lengths within the set but have infinite distance across the sets.[17] Since average path length is defined as a mean conditional on all finite paths, this is precisely where sign-switching may occur in practice. (Alatas, Banerjee, Chandrasekhar, Olken, and Hanna 2011) contains an example where this happens in Indonesian networks.

Spectral functions are network statistics that relate to the set of eigenvalues of matrices which represent the graph, such as the adjacency matrix. They are useful in characterizing properties of the network. The distribution of eigenvalues has applications to models of information diffusion and risk-sharing as well. The number of k-length walks that cycle back to the original node correspond to k-th moment of the eigenvalue distribution, denoted $\mu^k(G)$,

$$\mu^k(G) = n^{-1} \sum_{i_1,..,i_k \in V^k} A_{i_1 i_2}...A_{i_k i_1} = n^{-1}\mathrm{Tr}(A^k)$$

where $V^k = V \times ... \times V$ ((Barabasi and Albert 1999)). Given that the graph spectrum carries a great deal of information about the diffusive properties of a network, it is a useful regressor.[18]

There are several applications of spectral statistics in economic theory. For instance, the first eigenvalue of the adjacency matrix, $\lambda_{\max}(G)$, describes how well the graph diffuses information (e.g., (Bollobás, Borgs, Chayes, and Riordan 2010)).[19] In models of social learning (Golub and Jackson 2009; Golub and Jackson 2010a) show that the second eigenvalue of a weighted adjacency matrix is related to the time it takes to reach consensus; similar results are shown in (DeMarzo, Vayanos, and Zwiebel 2003). (Ambrus, Mobius, and Szeidl 2010) also characterize the risk-sharing capacity of a network as a function of the expansiveness of the network; it is well-known in network theory that this maps into the eigenvalues of a transformation of the adjacency matrix ((Chung 1997)). It is difficult to precisely characterize the behavior of these spectral regressors, though we present bounds on their behavior under sampling.

PROPOSITION 1.3.4. *For an arbitrary graph G, we have*

[17] One can check that a graph H with $d_2(H)/d(H) < 1$ cannot be connected. The sign-switching case requires at least some $d_2/d < 1$ which we note the researcher can immediately detect.

[18] For a discussion of how spectral distributions correlate with graphical properties, see Barabasi and Albert (2002).

[19] In a percolation process the threshold probability above which a giant component emerges is precisely $1/\lambda_{\max}$. For another intuition, if A is diagonalizable, then the dominant factor in $\left\|A^k\right\|$ is $\lambda^k_{\max}$.

1. $\mu^k(G^S) < \mu^k(G)$ *and* $\mathrm{E}\left[\mu^k(G^{|S})|G\right] = \sum_{j=2}^{k} \frac{(m-1)_j}{(n-1)_j}\eta_j < \mu^k(G)$, *where* η_j *is the number of sets of* j*-distinct nodes that are counted.*

2. $\lambda_{\max}(G^{|S}) \leq \lambda_{\max}(G^S) \leq \lambda_{\max}(G)$.

Since $\lambda_{\max}$ can be thought of as measuring the number of walks through the graph (and with missing edges there are fewer walks) we expect expansion bias in $\widehat{\beta}$ when using these regressors.[20] This means that networks will appear to be more diffusive than they actually are.

1.3.2 Regression of Outcomes on Network Neighbors' Outcomes

We discuss the impact of sampled networks on regressions of nodes' outcomes on network neighbors' outcomes. The models we consider are developed in (Bramoulle, Djebbari, and Fortin 2009) and (De Giorgi, Pellizzari, and Redaelli 2010) and naturally extend the models discussed in (Manski 1993) to a network setting. (Blume, Brock, Durlauf, and Ioannides 2011) contains an extensive review of the literature. The network allows for nodes to have overlapping but not identical peer groups.

The model is given by (1.3) and we are interested in $\beta_0 = (\rho_0, \gamma_0, \delta_0)$. There are two natural examples for how neighbors' outcomes ought to affect a node's outcome. First, every node's outcome may be affected by the average outcome of its neighbors.[21] Second, every node's outcome may be affected by the total sum of its neighbor's outcomes.[22] The reduced form is

$$y = \alpha\iota/(1-\rho_0) + \gamma_0 x + (\gamma\rho_0 + \delta_0)\sum_{k=0}^{\infty}\rho_0^k w^{k+1}x + \sum_{k=0}^{\infty}\rho_0^k w^{k+1}\epsilon.$$

Since a node's neighborhood outcome, wy, is the endogenous regressor, the reduced form suggests that extended neighborhood effects, powers $w^k x$ $(k \geq 2)$, can be used as instruments for wy. We focus on the instrument $Z = [\iota, x, wx, w^2x]$.[23] Setting $X = [\iota, wy, x, wx]$, the estimator is $(X'P_ZX)^{-1}X'P_Zy$.

Identification comes from intransitive triads.[24] If i and j are connected and j and k are connected, but i and k are not connected, then k affects i only through j. As such, x_k is used as an instrument for y_j's effect on y_i. We caution that this identification strategy convincingly works only when x is randomly assigned (e.g., Ngatia, 2011; Dupas 2010) as identification crucially depends on exogeneity of x.

[20] Whether there is expansion bias depends on how the eigenvalues shrink across the initial distribution. For instance, if the contraction is by translation, the regression slope would clearly not change. Numerical with simulated networks and empirical data provide evidence of expansion bias.

[21] We can write the model as $y_i = \alpha + \beta\mathbb{E}_{N_i}[y_j] + \gamma x_i + \delta\mathbb{E}_{N_i}[x_j] + \epsilon_i$ as $\mathbb{E}_{N_i}[y_j] = \sum_{j\in N_i} y_j/d_i = \sum_j y_j A_{ij}/d_i$.

[22] We discuss the first case, though clearly by mimicking the argument the results follow for the second.

[23] Other estimation strategies are suggested in the literature, on the basis of efficiency ((Bramoulle, Djebbari, and Fortin 2009); Lee et al., 2009). They require the validity of the instrument Z.

[24] (Bramoulle, Djebbari, and Fortin 2009) provide formal identification conditions.

We examine the estimation of (1.4) using $\bar{w} = w(G^S)$ or $\bar{w} = w(G^{|S})$ with instrument $Z_{\bar{G}} = [\iota, x, \bar{w}x, \bar{w}^2x]$. We show that the exclusion restriction is invalid when using sampled network data, even if the covariates are exogenous and the usual identification requirements are met if the full network data was available.[25]

PROPOSITION 1.3.5. *Assume* $\gamma_0\rho_0 + \delta_0 \neq 0$ *and* $w^2 \neq 0$, *so 2SLS is valid for* (1.3). *Then 2SLS with*

1. $w(G^S)$ *with* (y_i, x_i) *for all* $i \in V$ *generically yields* $\mathrm{E}[Z_{G^S} u_{G^S}] \neq 0$,[26]
2. $w(G^{|S})$ *with* (y_i, x_i) *for all* $i \in S$ *generically yields* $\mathrm{E}[Z_{G^{|S}} u_{G^{|S}}] \neq 0$,
3. $w(G^S)$ *with* (y_i, x_i) *for all* $i \in S$ *generically yields* $\mathrm{E}[Z_{G^S} u_{G^S}] \neq 0$,
4. $w(G^S)$ *with* (y_i, x_i) *for all* $i \in S$ *generically yields* $\mathrm{E}[Z_{G^S} u_{G^S}] = 0$,
5. $w(G^S)$ *with* (y_i, x_i) *for all* $i \in V$ *but restricting the second stage to* $i \in S$, *yields* $\mathrm{E}[Z_{G^S} u_{G^S}] = 0$.

Sampling induces an errors-in-variables problem, wherein the neighborhood effect is mismeasured since the neighborhoods themselves are misspecified. Though typically one uses instruments to address such a problem, here the instrument is correlated with the measurement error in the regressor, as the instrument involves powers of the mismeasured adjacency matrix. As such, the exclusion restriction is violated.

Figure 1-3 provides two examples where invalid instrumentation is generated. Figure 1-3(a,b) show that if j is sampled but i and k are not, the sampled network falsely suggests that k is a valid instrument for j's effect on i. Similarly, figure 1-3(c,d) show a case with the induced subgraph, where k instrumenting for j's effect on i will be invalid as the other channels through which k affects i are not accounted for due to sampling. In this case, the channel through l is omitted.

With G^S data, however, there is a simple analytical correction. For $i \in S$, notice that $[\bar{w}x]_i = [wx]_i$ and $[\bar{w}y]_i = [wy]_i$. Consequently, there is no measurement error in the second stage for these observations. As only the first stage contains measurement error, uncorrelated with the second stage residual, such an exercise satisfies the exclusion restriction.

Finally, we note that the results presented in this section also have implications for the opposite case, wherein the researcher incorrectly adds edges to the network architecture and then estimates peer effects using modified network data. For instance, if the researcher assigns friendship links among students who share classes, when in fact the peer effect channel is through friendship links only, then for analogous reasons the estimation will be biased.

[25] (De Giorgi, Pellizzari, and Redaelli 2010) are aware that measurement error may cause problems in this model and conduct a numerical robustness exercise.

[26] We say generically in the sense that given (G, x, β_0), only a finite set of $\psi \in [0, 1]$ satisfy $\mathrm{E}[Z_{G^S} u_{G^S}] = 0$.

1.3.3 A Model of Diffusion

Having discussed several examples of network-based regressions, we now turn to a model of diffusion examined in (Jackson and Rogers 2007) which we discuss in the context of our microfinance example.[27] The researcher is interested in estimating this diffusion model which satisfies equilibrium moment equations. There are two states: whether or not a household endorses microfinance in a weekly village gathering. A non-endorsing household with d_i links may choose to endorse with probability $\nu_0 d_i \sigma_i$ where ν_0 is a transmission parameter and σ_i is the fraction of i's neighbors that have decided to endorse. However, an endorsing household may naturally decide not to endorse, which can happen with probability δ_0. (Jackson and Yariv 2007) extend this model to a number of strategic environments.

The model is identified up to parameter $\beta_0 := \nu_0/\delta_0$, which is the transmission to recovery rate. Let $P(d)$ denote the degree distribution and $\rho(d)$ the share of nodes with degree d that endorse. Finally, $\bar{\rho}_* := \sum_d \rho(d)P(d)$ is the average endorsement rate in the network and the researcher observes $y := \bar{\rho} + \epsilon$, with ϵ an exogenous zero mean shock.

The second neighbors endorsement rate is given by $\sigma = (\mathrm{E}d)^{-1} \sum_d \rho(d)P(d) \cdot d$. (Jackson and Rogers 2007) use a mean-field approximation to derive a steady state equation,

$$\rho(d) = \frac{\beta_0 \sigma d}{1 + \beta_0 \sigma d}.$$

The equilibrium satisfies

$$\sigma(\beta_0) = (\mathrm{E}d)^{-1} \sum_d \frac{\beta_0 \sigma(\beta_0) d^2}{1 + \beta_0 \sigma(\beta_0) d} P(d). \tag{1.5}$$

By combining (1.5) with the definition of $\bar{\rho}$, we find that $h(G_r; \beta) := \sum_d \frac{\beta \sigma_r(\beta) d}{1+\beta \sigma_r(\beta) d} P(d) = \bar{\rho}$. Therefore the researcher can use as moments

$$m(y_r, G_r; \beta) := y_r - h(G_r; \beta),$$

and estimate β_0 via nonlinear least squares. (Jackson and Rogers 2007) show that an equilibrium with non-zero endorsement rate exists only if $\beta > \mathrm{E}d/\mathrm{E}d^2$. The ratio of squared degree to degree, similar to what we have encountered when studying graph span, again becomes an important feature of the network. We put $\zeta := \mathrm{E}d^2/\mathrm{E}d$. Note that the typical summand of $h(G_r, \beta)$ is monotone and convex in d. Therefore, stochastic dominance relations among various distributions $P(d)$ will play a central role. We show that due to sampling of networks the researcher will overestimate the transmission parameter. An intuition for this is provided by the case of the star subgraph. This form of subsampling leads to a degree distribution that will be first order stochastically dominated by the true distribution. Therefore, the sampled network seems as if it has poorer diffusive properties; to

[27] They study the SIS (susceptible, infected, susceptible) model of epidemiology, which they and others show have applications in a wide variety of economic contexts.

generate the same average endorsement rate, the parameter governing the diffusion process must be higher. In addition, we show that the diffusion with the true parameter β_0 occurring on the sampled network may have no non-zero equilibria. When β_0 is close enough to the threshold $1/\zeta(G)$, the partially sampled network will make threshold ratio $1/\zeta(\bar{G})$ rise and therefore β_0 may appear to be less than $1/\zeta(\bar{G})$.

PROPOSITION 1.3.6. *Assume we have a triangular array* (y_{Rr}, G_{Rr}) *with degree distributions* $\mathrm{P}_{Rr}(d)$ *and (i)* (1.5) *holds in expectation for each* r*, (ii)* β_0 *is such that there is a positive endorsement in every equilibrium, (iii)* $\mathcal{B}$ *is a compact subset of* $(0, \infty)$*, (iv)* (ϵ_r) *are iid zero mean finite variance disturbances, and (v)* $\limsup_{R\to\infty} \sup_{r\leq R} \sup_d |\mathrm{P}_{Rr}(d) - \mathrm{P}_{\infty r}(d)| = 0$.

1. *The parameter estimates exhibit expansion bias:* $\operatorname{plim} \widehat{\beta}(G^S) > \beta_0$ *and* $\operatorname{plim} \widehat{\beta}(G^{|S}) > \beta_0$.

2. *For all* r, β_0 *is outside the range generating positive endorsement rate in the estimated equilibrium, with probability approaching one, under the following additional assumptions. Put* $\delta_r := \beta_0 - 1/\zeta_r > 0$ *and assume*

   (a) for star subgraphs, $\liminf_{R\to\infty} \zeta_r > 1 + \psi$ *and* $\limsup_{R\to\infty} \delta_r < (1 - \psi) \cdot \frac{1-\zeta_r^{-1}(1+\psi)}{\zeta_r+(1-\psi^2)}$,

   (b) or for induced subgraphs, $\limsup_{R\to\infty} \delta_r < (1 - \psi) \cdot \frac{1+\zeta_r^{-1}}{\psi\zeta_r+(1-\psi)}$.

It is easy to see that for the star subgraph, an analytical solution to the bias is to use the degree distribution of the sampled nodes. However, this is a highly non-generic solution. The induced subgraph, for instance, does not allow this approach nor do other sampling schemes (e.g., randomly chosen edges, etc.) A natural question to ask is whether we may use the sampled degree distribution, such as $P^{|S}(d)$, to obtain $P(d)$. We note that this will not be straightforward to do, in general, because it generates an ill-posed inverse problem. The researcher is faced with an under-determined system; while we can describe how $P(d)$ maps into $P^{|S}(d)$ due to sampling, there appears to be no unique inverse. Graphical reconstruction, however, will provide a way to address the problem.

1.3.4 From Analytical Examples to Graphical Reconstruction

In this section we have analytically examined biases that emerge from sampled networks. We focused on three main network statistics: degree, graph span, and clustering. This choice was motivated by a number of applied questions concerning diffusion of information, network importance, risk-sharing, and social collateral. By analytically characterizing the biases, we were able to describe the mechanics of the non-classical measurement error and construct analytical corrections to eliminate the biases, under regularity conditions. The analytical study required us to focus on graph-level regressions; moreover, to be consistent, the analytical corrections focused on eliminating a slope effect of the biases, but needed to assume away or estimate a dispersion effect.

We also examined a model where a node's outcome depends on its neighbors' outcomes and demonstrated that a network-based instrumentation method violates the exclusion restriction when

the network is sampled. With certain data structures, we provided a simple solution. Furthermore, we extended our analysis to a GMM model of diffusion and pointed out how the estimated parameters would exhibit expansion bias.

In general, our discussion has been on a case-by-case basis in this section. We have mostly focused on graph-level regressions and have been only able to examine very tractable network statistics. Numerous network statistics such as betweenness centrality, eigenvector centrality, and the aforementioned spectral statistics do not permit easy analytical examination nor corrections. The next section provides a more general method to estimate the economic parameter. Though the method is not limited to graph level regressions nor tractable network statistics, it comes at the cost of requiring more data and putting more structure on the problem by assuming models.

1.4 Graphical Reconstruction Estimation

In this section, we discuss a two-step estimation procedure to consistently estimate economic parameters from linear regression and GMM models. The formal asymptotic theory is developed in Chandrasekhar (2012). In our asymptotic frame, both the size of each network and the number of networks grow. Every network is a draw from a distribution governed by its own parameter θ_{0r}. This will force us to control an incidental parameter problem. Clearly, we can nest the special case where every network is drawn from the same distribution, $\theta_{0r} = \theta_0$ for every r, and thereby assume away the incidental parameter problem.

We present an informal overview of our method in section 1.4.1 and detail the estimation procedure in section 1.4.2. Chandrasekhar (2012) derives the theory of the estimation procedure.

1.4.1 Overview

In our overview we describe our procedure for regression,

$$y_{ir} = \alpha_0 + w_{ir}(G_r)\beta_0 + \epsilon_{ir}.$$

We assume that the researcher has the following data. First, she has outcome data for every node in every graph, $\{y_{ir} : i = 1, ..., n,\ r = 1, ..., R\}$, such as whether household i in village r participates in microfinance.[28] Second, she has a set of partially observed graphs, $\{G_r^S : r = 1, ..., R\}$ or $\{G_r^{|S} : r = 1, ..., R\}$. Third, she has variables which are predictive in a network formation model $\{z_r : r = 1, ..., r\}$.[29] For instance, the researcher may have basic demographic characteristics such as religion, caste, household amenities, occupation or geographic location. This data structure is relatively innocuous and common in numerous applications. In development, when deciding how to draw a random sample to administer treatments, researchers usually conduct a listing in each

[28] In what follows it is not necessary for y_{ir} to be observed for every node, but it simplifies notation.

[29] E.g., $z_r = \{z_{ir} : i = 1, ..., n\}$ or $z_r = \{z_{ij,r} : i, j \in V\}$ where z_{ir} or $z_{ij,r}$ are covariates for nodes or pairs.

enumeration area. This requires obtaining a census of the economic units, which can be done directly (e.g., (Townsend 2007); Suri, 2011; (Banerjee, Chandrasekhar, Duflo, and Jackson 2011)) or indirectly by obtaining census information from the village representatives (e.g., (Macours 2003); (Takasaki, Barham Oliver, and Bradford 2000)).[30] It is well-known that obtaining GPS and basic demographic data during enumeration is cheap; the bulk cost of a network survey is the network module itself. For a different example, consider school networks where it is straightforward to obtain rosters and demographic data for all students. The full set of observed data is (y_r, A_r^{obs}, z_r), consisting of y_r the vector of outcome data, A_r^{obs} the observed part of the graph, and z_r the vector of network formation covariates. The missing data for each network is A_r^{mis} and recall $G_r = (A_r^{obs}, A_r^{mis})$.

Every network is thought of as a realization of a random network formation process, drawn from a distribution which depends on z_r and parameter $\theta_{0r} \in \Theta_r$. To estimate β_0 we use an argument based on conditional expectations. If θ_{0r} were known for all r, we could estimate a conditional expectation of $w_{ir}(G_r)$ given the observed data,

$$\mathcal{E}_{ir}(A_r^{obs}, z_r; \theta_{0r}) := \mathrm{E}\left[w_{ir}(G_r)|A_r^{obs}, z_r; \theta_{0r}\right].$$

By the properties of conditional expectation, using $\mathcal{E}_{ir}$ in the regression instead of w_{ir} yields consistent estimation of β_0. The least squares estimator is given by[31]

$$\widehat{\beta}_{\text{ols}} = \left(\sum_{r=1}^{R}\sum_{i=1}^{n} \mathcal{E}_{ir}(\widehat{\theta}_r)\mathcal{E}_{ir}(\widehat{\theta}_r)'\right)^{-1} \cdot \sum_{r=1}^{R}\sum_{i=1}^{n} \mathcal{E}_{ir}(\widehat{\theta}_r) y_{ir}.$$

A similar but more involved result is true for GMM. Notice $\widehat{\beta}_{\text{ols}}$ depends on $\widehat{\theta}_r$ for all r.

To control the estimation of $\widehat{\theta}_r$, we need to argue not only that it is consistent for θ_{0r}, but uniformly so. That is, $\sup_r \|\widehat{\theta}_r - \theta_{0r}\| = O_{\mathrm{P}}(a_R^{-1} \cdot R_n^{1/b})$, where a_R is the rate of convergence of $\widehat{\theta}_r$ to θ_{0r} for every r, and $b > 1$ is the number of moments that the network formation model has. This imposes a rate requirement on the problem which says that the network-formation parameter needs to be estimated fast enough: $\sqrt{nR} \cdot a_R^{-1} \cdot R^{1/b} \to 0$.

The consistency of $\widehat{\theta}_r$ follows from assumptions on the model of graph formation and the sampling procedure. With missing-at-random data, under assumptions on the graph model, a consistent estimator exists. Consider a model where an edge forms independently, conditional on covariates,

$$\mathrm{P}(A_{ijr} = 1|z_r; \theta_{0r}) = \Lambda\left(f(z_{ir}, z_{jr})'\theta_{0r}\right),$$

where $\Lambda(\cdot)$ is some link function (e.g., logistic or normal), z_i is a vector of covariates for vertex i, and f is a vector-valued function. For instance, f may be the difference between characteristics of two nodes $f(z_i, z_j) = \|z_i - z_j\|$. If the sampling procedure is orthogonal to the network formation,

[30] Researchers can either collect simple covariate data from all nodes or from representatives who carry information.

[31] For notational simplicity, assume the regressors are demeaned.

a random subset of the $\binom{n}{2}$ pairs of nodes is observed. Therefore, $\widehat{\theta}_r$ is consistent.

This model converges with $a_R = n$, since we have on the order of $n(n-1)/2$ observations. The requirement becomes $n^{-1/2}R^{1/2+1/b} \to 0$, so the number of networks must grow sufficiently slower than the number of nodes. In other models, the rate a_R may be different (e.g., $n/\log n$, n^τ for $\tau \in [1/2, 2)$, $\sqrt{n/\log n}$). If the rate is too slow, the requirement for node level regression may not be met, though usually the requirement for graph level regressions will be satisfied.

1.4.2 Estimation in Practice

We describe the estimation algorithm for linear regression.
ALGORITHM (Estimation of $\widehat{\beta}_{\text{ols}}$).

1. Use (z_r, A_r^{obs}) to estimate $\widehat{\theta}_r$ based on the assumed network formation model.
2. Estimate $\mathcal{E}_{ir}(A_r^{obs}, z_r; \theta_{0r}) := \mathrm{E}\left[w_{ir}(G_r)|A_r^{obs}, z_r; \theta_{0r}\right]$.
   (a) Given (z_r, A_r^{obs}), for simulations $s = 1, ..., S$, draw $A_{r,s}^{mis*}$ from $\mathrm{P}_{\widehat{\theta}_r}(A_r^{mis}|A_r^{obs}, z_r)$.
   (b) Construct $w_{ir}(G_{rs}^*)$ where $G_{rs}^* = (A_r^{obs}, A_{r,s}^{mis*})$.
   (c) Estimate $\widehat{\mathcal{E}}_{ir}(A_r^{obs}, z_r; \widehat{\theta}_r) := \frac{1}{S}\sum_{s=1}^{S} w_{ir}(G_{rs}^*)$.
3. Estimate $\widehat{\beta}_{\text{ols}}$ using data $\{(y_{ir}, \widehat{\mathcal{E}}_{ir}(A_r^{obs}, z_r; \widehat{\theta}_r)) : i = 1, ..., n,\ r = 1, ..., R\}$.

The GMM algorithm is similar, requiring a reweighting term. Chandrasekhar (2012) provides an overview of standard errors and estimation methods. In practice, clustering at the graph level in vertex-level regressions and using heteroskedasticity robust standard errors for network-level regressions perform well, though we have explored various bootstrapping procedures (available upon request).

1.5 Numerical Experiments

This section reports the results of numerical simulations that characterize the biases due to sampling as well as the behavior of the analytical and graph reconstruction estimators.

1.5.1 Simulation Setup

We specify a data-generating process for a set of random graphs and outcome data, and then carrying out the following steps.
ALGORITHM (Simulation).

1. Generation of data.
   (a) Draw R networks the network formation families (below).

(b) Generate outcome data from a model with β_0 and data-generating process $(y, \epsilon)|G; \beta_0$.

(c) For each graph G_r construct sampled graphs $\{G_{rb}^S, G_{rb}^{|S} : b = 1, ..., B\}$.

2. Estimation of $\widehat{\beta}$ using $\{G_{rb}^S, G_{rb}^{|S} : b = 1, ..., B\}$.

(a) Estimate $\widehat{\beta}_b(G^{|S})$ and $\widehat{\beta}_b(G^S)$.

(b) If applicable, estimate the adjusted estimator $\widetilde{\beta}_b(G^{|S})$ and $\widetilde{\beta}_b(G^S)$.

(c) Estimate the graphical reconstruction estimators.

3. Perform (1)-(2) for $\psi \in \{1/4, 1/3, 1/2, 2/3\}$.

We generate networks of $n = 250$ nodes using the following simple conditional edge independence model. We set parameters such that the average degree, clustering, path length, maximal eigenvalue and variance of the eigenvector centrality distribution from networks in our simulations mimics those moments in the empirical Indian networks data-set. Dividing the set of nodes into 6 approximately equally sized groups, we place those groups on a line, indexed from 1 to 6. The probability that an edge formed between two members within the same group is high. The probability that an edge formed between two members of two different groups declines in the cross-group distance, represented by the difference in the indexed location of those groups on the line. Formally, let $g(i)$ denote the group of vertex i. We set

$$\mathrm{P}(A_{ijr}|z_{ijr}) = z'_{ijr}\theta_{0r}.$$

θ_{0r} is a $\binom{6}{2}$-vector with elements $\theta_{0r,lm}$ with $1 \leq l < m \leq 6$ and z_{ijr} is a $\binom{6}{2}$-vector with $z_{ijr,lm}$ with $1 \leq l < m \leq 6$. $\theta_{0r,lm}$ is the probability that a member of group l is linked to a member of group m. The lm-component of z_{ijr} is a dummy for whether i and j are in groups l and m respectively, $z_{ijr,lm} = \mathbf{1}\{g(i) = l\}\mathbf{1}\{g(j) = m\}$. In order to generate $\theta_{0r,lm}$ we use a simple distance function, with $\theta_{0r,lm} = p_r\left(1 - |l - m|/6\right)$ where p_r is a uniform random variable chosen such that the average degree generated mimics the average degree from the empirical application.[32]

1.5.2 Regression of Outcomes on Network Characteristics

We simulate and estimate a model with heteroskedastic residuals, $y_{ir} = \alpha_0 + w_{ir}(G_r)\beta_0 + \sigma_0 \cdot u_{ir}$, where $(\alpha_0, \beta_0) = (1, 2)$ and

$$u_{ir} := \mathcal{N}\left(0, \sigma_{ir}/\sqrt{\hat{\mu}^2_{\sigma_{ir}}}\right), \ \sigma_{ir} := 3\frac{w_{ir} - w_{ir\,\min}}{w_{ir\,\max} - w_{ir\,\min}} + 0.2, \text{ and } \hat{\mu}^2_{\sigma_{ir}} := \mathbb{E}_{n,R}\left[\sigma^2_{ir}\right].$$

[32] We have conducted simulations for alternative formation models, such as one in which covariates are generated by an autoregressive process and the edge formation probability is governed by a logistic regression. Results are qualitatively and quantitatively similar.

This formulation creates a fan-like heteroskedasticity. We then can easily set the R^2 of the regression to approximately 0.3 by defining $\sigma_0^2 := (1/R^2 - 1) \cdot \mathbb{E}_{n,R}(\tilde{y}_{ir} - \bar{\tilde{y}}_{ir})^2$ for $\hat{y}_{ir} = \alpha_0 + w_{ir}(G_r)\beta_0$.

Columns 1-5 of Tables 1-1 and 1-2 show the estimation bias, in percentages, for regression parameters when using sampled network data for a variety of network statistic regressors. Table 1-1 shows the biases when estimating regressions at the network level while Table 1-2 shows the biases when estimating regressions at the node level.

At the network level, we consider average degree, graph clustering, graph span, average path length, and $\lambda_{\max}$. In addition, we show results for the standard deviation of the eigenvector centrality distribution and the spectral gap. The eigenvector centrality represents how important a node is in information transmission ((Jackson 2008c)) and the spectral gap of a graph characterizes how rapidly diffusion processes on networks spread ((Chung 1997)). The latter is closely related to the expansiveness of a network that (Ambrus, Mobius, and Szeidl 2010) show characterizes good risk-sharing properties. (Hochberg, Ljungqvist, and Lu 2007) use eigenvector centrality in applied work.

At the node level, we show results for the degree, clustering coefficient, and eigenvector centrality of a node. Moreover we consider two regressions which characterize how far a node i is from another node j. We select a random node j (corresponding to a randomly treated node in an experimental setting) and generate a regressor which is the path length from i to j. In addition, we partition the nodes into two subsets which communicate the most within themselves and least across the sets. We say i is on the same side of the spectral partition of j if they are in the same subset. This partition is related to the spectral gap ((Chung 1997)) and therefore has implications for the (Ambrus, Mobius, and Szeidl 2010) approach to characterizing risk-sharing.

Overall we find that sampling the network leads to significant biases. To illustrate this, we discuss the biases at 1/3 sampling for the graph and node level. At the graph level the maximum bias is 260% ($\lambda_{\max}$), the mean is 90.9%, and the minimum is 15%.[33] The biases include expansion bias in the cases of degree, maximal eigenvalue, spectral gap, and graph clustering (for the star subgraph). The node level regressions exhibit a similar pattern: the maximum bias magnitude is 91%, the mean is 63%, and the minimum is 7% (same side of the spectral partition).

In columns 6-10 of Table 1-1, we present the results using the analytical corrections from section 1.3.[34] We find that the adjusted regression estimators perform uniformly better than the unadjusted estimators. The biases are usually low. Overall, the analytical corrections improve 100% of the parameter estimates in the simulations. At 1/3 sampling, the mean reduction in bias percentage when comparing the analytical correction to the raw network statistic is 69pp with a median reduction of 69pp and a maximum of 243pp. Columns 6-10 of Table 1-2 applies the analytical corrections which were derived for the case of graph level regression, to node level analysis; the

[33] When we looking at maximum, mean, and minimum, we are interested in the magnitude of the biases, so our discussion focuses on the absolute value of the bias.

[34] In Table 1-8 of Appendix 1.C we show an example of an analytical correction that involves estimating $\widehat{\sigma}_v^2$.

results are not motivated by theory and of course are mixed.[35]

1.5.3 Regression of Outcomes on Network Neighbors' Outcomes

Table 1-3 presents the results for simulations for the model of equation (1.3) with $(\alpha_0, \rho_0, \gamma_0, \delta_0) = (1, 0.5, 2, 0.5)$. We use three specifications to demonstrate the emergence of biases in peer effects regression due to two distinct causes: correlation of the instrument with the errors-in-variables problem and a weak instrument/finite sample problem induced by sampling. The table presents the mean bias percentage as well as the standard error of the bias.

We present three methods of estimating peer effects with sampled data and one correction. First, we show an estimate of the peer effect model with the network given by the induced subgraph. Second, we present an estimation of the model where (y, x) are known for every node, but the network used is the star subgraph. Finally, we present the same specification but only allow the researcher to have covariates for surveyed nodes. Each specification exhibits biases.

We vary the number of networks and disturbance size across three models to study how the bias varies. In Model 1, we use one network with 250 nodes per simulation drawn from the aforementioned model and set the number of simulations to $50,000$. In Model 2, we use one network with 250 nodes per simulation and use $50,000$ simulations, but reduce the variance of the disturbance. Model 3 presents results a from $2,500$ simulations of 20 networks each with 250 nodes, drawn from the model.[36]

All three specifications show significant bias in the estimates of the endogenous and exogenous peer effects. Comparing Panels A and B of the first and second set of columns shows the biases are greater when there is more noise in the system. Moreover, comparing Panels A and B of columns 1-5 and 11-15 shows that increasing the number of graphs in the estimation from 1 to 20 only modestly reduces the bias due to sampling. There are still non-trivial biases which remain.

Overall, the analytical correction performs well. In the Models 2 and 3, the estimates are essentially unbiased across all sampling levels presented. Moreover, the analytical correction for Model 1 exhibits negligible bias for sampling rates of $2/3$ and $1/2$. However, biases emerge at very low sampling rates, $1/3$ and $1/4$, in the case of Model 1. Furthermore, as evidenced by the standard errors at $1/4$, the estimates are extremely unstable.

To measure whether there is a weak instruments problem, in Panel C we display a generalization of the concentration parameter of the first stage, allowing for interdependence in the variance following (Kleibergen 2007).[37] The intuition is that in these networks, even for the analytical correction there is measurement error in the instrument. Since the number of connections to

[35]The correction working for degree with G^S is mechanical since there is no mismeasurement for d_i with $i \in S$.

[36]The number of simulations was chosen to roughly equate the computation time, on the order of $n^4 \cdot$ # of simulations, for each of the three specifications.

[37]For a first stage $X = Z\pi + v$, we use the generalized concentration parameter $\mu^2 := \pi' \Sigma_\pi^{-1} \pi$ where $\widehat{\pi} = (Z'Z)^{-1} Z'X$ and $\Sigma_\pi = \text{var}(\widehat{\pi})$.

neighbors and second neighbors in a network is low, the amount of noise in the first stage increases.[38] Panel C shows that the concentration parameter is very low for the first stage estimates in Model 1, especially at low sampling levels. Moreover, once the number nodes in the network is high enough or the amount of independent data (20 networks) is high enough, the concentration parameter is extremely high. In these cases our analytical correction removes the bias entirely while biases remain with the sampled estimators.

1.5.4 A Model of Diffusion

We numerically study the (Jackson and Rogers 2007) model of diffusion and present the results in Table 1-5. In Panel A we use the aforementioned simulated network data to generate a model with $\beta_0 = e^{-2}$.[39] Columns 1-5 present evidence of severe expansion bias in the estimates $\widehat{\beta}$ when using sampled data. At 1/3 sampling, the transmission parameter is overestimated by 250% when we study the induced subgraph and 85% when we turn to the star subgraph.

1.6 Using the Results to Better Collect Data

In this section we discuss how researchers can adopt our framework to think about data collection. The question we are interested in is: given that a researcher faces a budget constraint and needs to trade off the sampling rate and the number of networks in her sample, is there a method by which she can assess the trade-off?

We suppose that a researcher is interested in estimating a coefficient in a regression of an outcome on a network statistic. Assume that the researcher has a project budget b and a pilot budget p. Each village has a fixed cost f associated with the survey as well as a variable cost c for sampling.[40] We assume that the cost to sample individuals is linear and therefore the cost to sample a ψ-sample of the village is $c\psi n$. Finally, let $\bar{R}$ be the maximum number of villages available to study.

We posit that the researcher is interested in minimizing mean-squared error (MSE) in the estimation of β_0.[41] The relevant program is[42]

$$\min_{\psi \in [0,1], R \leq \bar{R}} \text{MSE}(\psi, R) \text{ s.t. } (c\psi n + f) R \leq b. \tag{1.6}$$

At the optimum $\psi = (b/R - f)/(cn)$ and therefore we may consider the concentrated objective

[38] The extent to which this matters can be seen by noticing that the concentration parameter is 2 for $\psi = 1/4$, while if the signal to noise ratio had stayed the same in the first stage, the concentration parameter should have only decreased from 16 to 4.

[39] This choice was motivated by (Jackson and Rogers 2007) who numerically show this corresponds to a 20% steady-state rate of diffusion. This matches the microfinance take-up rate in our empirical application.

[40] This method can be applied to richer budgeting frameworks.

[41] Researchers can replace this with an objective function of their choosing.

[42] For formal asymptotics we may have to let $b = b_n$ grow such that $\frac{b_n}{\bar{R}n} \to k$ some positive constant.

function $\text{MSE}(R) = \text{MSE}(\psi(R), R)$. The researcher may estimate the MSE by fully sampling a small number of networks and hypothesizing β_0 and R^2 from the linear regression, in a manner analogous to performing power calculations by positing effect sizes and intra-cluster correlations before conducting a field experiment (e.g., (Duflo, Glennerster, and Kremer 2007)). A researcher first randomly selects k of the $\bar{R}$ graphs using the pilot budget, where $k = p/(cn + f)$. Then, using these k networks, the researcher conducts a numerical experiment, sampling them at different rates and applying graphical reconstruction to estimate the MSE. By doing this, she can select the optimal ψ and R.

ALGORITHM (Research Design).

1. Pick network statistics and network-based hypotheses to test.
2. Hypothesize β, R^2, and generate outcome variable.
3. Randomly sample $k = p/(cn + f)$ out of $\bar{R}$ villages and obtain entire networks.
4. For each $R \in \{\underline{R}, ..., \bar{R}\}$
   (a) Randomly draw, with replacement, R villages from the collection of k networks.
   (b) Estimate $\text{MSE}(R)$ using the sample and hypothesized parameter values.
      i. Sample each of the R village networks at rate $\psi(R) = (b/R - f)/(cn)$.
      ii. Apply graphical reconstruction to estimate $\widehat{\beta}_{\text{ols}}$ using outcome variable from (2).
      iii. Repeat 4(a) and 4(b).i-ii for B simulations.
5. Pick $R^* \in \text{argmin}_R \widehat{\text{MSE}}(R)$ *and pick* $\psi^* = \psi(R^*)$.

The algorithm enables the researcher to estimate the trade-off she faces, given her interest in specific network effects and the distribution of graphs in her region of study. We conduct a simulation exercise to demonstrate this procedure. We set $b = \$152,400$, $f = \$1200$, $c = \$12$, $n = 200$, $\bar{R} = 150$ and assume that the networks are drawn from the family described in section 1.5 and the empirical Indian networks.[43] We consider a grid of $R \in \{33, 40, 50, 60, 70\}$ and $\psi \in \{1, 0.7, 0.4, 0.2, 0.1\}$.

Figure 1-4 displays results for two node-level statistics, eigenvector centrality and clustering, as well as a network-level statistic, the maximal eigenvalue of the adjacency matrix (λ_{max}). We repeat the exercise for both our simulated network data as well as the Indian networks. The figure shows $\text{MSE}(\psi(R), R)$ for sampled networks and graphical reconstruction. It also displays a theoretical lower bound on MSE by plotting the MSE corresponding to using R graphs sampled at 100% instead of at $\psi(R)$. Of course, we find that MSE increases greatly as we move away from 100% sampling and use the raw sampled data. Next we turn to graphical reconstruction and focus on the star subgraph. Looking at the Indian networks, eigenvector centrality has the lowest MSE at 40%

[43] The numbers are motivated from (Banerjee, Chandrasekhar, Duflo, and Jackson 2011).

sampling while clustering has an optimum at 70%. For these statistics, the simulated networks give 100% as the optimum. Meanwhile, $\lambda_{\max}$ has the lowest MSE at 100% sampling with the Indian networks but 20% sampling is the optimum in the simulated networks.

Taken together, the results suggest that, first, performing graphical reconstruction is very important, even with model misspecification as the researcher will not know the true families generating the empirical networks. Second, the MSE-minimizing sampling rate depends greatly on parameters, the network family, and the statistic of interest. It is difficult, if not impossible, to say *ex ante* where the optimum lies, and systematic procedures that depend on the setting may be better than rules of thumb. Third, the results push against the prevailing habit of researchers to obtain more cluster-units (e.g., villages) at lower sampling rates when conducting cluster-level analysis. Our results suggest that, at times, just obtaining better data with fewer cluster-units may be worthwhile. Though it is not surprising that network-level statistics exhibit higher levels of MSE, as there are only R as opposed to nR observations, this says nothing about the trade-off between the sampling rate and number of villages.

1.7 Conclusion

Applied social network analysis often use graphs constructed from data collected from a partial sample of nodes. Even when nodes are selected randomly, the partial sampling induces non-classical measurement error and consequently biases estimates of regression coefficients and GMM parameters. Moreover, these biases are of unclear sign and magnitude. We analytically examine the biases in the estimation of a number of network-based regression and GMM models with applications to a variety of economic environments. To address the problem in general, we develop a consistent and asymptotically normal method to estimate the economic parameters using graphical reconstruction, while allowing for substantial heterogeneity across networks. Specifically, the method allows for every network in the sample to be generated by a different model.

We conclude that network-based applied work must proceed cautiously, paying close attention to network data quality. From an applied perspective, researchers should be careful to work either with specifications which provide conservative results when facing sampled data or implement bias correction procedures if possible. Moreover, researchers ought to address the bias problem *ex ante*, either by choosing a unit of study where more complete data is available or using graphical reconstruction to understand how mean-squared error may vary with the sampling rate. Undoubtedly, the performance of graphical reconstruction with empirical network data will only improve as the burgeoning literature on consistently estimable network formation models matures. To that end, from a theoretical perspective the lacuna in the literature is the absence of network formation models that both allow for higher-order dependencies in link formation and are also consistently estimable.

1.A Figures and Tables

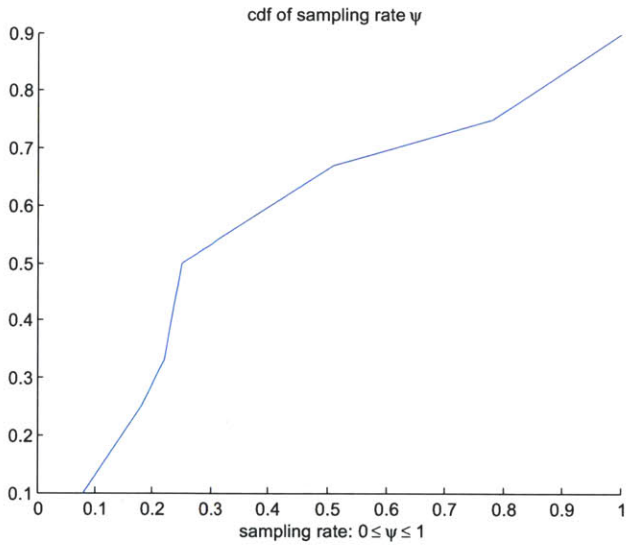


Figure 1-1: A CDF of the sampling rate in the literature. A detailed description of the literature survey is available from the author upon request.

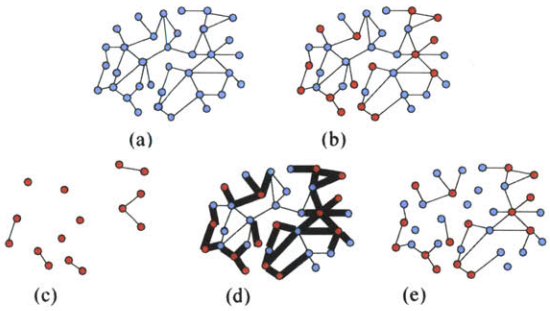


Figure 1-2: (a) $G = (V, E)$. (b) the sampled nodes, S, in red. (c) the induced subgraph, $G^{|S} = (S, E^{|S})$. (d) G, highlighting the sampled nodes and the edges that are induced if each sampled node reports all of its links from the census, E^S. (e) the star subgraph, G^S.

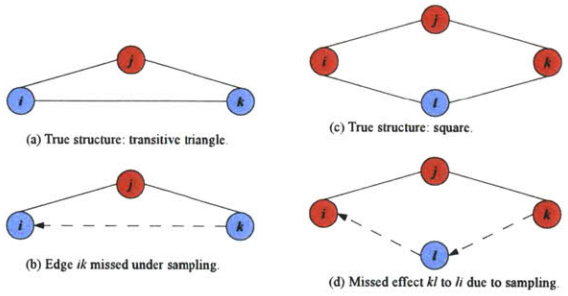


Figure 1-3: Red nodes are sampled. (a) and (c) show examples of true network structures. (b) shows the star subgraph from (a) when j is sampled. The effect of k on i is missed. (d) shows the induced subgraph from (c) when i, j, and k are sampled. The effect of k on i through l is missed.

Table 1-1: Graph Level Regressions (Simulated Network Graphs)

			Raw Network Statistics					*Analytic Corrections*					*Graphical Reconstruction*				
			1/4	1/3	1/2	2/3	1	1/4	1/3	1/2	2/3	1	1/4	1/3	1/2	2/3	1
Statistic	Type	Output	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]
Degree	G^S	Bias %	123.84%	92.13%	30.96%	12.66%	-0.01%	-1.97%	-1.97%	-1.24%	-1.03%	0.06%	-0.57%	-0.57%	-0.61%	-0.23%	-0.01%
	G^{IS}	Bias %	313.25%	245.06%	104.82%	55.62%	-0.61%	4.21%	2.56%	3.38%	8.23%	-0.01%	-7.44%	-5.50%	-1.26%	-0.87%	-0.61%
Clustering	G^S	Bias %	92.96%	69.73%	24.61%	10.51%	0.11%	1.41%	0.96%	-0.12%	-0.62%	-0.12%	-1.04%	-1.09%	-0.56%	0.02%	0.11%
	G^{IS}	Bias %	-27.15%	-14.69%	-1.78%	-0.02%	-0.14%	-26.71%	-14.71%	-2.01%	5.05%	-0.06%	-14.67%	-9.69%	-2.92%	-1.25%	-0.14%
Graph Span	G^S	Bias %	-45.27%	-36.97%	-23.48%	-11.56%	0.21%	-16.94%	-14.56%	-10.03%	-5.28%	-0.12%	-3.40%	-3.96%	-0.67%	-0.24%	0.35%
	G^{IS}	Bias %	-88.27%	-73.20%	-51.36%	-32.78%	0.09%	-30.67%	-23.44%	-16.16%	-10.43%	0.04%	-33.78%	-24.54%	-9.09%	-4.05%	0.54%
Avg Path Length	G^S	Bias %	-42.81%	-38.57%	-19.12%	-9.18%	0.38%						-5.35%	-5.81%	-1.62%	-0.79%	0.38%
	G^{IS}	Bias %	-103.98%	-81.71%	-53.39%	-33.75%	-0.01%						-43.21%	-33.10%	-13.48%	-6.15%	-0.01%
Max Eig	G^S	Bias %	79.55%	60.81%	24.20%	10.58%	0.00%						-0.57%	-0.91%	-0.22%	-0.18%	0.00%
	G^{IS}	Bias %	325.42%	259.54%	111.53%	58.55%	-0.24%						-7.35%	-5.57%	-1.59%	-0.65%	-0.24%
Eig. Cent. S.D.	G^S	Bias %	-51.56%	-50.03%	-30.73%	-16.48%	0.33%						-4.12%	-4.11%	-0.77%	0.69%	0.33%
	G^{IS}	Bias %	-92.91%	-92.26%	-84.07%	-72.43%	0.46%						-32.30%	-25.40%	-12.07%	-5.56%	0.46%
Spectral Gap	G^S	Bias %	-50.47%	-45.02%	-22.48%	-10.47%	0.12%						-38.81%	-32.66%	-13.76%	-6.15%	0.12%
	G^{IS}	Bias %	-101.45%	-101.55%	-71.14%	-50.59%	0.14%						-87.93%	-84.79%	-61.20%	-41.40%	0.14%

Notes: Bias % denotes $(\beta - \beta_0)/\beta_0$ x 100% and coverage percentage is available upon request. All reported mean bias % simulations have standard errors less than 3.07% with most having estimated standard errors of the mean bias % between 0.20% and 1.50% (percentage points). A complete table including simulation precision is available from the authors upon request. Estimation was performed using 50 homophily graphs with 200 nodes, 10 cliques of equal size, and expected degree of ~15, as described in the text. Each simulation was performed 100 times. To compute the conditional expectation in graphical reconstruction 40 simulations were performed. Coverage was computed using bootstrap standard errors. Heteroskedasticity-consistent standard errors yield similar results.

Table 1-2: Node Level Regressions (Simulated Network Data)

			Raw Network Statistics					*Experimental Corrections*					*Graphical Reconstruction*				
			1/4	1/3	1/2	2/3	1	1/4	1/3	1/2	2/3	1	1/4	1/3	1/2	2/3	1
Statistic	Type	Output	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]
Degree	G^s	Bias %	-60.60%	-56.56%	-32.33%	-13.66%	0.03%	-0.47%	-0.52%	-0.62%	-0.43%	-0.43%	-0.85%	-0.93%	-0.39%	-0.27%	0.03%
	G^{ls}	Bias %	106.09%	91.04%	53.03%	32.06%	-0.06%	-48.58%	-42.37%	-23.29%	-13.99%	0.36%	-7.58%	-6.36%	-2.00%	-1.07%	-0.06%
Clustering	G^s	Bias %	-80.68%	-75.74%	-44.20%	-17.09%	-0.01%	-89.75%	-85.63%	-55.18%	-25.47%	-0.06%	-1.55%	-1.35%	-0.63%	-0.34%	-0.01%
	G^{ls}	Bias %	-83.85%	-79.88%	-59.12%	-39.37%	-0.39%	-83.95%	-80.12%	-59.20%	-36.04%	0.02%	-14.40%	-9.62%	-3.75%	-2.01%	-0.39%
Path Length	G^s	Bias %	-68.03%	-55.98%	-22.32%	-9.35%	0.40%						-3.01%	-2.94%	-1.16%	-0.26%	0.40%
	G^{ls}	Bias %	-100.66%	-88.82%	-51.77%	-30.64%	-0.12%						-27.64%	-20.26%	-7.30%	-2.98%	-0.12%
Eig Centrality	G^s	Bias %	-60.74%	-54.46%	-30.00%	-14.56%	0.19%						-2.02%	-1.17%	-0.50%	-0.14%	0.19%
	G^{ls}	Bias %	-89.42%	-85.81%	-69.14%	-50.59%	0.32%						-4.38%	-3.81%	-3.04%	-1.23%	0.32%
Same Side	G^s	Bias %	-9.97%	-6.89%	-2.27%	-0.88%	-0.17%						-0.23%	0.04%	-0.10%	0.22%	-0.17%
	G^{ls}	Bias %	-45.60%	-32.78%	-7.66%	-3.90%	0.23%						0.17%	-0.13%	0.07%	-0.19%	0.23%

Notes: Bias % denotes $(\beta - \beta_0)/\beta_0$ x 100% and coverage is available upon request. All reported mean bias % simulations have standard errors less than 0.65% with most having estimated standard errors of the mean bias % between 0.10% and 0.20% (percentage points). A complete table including simulation precision is available from the authors upon request. Estimation was performed using 50 homophily graphs with 200 nodes, 10 cliques of equal size, and expected degree of ~15, as described in the text. Each simulation was performed 100 times. To compute the conditional expectation in graphical reconstruction 40 simulations were performed. Several standard errors were computed. Coverage was computed using block bootstrap standard errors: clustered standard errors yield very similar coverage rates.

Table 1-4: t-statistic Diagnosis (Simulated Network Data)

	# of Stats where $t_{GR}/t_{Naive} > 1$ out of tota				Mean of $t_{Reconstructed}/t_{Naive}$			
	1/4	1/3	1/2	2/3	1/4	1/3	1/2	2/3
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Panel A: Graph Level Regressions								
$G^{|S}$	6/7	7/7	7/7	7/7	5.32	2.15	1.61	1.29
G^{S}	5/7	7/7	6/7	3/7	1.61	1.41	1.26	1.08
Panel B: Node Level Regressions								
$G^{|S}$	5/6	5/6	6/6	6/6	4.09	2.74	2.21	1.83
G^{S}	6/6	6/6	6/6	6/6	3.23	2.94	2.46	1.93

Notes: The left panel displays the fraction of times the t-statistic increases when using graphical reconstruction as compared to using the raw sampled network statistic across network statistics. The right panel displays the average ratio of t-statistic under graphical reconstruction to t-statistic under the raw sampled network statistic across network statistics. We use 7 graph-level statistics and 6 node-level network statistics, identical to those used in Tables 1 and 2. The simulation data is the same as that used in Tables 1 and 2.

Table 1-5: Bias in Estimation of β_0 in Jackson and Rogers (2007) Model

	Raw Network Data				*Graphical Reconstruction*			
	1/4	1/3	1/2	2/3	1/4	1/3	1/2	2/3
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Panel A: Simulated Networks, Bias % in Estimation of β_0								
$G^{|S}$	329.0%	250.0%	104.0%	55.0%	1.0%	0.0%	0.0%	0.0%
G^{S}	117.0%	85.0%	28.0%	12.0%	0.0%	0.0%	0.0%	0.0%
Panel B: Indian Networks, Bias % in Estimation of β_0								
$G^{|S}$	263.0%	230.0%	103.0%	53.0%	8.0%	8.0%	7.0%	6.0%
G^{S}	129.0%	92.0%	31.0%	14.0%	6.0%	5.0%	3.0%	1.0%

Notes: Table presents bias in estimation of β_0 the transmission parameter in the Jackson-Rogers diffusion model described in section 3.3 of the text. Data generating process for the simulated networks is the same as in Table 2. Networks in panel B are same as those described in Table 6. We set β_0 to exp(-2) and perform each simulation 100 times.

Table 1-3: Peer Effects Regressions (Simulated Graphs)

	Model 1: 1 Graph; σ_ε=1; 50k Sims.					*Model 2:* 1 Graph; σ_ε=¼; 50k Sims.					*Model 3:* 20 Graphs; σ_ε=1; 2,500 Sims.				
	1/4	1/3	1/2	2/3	1	1/4	1/3	1/2	2/3	1	1/4	1/3	1/2	2/3	1
Panel A: 2SLS Mean Bias %	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]
G^{ls} m obs.	1606%	-146%	-56%	-40%	-1%	-107%	-78%	-56%	-37%	0%	-56%	-46%	-32%	-19%	0%
	[1682%]	[54%]	[2%]	[0%]	[0%]	[39%]	[2%]	[0%]	[0%]	[0%]	[1%]	[1%]	[0%]	[0%]	[0%]
G^s (y,x)-census, n obs.	-69%	-55%	-27%	-12%	-1%	-69%	-55%	-27%	-11%	0%	-35%	-22%	-5%	0%	0%
	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]	[0%]
G^s (y,x)-sample, m obs.	553%	404%	32%	54%	-2%	-81%	900%	288%	-267%	0%	432%	382%	140%	62%	0%
	[776%]	[442%]	[328%]	[51%]	[0%]	[302%]	[1848%]	[145%]	[443%]	[0%]	[147%]	[19%]	[1%]	[1%]	[0%]
G^s Analytic Correction; (y,x)-census, m obs.	18%	-19%	-3%	-2%	-1%	-1%	-1%	0%	0%	0%	-1%	0%	0%	0%	0%
	[178%]	[9%]	[1%]	[0%]	[0%]	[1%]	[0%]	[0%]	[0%]	[0%]	[1%]	[0%]	[0%]	[0%]	[0%]
Panel B: 2SLS Mean Bias %															
G^{ls} m obs.	-3692%	70%	-58%	-34%	1%	-30%	-62%	-53%	-36%	0%	-125%	-122%	-102%	-73%	0%
	[3602%]	[111%]	[5%]	[1%]	[1%]	[75%]	[3%]	[1%]	[0%]	[0%]	[1%]	[1%]	[1%]	[1%]	[0%]
G^s (y,x)-census, n obs.	-86%	-80%	-61%	-28%	1%	-83%	-78%	-58%	-28%	0%	-150%	-143%	-104%	-51%	0%
	[1%]	[1%]	[1%]	[1%]	[1%]	[0%]	[0%]	[0%]	[0%]	[0%]	[1%]	[1%]	[0%]	[0%]	[0%]
G^s (y,x)-sample, m obs.	-900%	-905%	-113%	-113%	2%	194%	-1631%	-614%	584%	0%	-913%	-800%	-295%	-129%	0%
	[1486%]	[935%]	[701%]	[108%]	[1%]	[606%]	[3642%]	[321%]	[961%]	[0%]	[304%]	[39%]	[3%]	[1%]	[0%]
G^s Analytic Correction; (y,x)-census, m obs.	-41%	38%	5%	2%	1%	2%	2%	0%	0%	0%	2%	0%	0%	0%	0%
	[361%]	[19%]	[1%]	[1%]	[1%]	[2%]	[1%]	[0%]	[0%]	[0%]	[1%]	[1%]	[1%]	[0%]	[0%]
Panel C: Concentration Parameter															
G^{ls} m obs.	1.2	1.8	3.7	6.6	15.6	15.7	26.6	60.9	111.6	272.0	17	28	51	74	111
G^s (y,x)-census, n obs.	7.0	8.2	10.6	13.1	15.6	118.2	139.6	182.1	227.3	271.2	131	136	128	119	112
G^s (y,x)-sample, m obs.	1.1	1.2	1.9	4.2	15.5	15.2	16.5	28.7	69.6	271.0	1	3	14	36	112
G^s Analytic Correction; (y,x)-census, m obs.	1.7	2.9	6.3	10.0	15.5	26.1	47.2	106.4	172.2	271.6	25	39	70	94	112

Estimators: The model is $y = \alpha + \rho_0 w(G)y + \gamma_0 x + \delta_0 w(G)x + \varepsilon$. We consider four estimators. Each estimator uses a sampled graph, a sample of (y,x), and a data sample for the 2-stage least squares (2SLS) estimation. The sampled graph is either G^{ls} or G^s, the sample of (y,x) is either for the sampled nodes or for an entire census, and the data sample for 2SLS is either data for the m sampled nodes or all available data (any unknown or missing data is replaced with zeros) for the census of n nodes. The "Analytic Correction" uses G^s, the census of (y,x), and 2SLS estimation on only the data for the m sampled nodes. The concentration parameter c is given as follows following Kleibergen (2007). For a first stage $X = Z\pi + v$, we use $c = \pi' S^{-1} \pi$ where $S := \text{var}(p)$ and $p = (Z'Z)^{-1}Z'X$.

Notes: Estimation is performed using homophily graphs with 250 nodes, 6 cliques of equal size, and expected degree of ~15. The simulation model parameters are $\rho 0$=0.5 and $\delta 0$=0.5. All bias percentages are computed relative to these parameter values. Standard errors of the simulation means are shown in brackets. For comparison, OLS biases at 100% sampling are [$\rho 0$; $\delta 0$] = [46%; -98%], [8%; -16%], and [28%; -63%], reading across the three columns.

Table 1-8: Bias for Network Level Regression for Average Degree (Simulated Graphs)

Estimator for Regression on Avg. Degree	Property	Edge Sampling Rate (1/4)^2	(1/3)^2	(1/2)^2	(2/3)^2	1
β Naïve Estimator	% Bias	1361.2%	756.8%	291.4%	123.0%	0.0%
	Coverage	0%	0%	0%	0%	96%
β Analytic Correction	% Bias	-8.7%	-4.8%	-2.2%	-0.9%	0.0%
	Coverage	39%	61%	81%	87%	95%
β Regularized Analytic Correction	% Bias	0.3%	0.3%	-0.2%	-0.1%	0.0%
(Using Analytic Variance of Correction)	Coverage	97%	96%	96%	96%	95%
β Reg. Analytic Correction, Bootstrap BC	% Bias	-0.4%	-0.1%	-0.4%	-0.1%	0.0%
(Using Analytic Variance of Correction)	Coverage	97%	96%	96%	96%	95%

Estimators: The naïve estimator uses the subgraph generated by independently sampling edges at the given sampling rates. The analytic correction estimator, as noted, has residual bias. The regularized analytic correction adjusts for the dispersion term using an analytic formula for the variance of the remaining measurement error left after performing the analytic correction, which is only correct in expectation.

Notes: The sampling rates have been chosen to be comparable with the edge count in the G^S sampling simulations. Coverage computed using bootstrapped standard errors for all estimators. R2 was set to 0.7. There were 200 simulations done using the standard setup: homophilic networks with 6 groups, graphs with 250 nodes, and 50 graphs per regression.

Table 1-6: Summary of Results from Numerical Simulations with Indian Village Networks

	Network Level Regressions				*Node Level Regressions*			
	1/4	1/3	1/2	2/3	1/4	1/3	1/2	2/3
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
Panel A: G^s								
Raw Network Statistic Median Bias	67.52%	48.62%	23.48%	9.88%	88.02%	70.10%	48.68%	25.33%
Reconstruction Median Bias	0.90%	0.56%	0.35%	0.10%	6.97%	4.62%	1.84%	0.73%
Reconstruction Reduces Bias	8/8	8/8	8/8	8/8	5/5	5/5	5/5	5/5
Raw Network Statistic Bias	67.52%	47.76%	23.48%	11.56%				
Analytic Correction Median Bias	14.58%	6.94%	2.33%	0.65%				
Analytic Corrections Reduce Bias	3/3	3/3	3/3	3/3				
Panel B: G^{ls}								
Raw Network Statistic Median Bias	88.27%	73.20%	51.36%	32.78%	90.94%	81.50%	60.25%	38.46%
Reconstruction Median Bias	13.50%	5.79%	2.28%	0.66%	28.43%	16.91%	7.98%	3.97%
Reconstruction Reduces Bias	7/8	8/8	8/8	8/8	5/5	5/5	5/5	5/5
Raw Network Statistic Bias	190.26%	132.95%	75.19%	41.41%				
Analytic Correction Median Bias	18.57%	13.62%	8.69%	5.41%				
Analytic Corrections Reduce Bias	2/2	2/2	2/2	2/2				

Notes: We ran all the regressions from Tables 1 and 2. We present the median bias across several network statistics at each sampling rate, where the network statistics are those analyzed in Tables 1 and 2. Networks are drawn from the data-set of Banerjee et al. (2011) where the graphs are treated as full networks and then sampled at rates 1/4, 1/3, 1/2, and 2/3. The network formation model used is a conditional edge-independence logistic regression model where covariates are distance in GPS coordinates, number of beds, number of rooms, number of latrines, and electricity provision. The full table is available upon request.

Table 1-7: Empirical Application: Diffusion of Microfinance

Panel A: Graph Level Regressions of Microfinance Take-up Rate of Village on Network Characteristics

	Eigenvector Centrality of Injection Point	Average Path Length	Conductance	Var. of Eig. Centrality Distribution
	[1]	[2]	[3]	[4]
G^s	1.6335	-0.0545	0.9659	58.6554
	[0.8715]	[0.0488]	[0.2671]	[40.5225]
Graphical Reconstruction	2.4244	-0.0928	1.4443	87.7406
	[1.0680]	[0.0597]	[0.3644]	[49.4477]

Panel B: Vertex Level Regressions of Microfinance Take-up Decision of a Household on Network Characteristics

	Eigenvector Centrality of Household	Inverse Distance to Injection Point
	[1]	[2]
G^s	0.5528	0.0221
	[0.1887]	[0.0277]
Graphical Reconstruction	0.662	0.0453
	[0.2091]	[0.0286]

Panel C: Regression of Household's Take-up Decision on Neighbors' Take-up Decisions

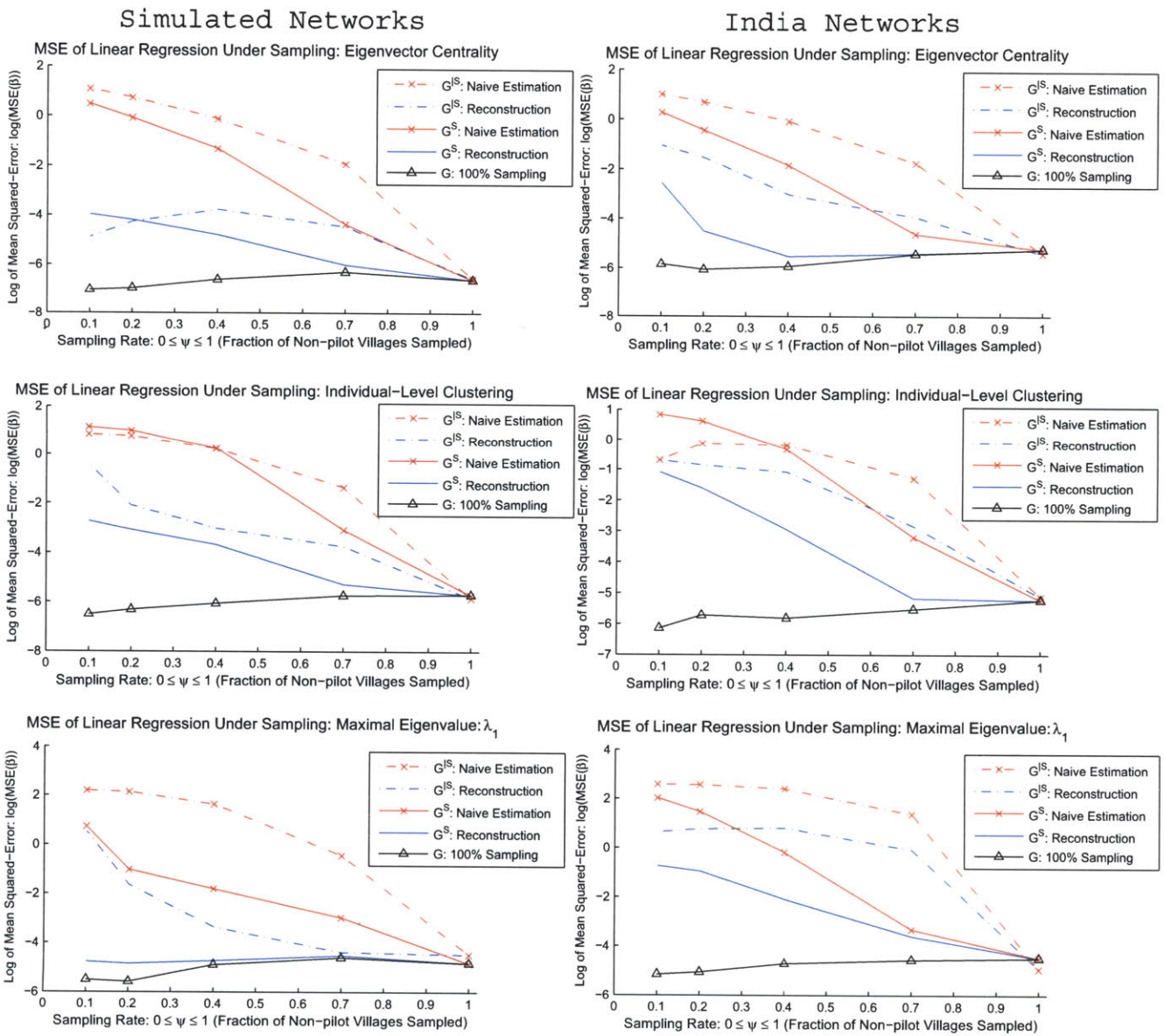
	ρ	γ	δ
	[1]	[2]	[3]
G^s, (y,x)-Census; n obs.	-0.027274	0.058066	0.016371
	[0.0221]	[0.0138]	[0.0115]
G^s, (y,x)-Sample; m obs.	0.047085	0.079207	-0.00090318
	[0.0098]	[0.0175]	[0.0047]
Analytic Correction, (y.x)-Census; m obs.	-0.071297	0.025769	0.025271
	[0.049]	[0.0206]	[0.0163]

Notes for Panel A: This panel provides estimates of regressions of the microfinance take-up rate of a village on its network characteristics. Each column presents a separate regression and all regressions control for number of households, fraction of households with savings access, and fraction of households that are SC/ST (scheduled caste/ scheduled tribe). The sample consists of 43 villages. The mean of the dependent variable is 0.1849. All regressions control for the number of nodes in the graph. Standard errors are heteroskedasticity robust.

Notes for Panel B: This panel provides estimates of regressions of whether a household decides to join microfinance on its network characteristics. The sample consists of 8441 households. The mean of the dependent variable is 0.1750. All regressions include village fixed effects. Standard errors are clustered by village. Inverse distance to injection point is the inverse of the minimal path length from the household in question to the set of initially informed households.

Notes for Panel C: This panel provides 2SLS estimates of regressions of whether a household joins microfinance on whether its neighbors join microfinance. The sample of (y,x) is either for just the sampled nodes or the entire census and the data sample for 2SLS is either data for the m sampled nodes or all available data for the census of n nodes. The analytic correction uses the induced subgraph, the census of (y,x), and 2SLS estimation on only the data for the m sampled nodes. Standard errors are clustered by village.

Figure 1-4: Budget Trade-Offs



* *The sampling rates used for these plots were 1, 0.7, 0.4, 0.2, and 0.1 which correspond to sampling a total of 33, 40, 50, 60, and 70 villages respectively (including 9 pilot villages).*

1.B Proofs for Section 1.3

1.B.1 Proofs for Section 1.3.1

Let (a_R) be a sequence of normalizing constants.

ASSUMPTION 1.B.1. *The mismeasured regressor is* $X_{rR}^* = \pi_R X_{rR} + v_{rR}$, $\lim_{R\to\infty} \pi_R = \pi$, $\operatorname{plim}_{R\to\infty} a_R^{-2}\mathbb{E}_R[v_r X_r] = 0$.

ASSUMPTION 1.B.2.

1. *The data consists of* $(y_{rR}, X_{rR})_{r\leq R}$ *with* $\mathrm{E}[\epsilon_R|X_R] = 0$ *and the model is* $y_{rR} = \alpha + X_{rR}\beta_0 + \epsilon_{rR}$.
2. $0 < \sigma_x^2 < \infty$ *where* $\sigma_x^2 := \operatorname{plim}_{R\to\infty} a_R^{-2}\mathbb{E}_R\left(X_r - \bar{X}_{Rr}\right)^2$.
3. $\sigma_v^2 := \operatorname{plim}_{R\to\infty} a_R^{-2}\mathbb{E}_R v_r^2 < \infty$.

LEMMA 1.B.1. *Under Assumption 1.B.1 and 1.B.2,* $\widehat{\beta} \xrightarrow{\mathrm{P}} \pi^{-1}\beta_0 \frac{\sigma_x^2}{\sigma_x^2+\pi^{-2}\sigma_v^2}$.

Proof. The proof is classical and follows from $\operatorname{plim}\widehat{\beta} = \operatorname{plim}\left(a_R^{-2}X_R^{*\prime}X_R^*\right)^{-1} a_R^{-2}\left(\pi_R X_R + v_R\right)' X_R\beta_0 = \frac{\beta_0}{\pi}\frac{\sigma_x^2}{\sigma_x^2+\pi^{-2}\sigma_v^2}$. □

Proof of Proposition 1.3.1(1). From Lemma 1.B.2 and 1.B.3, Assumption 1.B.1 holds,

$$\mathrm{E}[d(G_{rR}^{|S})|G_{rR}] = \underbrace{(\psi + \Theta(n^{-1}))}_{\pi_R}\underbrace{d(G_r)}_{X_{rR}} \text{ and } \mathrm{E}[d(G_r^S)|G_r] = \underbrace{(1-(1-\psi)^2 + \Theta(n^{-1}))}_{\pi_R}\underbrace{d(G_r)}_{X_{rR}}.$$

The assumptions of the proposition ensure Assumption 1.B.2 holds. The result follows from Lemma 1.B.1. □

In what follows, we let $t(G) := \sum_{i<j<k} A_{ij}A_{jk}$ denote the number of two-stars ($A_{ij}A_{jk} = 1$), and $p(G) := \sum_{i<j<k<l} A_{ij}A_{kl}$ denote the number of disjoint pairs in the graph.

LEMMA 1.B.2. *Under uniform random sampling of* m *out of* n *nodes, with* $m/n \to \psi$ *and as* $n \to \infty$,

$$\mathrm{E}[d(G^{|S})|G] = (\psi + \Theta(n^{-1}))d(G) \text{ and } \operatorname{var}(d(G^{|S})|G) = \Theta(n^{-1}d(G)) + \Theta(n^{-1}t(G)) + \Theta(n^{-2}p(G)).$$

Proof. Step 1: Let $\mathcal{S}$ be the set of all combinations of m vertices with σ a generic element. Then

$$\mathrm{E}[d(G^{|S})|G] = \sum_{\sigma\in\mathcal{S}} d(\sigma)\mathrm{P}(\sigma) = \binom{n}{m}^{-1}\sum_{\sigma\in\mathcal{S}} m^{-1}\sum_{i\wedge j\in\sigma} A_{ij}.$$

Each pair ij appears $|\{\sigma\in\mathcal{S} : i\wedge j\in\sigma\}| = \binom{n-2}{m-2}$ times, it follows that $\mathrm{E}[d(G^{|S})|G] = \frac{m-1}{n-1}\cdot\frac{1}{n}\sum_{i,j}A_{ij} = (\psi + \Theta(n^{-1}))d(G)$.

Step 2: Use $|\{\sigma\in\mathcal{S} : i_1\wedge\ldots\wedge i_k\in\sigma\}| = \binom{n-k}{m-k}$ and let $\epsilon_1 := \sum A_{ij}$, $\epsilon_2 := \sum\{A_{ij}A_{kl} : \ i\vee j\in\{k,l\}, \neg i\wedge j\in\{k,l\}\ \}$, and $\epsilon_3 := \sum\{A_{ij}A_{kl} : \ i\notin\{j,k,l\}, j\notin\{k,l\}, k\neq l\ \}$. Then $\epsilon_1 = 2|E|$, $\epsilon_2 = 8t(G)$, $\epsilon_3 = 8p(G)$, and

$$\mathrm{E}[d(G^{|S})^2|G] = \frac{1}{m^2}\frac{m!}{(n)_m}\sum_{j=1}^{3}\begin{pmatrix} n-(j+1) \\ m-(j+1)\end{pmatrix}\epsilon_j.$$

Some algebra yields $\mathrm{E}[d(G^{|S})^2|G] = \frac{1}{m}\frac{(m-1)}{(n-1)}\frac{2|E|}{n} + \frac{1}{m}\frac{(m-1)_2}{(n-1)_2}\frac{8t(G)}{n} + \frac{1}{m}\frac{(m-1)_3}{(n-1)_3}\frac{8p(G)}{n}$. Meanwhile, we can expand the square of the mean and obtain coefficients with the exact same summand terms $(\mathrm{E}[d(G^{|S})|G])^2 = \frac{1}{n^2}\left(\frac{m-1}{n-1}\right)^2\left(\sum_{i,j}A_{ij} + \sum A_{ij}A_{kl}\right)$. We have three sets of coefficients. Working with the coefficients on $\sum_{i,j}A_{ij}$, we have that

$$\frac{1}{m}\frac{(m-1)}{(n-1)}\frac{2|E|}{n} - \frac{1}{n}\left(\frac{m-1}{n-1}\right)^2\frac{2|E|}{n} = \frac{1}{n-1}\left(\frac{m-1}{m} - \frac{(m-1)^2}{n(n-1)}\right)d(G) = \left(n^{-1}(1-\psi^2) + \Theta(n^{-2})\right)d(G).$$

The second term is given by

$$\left(\frac{1}{m}\frac{(m-1)_2}{(n-1)_2} - \frac{1}{n}\left(\frac{m-1}{n-1}\right)^2\right)\frac{8t(G)}{n} = \frac{1}{n-1}\left(\frac{1}{m}\frac{(m-1)(m-2)}{(n-2)} - \frac{(m-1)^2}{n(n-1)}\right)8t(G) = \left(n^{-1}\psi(1-\psi) + \Theta(n^{-2})\right)8t(G).$$

Finally, we compute the last term $\left(\frac{1}{m}\frac{(m-1)_3}{(n-1)_3} - \frac{1}{n}\left(\frac{m-1}{n-1}\right)^2\right)\frac{8p(G)}{n} = \Theta(n^{-2})p(G)$, which completes the proof. □

LEMMA 1.B.3. *Under uniform random sampling of m out of n nodes, with $m/n \to \psi$ and as $n \to \infty$,*

$$\mathrm{E}[d(G^S)|G] = \psi(2 - \psi + \Theta(n^{-1}))d(G) \text{ and } \mathrm{var}\left(d(G^S)|G\right) = \Theta(n^{-1}d(G)) + \Theta(n^{-1}t(G)) + \Theta(n^{-2}p(G)).$$

Proof. Same as previous lemma, noting each pair ij appears $|\{\sigma \in \mathcal{S} : i \vee j \in \sigma\}| = 2\binom{n-1}{m-1} - \binom{n-2}{m-2}$ times in $\mathcal{S}$. □

Proof of Proposition 1.3.1(2). Step 1: For G^S, the expected number of triangles is given by $\mathrm{E}[\rho(G^S)|G] = (3\psi^2(1-\psi)+\psi^3+o(1))\rho(G)$ since with probability $3\psi^2(1-\psi)$ a transitive triangle can be counted with two nodes being sampled and ψ^3 with all three being sampled.[44] Meanwhile, $\mathrm{E}[\tau(G^S)|G] = (\psi(1-\psi)^2 + 3\psi^2(1-\psi) + \psi^3 + o(1))\tau(G)$ where the extra term comes from the fact that the exact center of the triple in question can be selected. In turn,

$$\frac{\rho(G^S)}{\tau(G^S)} = \frac{3\psi^2(1-\psi)+\psi^3+o(1)}{\psi(1-\psi)^2+3\psi^2(1-\psi)+\psi^3+o(1)} \cdot \frac{\rho(G)}{\tau(G)} = \frac{\psi(3-2\psi)}{1+\psi(1-\psi)} \cdot \frac{\rho(G)}{\tau(G)} + o(1).$$

For $G^{|S}$, since the sampling probability is $\psi^3 + o(1)$ for both the numerator and the denominator, there is no bias asymptotically in the estimate of clustering.

Step 3: From steps 1 and 2, Assumption 1.B.1 holds. The assumptions of the proposition ensure Assumption 1.B.2 holds. The result follows from Lemma 1.B.1. To check 1.B.2, let ϕ_ρ and ϕ_τ be the limiting scalings above. Notice

$$\left|\frac{\phi_\rho\rho_r}{\phi_\tau\tau_r} - \frac{\mathrm{E}\left[\rho(\bar{G}_r)|G_r\right]}{\mathrm{E}\left[\tau(\bar{G}_r)|G_r\right]}\right| \leq \left|\frac{\phi_\rho\rho_r\phi_\tau\tau_r - \mathrm{E}\left[\rho(\bar{G}_r)|G_r\right]\phi_\tau\tau_r}{\phi_\tau\tau_r\mathrm{E}\left[\tau(\bar{G}_r)|G_r\right]}\right| + \left|\frac{\phi_\rho\rho_r\mathrm{E}\left[\tau(\bar{G}_r)|G_r\right] - \phi_\rho\rho_r\phi_\tau\tau_r}{\phi_\tau\tau_r\mathrm{E}\left[\tau(\bar{G}_r)|G_r\right]}\right|.$$

Then we want to show $\mathrm{plim}\ \frac{1}{R}\sum_{r=1}^R \frac{\rho_r}{\tau_r}\left|\frac{\phi_\rho\rho_r - \mathrm{E}\left[\rho(\bar{G}_r)|G_r\right]}{\mathrm{E}\left[\tau(\bar{G}_r)|G_r\right]}\right| = 0$ and $\mathrm{plim}\ \frac{1}{R}\sum_{r=1}^R \frac{\phi_\rho\rho_r^2}{\phi_\tau\tau_r^2}\left|\frac{\mathrm{E}\left[\tau(\bar{G}_r)|G_r\right] - \phi_\tau\tau_r}{\mathrm{E}\left[\tau(\bar{G}_r)|G_r\right]}\right| = 0$. It is easy to check each summand converges to zero, e.g. $\frac{a_R\rho_r}{a_R\tau_r}\frac{a_R\left|\phi_\rho\rho_r - \mathrm{E}\left[\rho(\bar{G}_r)|G_r\right]\right|}{a_R\mathrm{E}\left[\tau(\bar{G}_r)|G_r\right]} = O(1)\cdot\frac{o(1)}{O(1)}$ since $a_R\mathrm{E}\left[\tau(\bar{G}_r)|G_r\right] \geq o(1) + \underline{\alpha} > 0$, and similarly for the second term. Notice that each summand is dominated by $\frac{\bar{\alpha}}{\underline{\alpha}}$ and a uniform constant. The result follows. □

1.B.2 Proofs for Section 1.3.1

Proof of Lemma 1.3.1. We show $\left|d(G^S) - (1-(1-\psi)^2)d(G)\right| = o_P(1)$. The arguments for $G^{|S}$ and $d_2(\cdot)$ are analogous. We have already shown that $\mathrm{E}\left[d(G^S)|G\right] = (1-(1-\psi)^2 + \Theta(n^{-1}))d(G)$in Lemma 1.B.3. Let χ_{ij} be an indicator of $i \vee j \in S$. Condition on the sequence of events $\mathcal{E}_n := \{d(G) \in (c_1a_{1n} \pm \epsilon_1a_{1n}) \cap d_2(G) \in (c_2a_{2n} \pm \epsilon_2a_{2n})\}$. By assumption of growth rates a_{1n} and a_{2n}, we these events happen with probability approaching one,

$$\mathrm{P}\left(\left|\frac{d(G)}{a_{1n}} - c_1\right| < \epsilon_1\right) \to 1 \text{ and } \mathrm{P}\left(\left|\frac{d_2(G)}{a_{2n}} - c_2\right| < \epsilon_2\right) \to 1.$$

Notice that the probability distribution and $d(G)$ are both implicitly indexed by n. Then using Chebyshev's inequality and letting $\varphi = (1-(1-\psi^2) + \Theta(n^{-1}))$,

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_i\sum_j A_{ij}\chi_{ij} - \varphi d(G)\right| > \epsilon \middle| \mathcal{E}_n\right) = \mathrm{P}\left(\left|\frac{1}{n}\sum_i\sum_{j>i} A_{ij}(\chi_{ij} - \varphi)\right| > \epsilon/2 \middle| \mathcal{E}_n\right) \leq \frac{4}{n^2\epsilon^2}\mathrm{var}\left(\sum_i\sum_{j>i} Z_{ij}|\mathcal{E}_n\right)$$

[44] We simply count, e.g., $\left\{3\binom{n-3}{m-2} + \binom{n-3}{m-3}\right\}\binom{n}{m}^{-1} = \frac{m(m-1)(m-2)}{n(n-1)(n-2)} = \psi^3 + o(1)$.

where $Z_{ij} = A_{ij}(\chi_{ij} - \varphi)$. If we condition on a graph $G \in \mathcal{E}_n$,

$$\operatorname{var}\left(\sum_i \sum_{j>i} Z_{ij}|G\right) = |E(G)| \operatorname{var}(Z_{ij}|G) + \sum_i \sum_{j>i} \sum_{k \neq i,j} \operatorname{cov}(Z_{ij}, Z_{jk}|G),$$

since $\operatorname{var}(Z_{ij}|G) = \varphi(1-\varphi)$ and $\operatorname{cov}(\chi_{ij}, \chi_{i'j'}|G) = \psi^3(1-\psi)$, with the covariance terms only entering when ij and $i'j'$ share a vertex. Recall that $d(G) = 2|E(G)|/n$, we have

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_i \sum_j A_{ij}\chi_{ij} - \varphi d(G)\right| > \epsilon \middle| \mathcal{E}_n, G\right) \lesssim \frac{1}{n}\left\{d(G)\varphi(1-\varphi) + d_2(G)\psi^3(1-\psi)\right\}$$

for every $G \in \mathcal{E}_n$ where the constant is uniform. Notice that $d(G) \in (c_1 a_{1n} \pm \epsilon_1 a_{1n})$ implies $d(G)/n \to 0$ because by assumption $a_{1n}/n \to 0$. The same is true for $a_{2n}/n \to 0$. This proves the result.

Next we look at $\left|d_2(G^S) - k(\psi)d_2(G)\right| = o_{\mathrm{P}}(1)$.

$$\mathrm{E}[n^{-1}\textstyle\sum_i \sum_{j\neq i} \sum_{k\neq i,j} A_{ij}A_{jk}\chi_{ij}\chi_{jk}|G] \quad = \quad n^{-1}\textstyle\sum_i \sum_{j\neq i} \sum_{k\neq i,j} A_{ij}A_{jk}\mathrm{E}[\chi_{ij}\chi_{jk}|g] = k(\psi)d_2(G).$$

This follows from $\mathrm{E}[\chi_{ij}\chi_{jk}|G] = 1-(1-\psi)^2[2-(1-\psi)]$. A similar argument to the above completes the proof. □

We relate the above to the statistical physics literature. To characterize the degree distribution we define a probability generating function $\Psi_0(x) = \sum_{k=0}^{\infty} p_k x^k$ such that $\Psi_0(1) = 1$, where $p_k := \mathrm{P}(k)$. Note that the expected degree is given by the derivative of the function evaluated at one, $\langle d \rangle = \sum_{k\in\mathbb{N}} k p_k = \Psi_0'(1)$. Assume nodes are sampled iid with probability ψ and let $\Psi_0^{|S}(x)$ be the generating function describing $G^{|S}$ and $\Psi_0^S(x)$ be that describing G^S.

LEMMA 1.B.4. $\Psi_0^S(x) = \psi\Psi_0(x) + (1-\psi)\Psi_0(1-\psi(1-x))$.

Proof. The proof for $\Psi_0^{|S}(x) = \Psi_0(1-\psi(1-x))$ comes from (Stumpf, Wiuf, and May 2005). We study $\Psi_0^S(x)$. Consider a random node of degree i. With probability ψ, the node is selected and thus the degree in the subgraph is i. With probability $(1-\psi)$, the node is not selected and thus the degree is $k \leq i$. The sampled probability is $p_{s,k}$, given by

$$p_{s,k} = \sum_{i\geq k}^{\infty} p_i \mathrm{P}(k|i) = \sum_{i\geq k} p_i \left[\psi \mathbf{1}\{i=k\} + (1-\psi)\binom{i}{k}\psi^k(1-\psi)^{i-k}\right] = \psi p_k + (1-\psi)\sum_{i\geq k} p_i \binom{i}{k}\psi^k(1-\psi)^{i-k} = \psi p_k + (1-\psi)p_k^*$$

where p_k* is the sampled probability under $G^{|S}$ (e.g., (Stumpf, Wiuf, and May 2005)). Thus, we have $\Psi_0^S(x) = \psi\Psi_0(x) + (1-\psi)\Psi_0^{|S}(x) = \psi\Psi_0(x) + (1-\psi)\Psi_0^S(1-\psi(1-x))$. □

Note that the implied expected degrees under this approximation are $\left\langle d(G^{|S}) \right\rangle = \psi \langle d \rangle$ and $\left\langle d(G^S) \right\rangle = \psi(2-\psi)\langle d \rangle$, which matches the previous derivations in lemmas 1.B.2 and 1.B.3.

LEMMA 1.B.5. $\left\langle d_2(G^{|S}) \right\rangle = \psi^2 \langle d_2 \rangle$ *and* $\left\langle d_2(G^S) \right\rangle = k(\psi)\langle d_2 \rangle$.

Proof. Follows by inspection. □

ASSUMPTION 1.B.3 (Graph Span Conditions). *Let (1)* $\mathrm{E}[\epsilon|\ell] = 0$, *(2)* $0 < \operatorname{plim}_{R\to\infty} \mathbb{E}_R \log^{-2}(\gamma d_{2r} d_r^{-1}) < \infty$ *for* $\gamma \in (0,1]$, *(3) the conditions of 1.3.1 hold with* $d(G), d_2(G) \in o_{\mathrm{P}}\left(n^{\gamma_1/2-1}\log^{\gamma_2/2} n\right)$ *for* $\gamma_1 < 2$ *and* γ_2.

Proof of Proposition 1.3.3. Recall $\zeta_r = d_2(G_r)/d(G_r)$ and let $\log\bar{\zeta}_r := \log\zeta_r + \log\gamma$. To sign the bias we are interested in

$$\lim_{R\to\infty} \frac{R^{-1}\sum \operatorname{cov}\left(\log^{-1}(d_2(\bar{G}_r)d(\bar{G}_r)^{-1}), \log^{-1}\zeta_r\right)}{R^{-1}\sum \operatorname{var}\left(\log^{-1}(d_2(\bar{G}_r)d(\bar{G}_r)^{-1})\right)}.$$

First observe that

$$R^{-1}\sum \operatorname{cov}\left(\log^{-1}(d_2(\bar{G}_r)d(\bar{G}_r)^{-1}), \log\zeta_r\right) = R^{-1}\sum \operatorname{cov}\left(\log^{-1}\bar{\zeta}_r, \log^{-1}\log\zeta_r\right) + o_{\mathrm{P}}(1).$$

This follows from $\left|\log^{-1}(d_2(\bar{G}_r)d(\bar{G}_r)^{-1}) - \log^{-1}\bar{\zeta}_r\right| = o_{\mathrm{P}}(1)$ which we can see by considering the numerator of the fraction and noting

$$\left|\log\bar{\zeta}_r - \log(d_2(\bar{G}_r)d(\bar{G}_r)^{-1})\right| \leq \left|\log d_2(G_r) - \log d(G_r) + \log\gamma - \log d_2(\bar{G}_r) + \log d(\bar{G}_r)\right| = o_{\mathrm{P}}(1),$$

by Lemma 1.3.1, where $\gamma = k(\psi)/\psi$ or ψ depending on G^S or $G^{|S}$, using the fact that $\log(\cdot)$ is Lipschitz on $\mathbb{R}_{\geq 1}$. Therefore we are interested in $\lim_{R\to\infty} \frac{R^{-1}\sum \mathrm{cov}\left(\log^{-1}\bar{\zeta}_r, \log^{-1}\zeta_r\right)}{R^{-1}\sum \mathrm{var}\left(\log^{-1}\bar{\zeta}_r\right)}$.

If $\gamma > \zeta_r^{-1}$ for every r, then $\mathrm{cov}\left(\log^{-1}\bar{\zeta}_r, \log^{-1}\zeta_r\right)$ is positive for every r by definition. In addition, for every r, since $\log\gamma < 0$,

$$\mathrm{cov}\left((\log\zeta_r + \log\gamma)^{-1}, \log^{-1}\zeta_r\right) < \mathrm{var}\left(\log^{-1}\zeta_r\right).$$

This shows $\lim_{R\to\infty} \frac{R^{-1}\sum \mathrm{cov}\left(\log^{-1}\bar{\zeta}_r, \log^{-1}\zeta_r\right)}{R^{-1}\sum \mathrm{var}\left(\log^{-1}\bar{\zeta}_r\right)} < 1$.

Next, assume that $\gamma < \zeta_r^{-1}$. Then $sign\left\{cov\left(\log^{-1}\bar{\zeta}_r, \log^{-1}\zeta_r\right)\right\}$ depends on the distribution of ζ_r; it cannot be signed. Therefore $\mathrm{plim}\, R^{-1}\sum \mathrm{cov}\left(\log^{-1}\bar{\zeta}_r, \log^{-1}\zeta_r\right)$ can take either sign. This is easily seen geometrically.

Finally, that the analytical corrections are consistent follows from the above argument in the first step, noting that we use $\log(\gamma d_2(\bar{G}_r)d(\bar{G}_r)^{-1})$ and therefore

$$\left|\log\zeta_r - \log(\gamma d_2(\bar{G}_r)d(\bar{G}_r)^{-1})\right| \leq \left|\log\bar{\zeta}_r - \log(d_2(\bar{G}_r)d(\bar{G}_r)^{-1})\right| + \left|\log\gamma - \log\gamma\right| = o_{\mathrm{P}}(1)$$

which completes the result. □

Proof of Proposition 1.3.4. Step 1: It is clear that $\mathrm{Tr}(A(G^S)^k) < \mathrm{Tr}(A^k)$, since we can partition

$$n^{-1}\sum_{i_1,..,i_k\in V^k} A_{i_1i_2}...A_{i_ki_1} = n^{-1}\sum_{\mathcal{A}} A_{i_1i_2}...A_{i_ki_1} + \frac{1}{n}\sum_{V^k\setminus\mathcal{A}} A_{i_1i_2}...A_{i_ki_1}$$

with $\mathcal{A} = \{i_1,...,i_k : \forall t\in[k],\ i_t \vee i_{t+\epsilon(t)} \in S,\ \epsilon(t)\in\{-1,1\}\}$ and $\sum_{V^k\setminus\mathcal{A}} A(G^S)_{i_1i_2}...A(G^S)_{i_ki_1} = 0$. Let $S_{k,\sigma}$ be the set of all k-sequences using elements from σ. Let η_j be the number of terms in the sum with j distinct nodes. Then notice $\frac{1}{n}\sum\eta_j = \mu^k$. For $A(G^{|S})$, we have

$$\mathrm{E}\left[\mu^k(G^{|S})|G\right] = m^{-1}\sum_{\sigma\in S}\sum_{i_1,...,i_k\in S_{k,\sigma}} A_{i_1i_2}...A_{i_ki_1}\mathrm{P}(\sigma) = m^{-1}\mathrm{P}(\sigma)\sum_{\sigma}\sum_{j=2}^{k}\binom{n-j}{m-j}\eta_j = m^{-1}\sum_{j=2}^{k}\frac{(m)_j}{(n)_j}\eta_j = n^{-1}\sum_{j=2}^{k}\frac{(m-1)_j}{(n-1)_j}\eta_j.$$

Step 2: $G^{|S}$ is a compression of both G^S and G, so $\lambda_k(G^{|S}) \leq \lambda_k(G), \lambda_k(G^S)$ by the Cauchy's interlacing theorem. Noticing $[A(G^S)]_{ij} \leq [A(G)]_{ij}$, $\lambda_{\max}(G^S) = \sup_{\alpha\in S^{n-1}}\alpha' A(G^S)\alpha \leq \sup_{\alpha\in S^{n-1}}\sum\alpha_i\alpha_j[A(G)]_{ij} = \lambda_{\max}(G)$. □

1.B.3 Proofs for Section 1.3.2

We use T to denote the row-stochastized adjacency matrix, $T_{ij} = A_{ij}/d_i$, instead of w, where d_i is the degree of node i to follow the literature (e.g., (Jackson 2008c)).

Proof of Proposition 1.3.5. Step 1: We show the argument for the case with G^S. The argument for $G^{|S}$ is similar, but omitted. Let

$$\bar{u} = M^0\left(\rho(T-\bar{T})y + \delta(T-\bar{T})x + \epsilon\right) = M^0u,$$

where $M^0 = I_n - \iota\iota'/n$. The instrument is $\bar{Z} = [\iota, x, \bar{T}x, \bar{T}^2x]$. It suffices to show $\mathrm{E}[\bar{Z}'M^0u] \neq 0$. We can write

$$\mathrm{E}\left[\bar{Z}'\bar{u}|x,G\right] = \mathrm{E}\left[\left(\iota'M^0u, x'M^0u, x'\bar{T}'M^0u, x'\bar{T}^2M^0u\right)'\middle| x,G\right].$$

The first two components mechanically have expectation zero. By the reduced form representation in section 1.3.2, we can write y as a function of x and powers of T. The third component requires considering terms of the form $x'\mathrm{E}\left[\bar{T}'M^0\left(\bar{T}-T\right)|x,G\right]\delta x$. For generic x, this is zero if and only if $\mathrm{E}\left[\bar{T}'M^0\left(\bar{T}-T\right)|x,G\right]=0$. If we show $\mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0\bar{T}|G\right]\right)\neq\mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0T|G\right]\right)$, the preceding equation does not hold. We pass the expectation by linearity, use a cyclic permutation and write

$$\mathrm{E}\left[\mathrm{Tr}(\bar{T}'M^0\bar{T})\right]=\mathrm{E}\left[\mathrm{Tr}(M^0\bar{T}\bar{T}')\right]=\mathrm{E}\left[\mathrm{Tr}(\bar{T}\bar{T}')\right]-n^{-1}\mathrm{E}\left[\mathrm{Tr}(\bar{T}'\iota\iota'\bar{T})\right].$$

Let $\langle\cdot,\cdot\rangle_F$ be the Frobenius inner product, $\langle A,B\rangle_F=\mathrm{Tr}(AB')$ and $\|\cdot\|_F$ the Frobenius norm, $\|A\|_F^2=\mathrm{Tr}(AA')$. In Lemmas 1.B.6, 1.B.7, 1.B.8, and 1.B.9 we compute the following four terms $\mathrm{E}\left[\|\bar{T}\|_F^2\right]$, $\mathrm{E}\left[\langle T,\bar{T}\rangle_F\right]$, $n^{-1}\mathrm{E}\left[\mathrm{Tr}(\bar{T}'\iota\iota'\bar{T})\right]$, and $n^{-1}\mathrm{E}\left[\mathrm{Tr}(\bar{T}'\iota\iota'T)\right]$ which we then use to complete the argument. We find

$$\mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0T|G\right]\right)=(1-n^{-1})\{\|T\|_F^2+\sum_i\xi_2(d_i,\psi)\}-n^{-1}\xi_4(\vec{d},\psi),\ \mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0\bar{T}|G\right]\right)=(1-n^{-1})\{\|T\|_F^2+\sum_i\xi_1(d_i,\psi)\}-n^{-1}\xi_3(\vec{d},\psi).$$

In Lemma 1.B.10 we show that $(n-1)\sum_i\{\xi_2(d_i,\psi)-\xi_1(d_i,\psi)\}-\{\xi_4(\vec{d},\psi)-\xi_3(\vec{d},\psi)\}\neq 0$ for all but finitely many $\psi\in(0,1)$, with an upper bound of $2\cdot\max_i d_i$ points, which completes the argument.

Step 2: We now show that the restriction of the set of observations in the second stage to $i\in S$ yields $\mathrm{E}[Z_{G^S}u_{G^S}]=0$. This follows from the fact that $\bar{T}y=Ty$ for all such $i\in S$, and therefore $\bar{u}=\epsilon$. The result follows from the fact that the instrument is correlated with Ty but orthogonal to ϵ, despite measurement error. □

LEMMA 1.B.6. $\mathrm{E}\left[\left\|\bar{T}\right\|_F^2\right]=\|T\|_F^2+\sum_i\xi_1(d_i,\psi)$, *where* $\xi_1(d_i,\psi):=\frac{1}{1-(1-\psi)^{d_i}}\sum_{r=1}^{d_i}\frac{1}{r}\binom{d_i}{r}\psi^r(1-\psi)^{d_i-r+1}-(1-\psi)/d_i$.

Proof. Observe that $\left\|\bar{T}\right\|_F^2=\sum_{i=1}^n\sum_{k=1}^n\bar{T}_{ik}^2=\sum_{i:\bar{d}_i>0}\bar{d}_i^{-1}$ since $\bar{T}_{ik}^2=\bar{d}_i^{-2}$ is greater than zero exactly $\bar{d}_i$ times. Note that $\mathrm{E}\left[\bar{d}_i^{-1}|i\notin S,\bar{d}_i>0\right]$ is the conditional expectation of the first negative moment of a binomial $\mathrm{Bin}(d_i,\psi)$, namely[45]

$$\mathrm{E}\left[\bar{d}_i^{-1}|i\notin S,\bar{d}_i>0\right]=\frac{1}{1-(1-\psi)^{d_i}}\sum_{r=1}^{d_i}\frac{1}{r}\binom{d_i}{r}\psi^r(1-\psi)^{d_i-r}.$$

The result follows from the fact that

$$\psi\frac{1}{d_i}+(1-\psi)\frac{1}{1-(1-\psi)^{d_i}}\sum_{r=1}^{d_i}\frac{1}{r}\binom{d_i}{r}\psi^r(1-\psi)^{d_i-r}=\frac{1}{d_i}+(1-\psi)\frac{1}{1-(1-\psi)^{d_i}}\sum_{r=1}^{d_i}\frac{1}{r}\binom{d_i}{r}\psi^r(1-\psi)^{d_i-r}-\frac{(1-\psi)}{d_i}=\frac{1}{d_i}+\xi_{1i}.$$

Summing over the i nodes yields the result. □

LEMMA 1.B.7. $\mathrm{E}\left[\left\langle T,\bar{T}\right\rangle_F\right]=\|T\|_F^2+\sum_i\xi_2(d_i,\psi)$, *where* $\xi_2(d_i,\psi):=-d_i^{-1}(1-\psi)^{d_i+1}$.

Proof. We can wite $\left\langle T,\bar{T}\right\rangle_F=\sum_{i=1}^n\sum_{k=1}^nT_{ik}\bar{T}_{ik}=\sum_{i:\bar{d}_i>0}\bar{d}_i^{-1}d_i^{-1}\bar{d}_i=\sum_{i:\bar{d}_i>0}d_i^{-1}$. As $\mathrm{P}\left(\bar{d}_i>0\right)=1-(1-\psi)^{d_i+1}$, where none of the d_i neighbors nor i is sampled, $\mathrm{E}\left[\left\langle T,\bar{T}\right\rangle_F\right]=\sum_i\frac{1-(1-\psi)^{d_i+1}}{d_i}=\|T\|_F^2+\sum_i\xi_2(d_i,\psi)$. □

LEMMA 1.B.8. $\mathrm{E}\left[Tr(\bar{T}'\iota\iota'\bar{T})\right]=\|T\|_F^2+\sum_i\xi_1(d_i,\psi)+\xi_3(\vec{d},\psi)$, *where* $\xi_3(\vec{d},\psi)$ *is defined in the proof.*

Proof. Notice that

$$\mathrm{Tr}(\bar{T}'\iota\iota'\bar{T})=\sum_i\left|\left[\iota'\bar{T}\right]_i\right|^2=\sum_i\left(\sum_k\bar{T}_{ki}^2+2\sum_{l>k}\sum_k\bar{T}_{ki}\bar{T}_{li}\right)=\sum_{k:\bar{d}_k>0}\bar{d}_k^{-1}+2\sum_{l>k}\sum_k\sum_i\bar{T}_{ki}\bar{T}_{li}=\left\|\bar{T}\right\|_F^2+2\sum_{l>k:\bar{d}_l>0}\sum_{k:\bar{d}_k>0}\frac{\left|\bar{N}_k\cap\bar{N}_l\right|}{\bar{d}_k\bar{d}_l}.$$

[45] See, e.g., (Stephan 1945).

With probability ψ^2 both $k \in S$ and $l \in S$, so $\bar{d}_k^{-1}\bar{d}_l^{-1}\left|\bar{N}_k \cap \bar{N}_l\right| = d_k^{-1}d_l^{-1}\left|N_k \cap N_l\right| =: c(k,l)$. With probability $\psi(1-\psi)$ we have $\zeta_4(k,l)$ and with the same probability we have $\zeta_4(l,k)$, where ζ_4 is defined in Lemma 1.B.9 below. Finally, with probability $(1-\psi)^2$

$$\zeta_3(k,l) := \sum_{r=1}^{N_l - |N_k\cap N_l|} \sum_{t=1}^{N_k - |N_k\cap N_l|} \sum_{s=1}^{|N_k\cap N_l|} \binom{N_l - |N_k \cap N_l|}{r}\binom{N_k - |N_k \cap N_l|}{t}\binom{|N_k \cap N_l|}{s} \frac{\psi^{s+t+r}(1-\psi)^{|N_k \cup N_l| - s - t - r}}{(t+s)(r+s)}.$$

Then $\mathrm{E}\left[\mathrm{Tr}(\bar{T}'\iota\iota'\bar{T})\right] = \|T\|_F^2 + \sum_i \xi_1(d_i,\psi) + \xi_3(\vec{d},\psi)$ where

$$\xi_3(\vec{d},\psi) := 2\sum_{l>k}\sum_{k}\left\{\psi^2 c(k,l) + \psi(1-\psi)\zeta_4(l,k) + \psi(1-\psi)\zeta_4(k,l) + (1-\psi)^2\zeta_3(k,l)\right\}.$$

□

LEMMA 1.B.9. $\mathrm{E}\left[Tr(\bar{T}'\iota\iota'T)\right] = \|T\|_F^2 + \sum_i \xi_2(d_i,\psi) + \xi_4(\vec{d},\psi)$, *where* $\xi_4(\vec{d},\psi)$ *is defined below.*

Proof. We have

$$\mathrm{Tr}(\bar{T}'\iota\iota'T) = \sum_i [\bar{T}'\iota]_i \left[\iota'T\right]_i = \sum_i \left(\sum_k \bar{T}_{ki}T_{ki} + 2\sum_{l>k}\sum_k \bar{T}_{ki}T_{li}\right) = \left\langle T, \bar{T}\right\rangle_F + 2\sum_{l>k}\sum_{k:\bar{d}_k>0}\sum_i \bar{T}_{ki}T_{li}.$$

The first term has already been controlled in Lemma 1.B.7. To compute the second term, observe that with probability ψ, $k \in S$ and therefore and in this case $\sum_i \bar{T}_{ki}T_{li} = c(k,l)$. Wwith probability $1-\psi$, $k \notin S$, and as such the conditional expectation is given by

$$\zeta_4(k,l) := \sum_{t=1}^{N_k - |N_k\cap N_l|} \sum_{s=1}^{|N_k\cap N_l|} \binom{|N_k \cap N_l|}{s}\binom{N_k - |N_k \cap N_l|}{t} \frac{\psi^{s+t}(1-\psi)^{N_k - s - t}}{(s+t)d_l}.$$

Therefore $\mathrm{E}\left[\mathrm{Tr}(\bar{T}'\iota\iota'T)\right] = \|T\|_F^2 + \sum_i \xi_2(d_i,\psi) + \xi_4(\vec{d},\psi)$ where $\xi_4(\vec{d},\psi) = 2\sum_{l>k}\sum_k \{\psi c(k,l) + (1-\psi)\zeta_4(k,l)\}$.
□

LEMMA 1.B.10. *Given a graph with non-degenerate coefficients above, for only a finite number of* $\psi \in (0,1)$ *can* $Tr\left(\mathrm{E}\left[\bar{T}'M^0\bar{T}|G\right]\right) \neq Tr\left(\mathrm{E}\left[\bar{T}'M^0T|G\right]\right)$.

Proof. Let $f(\psi,G) := (n-1)\sum_i \{\xi_2(d_i,\psi) - \xi_1(d_i,\psi)\} - \left\{\xi_4(\vec{d},\psi) - \xi_3(\vec{d},\psi)\right\}$. Showing $\mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0\bar{T}|G\right]\right) \neq \mathrm{Tr}\left(\mathrm{E}\left[\bar{T}'M^0T|G\right]\right)$ is equivalent to showing $f(\psi,G) \neq 0$. By the definitions of ξ_k, $f(\psi,G)$ is a rational function in ψ, with a numerator polynomial of degree bounded by $2\cdot\max_i d_i$ and coefficients given by G. As we assume that the graph does not yield coefficients so that the rational function is degenerate at zero, by the fundamental theorem of algebra, there are at most $2\cdot\max_i d_i$ roots of the numerator polynomial in ψ. This bounds the number of sampling rates $\psi \in (0,1)$ that would exactly satisfy the exclusion restriction for G. □

1.B.4 Proofs for Section 1.3.3

As before let $\langle d\rangle := \mathrm{E}d$, $\langle d^2\rangle := \mathrm{E}d^2$, and $\zeta := \langle d^2\rangle/\langle d\rangle$.

Proof of Proposition 1.3.6. Step 1: Let β^* solve $\bar{\rho} = \sum_d \frac{\beta^*\sigma(\beta^*)d}{1+\beta^*\sigma(\beta^*)d}\bar{\mathrm{P}}(d)$ where $\bar{\mathrm{P}}(d)$ is a sampled degree distribution. By (1.5) and that $\frac{\beta\sigma(\beta)d}{1+\beta\sigma(\beta)d}$ is strictly increasing in d when $\beta > 0$, we have $\beta^* > \beta_0$ provided first order stochastic dominance of $\mathrm{P}(d)$ over $\bar{\mathrm{P}}(d)$. For G^S, for every d the count of nodes with at most degree d is weakly increasing under sampling; first order stochastic dominance follows. For $G^{|S}$ the argument is more delicate and relies on this being true in the limit. We use $\limsup_{R\to\infty}\sup_{r\leq R}\sup_d |\mathrm{P}_{Rr}(d) - \mathrm{P}_{\infty r}(d)| = 0$ which we did not need for

the star subgraph, which implies $\limsup_{R\to\infty}\sup_{r\leq R}\sup_d\left|\mathrm{P}^{|S}_{Rr}(d)-\mathrm{P}^{|S}_{\infty r}(d)\right|=0$. Let $F_{|S}(x)$ be the CDF for $\mathrm{P}^{|S}_{\infty}$ and $F(x)$ for P_∞.

$$\begin{aligned}F_{|S}(x) &= \sum_{d\leq x}\sum_{i\geq d}\mathrm{P}(i)\binom{i}{d}\psi^d(1-\psi)^{i-d}=\sum_{i=1}^{\infty}\mathrm{P}(i)\sum_{d\leq i\wedge x}\binom{i}{d}\psi^d(1-\psi)^{i-d}\\ &=\sum_{i=1}^{x}\mathrm{P}(i)\cdot F_{Bin(i,\psi)}(i)+\sum_{i=x+1}^{\infty}\mathrm{P}(i)\cdot F_{Bin(i,\psi)}(x)=\sum_{i=1}^{x}\mathrm{P}(i)+\sum_{i=x+1}^{\infty}\mathrm{P}(i)\cdot F_{Bin(i,\psi)}(x)\geq\sum_{d\leq x}\mathrm{P}(d)=F(x),\end{aligned}$$

which confirms the stochastic dominance. The usual argument for GMM consistency shows $\operatorname{plim}\widehat{\beta}>\beta_0$ since in the limit β^* for every graph is greater than β_0, proving the result.

Step 2: By (Jackson and Rogers 2007), in graph r infection can spread only if $\beta_0>\zeta_{rR}^{-1}$. By arguments analogous to those in Lemma 1.3.1,

$$\mathrm{E}[d(G^{|S})]/\mathrm{E}[d^2(G^{|S})]=\frac{\psi\mathrm{E}d}{\psi^2\mathrm{E}d^2+(1-\psi)\psi\mathrm{E}d}+o(1)=\zeta^{-1}+(1-\psi)\cdot\underbrace{\frac{1+\zeta^{-1}}{\psi\zeta+(1-\psi)}}_{\text{Positive}}+o(1)\text{ and}$$

$$\mathrm{E}[d(G^{S})]/\mathrm{E}[d^2(G^{S})]=\frac{\psi(2-\psi)\mathrm{E}d}{\psi\mathrm{E}d^2+\psi(1-\psi^2)\mathrm{E}d}+o(1)=\zeta^{-1}+(1-\psi)\underbrace{\left\{\frac{1-\zeta^{-1}(1+\psi)}{\zeta+(1-\psi^2)}\right\}}_{\text{Positive if }\zeta>(1+\psi).}+o(1).$$

The result follows since, by assumption on δ_{rR}, $\beta_0<\zeta^{-1}(\bar{G}_{rR})$ w.p.a.1 for every r, for $\bar{G}$ either $G^{|S}$ or G^S. □

1.C Discussion of Analytical Corrections with Regularization

1.C.1 A Regularized Estimator

We show a simple example where our analytical correction is consistent and does not require us to assume $\sigma_v^2\to 0$. In section 1.3, for degree and graph clustering, we have shown that $\widetilde{\beta}=\Phi\widehat{\beta}\xrightarrow{\mathrm{P}}\beta_0\frac{\sigma^2_{\langle x\rangle}}{\sigma^2_{\langle x\rangle}+\Phi^{-2}\sigma_v^2}$ where Φ is a deterministic function of sampling rate ψ. We now show we can construct estimates $\widehat{\sigma}^2_{\langle x\rangle}$ and $\widehat{\sigma}^2_v$ and therefore develop a consistent analytical correction β^*. We present a simple example to illustrate the argument. Our example is average degree and for simplicity assume the researcher samples every edge independently with probability ψ. It is easy to see $\mathrm{E}\left[d(\bar{G})|G\right]=\psi d(G)$, where $\bar{G}$ is the sampled graph. Putting $v=d(\bar{G})-\psi d(G)$, we can analytically compute $\operatorname{var}(v|G)=\frac{2}{n}\psi(1-\psi)d(G)$.

LEMMA 1.C.1. $\operatorname{var}(v|G)=2n^{-1}\psi(1-\psi)d(G)$.

Proof. Let $T=\binom{n}{2}$ and e index edges going from $e=1,...,\binom{n}{2}=T$. Notice average degree is $d(G)=\frac{2}{n}\sum_e A_e$. Then the $v=\frac{2}{n}\sum_e\chi_e A_e-\psi\frac{2}{n}\sum_e A_e$. Notice

$$\begin{aligned}\mathrm{E}\left[\frac{2}{n}\textstyle\sum_e\chi_e A_e\right]^2 &= \mathrm{E}\left[4n^{-2}\left\{\sum_e\chi_e A_e+2\sum_{e<e'}\chi_e\chi_{e'}A_eA_{e'}\right\}\right]=4n^{-2}\left\{\psi\sum_e A_e+\psi^2 2\sum_{e<e'}A_eA_{e'}\right\}\\ &= 2\psi n^{-1}d(G)+4n^{-2}\psi^2 2\textstyle\sum_{e<e'}A_eA_{e'}.\end{aligned}$$

Meanwhile, average degree squared is $\left(\frac{2}{n}\sum_e A_e\right)^2=4n^{-2}\left\{\sum A_e+2\sum_{e<e'}A_eA_{e'}\right\}$, which is useful since $(\mathrm{E}[v|G])^2=\left(\psi\frac{2}{n}\sum_e A_e\right)^2=\psi^2 d(G)^2$. Therefore,

$$\mathrm{E}\left[v^2|G\right]=2\psi n^{-1}d(G)+\psi^2\left\{4n^{-2}2\textstyle\sum_{e<e'}A_eA_{e'}\right\}=2n^{-1}\psi(1-\psi)d(G)+\psi^2 d(G)^2.$$

It follows that $\operatorname{var}(v|G)=2n^{-1}\psi(1-\psi)d(G)+\psi^2 d(G)^2-\psi^2 d(G)^2=2n^{-1}\psi(1-\psi)d(G)$. □

With the analytical formula for the variance of v, we can compute $\beta^\star$. Let $X := \psi \cdot (d(G_1), ..., d(G_R))'$ denote the scaled vector of true (unobserved) average degrees, $Z := (d(\bar{G}_1), ..., d(\bar{G}_R))$ the observed vector of sampled degrees, and $V := (v_1, ..., v_R) = Z - X$. Then it is clear that $\beta^\star := (Z'Z - V'V)^{-1} Z'y$ is consistent for β_0. By estimating $\Sigma_v := \text{plim}\, V'V/R$, we have $\beta^\star := \left(Z'Z - R\widehat{\Sigma}_v\right)^{-1} Z'y \xrightarrow{\text{P}} \beta_0$. Under mild regularity conditions on the growth of average degree, we may therefore estimate $\beta^\star$. Estimation can be improved by performing a bootstrap bias correction.

1.C.2 Numerical Evidence

Table 1-8 displays simulation results for network level regression of average degree with simulated outcomes, precisely of the form presented in Table 1-1. We choose ψ to make the expected edge count comparable to that of the induced subgraph. The results confirm the fact that the naive estimator exhibits significant biases, the analytical correction vastly reduces the biases but may still retain some residual bias which can further be mitigated by applying regularization.

Chapter 2

Graphical Reconstruction with Heterogenous Sampled Networks

2.1 Introduction

Applied research in social networks is often plagued by partially observed data. Applied researchers often use network data collected from a partial sample of nodes to estimate economic parameters in a regression or GMM model. Chandrasekhar and Lewis (2012) show that partial sampling leads to non-classical measurement error and therefore bias in estimates of the regression coefficients or GMM parameters. This paper develops a two-step estimation procedure using graphical reconstruction, which uses the observed part of the data to probabilistically reconstruct the missing part and then consistently and asymptotically normally estimate the economic parameter of interest. Specifically, we are interested in an environment in which the researcher has many networks, each with many nodes, and the networks are all generated by heterogenous formation processes.

For concreteness, I focus on a running example throughout the paper: the diffusion of microfinance in 43 villages in rural Karnataka, India ((Banerjee, Chandrasekhar, Duflo, and Jackson 2011)). A microfinance institution (MFI) based in Bengaluru expanded into these villages. Upon entering a village, the MFI informed certain households about its intentions and asked them to encourage others to join. The researcher wants to estimate how networks affect the diffusion of microfinance participation through these villages.

Consider the case where a researcher wants to perform a network-level regression of the rate of microfinance participation in a village on the average path length of the network.[1] Without the entire network, the researcher falsely codes some existing links between individuals as if they do not exist. Graphical reconstruction builds on the simple idea that replacing every regressor for each village with a conditional expectation of the regressor delivers a consistent estimate of the regression coefficient. In our example, instead of using the mismeasured average path length of

[1]The average path length is the mean of all shortest paths between all pairs of nodes in a network.

each network, the researcher ought to use the conditional expectation of the average path length, given the observed data. This requires integrating over all the missing data, given the observed information and sampling scheme, as opposed to treating missing links as if they did not exist. Furthermore, because different village networks likely formed separately, researchers ought to respect this heterogeneity in their analysis. By treating every network as an independent, but not identically distributed, random variable, I estimate the conditional expectation of the average path length in every network and consistently estimate the regression coefficient.

In practice, the researcher will have to estimate the distribution of missing links. I propose a natural two-step procedure. In the first stage, the researcher fits a potentially different model of network formation to each network in the sample by making use of the observed data. Having done so, the researcher uses the network formation models to take draws of networks from their respective distributions, conditional on the observed information. Using these draws, the researcher estimates the conditional expectation of the regressor or moment in a GMM setting. In the second stage, these conditional expectations are used in the usual way to estimate the economic parameter of interest. (Conley and Udry 2010) perform a robustness exercise where they estimate missing neighborhood data in their regression model, which is an instance of graphical reconstruction.[2]

This two-step procedure is useful for several reasons. First, it allows the researcher to capture realistic heterogeneity by estimating a different model of network formation for every network. Second, the theoretical frame is general, and I establish sufficient conditions ensuring that a desired class of network formation models can be used in graphical reconstruction. To build intuition, I make the analogy with panel data. Every network (individual in a panel) is independent, but the edges within a network (outcomes for an individual across time periods) exhibit dependence. Under regularity conditions, a large network, similar to a large time series, may contain enough information such that the researcher can use the observed data to accurately estimate the distribution which generates the network formation process. The technical challenge is to control an incidental parameter problem, where a parameter for every network must be estimated.[3] Third, in numerical experiments it performs well. Even at a sampling rate of 1/3, the median bias is 5.7% for network-level regressions and 1.4% for node-level regressions. The median reduction in bias is 62%. Each of the 96 estimated parameters exhibits reduction in bias when the reconstruction estimator is applied. Fourth, in addition to regression of economic outcomes on network statistics, the methodology can be applied to GMM models and even indexed GMM models where I may have a family of moment functions indexed by some parameter which presents technical challenges. Covering these cases is essential to network analysis as natural models, such as stopping time models for diffusion, may carry an index.

Given that estimation by graphical reconstruction requires integrating over missing links, I

[2]The present paper develops a general theoretical framework, along with asymptotic analysis, that nests this strategy. I believe that estimates from graphical reconstruction ought to be used not only for robustness checks but also as estimates in their own right that exhibit substantially less bias.

[3]This is similar to non-linear fixed effects in panel settings ((Hahn and Kuersteiner 2004; Hahn and Newey 2004)).

assume that the research has, in addition to the sampled networks and outcomes, covariates at the node (or pair of nodes) level that will be predictive in the network formation models. Examples of such covariates include GPS coordinates, ethnicity, and caste, which are often readily available in development applications. For a non-development example, consider school networks, where it is straightforward to obtain school rosters and demographic data for the entire collection of students.

There are several literatures across a number of fields including economics, epidemiology, statistics, sociology, and computer science that have noted problems due to partial network data. An extensive literature review is contained in Chandrasekhar and Lewis (2012). While augmented likelihood techniques for missing data are well-known in econometrics and statistics, I note that a collection of networks provides the researcher with a unique opportunity to set up the reconstruction problem in a manner which respects the substantial heterogeneity across networks. That is, a number of technical assumptions needed to control incidental parameter problems (e.g., nonlinear panel with fixed effects, (Hahn and Kuersteiner 2004; Hahn and Newey 2004)) become very palatable in the network context, given that each network carries within it tremendous amounts of information. Consequently, graphical reconstruction focuses on conditional expectations of network regressors or moments to consistently estimate economic parameters when graphs are drawn from heterogenous network formation models. This environment generates distinct technical challenges.

The rest of the paper is organized as follows. Section 2.2 establishes the framework. Section 2.3 discusses the estimation procedure and the asymptotic theory. In Section 2.4 I discuss several examples of network formation models and how they might be used in graphical reconstruction. Section 2.5 concludes. All proofs are in the appendices.

2.2 Framework

2.2.1 Notation and Setup

A network or a *graph* is a pair $G = (V, E)$ consisting of a set V of *nodes* and a set E of *edges*, with $n := |V|$. Nodes i and j are either connected or unconnected (the graph is unweighted) and if i is connected to j, then j is connected to i (the graph is undirected). Most of what follows in this paper is applicable to directed and weighted graphs, though following the bulk of the applied research I restrict our attention to the undirected, unweighted case. A graph with n nodes is a member of the set of all undirected, unweighted graphs, denoted by $\mathcal{G}_n$.

A graph is represented by its *adjacency matrix*, $A := A(G)$. It is a matrix of 0s and 1s that depicts whether two nodes are connected, where $A_{ij} = \mathbf{1}\{ij \in E\}$ with the convention that $A_{ii} = 0$. I denote the *neighborhood* of i, the set of nodes it is connected to, by $N_i := \{j \in V : ij \in E\}$. Researchers are interested in economic models where an economic behavior or outcome is predicted by network statistics. I let $w(G)$ represent a d_w-dimensional vector of these network statistics. Since the data set may contain multiple networks, I use R to denote the number of graphs. The researcher is interested in *economic parameter* β_0.

2.2.1.1 Sampling

Typically researchers obtain one of two types of sampled network data. First, the researcher may survey a set of m nodes and ask each node about the social connections with the other $m-1$ nodes in that data set. I call this the *induced subgraph*, as it restricts the network among those who are sampled. Second, the researcher may have a list of the nodes in the network (e.g., a household census list in a village). From this list, a sample of m nodes may be surveyed. These nodes can name their social connections, not only to other $m-1$ surveyed nodes, but connections to anyone from the list of n nodes. I call this the *star subgraph*.

Let S be the set of surveyed nodes, randomly chosen from V, with $m = |S|$. Let $m = \lfloor \psi n \rfloor$, where ψ is the *sampling rate*. The researcher obtains a subgraph of the graph in question. There are two potential resulting networks: the induced subgraph $G^{|S} = (S, E^{|S})$, which consists of the sampled nodes and the edges restricted to the set of surveyed nodes ($E^{|S}$), and the star subgraph $G^S = (V, E^S)$, where E^S are edges such that at least one of the nodes is in S.

Figure 2-1 provides an illustration of the problem that this paper intends to address. Figure 2-1(a) displays G, the target network, Figure 2-1(c) shows the induced subgraph and Figure 2-1(e) depicts the star subgraph. I will also write $A = (A^{obs}, A^{mis})$ to denote the observed and missing part of the adjacency matrix, which are random variables under the sampling procedure. Although this framework idealizes the random sampling used in many applied contexts, our setting can easily be extended to other sampling methods such as independent edge sampling or snowball sampling.[4]

2.2.2 Econometric Models

The researcher intends to study economic behavior on R networks, $\{G_r : r = 1, ..., R\}$. For simplicity, I assume every network has n nodes. An economic process has taken place on every network and can be described by an econometric model depending on an *economic parameter* β_0. Returning to the microfinance example, information about microfinance has been introduced to certain households in every village and households decide to participate as the information propagates throughout the villages. Our goal is to estimate an economic parameter. I could easily do so if the networks were fully observed. The general framework for analyzing such models is to presume that a conditional moment restriction is satisfied,

$$\mathrm{E}\left[m(y, w(G); \beta_0)|G\right] = 0. \tag{2.1}$$

where $y \in \mathbb{R}^{d_y}$ is an outcome random variable, $m(\cdot, \cdot; \cdot)$ is a moment function, $w(\cdot)$ is function on $\mathcal{G}_n$, and $\beta \in \mathcal{B}$ is a parameter with true value β_0.

Examples include discrete choice models, stopping time models (e.g., (Iyer and Puri ming)), quantile regression (e.g., Angelucci et al., 2010), and even network-based matching models (e.g.,

[4]Our graphical reconstruction solutions often apply to missing-at-random samples, where the probability of graph information being missing is independent of the missing data itself ((Rubin 1976)).

(Aral and Walker ming; Banerjee, Chandrasekhar, Duflo, and Jackson 2011)). More generally, our results apply to indexed GMM models with parameter $\beta_0(u)$ where $u \in \mathcal{U}$ (e.g., time in a stopping time model or quantile in quantile regression). Partial sampling will generally generate biases as the moment will be a nonlinear function of the network statistic, so the estimated parameter will be inconsistent.

For a simple example, imagine that a researcher would like to study a regression of a household's decision to join microfinance on its centrality in the network. Theory suggests that central nodes will be more likely to learn the new information such as the presence of microfinance in the village. Then the researcher may run[5]

$$y_{ir} = \alpha + w_i(G_r)\beta_0 + \epsilon_{ir} \tag{2.2}$$

and the regression has nR observations.[6] Clearly, similar regressions can be run at the graph level (cross-village regressions) or even at the edge level (dyadic regressions).

Using sampled networks, the researcher runs regressions of the form

$$y_{ir} = \alpha + w_i(\bar{G}_r)\beta + u_{ir},$$

where $\bar{G}$ is either $G^{|S}$ or G^S, depending on the sampling scheme. In general, the measurement error will not be classical and may result in attenuation bias, expansion bias, or even sign switching. Chandrasekhar and Lewis (2012) provide examples of regression models of common and economically meaningful network statistics where such biases exist.

Finally, I employ the following notation throughout the paper. $\mathrm{E}[\cdot]$ denotes expectation, $\mathbb{E}_n[\cdot]$ the empirical expectation,[7] $\|\cdot\| = \|\cdot\|_2$ the ℓ^2-norm, $\|\cdot\|_\infty$ the sup-norm, and $\ell^\infty(U)$ the space of bounded functions on $\mathcal{U}$. Also, $f_n \in \Theta(g_n)$ means $\exists k_1, k_2 > 0, n_0$ such that $\forall n > n_0$ $|g_n| \cdot k_1 \leq |f_n| \leq |g_n| \cdot k_2$, and the falling factorial is given by $(n)_j = n(n-1)...(n-j+1)$.

2.3 Estimation by Graphical Reconstruction

In this section, I discuss a two-step estimation procedure to consistently estimate economic parameters from linear regression and GMM models. In our asymptotic frame, both the size of each network and the number of networks grow. Every network is a draw from a distribution governed by its own parameter θ_{0r}. This will force us to control an incidental parameter problem. Clearly, I can nest the special case where every network is drawn from the same distribution, $\theta_{0r} = \theta_0$ for every r, and thereby assume away the incidental parameter problem.

I present an informal overview of our method in section 2.3.1. In section 2.3.2, I present the asymptotic distribution of $\widehat{\beta}$ under high level assumptions on $\widehat{\theta}_r$. I discuss low level conditions for $\widehat{\theta}_r$ by studying several key classes of network formation models, which also shed light on the limits

[5] A vector of demographic covariates may be included, though I omit it for simplicity.

[6] With missing data, there will be $O(nR)$ observations. For instance with $G^{|S}$, one has $mR = \psi nR$ observations.

[7] For $a = (a_1, ..., a_n)$, $\mathbb{E}_n[a_i] = \frac{1}{n}\sum_i a_i$. Similarly, $\mathbb{E}_R[a_r] = \frac{1}{R}\sum_r a_r$ and $\mathbb{E}_{n,R}[a_{i,r}] = \frac{1}{n}\sum_r \sum_i a_{i,r}$.

of our approach, in section 2.3.3. Section 2.4.1 reflects on the interplay between network formation models and graphical reconstruction.

2.3.1 Informal Overview

In our overview I describe our procedure for regression,

$$y_{ir} = \alpha_0 + w_{ir}(G_r)\beta_0 + \epsilon_{ir}.$$

I assume that the researcher has the following data. First, she has outcome data for every node in every graph, $\{y_{ir} : i = 1, ..., n,\ r = 1, ..., R\}$, such as whether household i in village r participates in microfinance.[8] Second, she has a set of partially observed graphs, $\{G_r^S : r = 1, ..., R\}$ or $\{G_r^{|S} : r = 1, ..., R\}$. Third, she has variables which are predictive in a network formation model $\{z_r : r = 1, ..., r\}$.[9] For instance, the researcher may have basic demographic characteristics such as religion, caste, household amenities, occupation or geographic location. This data structure is relatively innocuous and common in numerous applications. In development, when deciding how to draw a random sample to administer treatments, researchers usually conduct a listing in each enumeration area. This requires obtaining a census of the economic units, which can be done directly (e.g., (Townsend 2007); Suri, 2011; (Banerjee, Chandrasekhar, Duflo, and Jackson 2011)) or indirectly by obtaining census information from the village representatives (e.g., (Macours 2003); (Takasaki, Barham Oliver, and Bradford 2000)).[10] It is well-known that obtaining GPS and basic demographic data during enumeration is cheap; the bulk cost of a network survey is the network module itself. For a different example, consider school networks where it is straightforward to obtain rosters and demographic data for all students. The full set of observed data is (y_r, A_r^{obs}, z_r), consisting of y_r the vector of outcome data, A_r^{obs} the observed part of the graph, and z_r the vector of network formation covariates. The missing data for each network is A_r^{mis} and recall $G_r = (A_r^{obs}, A_r^{mis})$.

Every network is thought of as a realization of a random network formation process, drawn from a distribution which depends on z_r and parameter $\theta_{0r} \in \Theta_r$. To estimate β_0 I use an argument based on conditional expectations. If θ_{0r} were known for all r, I could estimate a conditional expectation of $w_{ir}(G_r)$ given the observed data,

$$\mathcal{E}_{ir}(A_r^{obs}, z_r; \theta_{0r}) := \mathrm{E}\left[w_{ir}(G_r)|A_r^{obs}, z_r; \theta_{0r}\right].$$

By the properties of conditional expectation, using $\mathcal{E}_{ir}$ in the regression instead of w_{ir} yields con-

[8] In what follows it is not necessary for y_{ir} to be observed for every node, but it simplifies notation.

[9] E.g., $z_r = \{z_{ir} : i = 1, ..., n\}$ or $z_r = \{z_{ij,r} : i, j \in V\}$ where z_{ir} or $z_{ij,r}$ are covariates for nodes or pairs.

[10] Researchers can either collect simple covariate data from all nodes or from representatives who carry information.

sistent estimation of β_0. The least squares estimator is given by[11]

$$\widehat{\beta}_{\text{ols}} = \left(\sum_{r=1}^{R} \sum_{i=1}^{n} \mathcal{E}_{ir}(\widehat{\theta}_r) \mathcal{E}_{ir}(\widehat{\theta}_r)' \right)^{-1} \cdot \sum_{r=1}^{R} \sum_{i=1}^{n} \mathcal{E}_{ir}(\widehat{\theta}_r) y_{ir}.$$

A similar but more involved result is true for GMM. Notice $\widehat{\beta}_{\text{ols}}$ depends on $\widehat{\theta}_r$ for all r.

To control the estimation of $\widehat{\theta}_r$, I need to argue not only that it is consistent for θ_{0r}, but uniformly so. That is, $\sup_r \|\widehat{\theta}_r - \theta_{0r}\| = O_{\text{P}}(a_R^{-1} \cdot R^{1/b})$, where a_R is the rate of convergence of $\widehat{\theta}_r$ to θ_{0r} for every r, and $b > 1$ is the number of moments that the network formation model has. This imposes a rate requirement on the problem which says that the network-formation parameter needs to be estimated fast enough: $\sqrt{nR} \cdot a_R^{-1} \cdot R^{1/b} \to 0$.

The consistency of $\widehat{\theta}_r$ follows from assumptions on the model of graph formation and the sampling procedure. With missing-at-random data, under assumptions on the graph model, a consistent estimator exists. Consider a model where an edge forms independently, conditional on covariates,

$$\text{P}(A_{ijr} = 1 | z_r; \theta_{0r}) = \Lambda \left(f(z_{ir}, z_{jr})' \theta_{0r} \right),$$

where $\Lambda(\cdot)$ is some link function (e.g., logistic or normal), z_i is a vector of covariates for vertex i, and f is a vector-valued function. For instance, f may be the difference between characteristics of two nodes $f(z_i, z_j) = \|z_i - z_j\|$. If the sampling procedure is orthogonal to the network formation, a random subset of the $\binom{n}{2}$ pairs of nodes is observed. Therefore, $\widehat{\theta}_r$ is consistent.

This model converges with $a_R = n$, since I have on the order of $n(n-1)/2$ observations. The requirement becomes $n^{-1/2} R^{1/2+1/b} \to 0$, so the number of networks must grow sufficiently slower than the number of nodes. In other models, the rate a_R may be different (e.g., $n/\log n$, n^τ for $\tau \in [1/2, 2)$, $\sqrt{n/\log n}$). If the rate is too slow, the requirement for node level regression may not be met, though usually the requirement for graph level regressions will be satisfied.

It is worth illustrating why respecting the heterogeneity across network formation models in this fixed effects fashion is worthwhile. Figure 2 presents two bar graphs constructed from the following data. I considered collections of 50 networks drawn from a network formation family (either simulated networks or empirical Indian village networks). Using these networks I generated simulated outcome data of a regression y_{ir} on $w_{ir}(G)$ where w is some network statistic of interest (e.g. eigenvector centrality of node i or clustering of node i). (I repeated this across 96 specifications of different network statistics and sampling rates.) Then I estimated the bias when G^S or $G^{|S}$ are used instead of G in the regression. I then conducted graphical reconstruction using a simple network formation model, detailed in Chandrasekhar and Lewis (2012). This is done in two ways. First, I allow each of the villages to be generated by their own network formation processes and second I constrain them to come from the same model. The bar graph presents the percentage point reduction in bias from using graphical reconstruction averaged across the 96 specifications.

[11] For notational simplicity, assume the regressors are demeaned.

It is clear that that respecting the heterogeneity in the network formation process is essential to reducing the bias.

2.3.2 Theory for $\hat{\beta}$

I begin by establishing that $\widehat{\beta}$ is consistent and asymptotically normal. The main theorem is stated in section 2.3.2.2, under regularity conditions, including simple high level assumptions about the behavior of $\widehat{\theta}_r$, which I will verify in section 2.3.3. In section 2.3.2.1 I discuss the regularity conditions in depth.

I have already introduced the regression environment. I consider the GMM environment of (2.1). Relative to regression, in GMM the value of y affects the conditional expectation of w. Observe (2.1) implies an unconditional moment restriction holds:

$$0 = \mathrm{E}m(X;\beta_0) = \sum_{G\in\mathcal{G}_n} \mathrm{E}\left[m(X;\beta_0)|G\right]\mathrm{P}_{\theta_0}(G)$$

where $X = (y, w(G))$. Let x_r denote the triple of observed data, $x_r := (y_r, A_r^o, z_r)$. By iterated expectations, the conditional function

$$\mathcal{E}_{ir}(x_r;\beta_0,\theta_0) := \mathrm{E}\left[m(X_{ir};\beta_0)|x_r;\beta_0,\theta_0\right] \tag{2.3}$$

satisfies $\mathrm{E}\mathcal{E}_{ir}(x_r;\beta_0,\theta_0) = 0$. Given an observed data series $\{(X_{ir}, z_{ir}) : i = 1,...,n,\ r = 1,...,R\}$ and an estimator $\widehat{\theta}_r$ of θ_{0r}, the estimator is

$$\widehat{\beta}_{\text{gmm}} := \underset{\beta\in\mathcal{B}}{\text{argmin}} \left(\mathbb{E}_{n,R}\mathcal{E}_{ir}(x_r;\beta,\widehat{\theta}_r)\right)' \widehat{W} \left(\mathbb{E}_{n,R}\mathcal{E}_{ir}(x_r;\beta,\widehat{\theta}_r)\right)$$

where $\widehat{W}$ is a consistent estimator of W.[12]

In order to compute the conditional moment in (2.3) I need to be able to integrate with respect to a conditional probability for every graph in our sample, $\mathrm{P}_{\beta_0,\theta_0}(A_r^{mis}|x_r)$. Computing the expectation requires a reweighting term,

$$\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r}) = \sum_{A_r^m} m(X_{ir};\beta_0)\mathrm{P}_{\beta_0,\theta_{0r}}(A_r^{mis}|x_r),$$

with $\mathrm{P}_{\beta_0,\theta_{0r}}(A_r^{mis}|x_r) \propto f_{\beta_0}(y_r|G_r)\mathrm{P}_{\theta_{0r}}(A_r^{mis}|A_r^{obs}, z_r)$. To be able to utilize this approach, the researcher must make assumptions on the distribution of y given G.[13]

[12] In the case of maximum likelihood where $\mathcal{E}$ is the conditional score, $W = I$.

[13] With an index $u \in \mathcal{U}$, $\widehat{\beta}(u) := \text{argmin}_{\beta\in\mathcal{B}} \left(\mathbb{E}_{n,R}\mathcal{E}_{ir}(x_r;\beta,\widehat{\theta}_r,u)\right)' \left(\mathbb{E}_{n,R}\mathcal{E}_{ir}(x_r;\beta,\widehat{\theta}_r,u)\right)$.

2.3.2.1 Regularity Conditions

The main results are presented in section 2.3.2.2, to which the reader may skip ahead if desired. In this section I discuss the regularity conditions on which the results depend. Let $\mathrm{P}(G_r|z_r;\theta_r)$ be the distribution of the graph G_r given covariates z_r.

ASSUMPTION 1 (Random Graph Model and First Stage Estimation).

1. *$\forall r$, Θ_r is a compact subset of $\mathbb{R}^{d_\theta}$; G_r is a $\mathcal{G}_n$-valued random graph with $\mathrm{P}(G_r|z_r;\theta_r) \in C^2(\Theta_r)$ at every $(G,z) \in \mathcal{G}_n \times \mathcal{Z}$; $\bar{H}_{r,R} := \sup_z \max_{G,\theta_r} \left|\frac{\partial}{\partial\theta}\mathrm{P}(G|z;\theta)\right|$, $\sup_R \sup_r \bar{H}_{R,r} < \infty$.*
2. *The first stage estimation satisfies for some sequence of normalizing constants (a_R), $b > 1$, and $r \leq R$, $a_R \cdot (\widehat{\theta}_r - \theta_{0r}) = O_{\mathrm{P}}(1)$ and $\sup_{r\leq R} \|\widehat{\theta}_r - \theta_{0r}\| = O_{\mathrm{P}}(a_R^{-1} \cdot R^{1/b})$.*
3. *For node level analysis $a_R^{-1} \cdot \sqrt{nR^{1+2/b}} \to 0$ and for graph level analysis $a_R^{-1} \cdot \sqrt{R^{1+2/b}} \to 0$.*
4. *$\beta_0(u)$ is an interior point of $\mathcal{B}$, a compact subset of $\mathbb{R}^{d_\beta}$, for every $u \in \mathcal{U}$.*

Condition 1 ensures that the random graph family is smooth enough in the parameter, so small deviations from the true parameter do not result in very different probability distributions. Condition 2 is a high-level condition on the first stage estimation which I will microfound in section 2.3.3. It guarantees that I can uniformly replace the estimated network formation parameter for every graph in the sequence with its true value. Condition 3 is a rate requirement which relates the rate of estimation of the network formation process to the rate of estimating the economic model of interest. Condition 4 is a standard interiority condition. Let h denote a random variable.

DEFINITION 2.3.1. *A sequence of measurable (potentially matrix-valued) functions $\{\phi_{i,r}(h_{ir};\alpha) : i = 1,..,n_R,\ r = 1,...,R\}$ satisfies an* envelope condition *over $\alpha \in \mathcal{A}$ if there exist measurable functions $L_{i,r}(h_{ir})$, with $\|\phi_{i,r}(h_{ir};\alpha)\| \leq L_{i,r}(h_{ir})$ for every h_{ir} and α, and $\sqrt{n}\mathbb{E}_n L_{i,r,R}$ has uniformly integrable vth moment for $v \geq 2$.*

DEFINITION 2.3.2. *A sequence of measurable (potentially matrix-valued) functions $\{\phi_{i,r}(h_{ir};\alpha) : i = 1,..,n_R,\ r = 1,...,R\}$ is* Lipschitz continuous *in $\alpha \in \mathcal{A}$ if there exist measurable functions $M_{i,r}(h_{ir})$ with $\|\phi_{i,r}(h_{ir};\alpha) - \phi_{i,r}(h_{ir};\bar{\alpha})\| \leq M_{i,r}(h_{ir})\|\alpha - \bar{\alpha}\|$ for every h_{ir} and $\alpha, \bar{\alpha} \in \mathcal{A}$, and $\sqrt{n}\mathbb{E}_n M_{i,r,R}$ has uniformly integrable vth moment for $v \geq 2$.*

In addition, I use $\mathcal{I}_{h|x}(\alpha) := \mathrm{E}\left[\frac{\partial}{\partial\alpha'}\log f(h|x;\alpha)\frac{\partial}{\partial\alpha}\log f(h|x;\alpha)|x;\alpha\right]$ to denote the conditional information matrix with random variable $h|x$, density or pmf $f(h|x)$, and parameter α.

Turning to the economic model, observe that the network statistic $w(G)$ may be growing or shrinking in n. For instance, the eigenvector centrality declines as it is a unit norm object. The degree of a node may be $\Theta(1)$, $\Theta(\log n)$, or $\Theta(n)$ depending on the graph family. In what follows, in regression I assume that the model is such that all regressors are rescaled at the appropriate rate: if they exhibit growth or shrinkage at b_n, I assume that the models are specified using $\tilde{w} := b_n^{-1}w$ as regressors. Similarly, in GMM I assume that the moments and network statistics, both of which may depend on R, are appropriately rescaled. For two reasons I present regularity conditions for

least squares, GMM, and GMM with an index separately, though they essentially can be nested. First, least squares does not require assuming the joint distribution of y and G. Second, GMM conditions are more transparent than the more general case where parameters carry an index. After presenting the assumptions I discuss what they mean for networks.

ASSUMPTION 2 (Linear Regression).

1. $\mathrm{E}[\epsilon|w] = 0$, $\mathrm{E}[\epsilon\epsilon'|w] = \Omega$, *p.d. with* $\sup_R \lambda_{\max}(\Omega) < \infty$
2. $\mathrm{E}\left[\|w_{ir}\|^k |x_r; \theta_r\right]$ *and* $\left\|\mathcal{I}_{w_{ir}|x_r}(\theta_r)\right\|$ *for* $k = 1, 2$ *satisfy the envelope condition with* $L_{i,r}(x_r)$.
3. $\sup_{R\geq 1} \sup_{r\leq R} \mathrm{var}\left(\sqrt{n}\mathbb{E}_n w_{ir}(G_r)\right) < C_1 < \infty$ *and* $\inf_{R\geq 1} \inf_{r\leq R} \mathrm{var}\left(\sqrt{n}\mathbb{E}_n \mathcal{E}_{ir}(x_r; \theta_{0r})\right) > C_0 > 0$, *uniformly over the array.*

Define $g_R(\beta) := \mathbb{E}_{n,R} \mathrm{E} m(y_{ir}, w_{ir}(G_r); \beta)$ and $f(m|x; \beta, \theta)$ be known up to parameters.

ASSUMPTION 3 (GMM).

1. $\widehat{W} = W + o_{\mathrm{P}}(1)$, W *is p.s.d. and the model satisfies* $\lim_{R\to\infty} W g_R(\beta) = 0$ *only if* $\beta = \beta_0$.
2. *The limits* $\lim_{R\to\infty} \mathbb{E}_{n,R}\left[\mathrm{E}\mathcal{E}_{ir}(x_r; \theta_r, \beta)\right]$ *and* $\lim_{R\to\infty} \mathbb{E}_{n,R}\left[\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r; \theta_r, \beta)\right]$ *exist uniformly over* $\mathcal{B} \times \prod_{r\in\mathbb{N}} \Theta_r$.
3. $\left\|\mathcal{I}_{m_{ir}|x_r}(\beta, \theta_r)\right\|$, $\mathrm{E}\left[\left\|\frac{\partial}{\partial\beta'} m(X_{ir}; \beta)\right\| |x_r; \theta_r, \beta'\right]$, *and* $\mathrm{E}\left[\|m(X_{ir}; \beta)\|^k |x_r; \theta_r, \beta'\right]$ *for* $k = 1, 2$ *satisfy the envelope condition with envelope* $L_{i,r}(x_r)$.
4. $m(X; \beta)$ *is continuously differentiable on the interior of* $\mathcal{B}$ *for every* $X \in \mathcal{X}$ *and both* $m(X; \beta)$ *and* $\frac{\partial}{\partial\beta'} m(X; \beta)$ *satisfy the Lipschitz condition with constant* $M_{i,r}(X_{ir})$, *where* $\mathrm{E}[M_{i,r}(y_{ir}, w_{ir})|x_r] \leq L_{i,r}(x_r)$.
5. $\sup_{R\geq 1} \sup_{r\leq R} \mathrm{var}\left(\sqrt{n}\mathbb{E}_n[m(X_{ir}; \beta_0)]\right) < C_1 < \infty$ *and* $\inf_{R\geq 1} \inf_{r\leq R} \mathrm{var}\left(\sqrt{n}\mathbb{E}_n[\mathcal{E}_{ir}(x_r; \beta_0, \theta_{0r})]\right) > C_0 > 0$, *uniformly over the array.*

Finally define function classes $\mathcal{F}_R^* := \{\sqrt{n}\mathbb{E}_n m(X_{ir}; \beta(u), u) : (\beta, u) \in \mathcal{B} \times \mathcal{U}\}$ and for $\beta', u' \in \mathcal{B}\times\mathcal{U}$, $\mathcal{H}_{R|\beta',u'} := \{\sqrt{n}\mathbb{E}_n \mathrm{E}[m(X_{ir}; \beta', u')|x_r; \beta, u] : (\beta, u) \in \mathcal{B} \times \mathcal{U}\}$. This nests the above with $\mathcal{U} = \{u\}$.

ASSUMPTION 4 (Indexed GMM).

1. *The maps* $m(X_{ir}; \beta, u)$ *are measurable and is continuous at each* (β, u) *with probability one and* $\mathrm{P}(A_r^m|x_r; \beta, u)$ *is continuous at each* (β, u) *with probability one.*
2. $\mathrm{E}\mathcal{E}_{ir}(\beta(u), u)$ *is continuously differentiable at* $\beta_0(u)$ *uniformly in* $\mathcal{U}$ *and* $\frac{\partial}{\partial\beta'}\mathrm{E}\mathcal{E}_{ir}(\beta, u)$ *is uniformly non-singular at* $\beta_0(u)$ *over* $\mathcal{U}$.
3. $\|\mathrm{E}[m(X_{ir}; \beta, u)|x_r; \beta, u]\| \leq L_{ir}(x_r)$, $\sup_R \sup_r \mathrm{E}\left(\mathbb{E}_n L_{i,r}(x_r)\right)^{2+\delta} = O(1)$ *for some* $\delta > 0$.
4. *The following uniform entropy integral condition holds:*

$$\int_0^\infty \sup_{Q\in\mathcal{Q}} \sqrt{\log N\left(\epsilon \left\|\bar{F}_R^*\right\|_{Q,2}, \mathcal{F}_R^*, L_2(Q)\right)} d\epsilon + \sup_{(\beta',u')\in\mathcal{B}\times\mathcal{U}} \int_0^\infty \sup_{Q\in\mathcal{Q}} \sqrt{\log N(\epsilon \left\|\bar{H}\right\|_{Q,2}, \mathcal{H}_{R|\beta',u'}, L_2(Q))} d\epsilon < \infty.$$

Assumptions 2, 3, and 4 are similar, so I discuss the GMM case. Assumption 3.1 is a standard identification condition. Assumption 3.2 is standard (e.g., (Andrews 1994)) and Assumption 3.3 places uniform restrictions on higher moments of the conditional moment, slope of the moment, and information matrix allowing weak laws of large numbers to be applied. Assumption 3.4 allows these convergences to be uniform over the parameter space.

Assumption 3.5 is what allows us to pass a central limit theorem to the conditional random variable if the unconditional satisfies one.[14] It is reasonable in practice because I use independence across graphs and simply a uniform boundedness condition within graph. This is substantially weaker than having to assume a within-graph central limit theorem for m_{ir}, which would depend on the idiosyncrasies of the network formation model and network statistics. However, it comes at the cost of requiring data from multiple networks. I make this assumption because currently the statistics of networks literature has not characterized within-graph node characteristic interdependencies (e.g., the correlation of eigenvector centrality between nodes for various random graph families). Assumption 4.4 allows for a functional central limit theorem to be applied. Intuitively, the condition on $\mathcal{F}_R^*$ would have to be assumed for the result to be true even without missing data and $\mathcal{H}_{R|\beta',u'}$ controls how much the conditional expectation changes as (β, u) change.

2.3.2.2 Asymptotic Distribution

In this section I show that $\widehat{\beta}_{\text{ols}}, \widehat{\beta}_{\text{gmm}}$, and $\widehat{\beta}(u)$ are consistent and asymptotically Gaussian. I define covariance matrices which characterize the asymptotic variance. For linear regression,

$$H_{\text{ols}} := \lim_{R\to\infty} \mathbb{E}_{nR}\left[\mathrm{E}\mathcal{E}_{ir}\mathcal{E}_{ir}'\right] \text{ and } V_{\text{ols}} := \lim_{R\to\infty} \mathbb{E}_R\left[\text{var}\left(\sqrt{n}\mathbb{E}_n\left[\mathcal{E}_{ir}\epsilon_{ir} + \mathcal{E}_{ir}(w_{ir} - \mathcal{E}_{ir})'\beta_0\right]\right)\right],$$

and for GMM,

$$M := \lim_{R\to\infty} \mathbb{E}_{n,R}\left[\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})\right],\ \Omega := \lim_{R\to\infty} \mathbb{E}_R\left[\text{var}\left(\sqrt{n}\mathbb{E}_n\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})\right)\right],$$
$$H_{\text{gmm}} := M'WM \text{ and } V_{\text{gmm}} := M'W\Omega W'M.$$

THEOREM 2.3.1 (Asymptotic Distribution). *Under Assumptions 1,*

1. *Assumption 2 implies* $\sqrt{nR}(\widehat{\beta}_{\text{ols}} - \beta_0) \rightsquigarrow \mathcal{N}\left(0, H_{\text{ols}}^{-1}V_{\text{ols}}H_{\text{ols}}^{-1}\right)$.
2. *Assumption 3 implies* $\sqrt{nR}(\widehat{\beta}_{\text{gmm}} - \beta_0) \rightsquigarrow \mathcal{N}\left(0, H_{\text{gmm}}^{-1}V_{\text{gmm}}H_{\text{gmm}}^{-1}\right)$.
3. *Assumption 4 implies* $\sqrt{nR}(\widehat{\beta}(\cdot) - \beta_0(\cdot)) \rightsquigarrow -\dot{\Psi}_{\beta,u}(\beta_0(\cdot),\cdot)^{-1}\mathcal{Z}(\beta_0(\cdot),\cdot)$, *in* $\ell^\infty(\mathcal{U})$, *a mean-*

[14] Since this paper focuses on the effect of sampling on network analysis and not on regression or GMM models on graphs, I make the assumption that the underlying model satisfies reasonable regularity conditions if the full networks were observed and focus on the effect of sampling and graphical reconstruction.

zero Gaussian process with covariance function

$$\Omega(u,\tilde{u}) := \lim_{R\to\infty} \mathbb{E}_R \left[n^{-1} \sum_{i,j} \mathrm{E}\left[\mathcal{E}_{ir}(x_r;\beta_0(u),u)\mathcal{E}_{jr}(x_r;\beta_0(\tilde{u}),\tilde{u})\right]\right]$$

where $\dot{\Psi}_{\beta,u}(\beta,u) = \lim_{R\to\infty} \mathbb{E}_{n,R}\mathrm{E}\frac{\partial}{\partial\beta}\mathcal{E}_{ir}(\beta,u)$.

Intuitively, if I can uniformly replace $\widehat{\theta}_r$ with θ_{0r}, since conditional expectations are centered correctly and, under regularity conditions, also satisfy central limit theorems if the unconditioned random variables do, the estimator is consistent and normal. While I wrote the theorem for vertex-level analysis, similar results with modified regularity conditions extend to regressions at the graph-level, edge-level, vertex-triples, etc. Each will allow for different amounts of interdependency in the graph formation process. To be concrete, under the above normalizing assumptions, graph level regression converges at $\sqrt{R}$ while edge level regression converges at $\sqrt{\binom{n}{2}R} = n\sqrt{R}$.

To build further intuition, let us consider what could go wrong. First, for GMM, if one estimates the conditional expectation without reweighting, unless the model was additively separable, $\widehat{\beta}_{\text{gmm}}$ would be inconsistent. Second, there are several reasons why uniform estimation may fail: the size of the networks relative to the number of networks may be too small, the network formation process may have $\dim(\Theta_r)$ exploding too fast, and the level of interdependency in the random graph processes may be too high. I provide a more detailed discussion in section 2.4.1.

I provide an overview of standard errors and estimation methods in Appendix 2.C. In practice, clustering at the graph level in vertex-level regressions and using heteroskedasticity robust standard errors for network-level regressions perform well, though I have explored various bootstrapping procedures (available upon request).

2.3.3 Theory for $\hat{\theta}_r$

In this section I discuss the uniform estimation of the network formation model parameters. I am interested in the joint convergence of $\sup_r \left\|\widehat{\theta}_r - \theta_{0r}\right\|$ in the sense of Assumption 1.3. The literature on consistently estimable network formation models is young and limited. Most models of network formation lack asymptotic frames (see, e.g., exponential random graphs models (ERGMs)). That is, larger networks do not lead to tighter parameter estimates. In general there are no known results characterizing consistency for ERGMs and is currently the topic of research in (Chandrasekhar and Jackson 2011). In addition, (Goldsmith-Pinkham and Imbens 2011) and (Kolotilin 2011), among others, are working papers currently developing consistently estimable random graph models. There are a few classes of models known to be consistent and I discuss several as examples below. Given how new this literature is, it is useful to reflect on a simple, checkable sufficient conditions for joint convergence so that one could check new models as they develop. After this, I discuss four common classes of network formation models and check the condition that can be used in graphical reconstruction. The examples have been chosen to provide intuition about different problems that

may arise.

I have a collection of network formation models which maximize criterion functions, $\theta_{0r} = \arg\max_\theta Q_{(r)}(\theta_{0r})$. I estimate these parameters with a collection of empirical criterion functions, $\widehat{Q}_{(r)}(\theta_r)$, with $\widehat{\theta}_r = \arg\max_\theta \widehat{Q}_{(r)}(\theta_r)$. The lemma is analogous to (Hahn and Newey 2004).

LEMMA 2.3.1. *Let* $V_{(r)}(\theta_r) := \nabla_\theta Q_{(r)}(\theta_r)$ *and* $\widehat{V}_{(r)}(\theta_r) := \nabla_\theta \widehat{Q}_{(r)}(\theta_r)$. *Assume the following.*

1. $\forall r$, $Q_{(r)}(\theta_r)$ *has unique maximum* θ_{0r}; Θ_r *is compact;* $Q_{(r)}(\theta) \in C^2(\Theta)$; $\sup_\theta |\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)| = o_{\mathrm{P}}(1)$.
2. *The criterion functions uniformly converge in the sense that for some* $v > 0$
$$\mathrm{P}\left(\sup_{r \le R} \sup_{\theta \in \Theta_r} \left|\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)\right| \ge \eta\right) = o(a_R^{-v}).$$
3. *There exists a sequence of constants* (a_R) *such that (i) for all* r, $a_R \cdot \widehat{V}_{(r)}(\theta_{0r}) = O_{\mathrm{P}}(1)$; *(ii) for some* $b > 1$, $\sup_{r \le R} \mathrm{E}\left\|a_R \cdot \widehat{V}_{(r)}(\theta_{0r})\right\|^b < \infty$.
4. $\nabla \widehat{V}_{(r)}(\theta_r)$ *satisfies a Lipschitz condition with coefficient* B_r, $\sup_r \|B_r\| = O_{\mathrm{P}}(1)$.
5. *The Hessian satisfies* $\sup_r \left\|\nabla \widehat{V}_{(r)}(\theta_r) - \nabla V_{(r)}(\theta_r)\right\| = o_{\mathrm{P}}(1)$.

Then $a_R \cdot R^{-1/b} \cdot \sup_{r \le R} \left\|\widehat{\theta}_r - \theta_{0r}\right\| = O_{\mathrm{P}}(1)$.

This comes from a usual first order expansion argument. Condition 1 adds extra smoothness to a standard assumption for consistency. Condition 2 requires that all the criterion functions $\widehat{Q}_{(r)}(\theta)$ uniformly lie in an η-"sleeve", $[Q_{(r)}(\theta) - \eta, Q_{(r)}(\theta) + \eta]$; in practice this is argued by applying union bounds and controlling interdependencies across summands in the objective function. Condition 3 provides a rate of convergence of the first-order term and a moment requirement. Condition 4 requires an envelope condition for the third derivative of the objective. Condition 5 requires uniform convergence of the Hessian. Below, Lemma 2.3.1 holds under low-level assumptions.

2.4 Examples

I discuss four common classes of network formation models that may be used in graphical reconstruction. The examples are intended to provide a broad view of the assumptions that need to be made. The first example is the most parsimonious and assumes that the probability that an edge forms is independent given covariates. The second example allows for more flexibility in describing the structure of the network by having a rapidly growing a number of parameters. Despite this, network level reconstruction meets sufficient conditions, though node level reconstruction does not. The third example considers an exponential random graph model wherein the network exhibits greater interdependency among links, albeit in a very limited manner. The final example allows for network effects, where the probability of an edge forming depends on the probabilities of other edges as well as covariates.

Class 1: Conditional Edge Independence Models

I begin by considering a class of models in which edges form independently, given covariates. This is the most common class of model used in the literature (see e.g., Jackson 2008, (Christakis, Fowler, Imbens, and Kalyanaraman 2010), (Goldsmith-Pinkham and Imbens 2011), and (Santos and Barrett 2008)). Let Ξ be a set consisting of all pairs ij. Ξ is implicitly indexed by n and has $n(n-1)/2$ elements. I denote an element $s \in \Xi$ and, when referencing explicitly which pair it corresponds to, I write $s = s_{ij}$. Let z_s denote a covariate for the pair of nodes s_{ij}. Examples include whether two villagers are of the same caste, the distance between their households, etc. The probability that an edge forms in graph r is

$$\mathrm{P}\left(A_{sr} = 1 | z_{sr}; \theta_{0r}\right) = \Phi\left(z_{sr}'\theta_{0r}\right) \tag{2.4}$$

where $\Phi(\cdot)$ is some link function. This framework allows us to consider undirected graphs, directed graphs, and models in which nodes have to agree for a link to form. The undirected case is clear. If the graph formation model is directed, then Ξ consists of all $n(n-1)$ ordered pairs of ij. When the model is undirected but both nodes need to agree, one may use a model such as $A_{ijr} = \mathbf{1}\{z_{ij}'\theta_{0r} - \epsilon_{ijr} \geq 0\} \cdot \mathbf{1}\{z_{ji}'\theta_{0r} - \epsilon_{jir} \geq 0\}$ with link function $\Phi\left(z_{sr}'\theta_{0r}\right) := \Psi(z_{ij}'\theta_{0r})\Psi(z_{ji}'\theta_{0r})$, where $\Psi(\cdot)$ is the cdf of ϵ.[15]

I maximize the log-likelihood, $|\Xi|^{-1}\sum_{s\in\Xi} q(X_s;\theta_r)$, $X_s = [A_s, z_s']$, with summand

$$q(X_s;\theta_r) = A_{sr}\log\Phi(z_{sr}'\theta_r) + (1 - A_{sr})\log\left(1 - \Phi(z_{sr}'\theta_r)\right).$$

For joint convergence, I require that Φ is such that the following hold.

ASSUMPTION 5 (Joint Convergence). Let $Q_{(r)}(\theta_r) := \mathrm{plim}_{n\to\infty}|\Xi|^{-1}\sum_{s\in\Xi}\mathrm{E}q(X_s;\theta_r)$.

1. $\forall\eta > 0$, $\inf_{r\leq R}\left[Q_{(r)}(\theta_{0r}) - \sup_{\theta:\|\theta-\theta_{0r}\|>\eta} Q_{(r)}(\theta)\right] > 0$.
2. $D^{|v|}q(X_{sr};\theta)$ *satisfies a Lipschitz condition with* $B(X_{sr})$ *, for multi-index* $|v| \geq 2$.
3. 2^{b-1} *moments exist for envelope* $B(z_{sr}) \geq \|\partial Q_{(r)}(\theta_r)/\partial\theta_r\|$.

Condition 1 is standard for identification, 2 requires sufficient smoothness, and 3 requires that the score functions have well-behaved envelopes. Many assumptions about $\Phi(\cdot)$ ensure that Assumption 5 holds (e.g., if the link function is logistic).

To be able to apply Lemma 2.3.1, I need to control the interdependence in the covariates z_{sr}. I assume that the set of nodes itself has an embedding into an integer lattice, $\Lambda \subset \mathbb{Z}^d$. Let $t \in \Lambda$ denote a generic element, and when referencing the corresponding node I write $t = t_i$. To build intuition imagine that the nodes are embedded in $\mathbb{Z}^2$ as analogous to geographic placement; households in a village are placed on a grid on the ground and certain households are closer to

[15] A more realistic model may have ϵ_{ij} and ϵ_{jr} being jointly normal.

others. This closeness determines the covariance of their other characteristics. Then, every node is given a random covariate z_i, $t_i \in \Lambda$. The pair level covariate z_{ij} is given by $z_{ij} = f(z_i, z_j)$ for some function $f(\cdot,\cdot)$. The interdependency in the node level covariates will translate to interdependencies among the edge level covariates which is what I will ultimately use in our argument.[16]

I define a distance (pseudo-metric) $d_\Xi(\cdot,\cdot)$ over the set of pairs, given by Ξ, where two pairs ij and kl's distance is said to be the minimum coordinate-wise distance between an element of the first pair and an element of the second pair. Specifically, for $s_{ij}, s_{kl} \in \Xi$ and $d_\Lambda(t_i, t_j) := \|t_i - t_j\|_1$,

$$d_\Xi(s_{ij}, s_{kl}) := d_\Lambda(t_i, t_j) \wedge d_\Lambda(t_i, t_l) \wedge d_\Lambda(t_j, t_k) \wedge d_\Lambda(t_k, t_l).$$

To describe interdependence in the data, I define a mixing coefficient. Let $D \subset \mathbb{Z}^d$ be an integer lattice and to each $s \in D$ I associate a random variable z_s. Then $\{z_s : s \in D\}$ is a random field and I are interested in controlling the dependence of z_s and $z_{s'}$. Let $\mathcal{A}_\Omega$ be the σ-algebra generated by a random field $\{z_s : s \in \Omega\}$. I define the mixing coefficient

$$\alpha_{k,l}(m) := \sup\left\{\left|\mathrm{P}\left(A_1 \cap A_2\right) - \mathrm{P}\left(A_1\right)\mathrm{P}\left(A_2\right)\right| : A_i \in \mathcal{A}_{\Omega_i}, |\Omega_1| \leq k, |\Omega_2| \leq l, d\left(\Omega_1, \Omega_2\right) \geq m\right\}$$

where $d(\Omega_1, \Omega_2) = \min_{x,y \in \Omega_1 \times \Omega_2} \|x - y\|_1$. I will need to assume that the level of interdependence goes to zero as the distance between the two subsets goes to infinity.[17]

ASSUMPTION 6 (Mixing Conditions). *$\forall r$, $z_r := \left\{z_{ir} : t_i \in \Lambda \subset \mathbb{Z}^d\right\}$ is a stationary mixing random field, $z_{ijr} := f(z_{ir}, z_{jr})$ satisfies $\sup_r \mathrm{E}\|z_{ijr}\|^{p+\delta} \lesssim \sup_r \mathrm{E}\|z_{1r}\|^{p+\delta}$, and $\sup_r \mathrm{E}\|z_{1r}\|^{p+\delta} < \infty$, with (i) $\sup_r \alpha^r_{2,\infty}(m) \leq Ca^m$ for $a \in (0,1)$ or (ii) $\sup_r \alpha^r_{2,\infty}(m) = o(m^{-d})$.*

While hard to verify, this assumption is analogous to those made in time series and spatial econometrics contexts. I require that the random fields z_r satisfy uniform mixing requirements where, as the distance between the sites of two random variables increase, the level of interdependency decays fast enough. The assumption on f is not very restrictive. The most natural example is a covariate based on the difference in characteristics of nodes i and j: $z_{ij} = \|z_i - z_j\|$. It is easy to see that $\mathrm{E}\|z_{ij}\|^k \leq 2^k \mathrm{E}\|z_1\|^k$ by the binomial theorem and stationarity. Because A_{sr} is a measurable function of z_{sr}, it will inherit stationarity and mixing properties and therefore so will X_{sr}.

PROPOSITION 2.4.1. *Assumptions 5 and 6 imply the conditions of Lemma 2.3.1.*

Since there are effectively $n(n-1)/2$ edge formation outcomes being estimated, usually $a_R = n$. Compared to slower rates, this enables us to have relatively more graphs for a given level of network

[16] We will have $\left|\mathrm{P}\left(A_{ij} \cap A_{kl}\right) - \mathrm{P}\left(A_{ij}\right)\mathrm{P}\left(A_{kl}\right)\right| \leq Ca^m$ for pairs ij, kl where $i, j \neq \{k, l\}$ and $d(ij, kl) = m$ since $\Lambda_1 = \{i, j\}$ and $\Lambda_2 = \{k, l\}$ generate sub-algebras. The only time when I cannot control the probabilities like this is for the $2(n-1)$ terms where two distinct pairs share one index, but there are only $O(n)$ of these and $|\Xi| = O(n^2)$.

[17] The triangular array notation is cumbersome, see (Jenish and Prucha 2009), but formally is $\{z_{s,R} : s \in D_R,\ R \in \mathbb{R}\}$ a triangular array defined on a sequence of probability spaces where D_R is a finite subset of D and

$$\alpha_{k,l}(m) := \sup_{R \geq 1} \sup\left\{\left|\mathrm{P}^R\left(A_1 \cap A_2\right) - \mathrm{P}^R\left(A_1\right)\mathrm{P}^R\left(A_2\right)\right| : A_i \in \mathcal{A}^R_{\Omega_i},\ \left|\Omega^R_1\right| \leq k,\ |\Omega_2| \leq l,\ \Omega_i \subset D_R,\ d\left(\Omega_1, \Omega_2\right) \geq m\right\}.$$

size while still satisfying the rate requirement of Assumption 1.3.[18]

Until now I have not discussed the role of random sampling. It is easy to see with random sampling of nodes (G^S or $G^{|S}$) or random sampling of pair data A_{ij} that the criterion function $Q_{(r)}(\theta) := \lim_{n\to\infty} \left|\Xi^S\right|^{-1} \sum_{s\in\Xi} \mathrm{E}[q(X_{sr};\theta)\mathbf{1}\{s \in \Xi^S\}]$ is minimized at true parameter θ_{0r}. For instance, under the star subgraph I have $\mathrm{E}[q(X_{sr};\theta)\mathbf{1}\{s \in \Xi^S\}] = (1-(1-\psi)^2)\mathrm{E}q(X_s;\theta_{0r})$ while $\left|\Xi^S\right| = (1-(1-\psi)^2)\binom{n}{2}$. More generally, if the sampling procedure is known, then in such a model augmenting the likelihood to account for the sampling will produce consistent estimates.

Class 2: Group Models

By allowing for an increasing number of parameters, a network formation model may be able to better and more flexibly describe the random graph process. Models of this vein are discussed in (Bickel and Chen 2009), among others, who provide a discussion of what they call a nonparametric view of network models. Our specific example comes from (Chatterjee, Diaconis, and Sly 2010), who study an environment in which the degree distribution is the sufficient statistic for graph formation: given $(d_1, ..., d_n)$, one estimates a formation model.[19] Following (Diaconis and Freedman 1984) they show the network is described by

$$\mathrm{P}(A_{ij} = 1) \propto \exp(\theta_{0i} + \theta_{0j}),$$

which is a model that allows the number of parameters to grow at a $\Theta(n)$ rate.

I extend this framework to our environment and assume there are k_n categories of nodes. For instance, if graph formation depends on two characteristics, gender (male/female) and education (high/low), there are four such categories. By allowing k_n to grow rapidly with n, I can capture substantial variation in the formation of the network. I allow $k_n = \Theta(n)$. Define an equivalence class of nodes: if i and j are in the same class, then they have the same parameter, $\theta_{0i} = \theta_{0j}$. In our example, two individuals in the same category (e.g., female and high education) are governed by the same parameter $\theta_{0,female,high}$. If I have q_n characteristics with a uniformly bounded values (e.g., two genders, a bounded number of education levels), the number of categories can grow at $q_n = \Theta(\log n)$, which yields $k_n = \Theta(n)$. One can think of this model as having group fixed effects with a growing number of groups. It turns out that with probability approaching one, $\sup_{r\leq R} \left\|\widehat{\theta}_r - \theta_{0r}\right\|_\infty \lesssim \sqrt{\log n/n}$ which is a very slow rate, though expected given how rapidly I are increasing the parameter dimension.

PROPOSITION 2.4.2. *Let the maximum coordinate value of $\theta_r \in \Theta_r$ be uniformly bounded over all r, $R = o(n \cdot \log^{-1} n)$, and $k_n = \Theta(n)$, $k_n < n$. Then, under stratified random sampling with either*

[18] One can modify this model to change a_R (e.g., (Christakis, Fowler, Imbens, and Kalyanaraman 2010) and (Goldsmith-Pinkham and Imbens 2011)) by introducing truncation terms which make certain nodes i and j have essentially zero probability of being linked. A simple example uses a covariate $-|i-j|$ with nodes $i \in \mathbb{N}$ yields $a_R = \sqrt{n}$.

[19] Conditional on the degree distribution there is no information about the model from the actual network data.

the induced or star subgraph and Assumptions 2, 3, or 4, the conclusion of Theorem 2.3.1 holds.
This example shows that even when I are adding parameters at rate n, graphical reconstruction is possible in network level analyses. Here $a_R = \sqrt{n/\log n}$ and therefore $a_R^{-1} \cdot \sqrt{R^{1+2/b}} \to 0$. Meanwhile, the sufficient condition is not met for vertex level analysis as $a_R^{-1} \cdot \sqrt{nR^{1+2/b}} \to \infty$. This example provides an illustration of both the strengths and limitations of graphical reconstruction by testing the limits as I add dimensions at the same rate as the number of nodes. Certainly more slowly growing parameter dimension will only help.

Class 3: A Model with Network Effects (P-ERGM)

Our third example studies a graph formation model with network effects where the edge-formation probability depends on other edges above and beyond just through the correlated covariate effect. In general, the principle difficulty in studying network formation models is balancing the interdependency while still allowing consistent estimation of parameters. As previously discussed, most off-the-shelf network formation models lack asymptotic theory. Here I discuss the consistency of graphical reconstruction with a simple model that incorporates network effects, using a simple example from Chandrasekhar and Jackson (2012). We consider a specialized example partitioned exponential random graph model (P-ERGM) wherein a network is partitioned into two sets of links. One set counts triangles, and there are $\Omega\left(n^{\gamma_1}\log^{\gamma_2} n\right)$ of these for $\gamma_1 \geq 1, \gamma_2 \geq 0$ with at least one strict. The other predicts links of which there are order n^2 of potential links. Assume that in the links form with probability p_n and the triangles with probability q. This corresponds to an exponential random graph model

$$\mathrm{P}\left(S(G);\theta_{0r}\right) \propto \binom{\binom{n}{2} - 3a_{n,T}}{S_1(G)} \binom{a_{n,T}}{S_2(G)} \times \exp\left(\theta_{0r,2}S_2(G) + \left(\theta_{0r,1} - \log n\right) S_1(G)\right)$$

where $\theta_{0r,1} = \exp(\delta)$, $S_2(G) = \#\left\{A_{ij}A_{jk}A_{ki} = 1 : i < j < k\right\}$, and $S_1(G) := |E| - 3S_2(G)$. It can be shown that the network formation parameters can be consistently and asymptotically normally estimated for a given graph r.

PROPOSITION 2.4.3. *Let the maximum coordinate value of $\theta_r \in \Theta_r$ be uniformly bounded over all r and $nR^{1+2/b}/a_{n,T} \to 0$. Then the conditions of Lemma 2.3.1 hold.*

Such a model can be used easily with sampled network data. Let $q_r := \exp(\theta_{0r,2})$ be the probability that a given triangle forms out of the $a_{n,T}$ possible triangles in that part of the partition. Notice that asymptotically all triangles in this graph form from this part, and therefore, we can adjust for the sampling scheme and simply count triangles to estimate q_r. Specifically, $\widehat{q_r} = (3\psi^2(1-\psi) + \psi^3)^{-1} \cdot \#\text{Triangles}(G_r^S)/a_{n,T}$ delivers a consistent estimate. Similarly, letting $p_{r,n} := \exp(\theta_{0r,1})/n$ be the implied probability that a link forms on the non-triangle part of the partition, we can compute this directly with induced graph data.

Given our estimates of $\widehat{\theta_r}$, drawing from the distribution is straightforward. We can simply use importance sampling by taking a set of $\{G_{r,s}^* : s = 1, ..., S\}$, where $G_{r,s}^* = (A_r^{obs}, A_r^{mis,s})$ across S

simulations. Then we can compute the implied probability under the model $\mathrm{P}\left(S\left(G_{r,s}^{*}\right);\widehat{\theta}_{r}\right)$ and construct the requisite conditional probability by $\mathrm{P}\left(G_{r,s}^{*}|A_{r}^{obs}\right)=\mathrm{P}\left(G_{r,s}^{*}\right)/\sum_{s'}\mathrm{P}\left(G_{r,s'}^{*}\right)$.

Class 4: Models with Network Effects

In this example I illustrate the consistency of graphical reconstruction with a simple model that exhibits network effects, taken from the dynamic binary choice literature ((De Jong and Woutersen 2010)). Similar to (Christakis, Fowler, Imbens, and Kalyanaraman 2010), I assume that pairs of nodes meet in a sequence, and for simplicity I assume the sequence is known. Consequently there are $\binom{n}{2}$ periods in which pairs meet. Assume that the probability that i and j, meeting at period s, form a link is given by

$$\mathrm{P}\left(A_{s,r}=1|z_{sr},A_{r}^{s-1};\theta_{0r}\right)=\Phi\left(\sum_{k=1}^{l}\theta_{0r,k}A_{s-k,r}+\theta_{0r,l+1}'z_{sr}\right)$$

where A_r^{s-1} is the structure of the graph that has formed through period $s-1$. The probability i and j form an edge depends on covariates as well as a finite collection of previously formed edges.

ASSUMPTION 7 (Mixing Conditions). *(1) For each r, $z_r := \{z_{sr} : s = 1, ..., \binom{n}{2}\}$ is a stationary mixing random sequence satisfying $\sup_r \mathrm{E}\|z_{1r}\|^{p+\delta} < \infty$, with (i) $\sup_r \alpha^r(m) \leq Ca^m$ for $a \in (0,1)$ or (ii) $\sup_r \alpha^r(m) = o(m^{-d})$; (2) $\epsilon_{sr}|A_r^{s-1}, z_r^{s-1}$ are iid $\mathcal{N}(0,1)$.*

PROPOSITION 2.4.4. *Assumption 7 implies the conditions of Lemma 2.3.1.*

In this example, randomly sampled data such as the induced subgraph, star subgraph, or even randomly sampled edges are easy to accommodate. The key take-away here is not that this example is a particularly good model of network formation, but rather that it demonstrates how interdependencies in edge formation can be accommodated.

2.4.1 Discussion of Boundaries

In this section I have developed a general method to consistently estimate the economic parameter using graphical reconstruction. The method allows the researcher to estimate network effects using a general set of network statistics, such as eigenvector centrality, where no analytical solutions are available. I now discuss how the effectiveness of graphical reconstruction may vary with network formation models, statistics of interest, and misspecification.

First, we may be interested in how misspecification of the network formation model affects graphical reconstruction. In practice, we do not know the family of models which generated the empirical networks. Clearly, misspecification is problematic only to the extent of the covariance between the conditional expectation of the misspecified model and its deviation from the true model. While this is not easy to analytically characterize, it does suggest that the model one needs relates to the network statistic one is interested in studying. For instance, graphical reconstruction with Erdos-Renyi style models may be sufficient to study questions pertaining to the degree distribution,

but may perform poorly if one is interested in clustering. Numerical simulations confirm precisely this intuition, suggesting that chosen models ought to be a function of the statistic of interest.

Second, if the model is consistently estimable, one may ask whether a dense versus sparse version of the model will exhibit better convergence properties. A dense graph has expected average degree growing asymptotically and a sparse graph has bounded degree. While dense graphs have more edge-decisions than the sparse counterparts, it turns out that such a statement cannot be made.[20]

Third, models of conditional edge independence where I flexibly introduce covariates can capture substantial co-variation in network formation. (Bickel and Chen 2009) present a discussion of the literature where any probability distribution on an infinite undirected graph can be represented as edges emerging conditionally independently given some latent variables and discuss how group based models are natural parametric models that approximate the nonparametric models which can capture any network formation process. In practice I lack these latent variables and therefore conditional edge independent models do poorly precisely for statistics such as clustering, raising the necessity for models with network effects.

2.5 Conclusion

Applied social network analysis often use graphs constructed from data collected from a partial sample of nodes. Even when nodes are selected randomly, the partial sampling induces non-classical measurement error and consequently biases estimates of regression coefficients and GMM parameters. To address the problem in general, I develop a consistent and asymptotically normal method to estimate the economic parameters using graphical reconstruction, while allowing for substantial heterogeneity across networks. Specifically, the method allows for every network in the sample to be generated by a different model. Undoubtedly, the performance of graphical reconstruction with empirical network data will only improve as the burgeoning literature on consistently estimable network formation models matures. To that end, from a theoretical perspective the lacuna in the literature is the absence of network formation models that both allow for higher-order dependencies in link formation and are also consistently estimable.

[20]Consider two examples with $\mathrm{E}d(G) = \Theta(1)$. The first example in general is an Erdos-Renyi graph with $p_n = \delta n^{-\gamma}$, $\gamma \in [1, 2)$, which is sparse when $\gamma = 1$. The second is a block model where a node has d links in expectation, there are k_n blocks with m nodes per block, and nodes only connect to other nodes in the same block with iid probability $p = d/m$. The first model has $\widehat{p} - p = O_{\mathrm{P}}(n^{-1-\gamma})$ and the latter has $\widehat{p} - p = O_{\mathrm{P}}(n^{-1/2})$. The dense Erdos-Renyi model has $\widehat{p} - p = O_{\mathrm{P}}(n^{-1})$.

2.A Figures and Tables

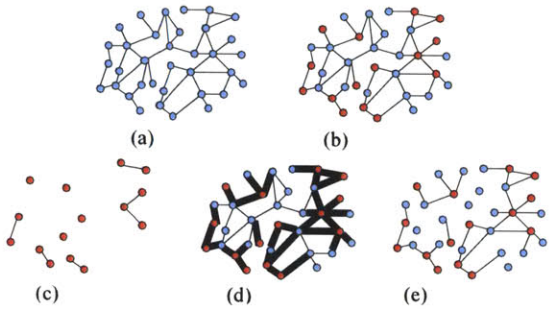


Figure 2-1: (a) $G = (V, E)$. (b) the sampled nodes, S, in red. (c) the induced subgraph, $G^{|S} = (S, E^{|S})$. (d) G, highlighting the sampled nodes and the edges that are induced if each sampled node reports all of its links from the census, E^S. (e) the star subgraph, G^S.

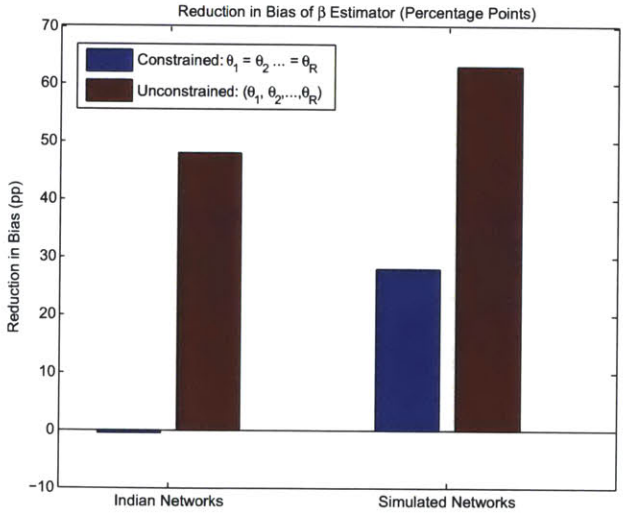


Figure 2-2: The percentage-point reduction in bias when using graphical reconstruction and constraining $\theta_1 = ... = \theta_R$ versus allowing for heterogenous θ_r, for both simulated network data (following Chandrasekhar and Lewis, 2012) and empirical Indian village network data (Banerjee et al., 2012).

2.B Proofs for Section 2.3

2.B.1 Useful Results

In what follows, for $p \times q$ matrix D, let $\|\cdot\| := \|\cdot\|_2$ be the matrix norm induced by vector norm $\|\cdot\|$, with $\|D\| := \max_{x \in S^{q-1}} \|Dx\|$.

LEMMA 2.B.1. *Assume for all $\theta_r \in \Theta_r$, $\zeta_{ir}(\theta_r, u)$ are $q \times kq$ matrix (or vector) valued functions satisfying $\sup_{u \in \mathcal{U}} \sup_{i,r} \|\zeta_{ir}(\theta_r, u)\| \leq B_{ir}$ with $\limsup_{R \to \infty} \mathbb{E}_{nR}[\mathrm{E} B_{ir}] < \infty$ and Assumption 1.3 hold. Then*

$$\sup_{u \in \mathcal{U}} (nR)^{1/2} \left\| \mathbb{E}_{n,R} \left[\zeta_{ir}(\bar{\theta}_r, u)(I_q \otimes (\widehat{\theta}_r - \theta_{0r})) \right] \right\| = o_{\mathrm{P}}(1)$$

for $\bar{\theta}_r$ on the line between $\widehat{\theta}_r$ and θ_{0r}.[21]

Proof of Lemma 2.B.1. This follows from

$$\sup_{u \in \mathcal{U}} (nR)^{1/2} \mathbb{E}_{n,R} \left[\left\| \zeta_{ir}(\bar{\theta}_r, u) \right\| \left\| I_q \otimes (\widehat{\theta}_r - \theta_{0r}) \right\| \right] \leq (\mathbb{E}_{nR}[B_{ir}]) \cdot \sqrt{nR} \sup_{r \leq R} \left\| \widehat{\theta}_r - \theta_{0r} \right\| \leq \mathbb{E}_{nR}[B_{ir}] \cdot O_{\mathrm{P}} \left(a_n^{-1} \cdot \sqrt{nR^{1+2/b}} \right) = o_{\mathrm{P}}(1)$$

as the mean of the envelopes converges and the rates obey the assumed relationship. □

Proof of Lemma 2.3.1. For every network $a_n \cdot (\widehat{\theta}_r - \theta_{0r}) = -\left(\nabla_\theta \widehat{V}_{(r)}(\bar{\theta}_r) \right)^{-1} \cdot a_n \cdot \widehat{V}(\theta_{0r})$. By the Lipschitz condition 4 of the Lemma and Lemma 2.B.2, $\sup_r \left\| \nabla_\theta \widehat{V}_{(r)}(\bar{\theta}_r) - \nabla_\theta \widehat{V}_{(r)}(\theta_{0r}) \right\| = o_{\mathrm{P}}(1)$, which can be seen from

$$\sup_r \left\| \nabla_\theta \widehat{V}_{(r)}(\bar{\theta}_r) - \nabla_\theta \widehat{V}_{(r)}(\theta_{0r}) \right\| \leq \sup_r \|B_r\| \cdot \sup_r \left\| \widehat{\theta}_r - \theta_{0r} \right\| = o_{\mathrm{P}}(1).$$

By condition 5 of the Lemma $\sup_r \left\| a_n \cdot (\hat{\theta}_r - \theta_{0r}) + \left(\nabla_\theta V_{(r)}(\theta_{0r}) \right)^{-1} \cdot a_n \cdot \widehat{V}(\theta_{0r}) \right\| = o_{\mathrm{P}}(1)$. By condition 3, we have that $a_n \cdot \sup_r \|\widehat{V}_{(r)}(\theta_r)\| = O_{\mathrm{P}}(R^{1/b})$, since $\mathrm{E}[\|\sup_r a_n \cdot \widehat{V}_{(r)}(\theta_{0r})\|^b]^{1/b} \leq R^{1/b} \sup_r \|a_n \cdot \widehat{V}_{(r)}(\theta_{0r})\|_{\ell_b}$ and therefore $a_n \cdot \sup_r \left\| \widehat{\theta}_r - \theta_{0r} \right\| = O_{\mathrm{P}}(R^{1/b})$. If instead of b moments we assume all moments exist, then $\sqrt{\log(R+1)}$ replaces $R^{1/b}$ by a standard Orlicz inequality (e.g., (Van der Vaart and Wellner 1996)). □

LEMMA 2.B.2. *Conditions 1 and 2 of Lemma 2.3.1 imply $\sup_r \left\| \widehat{\theta}_r - \theta_{0r} \right\| = o_{\mathrm{P}}(1)$.*

Proof of Lemma 2.B.2. Arguing along the lines of Theorem 5 of Supplementary Appendix I of (Hahn and Newey 2004), among others, pick $\eta > 0$, define $\epsilon := \inf_{r \leq R} \left[Q_{(r)}(\theta_{0r}) - \sup_{\theta_r : \|\theta_r - \theta_{0r}\| > \eta} Q_{(r)}(\theta_r) \right]$, and condition on the event $\left\{ \sup_{r \leq R} \sup_\theta \left| \widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta) \right| < \frac{\epsilon}{3} \right\}$ which has probability $1 - o(a_n^{-v})$ by Condition 2. If we look for $\widehat{\theta}_r$ outside an η-radius ball of the true parameter, we have $\sup_{\theta_r : \|\theta_r - \theta_{0r}\| > \eta} \widehat{Q}_{(r)}(\theta_r) < \sup_{\theta_r : \|\theta_r - \theta_{0r}\| > \eta} Q_{(r)}(\theta_r) + \frac{\epsilon}{3} < Q_{(r)}(\theta_{0r}) - \frac{2\epsilon}{3} < \widehat{Q}_{(r)}(\theta_{0r}) - \frac{\epsilon}{3}$, contradicting $\widehat{Q}_{(r)}(\widehat{\theta}_r) \geq \widehat{Q}_{(r)}(\theta_{0r})$, implying $\left\| \widehat{\theta}_r - \theta_{0r} \right\| < \eta$ for all networks. □

The next lemma is useful throghout and we make explicit the dependence on R in $\mathrm{P}^R(\cdot)$ and E^R here.

LEMMA 2.B.3 (Extended Vitali Convergence). *Let $\{Z_R : R \in \mathbb{N}\}$ be L_R^1-integrable functions on a sequence of measure spaces indexed by R. (1) $Z_R \xrightarrow{\mathrm{P}} 0$ and (2) Z_R is uniformly integrable, $\sup_{R \geq 1} \mathrm{E}^R [|Z_R| \cdot \mathbf{1}\{|Z_R| > c\}] \to 0$ as $c \to \infty$, imply $\mathrm{E}^R [Z_R] \to 0$.*

Proof. The argument is analogous to the proofs of Theorem 10.3.5 and 10.3.6 in (Dudley 2002). Let $Z_R \geq 0$ w.l.o.g. First, observe (2) implies that for every $\epsilon > 0$ there exists $\delta > 0$ such that for each A_R with $\mathrm{P}^R(A_R) < \delta$,

[21] Each component of $\bar{\theta}_r$ may be at a different intermediate point.

$\mathrm{E}^R[Z_R \cdot \mathbf{1}\{A_R\}] < \epsilon$ for all R. To see this, by uniform integrability, given $\epsilon > 0$ and pick $\delta < \epsilon/(2K)$ where K is large enough such that $\mathrm{E}^R[Z_R \cdot \mathbf{1}\{Z_R > K\}] < \epsilon/2$. Then

$$\mathrm{E}^R[Z_R \cdot \mathbf{1}\{A_R\}] \leq \mathrm{E}^R[Z_R \cdot \mathbf{1}\{A_R\} \cdot \mathbf{1}\{Z_R \leq K\}] + \mathrm{E}^R[Z_R \cdot \mathbf{1}\{A_R\} \cdot \mathbf{1}\{Z_R > K\}] < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

as $\mathrm{P}^R(A_R) < \delta$. Second, given $\epsilon > 0$, let $A_R := \{Z_R > \epsilon\}$. By (1), $\mathrm{P}^R(A_R) \to 0$ as $R \to \infty$. As such, for R large enough $\mathrm{P}^R(A_R) < \epsilon/(2K)$. Therefore $\mathrm{E}^R[Z_R] = \mathrm{E}^R[Z_R \cdot \mathbf{1}\{Z_R \leq \epsilon\}] + \mathrm{E}^R[Z_R \cdot \mathbf{1}\{A_R\}] \leq 2\epsilon$. □

2.B.2 Proof of Theorem 2.3.1

The argument is straightforward and follows by expanding around $\widehat{\theta}$ uniformly and then checking that conditional expectations preserve asymptotic normality.

Proof of Theorem 2.3.1.1. Consistency is clear so we directly check normality. Let $u_{ir} = \epsilon_{ir} + (w_{ir}(G_r) - \mathcal{E}_{ir}(x_r;\widehat{\theta}_r))\beta_0$ and $H_R(\widehat{\theta}) := (nR)^{-1}\mathcal{E}(x;\widehat{\theta})'\mathcal{E}(x;\widehat{\theta})$. As usual $\sqrt{nR}\left(\widehat{\beta} - \beta_0\right) = H_R(\widehat{\theta})^{-1} \cdot (nR)^{-1/2}\mathcal{E}(x;\widehat{\theta})'u$.**Step 1:** To show $H_R(\widehat{\theta}) - H_{R,0}(\theta_0) = o_{\mathrm{P}}(1)$, where $H_{R,0}(\theta_0) := \mathbb{E}_{nR}[\mathrm{E}[\mathcal{E}_{ir}(x_r;\theta_{0r})\mathcal{E}_{ir}(x_r;\theta_{0r})']]$. By a term-by-term expansion,

$$H_R(\widehat{\theta}) = H_R(\theta_0) + \mathbb{E}_{nR}\left[\dot{\Psi}_{ir}(\bar{\theta}_r)\left(I_p \otimes (\widehat{\theta}_r - \theta_{0r})\right)\right]$$

where $\dot{\Psi}_{ir}(\bar{\theta}_r) = \frac{\partial}{\partial\theta'}\left\{\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\mathcal{E}_{ir}(x_r;\bar{\theta}_r)'\right\}$ is a $p \times pk$ matrix of derivatives w.r.t. θ_{rk}, which exists by Assumption 1. By Lemma 2.B.5 the second term $o_{\mathrm{P}}(1)$ since $\left\|\dot{\Psi}_{ir}(\bar{\theta}_r)\right\| \lesssim \left\|\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\right\| \left\|\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\right\| \lesssim L_{i,r}(x_r)^2$, which by Assumption 2.2 follows from

$$\left\|\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\right\| \leq L_{i,r}(x_r) \text{ and } \left\|\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\right\| \leq \sqrt{\mathrm{E}\left[\|w_{ir}\|^2 | x_r;\bar{\theta}_r\right]} \cdot \sqrt{\mathrm{E}\left[\left\|\mathcal{I}_{w_{ir}|x_r}(\bar{\theta}_r)\right\| | x_r;\bar{\theta}_r\right]} \leq L_{i,r}(x_r).$$

That $H_R(\theta_0) - H_{R,0}(\theta_0) = o_{\mathrm{P}}(1)$ follows from Vitali convergence.

Step 2: We show $(nR)^{-1/2}\mathcal{E}(x;\widehat{\theta})'u \rightsquigarrow \mathcal{N}(0,V)$. Let $f_{ir}(\theta_r;\beta) := \mathcal{E}_{ir}(x_r;\theta_r)(w_{ir}(G_r) - \mathcal{E}_{ir}(x_r;\theta_r))'\beta$. Then

$$\mathcal{E}(x;\widehat{\theta})'u/\sqrt{nR} = f(\theta_0;\beta_0)/\sqrt{nR} + \mathcal{E}(x;\theta_0)'\epsilon/\sqrt{nR} + \sqrt{nR}\mathbb{E}_{nR}\left[\dot{\Phi}_{ir}(\bar{\theta}_r)\left(I_p \otimes (\widehat{\theta}_r - \theta_{0r})\right)\right]$$

where $\dot{\Phi}_{ir}(\bar{\theta}_r) = \frac{\partial}{\partial\theta_r}\left\{f_{ir}(\bar{\theta}_r;\beta_0) + \mathcal{E}_{ir}(x_r;\bar{\theta}_r)\epsilon_{ir}\right\}$. Clearly $\mathrm{E}[f_{ir}(\theta_{0r},\beta_0)] = 0$ and $\mathrm{E}[\mathcal{E}_{ir}(x_r;\theta_{0r})\epsilon_{ir}] = 0$.[22] Let $g_{ir}(\theta_{0r};\beta_0) := f_{ir}(\theta_{0r};\beta_0) + \mathcal{E}_{ir}(x_r;\theta_{0r})\epsilon_{ir}$. That $\sqrt{nR}\mathbb{E}_{n,R}g_{ir}(\theta_{0r},\beta_0) \rightsquigarrow \mathcal{N}(0,V)$, where $V = \lim \mathbb{E}_R\mathrm{var}\left(\sqrt{n}\mathbb{E}_n g_{ir}(\theta_{0r},\beta_0)\right)$, follows from Lemma 2.B.5 which can be applied by Assumption 2.3.

Finally, the $\dot{\Psi}_{ir}(\bar{\theta}_r)$ term is controlled by Lemma 2.B.1 using $\left\|\frac{\partial}{\partial\theta'}f_{ir}(\bar{\theta}_r;\beta_0)\right\| \lesssim_{\mathrm{P}} L_{i,r}(x_r)^2$, which follows from

$$\left\|\frac{\partial}{\partial\theta'}f_{ir}(\bar{\theta}_r;\beta_0)\right\| \leq \sup_{\beta\in\mathcal{B}}\|\beta\| \cdot \left\|\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\right\| \left\|w_{ir}(G_r) - \frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(x_r;\bar{\theta}_r)\right\| \lesssim L_{i,r}(x_r)\|w_{ir}(G_r)\| + L_{i,r}(x_r)^2 \lesssim_{\mathrm{P}} L_{i,r}(x_r)^2$$

as $\|w_{ir}(G_r)\| \lesssim_{\mathrm{P}} \mathrm{E}[\|w_{ir}(G_r)\|] \leq \mathrm{E}[L_{i,r}(x_r)]$. □

Next, we turn to GMM. The argument is conceptually similar and depends on uniform expansions. The main difference is control of score functions. Let us define the conditional score[23] with respect to β and the conditional score with respect to θ_r as $S_{m_{ir}|x_r}(\beta;\theta_r) := \frac{\partial}{\partial\beta'}\log f(m_{ir}|x_r;\beta,\theta_r)$ and $S_{m_{ir}|x_r}(\theta_r;\beta) = \frac{\partial}{\partial\theta_r'}\log f(m_{ir}|x_r;\beta,\theta_r)$. The corresponding information matrices are

$$\mathcal{I}_{m_{ir}|x_r}(\beta;\theta_r) = \mathrm{E}\left[S_{m_{ir}|x_r}(\beta;\theta_r)S_{m_{ir}|x_r}(\beta;\theta_r)'|x_r;\beta,\theta_r\right] \text{ and } \mathcal{I}_{m_{ir}|x_r}(\theta_r;\beta) = \mathrm{E}\left[S_{m_{ir}|x_r}(\theta_r;\beta)S_{m_{ir}|x_r}(\theta_r;\beta)'|x_r;\beta,\theta_r\right].$$

[22] The former by iterated expectations and the latter by assumption on ϵ.

[23] The right-hand side of the expression abuses notation. Given that graphs occupy a discrete space, with continuously distributed disturbances in the model we may formally have to decompose the measure into its absolutely continuous part and its pure point part (e.g., Lebesgue decomposition theorem). We rule out the singular part by assumption.

Notice that $\left\|\mathcal{I}_{m_{ir}|x_r}(\beta;\theta_r)\right\| \leq \left\|\mathcal{I}_{m_{ir}|x_r}(\beta,\theta_r)\right\|$ and $\left\|\mathcal{I}_{m_{ir}|x_r}(\theta_r;\beta)\right\| \leq \left\|\mathcal{I}_{m_{ir}|x_r}(\beta,\theta_r)\right\|$, since each is a projection of the larger information matrix, by Cauchy's interlacing theorem. We also use a shorthand $\mathcal{E}_{ir}(\beta,\theta_r)$ for $\mathcal{E}_{ir}(x_r;\beta,\theta_r)$.

LEMMA 2.B.4. *Under Assumptions 1 and 3,* $\widehat{\beta}_{\text{gmm}} \xrightarrow{\text{P}} \beta_0$.

Proof. In four steps we check conditions 1-4 of (Andrews 1994), Theorem A-1. **Step 1:** The first part is clear by Assumption 3.2, since $\text{E}\left[m(y_{ir},w_{ir};\beta)|x_r;\theta_r,\beta\right] = \mathcal{E}_{ir}(\beta,\theta_r)$. That $\mathcal{E}_{ir}(\beta,\theta_r)$ satisfies a uniform law of large numbers, $\sup_{(\beta,\theta)\in\mathcal{B}\times\prod_{r\in\mathbb{N}}\Theta_r}\left\|\mathbb{E}_{nR}\left\{\mathcal{E}_{ir}(\beta,\theta_r) - \text{E}\mathcal{E}_{ir}(\beta,\theta_r)\right\}\right\| = o_{\text{P}}(1)$, follows from a pointwise convergence, which is clear, and stochastic equicontinuity. Stochastic equicontinuity follows from a Lipschitz condition. Define the following.

$$\epsilon^1_{i,r}(\bar{\beta},\beta,\theta_r') := \text{E}\left[\frac{\partial}{\partial\beta'}m(y_{ir},w_{ir}(G_r);\bar{\beta})|z_r,A_r^o,y_r;\theta_r,\beta\right],\ \epsilon^2_{i,r}(\beta',\tilde{\beta},\theta_r') := \text{E}\left[m(y_{ir},w_{ir}(G_r);\beta')\cdot S_{m_{ir}|x_r}(\tilde{\beta};\theta_r')'|z_r,A_r^o,y_r;\theta_r,\tilde{\beta}\right],$$

$$\text{and } \epsilon^3_{i,r}(\beta',\bar{\theta}_r,\theta_r) := \text{E}\left[m(y_{ir},w_{ir}(G_r);\beta')\cdot S_{m_{ir}|x_r}(\bar{\theta}_r;\beta')'|z_r,A_r^o,y_r;\theta_r,\beta'\right].$$

By a Taylor expansion it follows that $\left\|\text{E}\left[m(y_{ir},w_{ir}(G_r);\beta')|z_r,A_r^o,y_r;\theta_r',\beta'\right] - \text{E}\left[m(y_{ir},w_{ir}(G_r);\beta)|z_r,A_r^o,y_r;\theta_r,\beta\right]\right\|$ is bounded by $\left\|\epsilon^1_{i,r}(\bar{\beta},\beta,\theta_r') + \epsilon^2_{i,r}(\beta',\tilde{\beta},\theta_r')\right\|\left\|\beta'-\beta\right\| + \left\|\epsilon^3_{i,r}(\bar{\beta},\theta_r')\right\|\left\|\theta_r'-\theta_r\right\|$. Then we have by Assumption 3.3,

$$\begin{aligned}
\left\|\epsilon^1_{i,r}(\bar{\beta},\beta,\theta_r')\right\| &\leq \text{E}\left[\left\|\frac{\partial}{\partial\beta'}m(y_{ir},w_{ir}(G_r);\bar{\beta})\right\||x_r;\theta_r',\beta\right] \leq L_{i,r}(x_r),\\
\left\|\epsilon^2_{i,r}(\beta',\tilde{\beta},\theta_r')\right\| &\leq \sqrt{\text{E}\left[\|m_{ir}(\beta')\|^2|x_r;\theta_r',\tilde{\beta}\right]}\cdot\sqrt{\left\|\mathcal{I}_{m_{ir}|x_r}(\tilde{\beta};\theta_r')\right\|} \leq L_{i,r}(x_r), \text{ and}\\
\left\|\epsilon^3_{i,r}(\beta',\bar{\theta}_r,\theta_r)\right\| &\leq \sqrt{\text{E}\left[\|m_{ir}(\beta')\|^2|x_r;\theta_r',\tilde{\beta}\right]}\cdot\sqrt{\left\|\mathcal{I}_{m_{ir}|x_r}(\bar{\theta}_r;\beta')\right\|} \leq L_{i,r}(x_r).
\end{aligned}$$

As such, with $d_{\mathcal{B}\times\prod_{r\in\mathbb{N}}\Theta_r}((\beta,\theta),(\beta',\theta')) := \|\beta-\beta'\| \vee \sup_{r\in\mathbb{N}}\|\theta_r-\theta_r'\|$,

$$\left\|\mathbb{E}_{n,R}\left\{\mathcal{E}_{ir}(\beta,\theta_r) - \mathcal{E}_{ir}(\beta',\theta_r')\right\}\right\| \leq \mathbb{E}_{n,R}L_{i,r}\cdot\sup_{r\leq R}\left(\left\|\beta'-\beta\right\| + \left\|\theta_r'-\theta_r\right\|\right) \leq \mathbb{E}_{n,R}L_{i,r}\cdot d_{\mathcal{B}\times\prod_{r\in\mathbb{N}}\Theta_r}((\beta,\theta),(\beta',\theta'))$$

Since $\sup_{R\geq 1}\mathbb{E}_{n,R}\left[\text{E}L_{i,r}\right] < \infty$, by (Andrews 1992) Lemma 2 it follows that stochastic equicontinuity holds and that $\mathbb{E}_{n,R}\text{E}\mathcal{E}_{ir}(\beta,\theta_r)$ is uniformly continuous in $(\beta,\theta)\in\mathcal{B}\times\prod_{r\in\mathbb{N}}\Theta_r$.

Step 2: The second and third part of the second condition are clear: by Assumption 3.1 $\widehat{W} - W = o_{\text{P}}(1)$ and by Assumption 1 $\left(\widehat{\theta}_1,...,\widehat{\theta}_R\right) \in \prod_{r=1}^R\Theta_r$. We need only show $\sup_{\beta\in\mathcal{B}}\left\|\lim_{R\to\infty}\mathbb{E}_{n,R}\text{E}\mathcal{E}_{ir}(\beta,\widehat{\theta}_r) - \lim_{R\to\infty}\mathbb{E}_{n,R}\text{E}\mathcal{E}_{ir}(\beta,\theta_{0r})\right\| = o_{\text{P}}(1)$ for $\theta_0\in\prod_{r\in\mathbb{N}}\Theta_r$. This follows from a Taylor expansion and the fact that

$$\sup_{\beta\in\mathcal{B}}\lim_{R\to\infty}\mathbb{E}_{n,R}\left\|\text{E}\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(\beta,\bar{\theta}_r)\cdot\left(I\otimes(\widehat{\theta}_r-\theta_{0r})\right)\right\| \leq \limsup_{R\to\infty}\mathbb{E}_{n,R}\text{E}L_{i,r}\cdot\sup_r\left\|\widehat{\theta}_r-\theta_{0r}\right\| = o_{\text{P}}(1)$$

where $\sup_r\left\|\hat{\theta}_r-\theta_{0r}\right\| = o_{\text{P}}(1)$ by Assumption 1.3 and $\limsup_{R\to\infty}\mathbb{E}_{R,R}\text{E}L_{i,r} < \infty$ by 3.3. This, in turn, results from the fact that $\left\|\text{E}\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(\beta,\bar{\theta}_r)\right\| \leq \text{E}L_{i,r}$. To see this, note $\left\|\text{E}\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(\beta,\bar{\theta}_r)\right\| \leq \text{E}\left[\left\|\text{E}\left[m_{ir}(\beta)S_{m_{ir}|x_r}(\beta;\bar{\theta}_r)|x_r;\beta,\bar{\theta}_r\right]\right\|\right]$. By Assumption 3.3, $\left\|\text{E}\frac{\partial}{\partial\theta'}\mathcal{E}_{ir}(\beta,\bar{\theta}_r)\right\| \leq \text{E}\left[\sqrt{\left(\text{E}\left[\|M_{i,r}(X_{ir})\|^2|x_r;\beta,\bar{\theta}_r\right]\right)}\cdot\sqrt{\left\|\mathcal{I}_{m_{ir}|x_r}(\beta;\bar{\theta}_r)\right\|}\right] \leq \text{E}L_{i,r}(x_r)$.

Step 3: This follows from $\text{E}\left\|\mathcal{E}_{ir}(\beta,\theta_r)\right\| \leq \text{E}L_{i,r}$ and $\limsup_{n\to\infty}\mathbb{E}_{n,R}\text{E}L_{i,r} < \infty$.

Step 4: By iterated expectations $g_n(\beta) = \mathbb{E}_{n,R}\text{E}\left[\text{E}\left[m(X_{ir};\beta)|x_r;\theta_{0r},\beta_0\right]\right] = \mathbb{E}_{n,R}\text{E}\eta_{ir}(\beta,\theta_{0r},\beta_0)$ where $\eta_{ir}(\beta,\theta_r,\beta) := \mathcal{E}_{ir}(\beta,\theta_r)$. By the identification condition, Assumption 3.1

$$W\lim_{R\to\infty}\mathbb{E}_{n,R}\text{E}\eta_{ir}(\beta,\theta_{0r},\beta_0) \neq 0 \text{ and } W\lim_{R\to\infty}\mathbb{E}_{n,R}\text{E}\mathcal{E}_{ir}(\beta,\theta_{0r}) \neq 0$$

for $\beta\neq\beta_0$, but $W\lim_{R\to\infty}\mathbb{E}_{n,R}\text{E}\mathcal{E}_{ir}(\beta_0,\theta_{0r}) = W\lim_{R\to\infty}\mathbb{E}_{n,R}\text{E}\eta_{ir}(\beta_0,\theta_{0r},\beta_0) = 0$. By positive semi-definiteness of W, letting $K'K = W$, observe that $0\neq Wf(\beta) = K'Kf(\beta)$ implies $Kf(\beta)\neq 0$. It follows that $Q(\beta,\theta_0,W) > Q(\beta_0,\theta_0,W)$ for any $\beta\neq\beta_0$. □

Proof of Theorem 2.3.1.2. Step 1: The estimator satisfies for $\widehat{\gamma}_{n,R}(\widehat{\beta};\widehat{\theta}) := \mathbb{E}_{n,R}\left[\mathcal{E}_{ir}(\widehat{\beta},\widehat{\theta}_r)\right]$,

$$\left[\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\widehat{\beta};\widehat{\theta})\right]' \widehat{W}\sqrt{nR}\widehat{\gamma}_{n,R}(\widehat{\beta};\widehat{\theta}) = o_{\mathrm{P}}(1).$$

A term-by-term expansion yields $\sqrt{nR}\widehat{\gamma}_{n,R}(\widehat{\beta};\widehat{\theta}) = \sqrt{nR}\widehat{\gamma}_{n,R}(\widehat{\beta};\theta_0^R)+\sqrt{nR}\mathbb{E}_{n,R}\left[\dot{\Psi}_{ir}(\bar{\theta}_r;\widehat{\beta})\left(I_q\otimes(\widehat{\theta}_r-\theta_{0r})\right)\right]$, where $\dot{\Psi}_{ir}(\bar{\theta}_r;\widehat{\beta}) = \epsilon^3_{i,r}(\widehat{\beta},\bar{\theta}_r,\bar{\theta}_r)$ has been controlled in the preceding lemma. Similarly $\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\widehat{\beta};\widehat{\theta}) = \frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\widehat{\beta};\theta_0^R) + \mathbb{E}_{n,R}\left[\dot{\Phi}_{ir}(\bar{\theta}_r;\widehat{\beta})\left(I_q\otimes(\widehat{\theta}_r-\theta_{0r})\right)\right]$,

$$\dot{\Phi}_{ir}(\bar{\theta}_r;\widehat{\beta}) = \mathrm{E}\left[\frac{\partial}{\partial\beta'}m(y_{ir},w_{ir};\widehat{\beta})S_{m_{ir}|x_r}(\bar{\theta}_r;\widehat{\beta})|A^o_r,y_r,z_r;\bar{\theta}_r,\widehat{\beta}\right] + \mathrm{E}\left[m(y_{ir},w_{ir};\widehat{\beta})S_{m_{ir}|x_r}(\widehat{\beta};\bar{\theta}_r)S_{m_{ir}|x_r}(\bar{\theta}_r;\widehat{\beta})|A^o_r,y_r,z_r;\bar{\theta}_r,\widehat{\beta}\right].$$

We can bound the first term by $(\mathrm{E}[\|\frac{\partial}{\partial\beta'}m(y_{ir},w_{ir};\widehat{\beta})\|^2|x_r;\bar{\theta}_r,\widehat{\beta}]\cdot\mathrm{E}[\|S_{m_{ir}|x_r}(\bar{\theta}_r;\widehat{\beta})\|^2|x_r;\bar{\theta}_r,\widehat{\beta}])^{1/2}$ and the second by

$$\sqrt{\mathrm{E}\left[\left\|\frac{\partial}{\partial\beta'}m(y_{ir},w_{ir};\widehat{\beta})\right\|^2|x_r;\bar{\theta}_r,\widehat{\beta}\right]}\sqrt{\sqrt{\mathrm{E}\left[\left\|S_{m_{ir}|x_r}(\widehat{\beta};\bar{\theta}_r)\right\|^4|x_r;\bar{\theta}_r,\widehat{\beta}\right]}\cdot\sqrt{\mathrm{E}\left[\left\|S_{m_{ir}|x_r}(\bar{\theta}_r;\widehat{\beta})\right\|^4|x_r;\bar{\theta}_r,\widehat{\beta}\right]}}$$

and therfore $\left\|\dot{\Phi}_{ir}(\bar{\theta}_r;\widehat{\beta})\right\| \le 2L_{i,r}(x_r)$. By Lemma 2.B.1, $\left[\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\widehat{\beta};\theta_0^R)\right]'\widehat{W}\sqrt{nR}\widehat{\gamma}_{n,R}(\widehat{\beta};\theta_0^R) = o_{\mathrm{P}}(1)$.

Step 2: A term-by-term expansion for the jth term yields $\sqrt{nR}\widehat{\gamma}^j_{n,R}(\widehat{\beta};\theta_0^R) = \sqrt{nR}\widehat{\gamma}^j_{n,R}(\beta_0;\theta_0^R)+\sqrt{nR}\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\tilde{\beta};\theta_0^R)\left(\widehat{\beta}-\beta_0\right)$, so it remains to be seen that $\left\|\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\tilde{\beta};\theta_0^R) - M\right\| = o_{\mathrm{P}}(1)$, where $M = \lim_{n\to\infty}\mathbb{E}_{nR}\left[\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})\right]$, and that $\sqrt{nR}\widehat{\gamma}_{n,R}(\beta_0;\theta_0^R) \rightsquigarrow \mathcal{N}(0,\Omega)$, where $\Omega := \lim_{n\to\infty}\mathbb{E}_R\left[\mathrm{var}\left(\sqrt{n}\mathbb{E}_n\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})\right)\right]$. For the derivative,

$$\begin{aligned}\left\|\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\tilde{\beta};\theta_0^R) - M\right\| \le{}& \left\|\frac{\partial}{\partial\beta'}\widehat{\gamma}_{n,R}(\tilde{\beta};\theta_0^R) - \mathbb{E}_{nR}\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\tilde{\beta},\theta_{0r})\right\| + \left\|\mathbb{E}_{nR}\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\tilde{\beta},\theta_{0r}) - \lim_{n\to\infty}\mathbb{E}_{nR}\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\tilde{\beta},\theta_{0r})\right\| \\ &+ \left\|\lim_{n\to\infty}\mathbb{E}_{nR}\mathrm{E}\frac{\partial}{\partial\beta'}\mathcal{E}_{ir}(x_r;\tilde{\beta},\theta_{0r}) - M\right\|.\end{aligned}$$

The first term is $o_{\mathrm{P}}(1)$ by a uniform law of large numbers for the derivative which exists by Assumption 3.4 and the second term is $o_{\mathrm{P}}(1)$ by as the limit exists uniformly over $\mathcal{B}$ by Assumption 3.2. The final term is $o_{\mathrm{P}}(1)$ as $\tilde{\beta}-\beta_0 = o_{\mathrm{P}}(1)$ by continuity.

To show $\sqrt{nR}\widehat{\gamma}_{n,R}(\beta_0;\theta_0^R) \rightsquigarrow \mathcal{N}(0,\Omega)$, by Assumption 3.5 we may apply Lemma 2.B.5 with $m(X_{ir};\beta_0) = Y_{ir}$, $\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r}) = Z_{ir}$, and $x_r = X_r$ in the notation of the lemma. It is clear that $\Omega := \lim_{n\to\infty}\mathbb{E}_R\left[\mathrm{var}\left(\sqrt{n}\mathbb{E}_n\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})\right)\right]$, exists under these conditions since for every r, $\lambda_{\max}\left(\mathrm{var}\left(\sqrt{n}\mathbb{E}_n m(X_{ir};\beta_0)\right)\right) \ge \lambda_{\max}\left(\mathrm{var}\left(\sqrt{n}\mathbb{E}_n\mathcal{E}_{ir}(x_r;\beta_0,\theta_{0r})\right)\right)$.

□

LEMMA 2.B.5. *Let $Y_{ir,R}$ be a triangular array of mean-zero random variables, $Z_{ir,R} := \mathrm{E}[Y_{ir,R}|X_r]$, and define*

$$\lambda_r := \sqrt{n}\mathbb{E}_n[Y_{ir}] \text{ and } \zeta_r := \sqrt{n}\mathbb{E}_n[Z_{ir}].$$

If (1) $\sup_R\sup_r\mathrm{var}(\lambda_r) < C_1$ and (2) $\inf_R\inf_r\mathrm{var}(\zeta_r) > C_0$, then $\sqrt{nR}\mathbb{E}_{n,R}[Z_{i,r}] \rightsquigarrow \mathcal{N}(0,V_Z)$.

Proof. Identical to the proof of step 1 in "Proof that Lemma 11 of (Chernozhukov, Fernandez-Val, and Melly 2009) applies".

□

Finally, we turn to GMM with an index, which clearly nests GMM with $\mathcal{U} = \{u\}$. The argument is nearly identical to the above, so we only focus on the difference below. The key distinction is that we need to check conditions such that a functional central limit theorem applies. This setup allows the moments to be not differentiable everywhere, per se, in the parameter, though their expectations indeed will be smooth. Therefore, we need to control the complexity of the function class. The usual argument in the iid case would make use of P-Donskerity of function classes, though

in our inid case we use Theorem 2.11.1 of (Van der Vaart and Wellner 1996), which is analogous. Let us put $\sqrt{n}$-sums

$$\zeta_{Rr}(\beta, u) = \sqrt{n}\mathbb{E}_n \mathcal{E}_{ir}(\beta, u)/\sqrt{n} \text{ and } \lambda_{Rr}(\beta, u) = \sqrt{n}\mathbb{E}_n m_{ir}(\beta, u)/\sqrt{n}$$

where $\mathcal{E}_{ir}(\beta, u)$ denotes $\mathcal{E}_{ir}(\beta, u, \theta_{0r})$. Our estimator is a root of $\widehat{\Phi}(\cdot, u, \widehat{\theta})$, $\widehat{\Phi}(\widehat{\beta}(u), u, \widehat{\theta}) := \mathbb{E}_{n,R}[\mathcal{E}_{ir}(\widehat{\beta}(u), u, \widehat{\theta}_r)] = 0$. Define $\Psi(\beta, u) := \lim_{R\to\infty} \mathbb{E}_R \mathrm{E}[\zeta_r(\beta, u)/\sqrt{n}] = \lim_{R\to\infty} \mathbb{E}_{nR}\mathrm{E}[\mathcal{E}_{ir}(\beta, u)]$ and $\widehat{\Psi}(\beta, u) = \mathbb{E}_{n,R}[\mathcal{E}_{ir}(\beta, u)]$. By definition $\Psi(\beta_0(u), u) = 0$. Under an identical argument presented above, using Lemma 2.B.1, we can uniformly write

$$\widehat{\Phi}(\widehat{\beta}(u), u, \widehat{\theta}) = \widehat{\Psi}(\widehat{\beta}(u), u) + o_{\mathrm{P}}(1).$$

Therefore, it sufficies to consider the behavior of the root of $\widehat{\Psi}(\cdot, u)$, which will converge in probability to the root of $\widehat{\Phi}(\cdot, u)$. Our proof focuses on the differences in the indexed case. The empirical estimator satisfies

$$\widehat{\Psi}(\widehat{\beta}(u), u) = \mathbb{E}_{n,R}[\mathcal{E}_{ir}(\widehat{\beta}(u), u)] = o_{\mathrm{P}}(1).$$

If we can apply Lemma 11 of (Chernozhukov, Fernandez-Val, and Melly 2009), we are done. It will be useful to define a function class

$$\mathcal{F}_R := \left\{\zeta_{Rr}(\beta, u) = \sqrt{n}\mathbb{E}_n[\mathcal{E}_{ir}(x_r; \beta, u)] : (\beta, u) \in \mathcal{B} \times \mathcal{U}\right\}.$$

Proof that Lemma 11 of (Chernozhukov, Fernandez-Val, and Melly 2009) applies. We need to check the three conditions of the lemma. Conditions 1 and 2 are straightforward to verify and have been mostly assumed in Assumption 4. To see that condition 3, $\sqrt{nR}\left(\widehat{\Psi} - \Psi\right) \rightsquigarrow \mathcal{Z}$ in $\ell^\infty(\mathcal{B}\times\mathcal{U})$, holds we apply Theorem 2.11.1 of (Van der Vaart and Wellner 1996) which allows a functional central limit theorem for the inid case with triangular arrays. This proceeds in three steps.

Step 1: To show $\mathbb{E}_R\mathrm{E}\left(\|\zeta_{Rr}(\beta, u)\|^2_{\mathcal{F}_R} \cdot \mathbf{1}\left\{\|\zeta_{Rr}(\beta, u)\|_{\mathcal{F}_R} > \eta\sqrt{R}\right\}\right) \to 0$ for every $\eta > 0$. Define

$$Z_R := \mathbb{E}_R \|\zeta_{Rr}(\beta, u)\|^2_{\mathcal{F}_R} \cdot \mathbf{1}\left\{\|\zeta_{Rr}(\beta, u)\|_{\mathcal{F}_R} > \eta\sqrt{R}\right\}$$

and show that $Z_R \xrightarrow{\mathrm{P}} 0$. A summand is non-zero only if $\|\zeta_{Rr}(\beta, u)\|_{\mathcal{F}_R} > \eta\sqrt{R}$. By Chebyshev's inequality

$$\max_{r\le R} \mathrm{P}\left(\|\zeta_{Rr}(\beta, u)\|^{2+\delta}_{\mathcal{F}_R} > \eta^{2+\delta}R^{1+\delta/2}\right) \le \frac{\max_{r\le R}\mathrm{E}\|\zeta_{Rr}(\beta, u)\|^{2+\delta}_{\mathcal{F}_R}}{\eta^{2+\delta}R^{1+\delta/2}}.$$

Observe $\|\zeta_{rR}\|_{\mathcal{F}_R} \le \sqrt{n}\mathbb{E}_n L_{i,r} =: L_{r,R}$. Under Assumption 4.3, as $\sup_{R\ge 1}\max_{r\le R}\mathrm{E}\left(\sqrt{n}\mathbb{E}_n L_{i,r}(x_r)\right)^{2+\delta} = O(1)$,

$$\sup_{r\le R}\mathrm{P}\left(\|\zeta_{Rr}\|_{\mathcal{F}_R} > \eta\sqrt{R}\right) \le \frac{\sup_{r\le R}\mathrm{E}\,|L_{r,R}|^{2+\delta}}{\eta^{2+\delta}R^{1+\delta/2}} = O\left(R^{-1-\delta/2}\right).$$

All summands are zero with probability at least $\left(1 - cR^{-1-\delta/2}\right)^R \to 1$, so $Z_R \xrightarrow{\mathrm{P}} 0$. Next, $Z_R \le \mathbb{E}_R L^2_{r,R} =: \bar{L}^2_R$. Since the envelope $\bar{L}^2_R$ is uniformly integrable, by Vitali convergence the result follows.

Step 2: Show $\sup_{\rho((\beta,u),(\beta',u'))<\delta_R}\mathbb{E}_R\mathrm{E}\left[\zeta_{Rr}(\beta, u) - \zeta_{Rr}(\beta', u')\right]^2 \to 0$ for every $\delta_R \downarrow 0$. Notice $\|(\beta, u) - (\beta', u')\| \to 0$ implies $\zeta_{Rr}(\beta, u) \xrightarrow{a.s.} \zeta_{Rr}(\beta', u')$, from Assumption 4.1. Next, apply the extended Vitali convergence theorem since $\lim(nR)^{-1}\sum_r \mathrm{E}(2\sum_i L_{i,r}(x_r))^2 < \infty$, meaning that $\mathbb{E}_R\mathrm{E}\left[\zeta_{Rr}(\beta, u) - \zeta_{Rr}(\beta', u')\right]^2 \to 0$ as $\delta_R \to 0$.

Step 3: To show that $\int_0^{\delta_R}\sqrt{\log N(\epsilon, \mathcal{F}_R, d_n)}d\epsilon \xrightarrow{\mathrm{P}} 0$ for every $\delta_R \downarrow 0$. First, $\mathcal{F}_R$ consists of measurable functions by 1 so we have

$$\int_0^{\delta_R}\sqrt{\log N(\epsilon, \mathcal{F}_R, d_n)}d\epsilon \le \int_0^\infty \sup_{Q\in\mathcal{Q}}\sqrt{\log N\left(\epsilon\|F_R\|_{Q,2}, \mathcal{F}_R, L_2(Q)\right)}d\epsilon$$

which follows from[24]

$$\int_0^{\delta_R} \sqrt{\log N\left(\epsilon, \mathcal{F}_R, d_n\right)} d\epsilon \leq \int_0^{\delta_R / \|F_R\|_{\mu_R}} \sqrt{\log N\left(\epsilon \|F_R\|_{Q,2}, \mathcal{F}_R, L_2(Q)\right)} d\epsilon \|F_R\|_{\mu_R} \leq \int_0^{\infty} \sup_{Q \in \mathcal{Q}} \sqrt{\log N\left(\epsilon \|F_R\|_{Q,2}, \mathcal{F}_R, L_2(Q)\right)} d\epsilon$$

by the standard argument ((Van der Vaart and Wellner 1996), Proof of Lemma 2.11.6). The strategy is to show show that the uniform entropy integral of $\mathcal{F}_R$ can be controlled by the uniform entropy integral of $\mathcal{F}_R^*$, the unconditional moment functions, added with the uniform entropy integral of $\sup_{\beta', u'} \mathcal{H}_{R|\beta',u'}$, functions which characterize how the probability distribution changes as (β, u) change.

To control $\mathcal{F}_R^*$, we begin by noting

$$\begin{aligned}
\left\|\zeta_{Rr}(\beta, u) - \zeta_{Rr}(\beta', u')\right\|_{Q,2} &= \sqrt{n} \left\|\mathbb{E}_n \mathrm{E}\left[m(X_{ir}; \beta, u) | x_r; \beta, u\right] - \mathbb{E}_n \mathrm{E}\left[m(X_{ir}; \beta', u') | x_r; \beta', u'\right]\right\|_{Q,2} \\
&\leq_{(i)} \sqrt{n} \left\|\mathrm{E}\left[\mathbb{E}_n\left\{m(X_{ir}; \beta, u) - m(X_{ir}; \beta', u')\right\} | x_r; \beta, u\right]\right\|_{Q,2} \\
&\quad + \sqrt{n} \left\|\mathbb{E}_n \mathrm{E}\left[m(X_{ir}; \beta', u') | x; \beta, u\right] - \mathbb{E}_n \mathrm{E}\left[m(X_{ir}; \beta', u') | x; \beta', u'\right]\right\|_{Q,2}
\end{aligned}$$

by the triangle inequality. Applying Jensen's inequality to the first term,

$$\left\|\mathrm{E}\left[\lambda_r(\beta, u) - \lambda_r(\beta', u') | x_r; \beta, u\right]\right\|_{Q,2} \leq \mathrm{E}\left[\left\|\lambda_r(\beta, u) - \lambda_r(\beta', u')\right\|_{Q,2} | x_r; \beta, u\right] \leq \left\|\lambda_r(\beta, u) - \lambda_r(\beta', u')\right\|_{Q,2}$$

since Q is a measure over X. By Assumption 4.4, $\int_0^{\infty} \sup_{Q \in \mathcal{Q}} \sqrt{\log N\left(\epsilon \|F_R^*\|_{Q,2}, \mathcal{F}_R^*, L_2(Q)\right)} d\epsilon < \infty$. Next, we control the second term in the triangle inequality (i) above. These are functions of the form $h(x_r; \beta, u)$, members of $\mathcal{H}_{R|\beta',u'}$ and it has envelope $\bar{H}(x_r) = \sqrt{n}\mathbb{E}_n L_{ir}(x_r)$ by Assumption 4.3. By Assumption 4.4, we have $\int_0^{\infty} \sup_{Q \in \mathcal{Q}} \sqrt{\log N(\epsilon \left\|\bar{H}\right\|_{Q,2}, \mathcal{H}_R, L_2(Q))} d\epsilon < \infty$. Combining the above steps, the result follows as we can bound $\int_0^{\infty} \sup_{Q \in \mathcal{Q}} \sqrt{\log N\left(\epsilon \|F_R\|_{Q,2}, \mathcal{F}_R, L_2(Q)\right)} d\epsilon$ by

$$\int_0^{\infty} \sup_{Q \in \mathcal{Q}} \sqrt{\log N\left(\epsilon \left\|F_R^*\right\|_{Q,2}, \mathcal{F}_R^*, L_2(Q)\right)} d\epsilon + \sup_{(\beta', u') \in \mathcal{B} \times \mathcal{U}} \int_0^{\infty} \sup_{Q \in \mathcal{Q}} \sqrt{\log N(\epsilon \left\|\bar{H}\right\|_{Q,2}, \mathcal{H}_{R|\beta',u'}, L_2(Q))} d\epsilon < \infty.$$

□

2.B.3 Conditional Edge Independent Models

Proof of Proposition 2.4.1. We check the conditions of Lemma 2.3.1. The first condition is clear by definition and Assumption 5. Condition 2 is shown in Lemma 2.B.7. We show directly $\sup_r \left||\Xi|^{-1/2} \mathbb{E}_\Xi[v(X_{rs}; \theta_{0r})]\right| = O_{\mathrm{P}}(R^{1/b})$ in Lemma 2.B.6, which is what condition (2) is used for in the proof of Lemma 2.3.1. The Lipschitz condition follows from a secondary expansion and the assumption of the existence of an envelope function. That is,

$$|\Xi|^{-1} \sum_{s \in \Xi} \{\nabla_\theta v(X_{rs}; \theta_r^*) - \nabla_\theta v(X_{rs}; \theta_{0r})\} = |\Xi|^{-1} \sum_{s \in \Xi} F_{rs} \left(I \otimes (\theta_r^* - \theta_{0r})\right)$$

where F_{rs} is a conformable matrix of derivatives evaluated at an intermediate value and $F_{rs} \leq B$, which can be done by Assumption 5.2. The Hessian condition hold by the envelope condition and Chebyshev's inequality. □

LEMMA 2.B.6. *Under Assumption 5.3,* $\sup_r \left|\frac{1}{\sqrt{|\Xi|}} \sum_{s \in \Xi} v(X_{rs}; \theta_{0r})\right| = O_{\mathrm{P}}(R^{1/b})$.

Proof. Let $T := |\Xi|$ and define $y_{rT} := \sum v(X_{rs}; \theta_{0r})/\sqrt{T}$. We assume y_{rT} has b^{th} moments, $\sup_{r \leq R} \mathrm{E}\|y_{rT}\|^b < \infty$, which will be shown below. That $\left(\mathrm{E}\|\sup_r y_{rT}\|^b\right)^{1/b} \leq R^{1/b} \sup_r \left(\mathrm{E}\|y_{rT}\|^b\right)^{1/b}$ implies $\sup_r y_{rT} = O_{\mathrm{P}}(R^{1/b})$. Notice $\mathrm{E}\left[\|y_{rT}\|^b\right] = \mathrm{E}\left[\left\|\sum_s v(X_{rs}; \theta_{0r})\right\|^b\right] T^{-b/2}$. We need only control $\mathrm{E}\left[\left\|\sum_s v(X_{rs}, \theta_{0r})\right\|^b\right] \leq \mathrm{E}\left[\left(\sum_s B(X_{rs})\right)^b\right]$.

[24] We have $\|F_R\|_{\mu_R} > \delta_R$ if the covering number is $1/\epsilon$, so the integrand is defined as zero for $\epsilon > 1$.

Observe $\mathrm{E}\left[\left(\sum_{s\in\Xi}B(X_{rs})\right)^b\right] = \mathrm{E}\left[\sum B(X_{rs})^b\right] + \mathrm{E}\left[\sum\prod_j B(X_{rs_j})^{\gamma_j}\right]$ where $\sum\gamma_j = b$. If 2^{b-1} moments exist for the envelopes, then this can be majorized into terms of $\prod \mathrm{E}[B(X_{rs_j})^{\delta_j}]$ where $\delta_j \leq 2^{b-1}$ by repeated application of Holder inequalities. □

LEMMA 2.B.7. *Under Assumptions 5 and 6,* $\mathrm{P}\left(\sup_r \sup_{\theta\in\Theta}\left|\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)\right| \geq \eta\right) = o\left(|\Xi|^{-1}\right)$.

Proof. The argument is along the lines of Lemma 2 of (Hall and Horowitz 1996) and Lemma 3 of Supplementary Appendix I of (Hahn and Newey 2004). We use Lemma 2.B.8. First we use a union bound over the R graphs and focus on $\sum_{r=1}^R \mathrm{P}\left(\sup_{\theta\in\Theta}\left|\widehat{Q}_{(r)}(\theta)(\theta) - Q_{(r)}(\theta)\right| \geq \eta\right)$. Next, considering a given graph, we choose $\epsilon > 0$ such that $2\epsilon \cdot \sup_r \mathrm{E}[B(X_{rs})] < \frac{\eta}{3}$. Divide Θ into subsets $\Theta_1, ..., \Theta_{M(\epsilon)}$ such that $\|\theta - \theta'\| \leq \epsilon$ when θ and θ' are in the same subset. A second union bound gives us $\sum_{j=1}^{M(\epsilon)} \mathrm{P}\left(\sup_{\theta\in\Theta_j}\left|\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)\right| \geq \eta\right)$.

Let θ_j denote a point in Θ_j. Noticing that,

$$\begin{aligned}\left|\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)\right| &< \left|\widehat{Q}_{(r)}(\theta_j) - Q_{(r)}(\theta_j)\right| + \left|\widehat{Q}_{(r)}(\theta) - \widehat{Q}_{(r)}(\theta_j) - Q_{(r)}(\theta) + Q_{(r)}(\theta_j)\right| \\ &\leq \left|\widehat{Q}_{(r)}(\theta_j) - Q_{(r)}(\theta_j)\right| + \frac{\epsilon}{|\Xi|}\left|\sum B(X_{rs}) - \mathrm{E}[B(X_{rs})]\right| + 2\epsilon\mathrm{E}[B(X_{rs})], \text{ and}\end{aligned}$$

$$\mathrm{P}\left(\sup_{\theta\in\Theta_j}\left|\widehat{Q}_{(r)}(\theta) - Q_{(r)}(\theta)\right| \geq \eta\right) \leq \mathrm{P}\left(\left|\widehat{Q}_{(r)}(\theta_j) - Q_{(r)}(\theta_j)\right| \geq \frac{\eta}{3}\right) + \mathrm{P}\left(|\Xi|^{-1}\left|\sum_{s\in\Xi}(B(X_{sr}) - \mathrm{E}B(X_{sr}))\right| \geq \frac{\eta}{3\epsilon}\right) = o(|\Xi|^{-k})$$

by Lemma 2.B.8, the result holds as $R = o(|\Xi|^k)$. □

LEMMA 2.B.8. *For each r, suppose $\{X_{sr} : s \in \Xi\}$ be covariates satisfying Assumption 6. Let $R = O\left(|\Xi|^h\right)$, and let h, k, p, γ, with k, p, γ defined below, satisfy $h + 1 \leq k < p/2 - \gamma pd$. Then $\forall\eta > 0$,*

$$\max_r \mathrm{P}\left(\left||\Xi|^{-1}\sum_{s\in\Xi}X_{sr}\right| > \eta\right) = o\left(|\Xi|^{-k}\right) \text{ and } \mathrm{P}\left(\max_r\left||\Xi|^{-1}\sum_{s\in\Xi}X_{sr}\right| > \eta\right) = o\left(|\Xi|^{-1}\right).$$

Proof. The argument follows Lemma 1 of (Hall and Horowitz 1996) and Lemma 2 of Supplementary Appendix I of (Hahn and Newey 2004).

Step 1: By Chebyshev's inequality

$$\mathrm{P}\left(\left||\Xi|^{-1}\sum_{s\in\Xi}X_{sr}\right| > \eta\right) \leq \frac{C\cdot\mathrm{E}\|X_{1r}\|^{p+\delta}\left(|\Xi|^{p/2}m^{dp} + |\Xi|^p\alpha_{2,\infty}^r(m)^{\frac{\delta}{p+\delta}}\right)}{\eta^p|\Xi|^p},$$

for $1 \leq m \leq C(p)|\Xi|$ where the second inequality follows from Lemma 2.B.9, which we can write under Assumption 6.[25] We can bound the right hand side by $\eta^{-p}C\cdot\mathrm{E}\|X_{1r}\|^{p+\delta}m^{dp}|\Xi|^{-p/2} + \eta^{-p}C\cdot\mathrm{E}\|X_{1r}\|^{p+\delta}\alpha_{2,\infty}^r(m)^{\frac{\delta}{p+\delta}}$ and for $m = |\Xi|^\gamma$ for some γ with $0 < \gamma \leq 1$, using the bound $\sup_r \alpha_{\infty,1}^r(m) \leq Ca^m$ on the mixing coefficient,

$$\max_r|\Xi|^k\,\mathrm{P}\left(\left||\Xi|^{-1}\sum_{s\in\Xi}X_{sr}\right| > \eta\right) \leq \eta^{-p}C\cdot\max_r\mathrm{E}\|X_{1r}\|^{p+\delta}\left(|\Xi|^{\gamma pd+k-p/2} + |\Xi|^k a^{\frac{\delta}{p+\delta}|\Xi|^\gamma}\right).$$

We have shown that for Assumption 6, $\max_r|\Xi|^k\,\mathrm{P}\left(\left|\frac{1}{|\Xi|}\sum_{s\in\Xi}X_{sr}\right| > \eta\right) = O\left(|\Xi|^{\gamma pd+k-p/2}\right) = o(1)$ if $\gamma pd + k < p/2$ which is the first result.

Step 2: By a union bound $|\Xi|\,\mathrm{P}\left(\max_r\left|\frac{1}{|\Xi|}\sum_{s\in\Xi}X_{sr}\right| > \eta\right) \leq |\Xi|\,R\cdot\max_r\mathrm{P}\left(\left|\frac{1}{|\Xi|}\sum_{s\in\Xi}X_{sr}\right| > \eta\right)$. If $R = O\left(|\Xi|^h\right)$, it follows that $|\Xi|\,\mathrm{P}\left(\max_r\left|\frac{1}{|\Xi|}\sum_{s\in\Xi}X_{sr}\right| > \eta\right) \leq O\left(|\Xi|^{h+1}\right)o\left(|\Xi|^{-k}\right)$. Since $h+1 \leq k$, $O\left(|\Xi|^{h+1}\right)o\left(|\Xi|^{-k}\right) =$

[25] We write the proof for $\sup_r \alpha_{\infty,2}^r(m) < Ca^m$, though the extension to $\sup_r \alpha_{\infty,2}^r(m) = o\left(m^{-d}\right)$ is straightforward and will merely result in more stringent requirements on r, k, γ, d. For Assumption 6(ii) we have

$$\max_r|\Lambda|^k\,\mathrm{P}\left(\left|\frac{1}{|\Xi|}\sum_{s\in\Xi}X_{sr}\right| > \eta\right) = O\left(|\Xi|^{d\gamma p+k-p/2} + |\Xi|^{k-\gamma d-\gamma\epsilon}\right)$$

for $\alpha_{1,\infty}^r(m) = O\left(m^{-d-\epsilon}\right) = o\left(m^{-d}\right)$. Then the requirement is $k < (p/2 + \gamma pd) \vee (d + \epsilon)$.

$o(1)$, $\mathrm{P}\left(\max_r \left|\frac{1}{|\Xi|}\sum_{s\in\Xi} X_{sr}\right| > \eta\right) = o\left(|\Xi|^{-1}\right)$ which proves the result. □

LEMMA 2.B.9. *Let $\{X_i : t_i \in \Lambda \subset \mathbb{Z}^d\}$ be a mean zero stationary random field satisfying Assumption 6 and $\{X_{ij} : ij \in \Xi\}$ covariates. Then for any positive integer r and for $1 \leq m < C(k)\cdot|\Xi|$, we have* $\mathrm{E}\left[\left(\sum_{ij\in\Xi} X_{ij}\right)^k\right] \leq C(r)\,\mathrm{E}\|X_1\|^{k+\delta}\left(|\Xi|^{k/2} m^{kd} + |\Xi|^k \alpha_{2,\infty}(m)^{\frac{\delta}{k+\delta}}\right)$.

Proof. The proof builds on (Lahiri 1992), with two differences: the first is an extension to mixing random fields,[26] and the second is that we are interested over moments of random variables on Ξ as opposed to Λ. The (Lahiri 1992) style of argument proceeds in four parts; we include the entire argument for completeness though the key differences are in the last two steps. First, we can control the first $k/2$ terms via a standard result. This will enable us to bound this part of the sum by a $|\Xi|^{k/2}$ rate. Second, for the remaining terms, we will divide the space into a set of all pieces with a well separated point τ on the lattice whose random variable X_τ has power 1 and is at least of distance m from any other point in the collection, and into its complement. Third, we will control this set using the mixing coefficient and fourth, by a counting argument we create an upper bound on the number of points in the complement. It is useful to note that for $ij \in \Xi$, for $1 \in \Lambda$, $\mathrm{E}\|X_{ij}\|^k \lesssim \mathrm{E}\|X_1\|^k$ by the assumption and stationarity. Moreover, we defined the pseudo-metric $d_\Xi(ij,kl) := d_\Lambda(i,k) \wedge d_\Lambda(i,l) \wedge d_\Lambda(j,k) \wedge d_\Lambda(j,l)$ where $d_\Lambda(x,y) := \|x-y\|_1$.

Step 1: For $k = 2h$, we can expand the term into a polynomial,

$$\left(\sum_{s\in\Xi} X_s\right)^k = \sum_{j=1}^k \sum_{\alpha_1,\ldots,\alpha_j} c(\alpha_1,\ldots,\alpha_j) \sum_{s_1,\ldots,s_j} \prod_{t=1}^j X_{s_t}^{\alpha_t}$$

where $t = 1,\ldots,j$ is an arbitrary index of a j-tuple $(s_1,\ldots,s_j) \subset \Xi$ and $c(\cdot)$ are coefficients. We can control the first $k/2$-tuples by a standard argument, e.g., Bhattacharya and Rao (1986), making use of $\mathrm{E}\|X_{ij}\|^k \lesssim \mathrm{E}\|X_1\|^k$,

$$\left|\sum_{j=1}^{k/2} \sum_{\alpha_1,\ldots,\alpha_j} c(\alpha_1,\ldots,\alpha_j) \sum_{s_1,\ldots,s_j} \mathrm{E}\left[\prod_{t=1}^j X_{s_t}^{\alpha_t}\right]\right| \leq C(k)\,|\Xi|^{k/2}\,\mathrm{E}\|X_1\|^k.$$

In what follows it suffices to show for fixed $j > k/2$ and $(\alpha_1,\ldots,\alpha_j)$,

$$\left|\sum_{s_1,\ldots,s_j} \mathrm{E}\left[\prod_{t=1}^j X_{s_t}^{\alpha_t}\right]\right| \leq C(k)\,\mathrm{E}\|X_1\|^{k+\delta}\left(|\Xi|^{k/2} m^{kd} + |\Xi|^k \alpha_{2,\infty}(m)^{\frac{\delta}{k+\delta}}\right).$$

Step 2: Next, we create a set that counts the sites s_τ where X_{s_τ} has power $\alpha_\tau = 1$ and s_τ is sufficiently far from the other s_t in the j-tuple. Let $u := j - k/2$. We put $A := \{t : \alpha_t = 1\}$ as the set of all points that have coefficient 1. Then let $\beta_0 = |A|$. This is a set that counts the number of indices that show up exactly one time. We want to show that this set is non-empty. Note that $1 \leq u \leq k/2$. Also, since $k = \sum_{t=1}^j \alpha_t \geq \beta_0 + 2(j-\beta_0)$, this means $2u \leq \beta_0 \leq k$ and therefore $(j - k/2) \leq \beta_0 \leq k/2$.

Then we partition the set of all j-tuples into B_m and B_m^c, with In Lahiri's notation, $\sum_2 \equiv \sum_{s_1,\ldots,s_j} =: \sum_3 + \sum_4$. To define the sets, put

$$B_m := \{(s_1,\ldots,s_j) : \inf_{l\neq t} d_\Xi(s_l,s_t) = d_\Lambda(i_{s_l},k_{s_t}) \wedge d_\Lambda(j_{s_l},k_{s_t}) \wedge d_\Lambda(i_{s_l},l_{s_t}) \wedge d_\Lambda(i_{s_l},l_{s_t}) > m \text{ for some } t \in A\}.$$

Now $\sum_3$ sums over the terms in B_m and $\sum_4$ over the terms in B_m^c.

Step 3: We want to control B_m. Fix $\tau \in A$. Then decompose $\prod_{t=1}^j X_{s_t}^{\alpha_t} = X_a X_b$, with $X_b = X_{s_\tau}$ and $X_a = \prod_{t\neq\tau} X_{s_t}^{\alpha_t}$. Then X_a is a random field with respect to $\sigma\left(X_{s_1},\ldots,\widehat{X}_{s_\tau},\ldots,X_{s_j}\right)$, where we use notation to indicate the omission of a term, $\left(a,\widehat{b},c\right) := (a,c)$, and X_b with respect to $\sigma(X_{s_\tau})$. These are of size $j-1$ and 1, respectively. By definition the distance between these two sets is at least m, so by applying (Lin, Lin, and Lu 1996) and using the fact that $\mathrm{E}[X_{ij}X_{kl}] \leq \|X_{ij}\|_{L^p(\mathrm{P})} \cdot \|X_{kl}\|_{L^q(\mathrm{P})}\, \alpha_{2,2}^{1-q^{-1}-p^{-1}}(d_\Xi(ij,kl))$, we have

$$\left|\mathrm{E}\left[\prod_{t=1}^j X_{s_t}^{\alpha_t}\right]\right| \leq |\mathrm{E}[X_a X_b]| \leq C \cdot \|X_a\|_{L^p(\mathrm{P})} \cdot \|X_b\|_{L^q(\mathrm{P})}\, \alpha_{2(j-1),2}^{1\ q^{-1}-p^{-1}}(m).$$

[26] (Hahn and Kuersteiner 2004) extend the Lahiri argument to the time series setting.

Taking $p = k + \delta$ and $\frac{1}{q} = \frac{k-1}{k+\delta}$, using stationarity and repeatedly applying the Holder inequality,

$$\left|\mathrm{E}\left[\prod_{t=1}^{j} X_{s_t}^{\alpha_t}\right]\right| \leq C \cdot \|X_a\|_{L^{k+\delta}(\mathrm{P})} \cdot \|X_b\|_{L^{(k-1)/(k+\delta)}(\mathrm{P})} \alpha_{2(j-1),2}^{\delta/(k+\delta)}(m) \leq C \cdot \mathrm{E}\|X_1\|^{k+\delta} \alpha_{2(j-1),2}^{\delta/(k+\delta)}(m).$$

Since $\alpha_{j,1}(m) \leq \alpha_{j+1,1}(m)$, $\left|\mathrm{E}\left[\sum_{B_m} \prod_{t=1}^{j} X_{s_t}^{\alpha_t}\right]\right| \leq C \cdot \mathrm{E}\|X_1\|^{k+\delta} |\Xi|^k \alpha_{2,\infty}(m)^{\frac{\delta}{k+\delta}}$.

Step 4: Finally, we control B_m^c. To get a (coarse) upper bound, first notice the maximum number of powers of 1 that can be placed is $2u$[27]. Construct a set $\Gamma \subset \{1, ..., j\}$, $|\Gamma| = 2u$ which will include all powers of one and perhaps some residual copies of terms with higher power. Then we will simply count

$$\bar{B}_m^c := \left\{(s_1, ..., s_j) : \inf_{l \neq t} d_\Xi(s_t, s_l) = d_\Lambda(i_{s_l}, k_{s_t}) \wedge d_\Lambda(j_{s_l}, k_{s_t}) \wedge d_\Lambda(i_{s_l}, l_{s_t}) \wedge d_\Lambda(i_{s_l}, l_{s_t}) \leq m \ \forall t \in \Gamma\right\}.$$

It will help us to define the *partners* of s_t as $\epsilon(t) \in \arg\inf_l d_\Xi(s_t, s_l)$. Then define $\mathcal{P}(\Gamma)$ as the set of partners of $t \in \Gamma$ and put $v = |\Gamma \cup \mathcal{P}(\Gamma)|$. This is the number of distinct elements of Ξ with power one or power greater than one whose copies form the residual members of Γ and their partners. By definition $2u \leq v \leq 4u \wedge j$ since $v \geq |\Gamma|$ and $v \leq 2|\Gamma|$ if each had a distinct partner.

We define an *m-unbroken set* as a collection of points in $\Gamma \cup \mathcal{P}(\Gamma)$ for which each member is within an m-distance under pseudo-metric d_Ξ. Put q as the number of m-unbroken sets. First, observe that there are less than $|\Xi|^q$ initial sites to place a seed for each of the unbroken sets.[28] Next, we have to place each of the $v - q$ terms. For a given ij there are less than $2(2m+1)^d$ elements of Ξ within an m-distance since $d_\Lambda(k,i) \wedge d_\Lambda(l,i) \wedge d_\Lambda(k,j) \wedge d_\Lambda(l,j) \leq m$. An upper bound is $O\left(m^{(v-q)d}\right)$. Finally, we can arbitrarily place the remaining $j - v$ elements yielding $|\Xi|^{j-v}$. This gives us $\left|\bar{B}_m^c\right| \lesssim |\Xi|^{q+j-v} m^{(v-q)d}$. The same counting exercise as in Lahiri (1992) gives us $j + q - v \leq h$ and $v - q \leq k$ which yields $|\Xi|^{q+j-v} m^{(v-q)d} \leq |\Xi|^h m^{kd}$. This concludes the proof. □

2.B.4 Groups Model

The proof of Proposition 2.4.2 is a corollary to the following lemma. We need to replicate the arguments from section 2.B.2 replacing Lemma 2.B.1 by Lemma 2.B.10. The model satisfies Assumption (2) below by (Chatterjee, Diaconis, and Sly 2010).

LEMMA 2.B.10. *Assume (1) for all $\theta_r \in \Theta_r$, $\zeta_r(\bar{\theta}_r, u)$ are $q \times k_n q$ matrix (or vector) valued functions, $\sup_{u \in \mathcal{U}} \sup_r \|\zeta_r(\theta_r, u)\| \leq B_r$ with $\operatorname{plim} \mathbb{E}_{nR}[B_{ir}] < \infty$, (2) for each r, $\sup_{i \leq k_n} \left\|\widehat{\theta}_{ir} - \theta_{0ir}\right\| \leq C(L)\sqrt{\frac{\log n}{n}}$ with probability at least $1 - cn^{-2}$, where L, c depend only on $\Theta_r \subset \mathbb{R}^{k_n}$ and ψ, (3) $\Theta_r \subset [a,b]^{k_n}$ for all r, and (4) $R = O(n^\gamma)$ with $\gamma < 1$. Then*

$$\sup_{u \in \mathcal{U}} R^{1/2} \left\|\mathbb{E}_R \zeta_r(\bar{\theta}_r, u)(I_q \otimes (\widehat{\theta}_r - \theta_{0r}))\right\|_2 \lesssim \sqrt{\frac{R \cdot \log n}{n}}$$

with probability approaching one. Specifically the probability is at least $\left(1 - cn^{-2}\right)^{R_n}$.

Proof of Lemma 2.B.10. Step 1: This follows from $\|\cdot\|_2 \leq \|\cdot\|_1$ and

$$\begin{aligned}
\sup_{u \in \mathcal{U}} R^{1/2} \left\|\mathbb{E}_R \zeta_r(\bar{\theta}_r, u)(I_q \otimes (\widehat{\theta}_r - \theta_{0r}))\right\|_1 &\leq \sup_{u \in \mathcal{U}} R^{1/2} \mathbb{E}_R \left[\left\|\zeta_r(\bar{\theta}_r, u)\right\|_1 \left\|(I_q \otimes (\widehat{\theta}_r - \theta_{0r}))\right\|_\infty\right] \\
&\leq \sup_{u \in \mathcal{U}} R^{1/2} \mathbb{E}_R \left[\left\|\zeta_r(\bar{\theta}_r, u)\right\|_1 \left\|\widehat{\theta}_r - \theta_{0r}\right\|_\infty\right] \\
&\leq \sup_{u \in \mathcal{U}} \mathbb{E}_R \left\|\zeta_r(\bar{\theta}_r, u)\right\|_1 \cdot \sqrt{R} \sup_{r \leq R} \left\|\widehat{\theta}_r \quad \theta_{0r}\right\|_\infty.
\end{aligned}$$

We have that $\sup_{u \in \mathcal{U}} \frac{1}{\sqrt{q}} \left\|\zeta_r(\bar{\theta}_r, u)\right\|_1 \leq \sup_{u \in \mathcal{U}} \left\|\zeta_r(\bar{\theta}_r, u)\right\|_2 \leq B_r$ and therefore

$$\sup_{u \in \mathcal{U}} R^{1/2} \left\|\mathbb{E}_R \zeta_r(\bar{\theta}_r, u)(I_q \otimes (\widehat{\theta}_r - \theta_{0r}))\right\|_1 \lesssim \sqrt{R} \sup_{r \leq R} \left\|\widehat{\theta}_r - \theta_{0r}\right\|_\infty + o_{\mathrm{P}}(1)$$

[27] Let x be the maximal number. Then $k = x + 2(j - x)$. Solving this yields the result.

[28] We leave ourselves open to certain double-counting; the bound is coarse.

by assumption on the envelope functions. It suffices to control $\sup_{r\leq R}\left\|\widehat{\theta}_r-\theta_{0r}\right\|_\infty$.

Step 2: Observe that $\sup_{r\leq R}\left\|\widehat{\theta}_r-\theta_{0r}\right\|_\infty \leq C(L)\sqrt{\frac{\log n}{n}}$ with probability at least $(1-cn^{-2})^{R_n}$, since the bound holds for each of the R (independent) terms with probability at least $1-cn^{-2}$. This approaches one for $\gamma<1$. □

2.B.5 DeJong and Woutersen (2011) Model

Proof of Proposition 2.4.4. We check that the conditions of Lemma 2.3.1 hold. First, by example 2 of (Hahn and Kuersteiner 2004), using Theorem 2 of (De Jong and Woutersen 2010) the model satisfies Conditions 1-7 of (Hahn and Kuersteiner 2004). Notice $X_{s,r}=[A_{s,r},A_{s-1,r},\ldots,A_{s-k,r},z_{sr}]$ satisfies condition 3 of (Hahn and Kuersteiner 2004). It is straightforward to verify conditions 1-5 of Lemma 2.3.1 hold. Condition 1 follows from Theorem 3 of (De Jong and Woutersen 2010) and checking that the probit model is twice continuously differentiable in θ. By an argument identical to Lemma 2.B.7, condition 2 holds. The advantage here is that we can directly apply Lemma 1 of (Hahn and Kuersteiner 2004) instead of Lemma 2.B.9, since we essentially are dealing with time-series type data. Since

$$q(z_{sr},A_r^{s-1};\theta_{0r})=A_s\log\Phi(\sum\theta_{0r,k}A_{s-k,r}+\theta'_{0r,l+1}z_{sr})+(1-A_s)\log(1-\Phi(\sum\theta_{0r,k}A_{s-k,r}+\theta'_{0r,l+1}z_{sr})),$$

we have

$$v(z_{sr},A_r^{s-1};\theta_{0r})=A_s\frac{\phi(\sum\theta_{0r,k}A_{s-k,r}+\theta'_{0r,l+1}z_{sr})}{\Phi(\sum\theta_{0r,k}A_{s-k,r}+\theta'_{0r,l+1}z_{sr})}-(1-A_s)\frac{\phi(\sum\theta_{0r,k}A_{s-k,r}+\theta'_{0r,l+1}z_{sr})}{1-\Phi(\sum\theta_{0r,k}A_{s-k,r}+\theta'_{0r,l+1}z_{sr})}$$

which is bounded in order by $1+\left|\sum\theta_{0r,k}A_{s-k,r}+\theta'_{0r,l+1}z_{sr}\right|$ using Feller's inequality. Thus the envelope depends on X_{sr} only, which satisfies the mixing condition 3 of (Hahn and Kuersteiner 2004). By compactness of Θ_r and the moment assumption on z, we can apply the argument of Lemma 2.B.6, with $|\Xi|=\binom{n}{2}$ and $v(X_{sr},A_r^{s-1};\theta_r)$ as above to show condition 3 holds. Conditions 4 and 5, which are a Lipschitz condition on the Hessian and a pointwise LLN (uniform across R) on the Hessian hold because for a probit model all finite order of derivatives of $\log\Phi(\cdot)$ exist and are continuous, on a compact parameter space they satisfy the Lipschitz condition and are uniformly bounded. □

2.C Overview of Estimation Algorithm and Standard Errors

We present a non-technical overview of the standard errors and estimation procedures used. A theoretical discussion of standard error estimation is beyond the scope of this paper. Here, we discuss standard errors and finite-sample simulation bias. A lengthier formal, technical discussion and simulations results are available from the authors upon request. Several approaches can be used for inference: heteroskedasticity-robust, clustered, block bootstrapped, first-stage bootstrapped, and importance sampled standard errors. Network-level inference can be performed using heteroskedasticity-robust standard errors under cross-network independence, even in the presence of sampling. Individual-level inference under sampling requires attention to within-network, cross-individual autocorrelation exacerbated by the reconstructed regressor's mismeasurement, what we call the reconstruction error. When an edge is missing in the computation of one individual's network statistic, that edge is also missing for all other individuals' network statistic computation. Clustered standard errors and their nonparametric block bootstrap analog both account for heteroskedasticity and within-graph autocorrelation. Additionally, a parametric bootstrap which simulates the autocorrelation from the reconstruction error can estimate the contribution of the sampling-induced autocorrelation to the reconstruction estimator's variance. The parametric structure of this bootstrap should allow these standard errors to be estimated on a smaller collection of networks than necessary for clustered standard errors.

While the graphical reconstruction estimator is formally a "two-step" estimator, super-consistency allows the researcher to ignore the first-stage estimation uncertainty. That said, first-stage uncertainty can be understood using a first-stage parametric bootstrap and importance sampling. The first-stage parametric bootstrap estimates the distribution of the first-stage parameters and encompasses the following effects: sample from the observed collection of networks with replacement, estimate each network's first-stage parameters using the observed data, simulate a

new collection of observed graphs using the first-stage estimates, and estimate the first-stage parameters using the simulated graphs. Repeat this process to obtain the bootstrapped variance of the first stage estimators to determine whether super-consistency is justified. Importance sampling can reduce the computational burden of bootstrapping the first-stage estimates by using the first-stage model's likelihood to rebalance a collection of reconstructed network statistics and greatly reduces the cost of accounting for any first-stage variability.

With or without first-stage superconsistency, the reconstruction estimator still uses a finite number of simulations, leaving another error term which we call the simulation error. A regularization to estimate and subtract out the variance in the regressors due to this noise can be employed.

Chapter 3

Informal Insurance, Social Networks, and Savings Access: Evidence from a Framed Field Experiment

3.1 Introduction

Individuals in developing countries face large amounts of risk (weather, health, prices, etc.). Interpersonal insurance can potentially play an important role in buffering this risk, but evidence suggests that it is incomplete: not all idiosyncratic risk is insured. Interpersonal insurance will be incomplete if individuals who receive positive income shocks are tempted to renege on their obligation to share with those who receive negative shocks—that is, if there is limited commitment (Ligon et al. 2002).[1] The ability of a group of individuals to overcome this lack of commitment may depend on the strength of social ties between them. Individuals more tightly linked within a network may be able to better sustain cooperative behavior despite not having access to contracts or other formal enforcement mechanisms.[2]

The amount of consumption smoothing individuals achieve will also depend on access to financial instruments such as savings. However, the sign and magnitude of savings' effect are not clear: although savings allows individuals to smooth uninsured risk, it may also restrict interpersonal insurance. Insurance will be restricted if people can use savings even after they renege[3] on their

[1]This is not the only explanation for incomplete insurance, although evidence suggests it is important. Other explanations include moral hazard, adverse selection, and hidden income. We consider the possibility of hidden income in another paper (Chandrasekhar et al. 2011).

[2]This may be for a variety of reasons including, but not limited to, altruism or repeated interactions. The aim of this paper is not unbundling these forces. Leider et al. (2009) and Ligon and Schechter (forthcoming) addresses this issue.

[3]Throughout the paper we refer to players implementing a lower transfer than the one they initially planned to make as "reneging" or "defecting." We use this terminology because it is standard in limited commitment models. However, to mimic the informal nature of real-life informal insurance systems, during our experiment we emphasized

insurance obligations, since the ability to save makes leaving the insurance agreement more tempting when income is high. This "crowdout" effect reduces the benefit of savings access and can even mean that overall, consumption smoothing is worse with access to savings than without, if interpersonal insurance is crowded out more than one-for-one (Ligon et al. 2000).

When commitment is limited, both savings' benefits (smoothing uninsured risk) and costs (crowding out insurance) may vary with individuals' position in social networks. The dimension of social network position we consider in this paper is network distance to other individuals.[4] Social connectedness—that is, being close to others on average—and access to formal savings may be either complements or substitutes. If savings access crowds out insurance unless social proximity is high enough to limit crowdout, then savings access will benefit those with high social connectedness (low average social distance) most. If, on the other hand, savings crowds out risk sharing among highly connected individuals while helping those with few social connections smooth risk intertemporally, saving will benefit those with low social capital most.

To shed light on these issues, we ran a framed field experiment in Indian villages, in which participants played risk-sharing games, each with a randomly assigned partner from the same village. Our design allowed us to switch features of the economic environment on and off: the ability to commit to contracts and whether individuals had access to a savings technology. Using detailed social network[5] data collected prior to the experiment, we are also able to assign players to play the games with partners of different social distance.[6] Players faced significant stakes and could not use transfers outside the game to smooth in-game risk[7], so the behavior we observe in the lab is informative about risk-sharing behavior outside the lab. By comparing how much risk sharing occurs between socially close vs. socially distant pairs, we can understand the role of social distance in enabling risk sharing. What's more, we can measure the extent to which the ability to commit to contracts and the ability to save affect the ability to share risk, and how these features interact with social proximity.

The risk-sharing games were structured as follows: in each round, one partner was randomly chosen to receive a large positive income (INR 250) while the other partner received no income (INR 0).[8] In every round, before income was realized, individuals decided on state-contingent transfers

---

to players the transfers they initially announced were not binding, and that they could change their mind after seeing their income draw. We present evidence below that players did not view these announcements as cheap talk.

[4]The social distance between i and j is given by the geodesic, or the shortest path between them through the social network. So, for example, if i is connected to k who is connected to j, the social distance between i and j is 2.

[5]The social network is an abstraction which describes a collection of relationships between individuals, in this case inhabitants of a village.

[6]We do not manipulate the social distance between a given pair. Instead, we assign randomly-selected pairs to interact in the experiment.

[7]While we cannot prevent transfers outside the lab, in section 3.3.2 we discuss why such transfers would not be effective in smoothing risk in this setting.

[8]Rs 250 is approximately \$5 at market exchange rates, or \$25 at purchasing power party-adjusted exchange rates. For comparison, the daily wage in Karnataka paid by the National Rural Employment Guarantee Act (NREGA) is Rs. 74.

to their partner for that round. The transfers allowed participants to smooth their consumption in the face of income risk. Every individual played all three versions of the game. In one version, individuals had access to full commitment contracts without savings, meaning that in each round an individual had to make the transfer that was decided on before the realization of income. However, the individuals had no ability to save in this version of the game. In two other versions of the game, we allowed for limited commitment instead. After seeing their realized income, participants had the opportunity to renege on the transfers they initially decided. The limited commitment games were played in two ways: without savings and with savings. We implemented the limited commitment with savings treatment by allowing participants to save their income across rounds, in addition to being able to make transfers with their partners.[9] This treatment allowed us to capture the interaction of limited commitment and access to savings. Across all versions, players had an incentive to smooth consumption because they knew that they would be paid their chosen consumption from just one round, randomly chosen from all the rounds they played.

In brief, our results are as follows. First, limited commitment matters: consumption is 22% more variable when players cannot commit ex ante to a risk-sharing agreement than when they can. Second, this effect varies with social distance in a way which shows that social proximity substitutes for commitment. For the socially closest pairs, limited commitment does not bind, i.e., consumption variability does not increase when commitment is removed. But, as social distance increases, limited commitment is increasingly important. Each unit of social distance causes an increase in the variability of consumption equal to roughly 7% of the full commitment level when commitment is removed. The most distant pairs, who are unconnected via the social network, see consumption variability increase by an amount equal to 80% of the full commitment level when commitment is removed.

Our third finding is that savings access does not appear to crowd out informal insurance. That is, transfers do not fall significantly when savings are available. However, players paired with a more distant partner use savings more than those with a socially close partner. This is because limited commitment leaves distantly-connected pairs with more risk that is not insured interpersonally, and savings are used to partially smooth this risk. Finally, we find that savings access improves welfare:[10] it allows individuals to intertemporally smooth some of the income risk that is not insured interpersonally, so individuals achieve greater consumption smoothing when savings are available.

These findings are relevant for several reasons. From a policy perspective, understanding if and how network relationships mitigate market incompleteness can inform attempts to harness these relationships through, e.g., microfinance (as in as in Feigenberg et al. 2011) or community-allocated aid (as in Alatas et al. 2012). Furthermore, access to formal savings and other banking

[9]Savings were observable by the partner. There is no private information in any of the games. We consider the issue of private information in another set of experiments, described in Chandrasekhar et al. (2011)

[10]Due to the fact that our experimental setup keeps expected consumption (nearly) constant across models, consumption smoothing can be used as a measure of welfare.

products is growing rapidly in developing countries, although the current level is quite low (Chaia et al. 2009). Therefore, it is important to understand the extent to which financial access affects informal insurance: if financial development undermines informal insurance, then better savings access should be coupled with improved formal safety nets, such as work guarantee programs or food aid. Moreover, our results inform the question of *who* benefits most from access to formal savings when commitment is absent: does savings benefit individuals whose income turns out to be high while hurting those whose income turns out to be low, or affect the lucky and unlucky equally? Does savings most benefit those with dense social networks or those who are socially isolated—or impact both groups equally? The answers to these questions are important for understanding how financial development will affect various dimensions of inequality.

To be sure, our experiment remains an abstraction from reality, and we do not claim that the precise magnitudes of the effects we identify would translate exactly to another setting. Our goal is to disentangle the roles of limited commitment, access to savings, and social distance. This allows us to study the relative effect of networks in mitigating limited commitment and mediating the effect of savings. The sign and significance of these effects, if not their precise magnitudes, are likely to generalize, particularly to other rural, developing country settings. Even real-world risk-sharing relationships are not simple objects in a dynamic risk-sharing contract, but are embedded in a broader set of repeated interactions and norms. By measuring the effect of social distance in our experiment, we can say something about how real-world risk sharing is sustained.

In our experiment we deliberately moved away from anonymity by going from a laboratory experiment to a framed field experiment.[11] Methodologically, this allows our results to shed light on the importance of accounting for social networks in laboratory settings. Our findings across social distance provide direct evidence against the standard exchangeability of actors assumed in many economic models. If we had used an anonymous experiment, we would likely have obtained misleading estimates of the size of the effects we study. Depending on how anonymous pairs interact vis-a-vis as socially distant pairs, an anonymous experiment may have overstated the role of limited commitment, since among non-anonymous pairs it is mitigated by social distance.[12] Also, an anonymous experiment may have mis-stated the usage and benefits of savings access, since savings usage is concentrated among socially distant pairs. These findings are relevant as laboratory and framed field experiments become increasingly common in development, labor, contract economics and other fields.

[11] We follow Harrison and List (2004), who define a framed field experiment as a laboratory experiment conducted with a nonstandard population (i.e., not university students) with "field context in either the commodity, task, or information set that the subjects can use" (p 1014). In our case, subjects can use information about the partners they are played with, as well as their experience with day-to-day risk sharing.

[12] Anonymous interactions may not be identical to those with a social distance= ∞. Players might instead believe they face a partner with average social distance. Whether the total average bias of anonymity is positive or negative depends on the slope of social distance and the experimental pool over which we integrate. Regardless, an anonymous experiment would not provide correct inference about the importance of limited commitment.

The causal interpretation of our findings relies on the ability to carefully control the economic environment and partners' social distance while holding the income process constant. None of this would be possible without conducting an experiment in a laboratory setting. Natural and field experiments can address the fact that access to savings is potentially correlated with many other factors that affect the sustainability of informal insurance, such as migration opportunities, wealth, social networks, or the nature of the income process. However, even exogenous variation in availability of, e.g., banks arising from a natural or field experiment would not isolate the effect of savings access that works through risk-sharing, because savings access may also allow investment, changing the income process.[13] Also, natural and field experiments are not able to manipulate the social distance of risk-sharing partners, so studying the causal effect of social proximity is not possible. In a lab setting, we are able to observe the same individual paired with partners of differing distance, and this allows us to distinguish the effect of social distance from the fact that those with more social ties may also be more cooperative, altruistic, etc.

Our results contribute to a recent literature which uses laboratory or framed experimental settings to answer questions relating to financial development. One part of this literature aims, as we do, to test predictions of the limited commitment model. Barr and Genicot (2008) and Barr, Dekker and Fafchamps (2008) use framed field settings in Zimbabwe, while Charness and Genicot (2009) use a university laboratory setting to test the limited commitment model. Attanasio et al (forthcoming) use a framed experiment in Colombia to study the role of risk preferences for group formation in the context of risk sharing. Glaeser et al. (2000), Leider et al. (2009) and Ligon and Schechter (2010, 2011) use framed field experiments to study the role of social ties in trust and dictator games. Several other papers use framed field experiments to study aspects of financial development which are related to our focus on savings. For instance, Giné et al. (2010) and Fischer (2010) use framed experiments in Peru and India, respectively, to test implications of joint-liability lending models.[14] Landmann et al. (2012) study the interaction between insurance and the ability to hide income using a framed experiment in the Philippines, but they study static, rather than dynamic contracts, their setup does not allow smoothing consumption over time with savings, and they do not study the effect of social distance. While the experimental literature on financial development is actively growing, to our knowledge ours is the only paper which explicitly combines interpersonal smoothing (transfers), intertemporal smoothing (savings), barriers to insurance (limited commitment) and social networks, in either a framed field or laboratory setting. This results in a game that is more realistic than settings with interpersonal or intertemporal smoothing only, and which uniquely allows studying the role of social ties in facilitating informal insurance.

The rest of the paper is organized as follows: Section 3.2 reviews the predictions of informal

[13]There is a growing body of evidence on this channel, from natural experiments (e.g., Burgess and Pande 2005, Kaboski and Townsend 2011); structural models (e.g., Gine and Townsend 2004); and field experiments (e.g., Dupas and Robinson 2009 and Brune et al. 2010).

[14]Joint liability, in turn, has been advocated as a solution to limited commitment vis-a-vis loan contracts (Ghatak and Guinnane 1999).

insurance with and without access to an intertemporal technology, and discusses how social networks affect informal insurance. Section 3.3 details our experimental protocol and data. Section 3.4 describes the tests we use to gauge how well each model fits the experimental data, and presents the results. Section 3.5 concludes. Figures, tables, and additional details are in the appendices.

3.2 Framework: Insurance without commitment

The theory of interpersonal consumption insurance without commitment (and without a savings technology) was developed by Coate and Ravallion (1993), and extended to a dynamic framework by Kocherlakota (1996) and Ligon et al. (2002). Ligon et al. (2000) show that access to savings may possibly make the village as a whole better off, by allowing better smoothing of originally uninsured individual and aggregate risk; or worse off, by increasing the temptation of lucky households to walk away. Here we review the predictions of two models—limited commitment without savings and limited commitment with savings which are retained after defection—to highlight the comparative statics that are predicted by each model of informal insurance, and the comparisons that will allow us to study the interaction of insurance and savings access. We also discuss how these models are affected by the presence of direct defection costs which are a function of social distance. A full characterization of these problems is provided in Appendix 3.B.

3.2.1 Limited commitment, no savings

The key feature of limited commitment models is that individuals cannot bind themselves to participate in the insurance agreement. As a result, an individual with a high income realization may prefer to renege on the agreement, rather than make the transfer to other insurance members that she previously agreed to. The benefit of reneging is the ability to keep more income today. The pecuniary cost is exclusion from or reduced access to insurance in the future. There may also be social sanctions and loss of the nonmonetary value of the relationship, which we discuss below. Thus, ceteris paribus, individuals expecting less future surplus from the insurance agreement will be more tempted to renege. The amount of future surplus an individual expects can be summarized by a single parameter, her "promised utility" (Spear and Srivastava 1987).

Kocherlakota (1996) and Ligon et al. (2002) characterized the optimal dynamic insurance contract subject to limited commitment: when an individual is tempted to renege, her current consumption and promised future surplus are increased to make her exactly indifferent between leaving and staying. Because the temptation to renege only arises when income is above average (requiring net transfers to be made to others in the insurance network), high incomes will be associated with increases in consumption. This generates a positive co-movement between consumption and income if the participation constraints bind. This is in contrast to the full insurance allocation in which, conditional on aggregate consumption, there is zero co-movement between consumption and income.

3.2.2 The role of savings

One of our goals in this paper is analyzing the welfare impact of introducing access to a savings technology in a limited commitment relationship.[15] Ligon et al. (2000) note that access to savings has a twofold impact on the constrained-efficient risk-sharing contract. On one hand, access to savings increases the utility that individuals enjoy after the violation of a contract, because they are no longer required to live hand-to-mouth in the absence of interpersonal transfers. By increasing the temptation to renege, this effect reduces the amount of interpersonal insurance which can be achieved in equilibrium. On the other hand, if full insurance is not feasible without access to a savings technology, savings can help to smooth over time the risk that cannot be spread interpersonally.[16] Overall, the effect of savings access on individuals' risk sharing and welfare is ambiguous and depends on the initial level of risk sharing. In order to illustrate this, Ligon et al. (2000) consider two extreme examples (both without aggregate risk). First, if full risk sharing is possible without savings, when the possibility of savings is introduced, full insurance may no longer be possible due to the tightened participation constraints; then, savings access reduces welfare. Second, it is possible that without savings almost no risk sharing is achieved. Then, access to savings allows individuals to smooth intertemporally some of the risk that they could not insure interpersonally, increasing welfare.

So far we have argued that *ex ante*, individuals may be better or worse off with access to savings. Further, there may be distributional effects. Namely, *ex post* unlucky (low-income) individuals, who are net recipients of insurance transfers, will be affected most adversely by a reduction in the amount of insurance.[17] Distributional effects are relevant for policy recommendations, since weak institutional capacity in many developing countries limits feasible transfers from "winners" to "losers," and governments and policy makers often put particular weight on the welfare of the poorest and most adversely affected.

3.2.3 The role of social proximity

The literature on risk-sharing and social networks often finds that, ceteris paribus, individuals are more likely to share risk with friends and family than with strangers (e.g., Hayashi et al. (1996), Fafchamps and Lund (2003), Angelucci and DeGiorgi (2009)). Individuals may behave differently when sharing risk with people at varying levels of social distance for many reasons. First, an individual sharing risk with a socially closer person versus socially farther person may

[15] Savings were fully observable by the partner. There is no private information in any of the games. We consider the issue of private information in another set of experiments, described in Chandrasekhar et al. (2011)

[16] In a setting with aggregate risk, savings may be helpful even if full insurance of idiosyncratic risk is attained without savings.

[17] With serially correlated income, an offsetting effect is possible: if there is little insurance even without savings, unlucky individuals may benefit the most from the ability to keep a buffer of savings, because today's low income suggests low income in the future and increases the motive to save (Deaton 1991). However, with i.i.d. incomes this possibility does not arise.

have different incentives. Individuals may be able to exact greater network-based punishments upon those closer to their social circles (Bloch, Genicot, and Ray 2008). For instance, they may meet socially closer individuals more often and therefore the threat of ostracism by a socially distant individual may be less severe. Second, there may be directed altruism. If, for instance, people closer on the network interact more often, and therefore place more weight on each others' consumption, this would induce differential levels of altruism as a function of the network location of the two partners, as emphasized by Leider et al. (2009). In general, there may be less at stake for people who are more socially distant.

There are many models that could be used to capture these forces. However, distinguishing them is not our goal in this paper. Instead, we take an admittedly reduced-form approach to introduce social networks into our framework. We want to capture the idea that punishments or other costs of defecting against socially closer partners may be greater. Therefore, as a reduced-form representation capturing all of the above possibilities, we assume that an individual who has reneged on the risk-sharing agreement with her partner pays a non-pecuniary cost $f(\cdot)$ that is larger, the greater the social proximity between the individual and her partner. That is, the cost is a decreasing function of social distance, γ, but is weakly positive for all distances:

$$f(\gamma) > f(\gamma') \ \forall \ \gamma < \gamma', \ \gamma, \gamma' \in \mathbb{N}, \text{ and } f(\gamma) \geq 0 \ \forall \ \gamma \in \mathbb{N}. \tag{3.1}$$

We assume that, if i reneges on his or her promises to j, a one-time cost $f(\gamma(i,j))$ is paid by i.[18] We discuss in Section 3.4 how the presence of the term $f(\gamma(i,j))$ affects risk-sharing, and how its importance can be tested empirically.

In this section we noted some key differences among the regimes we consider—full commitment, limited commitment without savings and limited commitment with savings—in terms of levels of average insurance and welfare, and possible distributional differences. We now describe our experimental setup, designed to mimic these regimes.

3.3 Experimental Details and Data

3.3.1 Setting

Our experiment was conducted in 34 villages in Karnataka, India. The villages range from 1.5 to 3 hours' drive from Bangalore. The average village, according to our census data, contains 164 households, comprising 753 individuals. These particular villages were chosen as the setting for our experiment because village censuses and social network data were previously collected on their inhabitants, as described below and in more detail in Banerjee at al. (2011). This gives us uniquely detailed data, not just on our experimental participants and their direct connections to

[18] This cost is conceptually similar to the costs $P_i(s)$ in Ligon et al. (2002); relative to their setting, we specify these costs to depend on i's social distance to his or her partner.

their partners, but also on indirect linkages between partners, e.g., through mutual friends. South India is an ideal setting for our experiment because South Indian villages have historically been characterized by a high degree of interpersonal risk-sharing, as demonstrated by Townsend (1994) and others for the ICRISAT villages, and because rural South India is currently experiencing rapid growth in the availability of savings, but from a low base.

In each village, 20 individuals aged 18 to 50 were recruited to take part in the experiment. As an incentive to attend, participants were paid a show-up fee of INR 20, and were told they would have the opportunity to win additional money. In total, 648 individuals participated in the experiment. The average age was 30, 56% of players were female, and the average education was 7th standard.[19] Over 98% of pairs in our sample could reach each other through the social network, meaning there exists a path of relationships connecting them. Among those who could reach each other, the average social distance was 3.67 and the median was 4, meaning that the members of a median pair were "friends of a friend of a friend of a friend." Tables 1a and 1b show summary statistics for the individuals and pairs that participated in the experiment.

3.3.2 Overall game structure

The purpose of our games was to harness the incentives to share income risk that exist in real life, but to do so in a way that can be implemented in an experimental session lasting a few hours. For external validity, individuals should have strong incentives to smooth risk and to think carefully about their choices. A timeline, illustrating the structure of each game and the full experiment, is shown in Figure 1.

Consumption smoothing has both intertemporal and interpersonal components. We create an interpersonal component by pairing individuals into groups of two (shown at left of the top timeline panel). In all games, the members of a pair can make transfers to each other. To simulate the intertemporal smoothing motive, individuals play many rounds during the experiment (18 rounds—six per game—on average), but are paid their "consumption" for one randomly-selected round.[20] To make this salient, income takes the form of tokens that represent INR 10 each, and each consumption realization is written on a chip and placed in a bag that the player keeps with him or her during the entire experiment. At the end of the experiment, an experimenter draws one chip from the bag, and the individual is paid the amount shown on the selected chip.

In our setting, this payment structure has the crucial implication that players cannot use transfers after/outside the experiment to insure the risk they faced during the experiment. First, outside transfers have the shortcoming that all players could get paid for rounds in which their income was

[19]The sample of villagers who took part in our games is not a random sample of the village as a whole: we informed local leaders that we would be coming to the village on a certain day, and looking for individuals to participate in a series of games. All comers aged 18-50 who could be located in the census data were considered for the experiment. Selection into the experiment poses no problems for internal validity, since all participants play all the games (with randomly chosen partners), and individual-fixed effects control for individual heterogeneity.

[20]This is standard in the literature, e.g., Charness and Genicot (2009) and Fischer (2010).

low, or vice versa. Additionally, while income was observable during the experiment, it was no longer fully observable outside the experiment, since selection of a round for payment and the actual payout were done in private. Finally, since each player was paired with three different partners, there was no guarantee of being paid for a round played with a particular partner. Thus, players had strong incentives to engage in risk sharing within the experiment–and the data show that they did so.

Incomes are risky: as in our theoretical setup,[21] there is a high income level (INR 250), and a low income level (INR 0). Moreover, to simulate the (possibly unequal) wealth individuals have at the time when they enter into an insurance relationship, before round 1 of each game one partner is randomly chosen to receive an endowment of INR 60; the other receives INR 30. The games are described in the context of a farmer who may receive high income because of good rains this season or low income because of drought. Discussions with participants indicate that they understood the risk they faced and the data show that both transfers and savings are used to smooth this risk.

To replicate an interaction that may extend indefinitely into the future, induce discounting and avoid a known terminal round, the game ends with 1/6 probability at the end of each period, determined by drawing a ball from a bag that has five red balls and one black ball. Participants are told before each game that the game will end when the black ball is drawn, and that therefore at any point when the game has not ended, it is expected to continue for 6 more rounds. Once a game ends, individuals are re-paired. The order of the games is randomized, and we control for game order in our regressions.

The options allowing players to decouple consumption from income vary by game. However, in all treatments, at the beginning of each round before incomes are realized (but after the endowment is realized in round 1), partners may decide on an income sharing "plan." (This is shown as the leftmost entry of the lower timeline panel in Figure 1.) That is, partner 1 chooses how much 1 will give 2, if 1 gets INR 250 and 2 gets 0 (τ_t^1), and 2 chooses how much 2 will give 1, if 2 gets INR 250 and 1 gets 0 (τ_t^2). This plan may be asymmetric ($\tau_t^1 \neq \tau_t^2$) and time-varying ($\tau_t^i \neq \tau_{t'}^i$). Communication between the partners was allowed while they made these decisions, to mimic real-life interactions, but one partner did not have veto power over the other's announced transfer.

The details of each treatment are as follows:

1. **Full commitment, no savings:** Partners announce an income sharing plan for the round.[22] Once incomes are realized, the experimenter implements the transfer that the lucky player announced ex ante. There is no opportunity for the lucky player to change her mind. Each individual then "consumes" by placing all of her tokens, net of any transfers, into a consumption cup. The experimenter removes the tokens, writes the consumption amount on a chip, and the chip is placed in the consumption bag. A random draw determines whether the game

[21] See Apprendix 3.B for details.

[22] For instance, this could be: "Player 1 will give Player 2 Rs. 100 if Player 1 gets the Rs. 250 payout, and if Player 2 gets the Rs. 250, she will give Player 1 Rs. 80."

will continue. If it continues, before the next round, partners make a new sharing plan (which can be the same or different than the prior one).

2. **Limited commitment, no savings:** Partners announce an income sharing plan as before. However, after seeing their income, the lucky individuals can reassess how much to transfer to their unlucky partners. (This is indicated by the timeline entry in a dotted box in Figure 1.) They may choose to transfer a different amount than the one announced *ex ante*, including transferring nothing. Individuals are told they will have the option to change their minds *ex post* before they decide their sharing rules. After any reassessment, the transfer is implemented, and each individual then places all her tokens, net of any transfers, into her consumption cup. The experimenter takes the tokens, writes the amount on a chip, and the chip is placed in the consumption bag.

3. **Limited commitment, with savings:** As in game 2, the lucky individual may renege on her announced transfer after seeing her income. In addition, each player has access to a "savings cup." Once transfers are made, players can consume tokens by placing them in the consumption cup, or save them by placing them in the savings cup. (The savings decision is indicated by the timeline entry in a dashed box in Figure 1.) Saved tokens are available to consume in later rounds, but are lost when the game ends. If an individual reneges on transfers she announced *ex ante*, she keeps her savings.

3.3.3 Regression specifications

Randomization ensures that many of our hypotheses of interest can be answered by simple comparisons of the mean of a particular outcome across games. We want to assess whether the limited commitment model is a good description of players' behavior, measure the effect of different treatments on the magnitude of interpersonal insurance and on welfare, and study how these cross-treatment differences respond to the strength of social proximity. Our main estimation specification take the following form for outcomes defined at the individual-game-round level:

$$\omega_{ivgr} = \alpha + \beta_g D_g + X'_{vg}\eta + \phi_i + Z'_{ig}\zeta + \varepsilon_{ivgr} \tag{3.2}$$

where ω_{ivgr} is an outcome for individual i in game g, round r; D_g is a game indicator (commitment, no commitment without savings, etc.), so β_g measures the average effect of that game on the outcome; X_{vg} includes characteristics of the game as played in that village (i.e., order-of-play and surveyor effects); ϕ_i is an individual-fixed effect, and Z_{ig} includes an indicator for whether i and i's partner j in game g are connected in the village social network, and, if connected, the distance between i and j.[23] The individual-game-round level outcomes we consider are absolute deviations of consumption from the overall average for that game, $|c_{ivgr} - \bar{c}_g|$; and savings, s_{ivgr}.

[23] We have also estimated specifications omitting individual-fixed effects and controlling for characteristics of the individual (education, wealth, and individual-level network characteristics measuring an individual's "importance" in

We also examine transfers, τ_{ivgr}, made from the lucky to unlucky individual in each round; for these regressions the sample is restricted to individual-game-round observations on lucky individuals.

The estimation errors (ε) in our regressions may be correlated across individuals within a given game in a particular village, due, for instance, to slight idiosyncrasies of game explanation, disruptions in the experiment venue, etc. Therefore all regression standard errors are clustered at the game-village level.

3.3.3.1 Capturing the effect of social distance

To examine how social distance affects lack of commitment when the outcomes are the amount the lucky partner transfers to the unlucky partner, and the absolute consumption deviations of both partners, we run regressions of the form:

$$\begin{aligned} \omega_{igr} &= \alpha + \beta_1 D_g + \eta_1 \rho(i,j) + \eta_2 \gamma(i,j) \\ &\quad + \delta_1 D_g \cdot \rho(i,j) + \delta_2 D_g \cdot \gamma(i,j) \\ &\quad + \phi_i + Z'_{ig}\zeta + \varepsilon_{igr} \end{aligned} \tag{3.3}$$

where ω_{igr} is an outcome for i in game g, round r, D_g is an indicator for a particular treatment, $\gamma(i,j)$ is the social distance between i,j (with infinite distances set to zero), and $\rho(i,j) = 1\left(\gamma(i,j) \neq \infty\right)$ is an indicator that i and j are reachable through the network. As outcomes, we consider absolute deviations of consumption from the overall average for that game, $\left|c_{igr} - \bar{c}_g\right|$, and savings, s_{igr}. We also examine transfers, τ_{igr}, made from the lucky to unlucky individual in each round.

The term ϕ_i is an individual-fixed effect, and Z_{ig} includes pair-level covariates. When the outcome is transfers made from the lucky to unlucky individual in each round, the sample is restricted to individual-game-round observations on lucky individuals.

The coefficient β_1 is the regression-adjusted effect of moving from full commitment to no commitment for unreachable (i.e., socially unconnected) pairs. The sum $\delta_1 + \delta_2$ is the differential no-commitment effect for closest pairs (of distance 1), and δ_2 is the additional differential effect for each additional unit of geodesic distance between the members of the pair. Since the reference category in these regressions is full commitment, the coefficient η_2 is the effect of increasing social distance by one unit in the full commitment treatment.

---

the network). As well, we can control for pair characteristics beyond geodesic distance, i.e., an indicator for being of the same caste and the difference in roof types (a proxy for wealth differences). These individual and pair-level characteristics enter with the expected signs and do not change the between-game comparisons we find in the baseline specifications. (Results available on request.)

3.3.4 Use of smoothing mechanisms

Because we want to use the results of our experiment to study how interpersonal and intertemporal consumption smoothing interact, we need to show that the players understand and are willing to use interpersonal transfers, and, when available, savings. Table 2, column 1 shows average transfers by game. Average transfers are INR 92.35 in the full-commitment treatment, about three-quarters of the INR 125 that would be associated with full insurance. (Recall that even if one individual always consumes more than the other due to a higher bargaining weight, average transfers will still equal half of aggregate income of INR 125 per round.[24]) Column 4 shows that average savings levels in the limited commitment with savings game are INR 22.65. Significant levels of transfers in savings and non-savings treatments, and use of savings when savings are available, suggest that meaningful consumption smoothing is occurring.

3.3.5 Measuring the degree of insurance

To examine the magnitude of interpersonal insurance, we examine average transfers made by individuals with high income realizations to those with low income realizations. The following lemma allows us to do so:

LEMMA 3.3.1. *Because Pareto weights are orthogonal to the in-game income process, under full risk-sharing average transfers will equal half of average income. If players insure, on average, fraction α of their idiosyncratic risk, average transfers will equal a fraction $\frac{\alpha}{2}$ of average income.*

(All proofs appear in Appendix 3.B.) This gives us a measure of the amount of interpersonal risk-sharing which does not rely on knowing the relative bargaining power or Pareto weights. Moreover, these comparisons do not rely on the assumption that individuals are on the Pareto frontier; merely that they are risk averse.

We can therefore interpret changes in transfers when moving from full commitment to limited commitment without savings as the change in interpersonal insurance due to binding participation constraints; and we can interpret changes in transfers when moving from limited commitment without savings to limited commitment with private savings as the change in interpersonal insurance due to savings access affecting participation constraints.

3.3.6 Measuring welfare implications

Examining transfers as an outcome tells us about the degree of interpersonal insurance. However, we are also interested in the implications for welfare. In particular: Is welfare higher (or lower) with savings access than without, and by how much? How much do binding participation constraints reduce welfare, relative to the full commitment case?

[24] Individuals receive extra income in the first round, in the form of the initial endowment. However, since this income is revealed before insurance agreements are made, the endowment should not be "insured." We test this prediction below.

In general, the effect of different treatments on welfare would be comprised of an effect on the level of consumption and an effect on the variability of consumption. However, because the income process was fixed across treatments, there will be no difference in average consumption between the full commitment (FCNS) and limited commitment (LCNS) games. Table 2, column 2 shows that this is indeed the case—average consumption is INR131 in both games.[25] Because savings are lost when the savings games end, consumption is very slightly lower in the LCWS games (INR 2).

In thinking about the external validity of the findings of this experiment in terms of insurance and welfare, three points are worth noting. First, the amounts of money involved are substantial. Average expected earnings in the experiment are about INR 130. To put this into perspective, the National Rural Employment Guarantee Act (which guarantees 100 days of work per year to anyone able and willing to do manual labor) pays a wage of INR 74 for a day's work in rural Karnataka. Thus, individuals face strong incentives to think carefully about how to maximize the benefit they can derive from playing the games, by aiming for consumption choices that are stable across rounds.

Second, great care was taken in designing the physicality of the games and the framing with which we presented them, in order to make them both easy to understand and similar to real life. In explaining the games to the participants, it was explained that the games that they play are much like the decisions they make in every day life. In each round they receive some income and (depending on the game) they are able to make decisions to consume, save for the future, or share money with their partner. Many players spontaneously noted the parallels between the games and real-life decisions.[26] Finally, using social network data, we can control for and study the interactions of the participants outside the experiment which may affect the incentives created by an experimenter. We discuss this below in detail.

3.3.7 Randomization and the role of social networks

3.3.7.1 Unique randomization

Our randomization was unique in that it stratified against the social network in real time in the field. To that end, we made use of a unique dataset containing information on all 34 villages in which our experiment was conducted. We have complete censuses of each of the villages as well as detailed social network data. The network data was collected for Banerjee et al. (2011) in which they conducted a survey about social linkages for a random subset of the population. For a village, the graph (or multi-graph), represents individuals as nodes with twelve dimensions

[25] Consumption is higher in round 1 of each game, where players receive Rs. 30 or Rs. 60 as an initial endowment. Because there were random variations in how long each game lasted, consumption is an insignificant Rs. 0.31 higher in the LCNS game than in FCNS.

[26] One player told us "The games were very interesting, especially for those who have some education... They help us think about how much we really should save and give to our friends in times of hardship." Furthermore, in two villages, after the experiment village leaders inquired about the possibility of having an microfinance institution come to their village, because they saw links between the games and the possibility of having actual savings.

of possible links between pairs of vertices. These dimensions include relatives, friends, creditors, debtors and advisors, among others. (See Appendix 3.C for details.) For our purposes, we work with an undirected, unweighted graph which takes the union of these dimensions, following Banerjee et al. (2011). In our villages, the multiple dimensions are highly correlated so the union network captures latent information. (Moreover, any weighting method would be rather *ad hoc* in nature.) Henceforth, we refer to this object as the social network of the village. Using this social network, we construct a variable $\gamma(i,j)$ that represents the length of the shortest undirected path between i and j. We refer to this as i and j's social distance. If i and j are connected directly (e.g., they are friends), their social distance is $\gamma(i,j) = 1$; if i is not connected to j directly, but is connected to some k, who is connected to j, $\gamma(i,j) = 2$, etc. Then, i and j are said to be reachable ($\rho(i,j) = 1$) if there exists any path from i to j. We provide a more detailed description of the construction of these variables in Appendix 3.C.

It should be emphasized that we are working with sampled networks—approximately 50% of households within each village completed the social network questionnaire. Links including the other, unsampled 50% will be observed only when one member of the dyad was sampled. This means that some ties between participants will be unobserved (e.g., if i is connected to j who is connected to k, the indirect tie between i and k will be missed if j is not surveyed). This has the effect of upward-biasing our measure of social distance, and attenuating our estimates of the effect of social distance, making our findings lower bounds on the true significance of social networks (Chandrasekhar and Lewis 2010).

Since most social networks exhibit small-world phenomena, even if a random subset of villagers took part in our experiments, randomly chosen *pairs* would tend to be fairly close in social distance. This tendency would be exaggerated if people tend to come to the experiment with their friends or relatives, which was the case for many people who took part in our experiment. Therefore, the distribution of social distances will be left-skewed, and simply randomly assigning partners would mean that more often than not, people would be paired with near-kin. This would limit the statistical power of our data to reveal how socially distant pairs play the games, yet the behavior of socially distant pairs is important to allow us to study how behavior across games changes with social distance. Therefore, to make the distribution of social distances between our pairs more uniform in our sample, we used the network data to oversample the right tail of the distance distribution. This was done in the field, once the experimental participants had been located in the village census data. Figure 2 shows the distributions of social distances for 3 villages: the full distribution and the distribution of assigned pairings in the experiment. The comparison between the full distribution and the distribution of assigned pairings reveals that we were successful in oversampling the right tail of the social distance distribution: the distribution of pairings used in the experiment has more mass at greater distances, particularly distances of 5 and 6, than the full distribution.

3.3.7.2 The role of social networks

The social distance data enable us to do two things. First, because we have exogenous variation in social distance[27], our results are informative for studying how limited commitment relationships and the insurance that they can support are affected by social proximity. Thus, we can examine how the ability to maintain high levels of interpersonal insurance in the absence of commitment is affected by the network relationships of risk-sharing partners, whether the extent to which savings crowds out interpersonal insurance depends on social distance, and whether socially distant pairs use savings with differing intensity due to different levels of uninsured risk. By estimating the gradients of these relationships with respect to social distance, we obtain estimates that are more portable to other settings, where levels of social distance vary. This would not be possible without random assignment of pairs to play different treatments, in which case social distance would be an omitted variable. If we had conducted an anonymous experiment, the average level of risk-sharing would not generalize to non-anonymous settings and we would not be able to estimate the effect of social distance.

Furthermore, with data on social distance we are able to estimate and control for the effect of relationships outside the experiment. In any non-anonymous experiment, as in non-experimental interactions, partners' actions may be affected by other dimensions of the relationship which extend beyond the game at hand. Including social distance in our regressions allows us to control for outside-the-game effects, something which, to our knowledge, is unusual for a framed experiment.

We now turn to discussing how the implications of the models map into testable predictions across the different versions of our experiment, and also how these testable predictions should be affected by social networks.

3.4 Testable implications and results

Our discussion of the testable implications of the models ("propositions"), and effects which are theoretically ambiguous but can be signed empirically ("empirical questions"), is divided into two parts. We begin in section 3.4.1 by discussing implications and questions that apply to the model without a role for social distance–or alternatively, averaging over all values of social distance in our dataset. We begin by showing evidence that limited commitment matters in reducing transfers and consumption smoothing. Then we demonstrate that information revealed before sharing plans are made is not insured, consistent with transfers being used for insurance motives, and show that players "punish" their partners when they deviate from their sharing plans, indicating that these plans are not merely cheap talk. In the final parts of this section, we show that the average impacts of savings access on consumption smoothing is positive, and then examine whether access to savings has distributional consequences for those who experience bad income draws vis-a-vis

[27] As noted above, we do not assign the social distance to a dyad: the social distance between i and j is given. Instead, among the potential dyads of participants in the experiment, we chose certain dyads to be paired.

those who receive good income draws, arguing that savings access benefits both fortunate and unfortunate players.

Then, in section 3.4.2 we discuss implications and questions for how transfers, consumption smoothing, defection rates and savings behavior varies across treatments for pairs with different levels of social distance. We first show that limited commitment's impacts on transfers and consumption smoothing are essentially eliminated among partners who are directly socially connected, and that the effect increases with social distance. Next we show that players save more when paired with socially distant partners, due to facing more uninsured risk, and finally, that defection occurs much more often among pairs that are socially distant.

3.4.1 Commitment, savings and smoothing

Comparing the limited commitment model (without savings) to the full commitment model yields the following implications and questions:

3.4.1.1 Transfers and consumption smoothing

To test the validity of the full and limited commitment models as a description of experimental subjects' behavior, we check whether the following comparative statics hold for transfers and consumption smoothing:

PROPOSITION 3.4.1. *When comparing full commitment no savings (FCNS) vs. limited commitment no savings (LCNS), if participation constraints bind, transfers will be lower and consumption variability higher under LCNS as compared to FCNS.*

Figure 3 shows levels of transfers by round of the game, plotted separately for each game. The level of transfers is higher under full commitment than under limited commitment in each round, as predicted by Proposition 3.4.1. No time trend is apparent, although transfers under FCNS are slightly lower in the first and second rounds, while under LCNS they are slightly higher.[28] This is consistent with players understanding the setup of each game from the beginning, and understanding the games' stationary structure.

Figure 4 shows averages of consumption absolute deviations across rounds of the game, again plotted separately for each game. In the no-savings treatments, FCNS and LCNS, consumption variability is markedly higher in round 1 than in later rounds–this reflects the fact that the initial endowment of INR 30 or 60 was not insured across players and, when savings was not available, was consumed in round one. We return below to the non-insurance of the initial endowment, which is consistent with our predictions. After round one, consumption variability is always lower under FCNS than under LCNS, consistent with Proposition 3.4.1.

Table 3 shows results testing Proposition 3.4.1 via regression, estimating equation (3.2), using

[28] All results are robust to dropping rounds 1 and 2. (Results available on request.)

as outcomes transfers and absolute deviations of consumption from its mean.[29] To understand the value of commitment, the relevant comparison is between the full commitment treatment (FCNS) and limited commitment treatments (LCNS and LCWS). Column 1 shows results for transfers. Consistent with Proposition 3.4.1, transfers are significantly lower in the two no-commitment treatments. Relative to the full commitment treatment, transfers are INR 9 (10%) lower under limited commitment-no savings, and INR 11.3 (12%) lower under limited commitment with savings, indicating reduced interpersonal consumption smoothing due to limited commitment.

Column 3 shows results for consumption smoothing. The measure of consumption variability we use is the absolute deviation of consumption in a given round from the player's total average consumption in that game.[30] Proposition 3.4.1 predicts that consumption smoothing worsens when commitment is removed. Consistent with this, moving from FCNS to LCNS leads to a INR 9 increase in the absolute deviation of consumption, significant at the 1% level. This effect is equal to almost 20% of the average absolute deviation in the FCNS game, an economically as well as statistically significant increase.

These comparisons are derived assuming that individuals are on the constrained Pareto frontier. However, if there is an additional cost or constraint to making interpersonal transfers (e.g., due to contemplation costs of calculating the appropriate transfer, an endowment effect which makes it unpleasant to surrender money one has won, etc.), then there may be less-than-full insurance even when participation constraints *per se* do not bind. We are able to estimate the extent of such costs of engaging in full risk sharing using the FCNS case. In the case that we see positive variance of consumption under FCNS, it suggests that forces other than participation constraints limit risk sharing. Even if individuals are not on the Pareto frontier parsimoniously defined by the limited commitment model, comparisons across the treatments are still informative. In the case that they are not on the Pareto frontier, while we would not be able to map our empirical findings into statements about parameters in a limited commitment problem (e.g., magnitudes of Lagrange multipliers on particular constraints), the comparison of the LCNS treatment versus the LCWS treatment will still help us to address the empirical questions that this paper proposes: namely, do individuals achieve better overall consumption smoothing with or without access to savings; and, how is this affected by social networks? Thus:

EMPIRICAL QUESTION 3.4.1. *Is there positive variability of consumption under FCNS?*

Table 3, column 3 shows that the average consumption deviation under full commitment is INR 40.9. Thus, individuals do not fully smooth consumption when there is no issue of limited commitment. This suggests that there is an additional cost or constraint to making interpersonal

[29] All p-values are from regressions controlling for individual-fixed effects, reachability and distance between partners, surveyor and team effects, and order and round of play, and adjusting for clustering at the village-by-game level. As expected, due to randomization, regression adjustment does not change the magnitude of the estimated effects, but does improve precision.

[30] Using squared deviations, variances and standard deviations yields similar results. The absolute deviation of consumption is in units of rupees and therefore easy to interpret.

transfers. However, the comparative statics that we discuss below are on the whole consistent with the hypothesis that individuals are on the constrained Pareto frontier, subject to this additional cost of engaging in full risk sharing (i.e., a cost not derived from the participation constraints of the limited commitment model).

3.4.1.2 The role of ex ante wealth

An implication of insurance (i.e., transfers made to smooth risk as opposed to transfers made for other reasons) is that a shock revealed before the insurance contract is signed cannot be insured, because its realization does not represent future risk that can be diversified away. If individuals are in fact making insurance agreements with each other, rather than simply sharing with each other due to altruism, social norms, etc., any shock revealed before pairs make their insurance agreements should not be insured, in any of our treatments. The initial endowment is such a shock. Therefore, if players share risk due to insurance motives, we should see that an individual's realization of the initial endowment feeds into individual consumption to a greater degree than subsequent income, which is realized after insurance agreements are made. In the case of fully self-interested, i.e. non-altruistic or other-regarding behavior, the initial endowment should feed fully into individual consumption.[31] Thus:

PROPOSITION 3.4.2. *If players share risk due to insurance motives, an individual's realization of the initial endowment before the insurance contract is signed should not be insured. As such, the high endowment individual should consume at least INR 30 more than the low endowment individual.*

Table 5, columns 1 and 4 test the prediction of Proposition 3.4.2 by regressing total consumption over the course of the game on an indicator for whether the player received the high or the low endowment (INR 60 vs. INR 30). The results reveal that individuals who receive the high endowment of INR 60 consume almost exactly INR 30 more than individuals who received only 30 INR; that is, the endowment shock is not insured at all, consistent with Proposition 3.4.2. The effect does not vary across games, suggesting that even in the game where individuals can commit, they do not use transfers to equalize *ex post* a shock that was revealed before the contract was made. This is strong evidence that players are using transfers out of true insurance motives and not out of pure or directed altruism (à la Leider et al. 2009 and Ligon and Schechter 2010), or as part of a larger risk-sharing agreement that might exist between the pair members (à la (Ambrus, Mobius, and Szeidl 2010), or Karlan et al. 2009(Karlan, Mobius, Rosenblat, and Szeidl 2009)), or due to demand effects from the experimental setup. Columns 2-3 and 5-6 show that the INR 30 increase in total game consumption comes partially from a drop in transfers made and partially from an increase in transfers received. Combined with the evidence that transfers are higher under full commitment than limited commitment, the evidence on the endowment shows that players

[31] In fact, the high-endowment player may consume even more than Rs. 30 more than the low-endowment player if the utility function exhibits decreasing relative risk aversion, because the low-endowment player would prefer smoother consumption, at the cost of lower consumption levels (Genicot, 2006).

insure risk when they can, but that inability to enforce contracts limits this.

3.4.1.3 Defection and punishment

Although the constrained-optimal insurance arrangement under limited commitment will not feature defection in equilibrium since every efficient insurance contract has an efficient continuation contract after every history (Ligon et al. 2002), in reality binding enforcement constraints may be manifested through players actually changing their minds, i.e., defecting.[32] Table 6 presents the results on defection probabilities, revealing that binding participation constraints manifest themselves through defection, i.e., players transferring a different (almost always lower) amount than they promised.[33] Defection, defined as transferring less than promised, occurs in 23% of limited-commitment rounds.[34] The left-hand panel of Figure 5a shows the distribution of defection amounts, when defection occurs. The average defection amount is INR 46, the 25th percentile is INR 20 and the 75th percentile is INR 75. Defection amounts ranging from INR 10 to INR 250 are observed. The right-hand panel of Figure 5a shows the distribution of defection amounts, expressed as a percentage of the initial promise. Conditional o defecting, the average defector transfers 43% less than promised. The 25th percentile is 20% and the 75th percentile is 55%.

Since we observe defection, we can also observe what type of post-defection responses individuals actually use, and what consequences they have for consumption smoothing. Table 7 shows how transfers are affected in the rounds following defection. Transfers are significantly reduced, by about INR 12, in the first round post-defection, returning to the level that prevailed before defection occurred after 4 rounds. We illustrate this graphically in Figure 5b. During the maximal punishment phase, transfers fall by roughly 15%. While the level of punishment imposed is a far cry from permanent reversion to autarky, the (off equilibrium) punishment assumed in most models of limited commitment, the fact that punishment is inflicted shows that players' statements about what they promised to transfer to their partner are not simply cheap talk—as does the fact that, 77% of the time in the limited commitment treatments, the promised transfer was actually made.

We now turn to examining how the effects of limited commitment and savings access vary for risk-sharing groups with different amounts of social proximity.

[32] In the context of an optimizing, forward-looking model, this can be modeled as additive error term, v, unforecastable by the individual (and unobserved by their partner), in the value of reneging on a particular promise. The probability of defection when y_H is realized is then the probability that v exceeds the surplus the lucky individual had anticipated when receiving y_H and making the promised transfer τ.

[33] In a small fraction of rounds, the lucky player transfers *more* than she promised. We do not consider this defection in our analysis. Results defining defection to have occurred only when the transferred amount was less than the promised amount by Rs. 20, the 25th percentile defection size, are similar (available on request).

[34] The participants in our games mention changing their minds about how much to transfer to their partner because after seeing their income, they were unwilling to transfer what they had initially promised, suggesting that binding participation constraints that were not perfectly forecast ex ante result in defection.

3.4.1.4 Savings, insurance and welfare

We now turn from focusing on the effect of commitment and the lack thereof, to examining the effects of introducing savings into a limited commitment environment. First, players do use savings when available: Figure 6 shows the level of savings in the limited commitment setting with savings access (LCWS). Average savings start at INR 30 and trend downward. This downward trend reflects lucky individuals' smoothing of the initial endowment.

The limited commitment model predicts that savings access crowds out interpersonal transfers, since the value of defection (i.e., opting out of the insurance agreement) with savings is higher than without savings, living hand to mouth (Ligon, Thomas, and Worrall 2000). Therefore:

PROPOSITION 3.4.3. *Participation constraints will be tightened by the introduction of savings (LCWS), crowding out interpersonal insurance. Hence, transfers under LCWS will be lower than under LCNS.*

Table 3, Column 1 shows that transfers in both LCNS and LCWS are significantly different from transfers in FCNS at the 1% level, and the point estimate suggests that transfers fall by almost 25% more under LCWS than LCNS (11.2 vs. 8.9), the reduction in transfers under LCNS is not significantly different than under LCWS ($p = .295$). Thus, we find only weak support for Proposition 3.4.3.

When comparing limited commitment no savings vs. limited commitment with savings, the comparison for consumption smoothing is theoretically ambiguous. On one hand, access to savings will tighten participation constraints. This implies that interpersonal transfers will be reduced (crowded out). On the other hand, savings access allows individuals to (partially) smooth income risk over time that cannot be smoothed interpersonally. Therefore, the net impact could be positive or negative. If the effect of tightening participation constraints outweighs the effect of savings access allowing intertemporal smoothing, then aggregate consumption smoothing will worsen and the variance of consumption will increase. If instead, interpersonal insurance is reduced by less than intertemporal smoothing is increased, the variance of consumption will decrease. This gives us:

EMPIRICAL QUESTION 3.4.2. *Does savings' beneficial effect of allowing intertemporal smoothing or its negative effect of tightening participation constraints dominate? Is average consumption smoothing better under LCNS or under LCWS?*

Table 3, Column 3 shows that moving from FCNS to LCWS increases the absolute deviation of consumption by INR 5 ($p < .01$), indicating that full commitment (FCNS) induces significantly more smooth consumption patterns than limited commitment with savings (LCWS): access to savings does not fully make up for the loss of insurance due to limited commitment. However, LCWS results in significantly smoother consumption than LCNS, by INR 4 ($p < .01$). That is, when savings access is introduced to a limited commitment environment, the ability to smooth some risk intertemporally that cannot be smoothed using interpersonal transfers outweighs any crowdout

of transfers. As noted above, conditional on sustaining the same level of consumption, consumption variability is a sufficient statistic for welfare implications. Therefore, we can interpret our results as saying that limited commitment (with or without savings) induces a welfare loss relative to the full commitment no savings case. However, the introduction of savings to the limited commitment game significantly *increases* welfare, relative to the limited commitment-no savings case.

3.4.1.5 Distributional consequences

As well as its effect for the average member of a risk-sharing network, savings access may have distributional consequences. When inter-household risk sharing is augmented by the ability to smooth risk across time, the average household may be better off (our results show they are), but households that experience especially poor luck may suffer more than they would under a mutual insurance-only system (Platteau 2000). Those suffering negative shocks may derive little direct benefit from savings, as they have no excess income to save, while receiving reduced transfers from more fortunate members of the village, who are willing to transfer less because reduced future insurance is less painful when they can rely on a buffer of savings:

Empirical question 3.4.3 therefore asks how the effect of access to savings differs across the distribution of ex post income realizations.

EMPIRICAL QUESTION 3.4.3. *Does savings' pro-insurance effect or its anti-insurance effect dominate for those with "bad luck," i.e. is consumption smoothing for those with low income realizations better under LCNS or under LCWS?*

Table 4 shows how the effects of limited commitment (with and without savings) differ according to how income is distributed between the members of a pair. We find that limited commitment binds more when income is unevenly distributed[35]: in games where one player has a realized income in the lowest tercile of the income distribution, that player's consumption smoothing is much worse in LCNS relative to FCNS—the absolute deviation of consumption increases by INR 16 (column 1). When both players' income realizations are in the middle tercile, the increase in absolute deviation of consumption when moving from FCNS to LCNS is only INR 4 (column 2).

Comparing the coefficients on LCNS and LCWS shows that the benefit of savings (in terms of consumption smoothing) is greatest in games where one player has a realized income in the lowest tercile of the income distribution: column 3 shows that the lucky individual's consumption is smoother in LCWS than LCNS: the absolute deviation falls by INR 9 ($p < .01$). Notably, the *unlucky* individual also benefits from savings access: their absolute deviation falls by INR 5.5 ($p = .07$). This is because there is little crowdout of transfers from the high-income partner, and access to savings allows intertemporal smoothing. These results are counter to the hypothesis that those individuals with the worst series of income realizations ("bad luck") would do worse when

[35]All series of income distributions are drawn from a process in which each player can expect to be lucky 50% of the time. However, due to random variation, the realized income distribution over the short games (averaging 6 rounds) will sometimes be quite unequal.

their partners have access to savings, because their more fortunate partners would prefer to save their income rather than repeatedly make transfers to the unlucky partner. Thus the answer to Empirical question 3.4.3 is that individuals across the income distribution benefit from savings access.[36]

3.4.2 The role of social networks

The results discussed so far indicate that the participants in the experiment are risk-averse and diversify away the risk they face in the experiment to the extent possible. However, the results also show that limited commitment impairs their ability to fully insure this risk. If pairs of players with closer social ties are more able to overcome the commitment problem, we should see that socially closer dyads are less affected by limited commitment than those who are socially distant. The fact that participants were randomly assigned to play with partners of varying social distances allows us to interpret such a finding as the causal effect of social distance in mitigating commitment problems.[37]

3.4.2.1 Proximity and limited commitment

The prediction of how social networks should matter in moving from full commitment to no commitment comes from the fact that closer social proximity lowers the utility individuals get from reneging on their risk-sharing relationship, because they pay a higher penalty for doing so. This implies that, *ceteris paribus*, in a pair with closer ties the players will face less temptation to renege for given income realizations. Hence moving from full commitment to limited commitment (without savings) should lead to a lower reduction in transfers and consumption smoothing in pairs with closer social proximity. This gives the following implications of the limited commitment model with social distance:

PROPOSITION 3.4.4. *Average transfers are lower under limited commitment (without savings), the more socially distant the pair.*

Figure 7 summarizes the fall in transfers (i.e., interpersonal insurance) across games, by social distance, in bar chart form. We group distances 1 and 2; distances 3 and 4; and distances 5 and greater (including unreachable pairs). When partners are of social distance one or two, the fall as commitment is removed is relatively slight, approximately INR 6. When partners are of distance three or four, the fall is somewhat more, INR 8.5. Finally, when partners are of distance five

[36] In a setting where individuals have heterogeneous income processes which are initially private information, so that individuals are learning about their partners' income process, it is possible that individuals with a series of low income realizations would see a larger drop in insurance going from the no savings to savings treatments than in the full information, i.i.d. income setting we consider.

[37] All of our results examining the effect of social distance are robust to controlling for partners' similarity in terms of wealth (proxied by the difference in their roof types) and caste. This suggests that the effects measured here for social distance are capturing social proximity per se, as opposed to homophily. (Results with wealth and caste controls available on request.)

or greater, the fall in transfers due to limited commitment is much greater, INR 13.4. This is what Proposition 3.4.4 predicts. (We discuss the portion of the figure corresponding to limited commitment with savings below.)

We can test Proposition 3.4.4 in a regression framework where social distance, conditional on reachability, enters linearly by estimating equation (3.3) on data from the FCNS and LCNS games only, where the outcome is transfers. Table 8 presents the results. Column 1 shows estimates of equation (3.2), the specification without the social distance interaction, on the sample FCNS and LCNS data, replicating the results of Table 3, column 1. Column 2 shows the specification corresponding to (3.3) where the outcome is transfers. The coefficient β_1 is the effect of moving from full commitment to no commitment for unreachable (i.e., socially unconnected) pairs. The estimated β_1 is 31.77, so for unreachable pairs, removing commitment is associated with a drop in transfers of almost one third. The sum $\delta_1 + \delta_2$ is the differential effect for closest pairs (of distance 1). We estimate $\delta_1 + \delta_2$ equal to positive 2.69 (not significantly different from zero), so for pairs of distance 1, the closest possible, the reduction in insurance due to limited commitment is essentially zero, and insignificant. Then, δ_2 is the additional differential effect for each additional unit of geodesic distance between the members of the pair. We estimate $\delta_2 = -2.996$, i.e. each additional unit of social distance increases the amount that transfers fall in response to limited commitment by INR 3 (significant at 10%).

The main effects of reachability (η_1) and social distance (η_2), are the effects of reachability and social distance in the full commitment treatment. The point estimate on the main effect of social distance is small and not statistically significant, indicating that under full commitment, players do not transfer more to a socially close partner than to a socially distant partner, conditional on being able to reach their partner through the network. In the limited commitment model with social sanctions, social distance only matters via the likelihood of binding participation constraints, so this finding is consistent with the model. However, the main effect of the reachability indicator, η_1, is significant and negative when the outcome is transfers, and significant and positive when the outcome is consumption smoothing, implying that unreachable pairs achieve more insurance than reachable pairs under full commitment. However, since only 1.3% of pairs are not reachable (see Table 1b), this may be purely due to sampling variation.

The limited commitment model also predicts that consumption smoothing will be more severely reduced when commitment is removed (in the absence of savings), the more socially distant is the pair:

PROPOSITION 3.4.5. *Consumption smoothing under limited commitment (without savings) is worse, the more socially distant the pair.*

Figure 8 summarizes the increase in the variability of consumption across games, by social distance, in a form analogous to Figure 7 for transfers. When partners are of social distance one or two and commitment is removed, consumption variability increases by approximately INR 4. When partners are of distance three or four, the increase is larger, INR 8. When partners are of

distance five or greater, consumption variability increases by INR 12 due to limited commitment. This is consistent with Proposition 3.4.5.

To test Proposition 3.4.5 via regression, we estimate (3.3) where the outcome is $|c_{igr} - \bar{c}_g|$, the absolute deviation of i's consumption in round r of game g from average consumption in game g. Table 8, Column 3 replicates the results of Table 3, column 3 on the sample FCNS and LCNS data. Column 4, shows that for unreachable pairs, removing commitment involves a INR 39 increase in the average absolute deviation of consumption, an increase of almost 150% over the full commitment mean of INR 27. However, for pairs of distance 1, there is no increase in consumption variability: the point estimate of $\beta_1 + \delta_1 + \delta_2$ is negative, small and insignificant. Though not statistically significant, the estimate of δ_2 suggests that each additional unit of social distance increases the amount of additional consumption variability in response to limited commitment by INR 7. Therefore the regression results are also consistent with Proposition 3.4.5.

In summary, these results confirm implications of the model of limited commitment with social networks derived in section 3.2: limited commitment leads to the largest reductions in interpersonal insurance and consumption smoothing for the most socially distant individuals, while social distance does not play a significant role when individuals can commit.

3.4.2.2 Proximity and savings access

To illustrate the role of social networks when moving from limited commitment without savings to limited commitment with savings, it is helpful to decompose the effect of savings into two parts: raising the value of autarky, and smoothing originally uninsured risk. The effect of raising the value of autarky is to make participation constraints bind more often, reducing the amount of interpersonal risk-sharing that can be sustained and crowding out transfers. This effect operates at all levels of social distance with two exceptions. First, for high social proximity it might be that, even with access to savings in autarky, participation constraints never bind. Second, for very high social distance, it may be that very little risk sharing is achieved in the absence of savings. Turning to the second effect, the scope for savings to smooth uninsured risk should increase with social distance by Proposition 3.4.5. Consequently, theory does not yield a sharp prediction about how the change in transfers between limited commitment without savings and limited commitment with savings should vary with social distance. Thus:

EMPIRICAL QUESTION 3.4.4. *How does the degree to which interpersonal transfers are crowded out by savings access vary with social distance?*

To answer this question empirically, we run regressions of the form of specification (3.3), but focus on data from the LCNS and LCWS games only—holding lack of commitment constant and varying access to savings. Table 9 shows the results. We do not observe significant social distance effects on how transfers and consumption variability respond to limited commitment. (The fall in transfers across games and social distance can be seen graphically in Figure 7.) The lack of a significant coefficient on the main effect when the outcome is transfers, and the significant negative

coefficient on the main effect of savings access when the outcome is consumption variability, replicate our earlier findings that, on average, savings access does not crowd out interpersonal insurance, and improves consumption smoothing throughout the ex post income distribution.

The sharp prediction that does emerge from the limited commitment model with social sanctions is that individuals in more socially distant pairs should use savings more, because limited commitment leads to more uninsured income variation for these pairs, and consequently, increases the scope for savings to smooth this uninsured risk:

PROPOSITION 3.4.6. *Socially distant pairs use savings more than socially close pairs.*

Because we only observe each individual playing one game with savings, Table 10 omits individual fixed effects to look at how use of savings varies with social distance. We see that the more distant i is from her partner, the more i uses savings, by INR 0.80 per unit of distance, significant at 1%. This offers further evidence that limited commitment binds more at greater social distances, and that savings allow these pairs to smooth some of the resulting uninsured risk.

We now turn to examining whether and how often players defect—that is, transfer less to their partner than they agreed when receiving high income. As noted above, if individuals' forecasts of the value of reneging are subject to an additive error term v, the probability of defection when high income is realized is then the probability of a binding participation constraint, times the probability that v exceeds the surplus the lucky individual had anticipated when receiving high income and making the promised transfer. The probability of a binding participation constraint is decreasing with social proximity, because socially closer pairs get more surplus, *ceteris paribus*, from maintaining their relationship. Therefore the likelihood of defection should be reduced by social networks, thus:

PROPOSITION 3.4.7. *Defection (lucky individuals reneging on the transfers they promised to their partners) should occur less in pairs with greater social proximity.*

Table 11 tests this in our data, showing how defection varies with social distance. To improve statistical power, we split pairs into high (median and above) and low (below-median) bins rather than using a linear measure of social distance. We find that Proposition 3.4.7 is confirmed. Under limited commitment without savings, defection occurs 10 percentage points more often when the pair members are distant than when they are close (significant at 10%). On the other hand, moving from limited commitment without savings to limited commitment with savings, there is neither an overall increase in defection, nor a differential increase by social distance. This is unsurprising because we find little evidence that access to savings tightens participation constraints on average.

3.5 Conclusion and future directions

Our results, from a unique framed field experiment, show that when individuals attempt to share risk, limited commitment matters: consumption smoothing, and welfare, are significantly lower

when players cannot commit ex ante to a risk-sharing agreement than when they can. Further, this effect varies with social distance: for the socially closest pairs, limited commitment does not bind but as social distance increases, limited commitment is increasingly important. Because interpersonal insurance leaves a significant amount of risk uninsured and savings access is found not to crowd out this insurance, savings access is beneficial, allowing individuals to smooth intertemporally some of the income risk that is not insured interpersonally.[38] Pairs that are more socially distant face more uninsured risk, and use savings more. If economic development weakens social ties between individuals, our results for socially distant pairs may be relevant in forecasting how well income risk can be insured and what role financial access might play in improving consumption smoothing.

These findings yield several new insights. First, the comparisons across social distance show that networks matter substantively in dynamic contracting environments, mitigating contracting frictions and facilitating efficient behavior. Moreover, the way the "game outside of the game," i.e., players' other relationships, enters into our experiment is analogous to how it affects real life interactions in a village, since all risk-sharing relationships among villagers are not isolated objects in a dynamic risk-sharing contract, but are embedded in a broader set of repeated interactions. Thus, the role we measure for social proximity is likely to translate to other settings, while not in exact magnitude, in sign and significance.

We introduced social distance to our experiment with a bundled approach, capturing altruism, reciprocity, etc. Unpackaging the bundle of forces in the social network was not the goal of this paper, but is an important avenue for future work. There are at least two strategies that might allow for this. First, one could use methods used by Karlan et. al (2009), Leider et al. (2009) and Ligon and Schechter (2012). This would require framed experiments in which partners engage in interpersonal transfers with varying levels of anonymity (e.g., one partner knows the other's identity, both know each other's identity, neither knows the other's identity) to distinguish between altruism and reciprocity. However, as one may imagine, this would be logistically difficult to execute directly in the field in rural villages while playing dynamic risk-sharing and savings games. Second, one could use methods of random assignment to networks (e.g., classroom or loan group assignment) or shock the network to generate exogenous variation, as in Feigenberg et al. (2011). However, this approach raises the question of whether the shocked network generates the same depth of social ties as pre-existing network links.

Given the important role of social networks, a natural question is whether and how networks

[38] Because we can use consumption variability as an ordinal measure of relative welfare, this comparison does not require assumptions about the utility function, beyond risk aversion: more consumption variability implies lower welfare. However, to get a sense of the magnitude of the welfare differences across the settings we study, we can plug our results for changes in consumption variability into a particular utility function. Assuming a CRRA utility function with a coefficient of relative risk aversion of 2 implies that the move from full commitment (no savings) to limited commitment (no savings) reduces average in-game welfare by 2.7%. Adding savings access to the limited commitment setting cuts the welfare loss in half, to 1.3% lower than under full commitment.

endogenously form to mitigate agency problems. For instance, do individuals choose to rely on socially close friends and relatives for insurance and credit, despite the likelihood of covariate shocks, in order to reduce the risk of opportunistic behavior? We are examining endogenous pair formation in ongoing work (Chandrasekhar et al. 2011).

3.A Figures and Tables

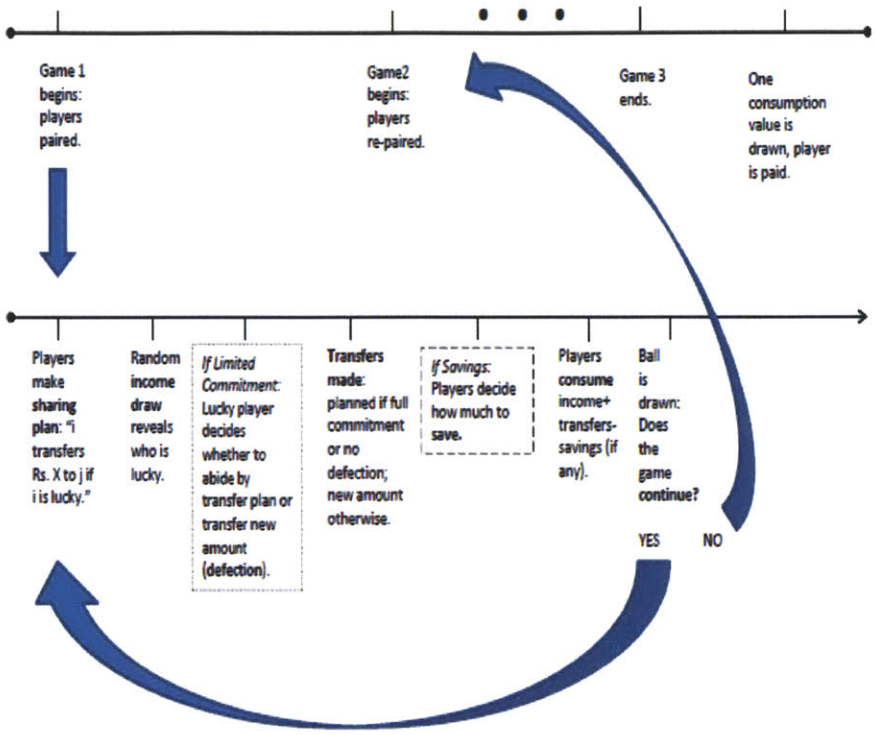


Figure 1: Experiment timeline

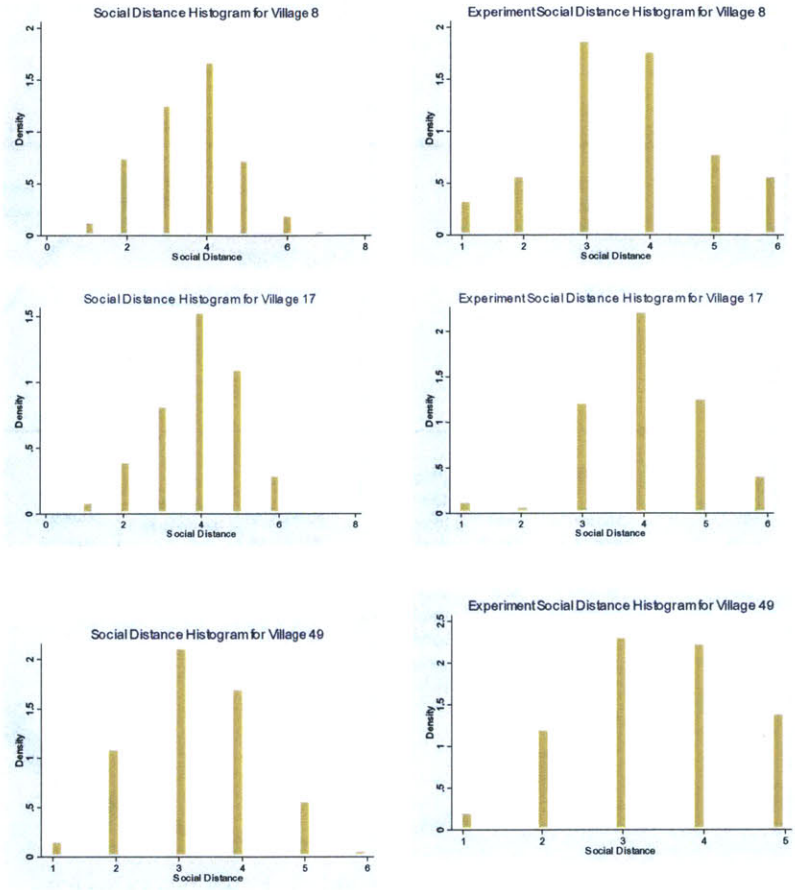


Figure 2: Sampling from the tail of the distribution

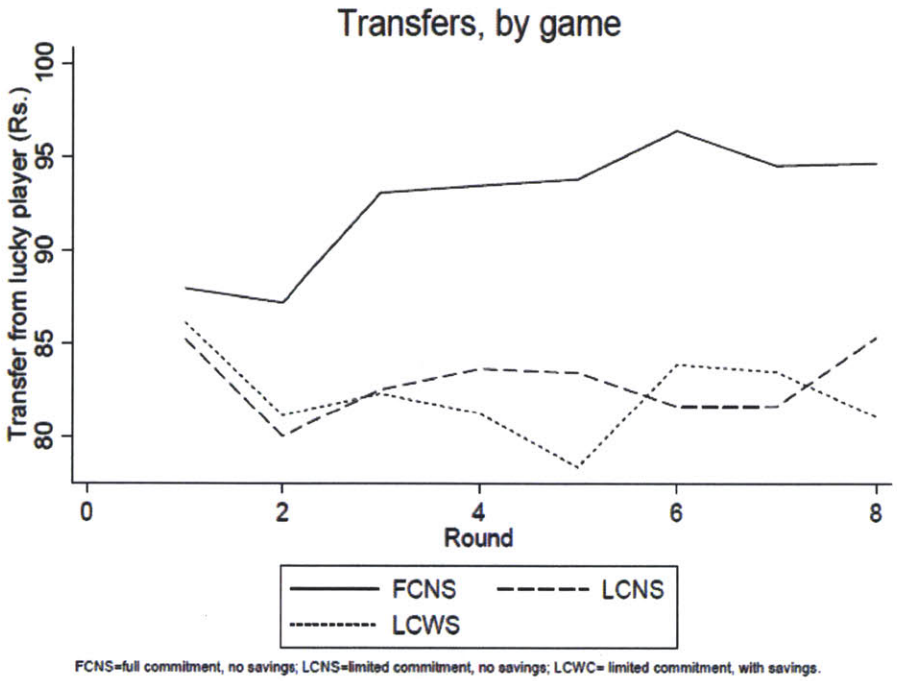


Figure 3: Transfers by game and round

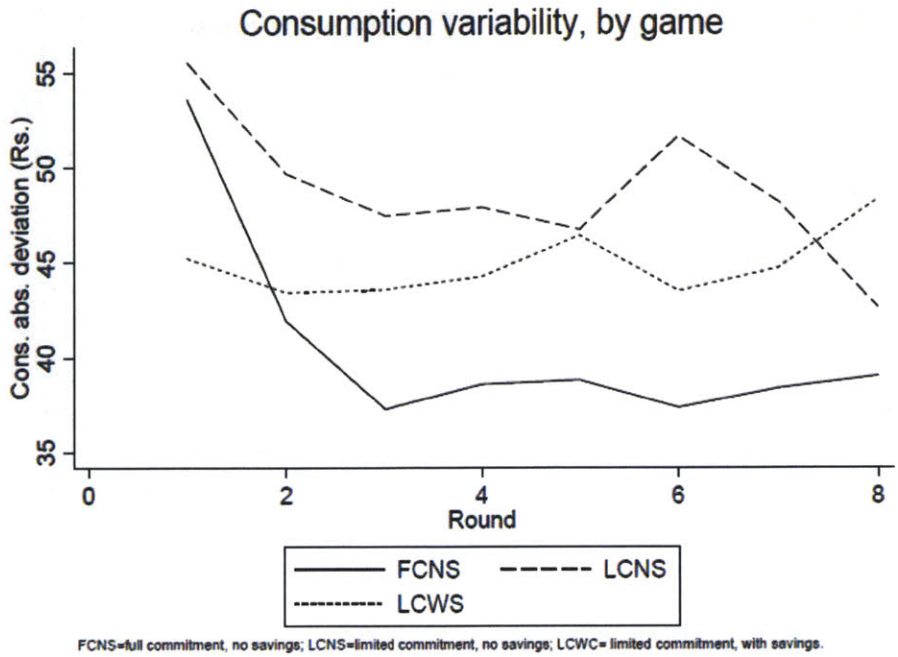


Figure 4: Consumption variability by game and round

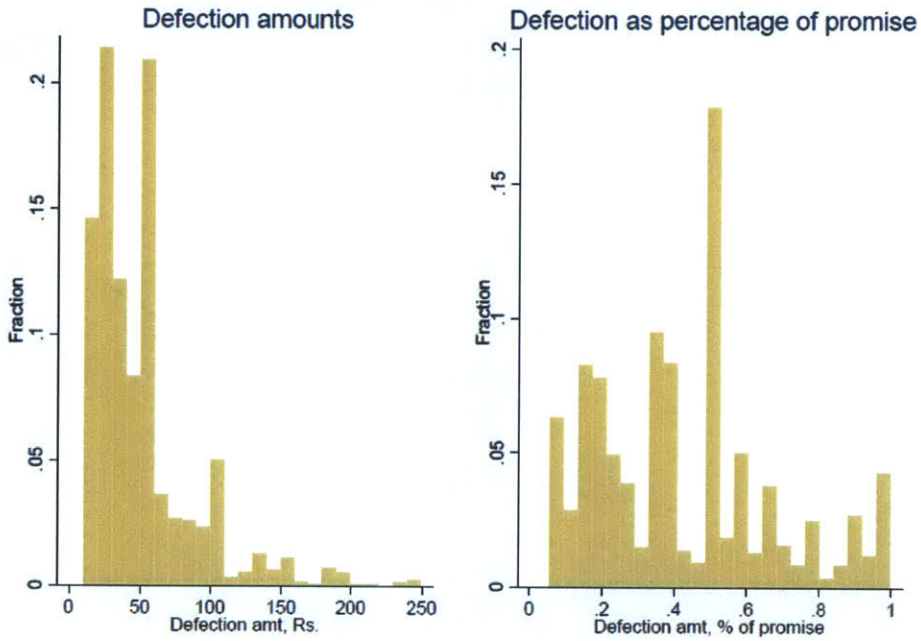


Figure 5a: Defection amounts, conditional on defection

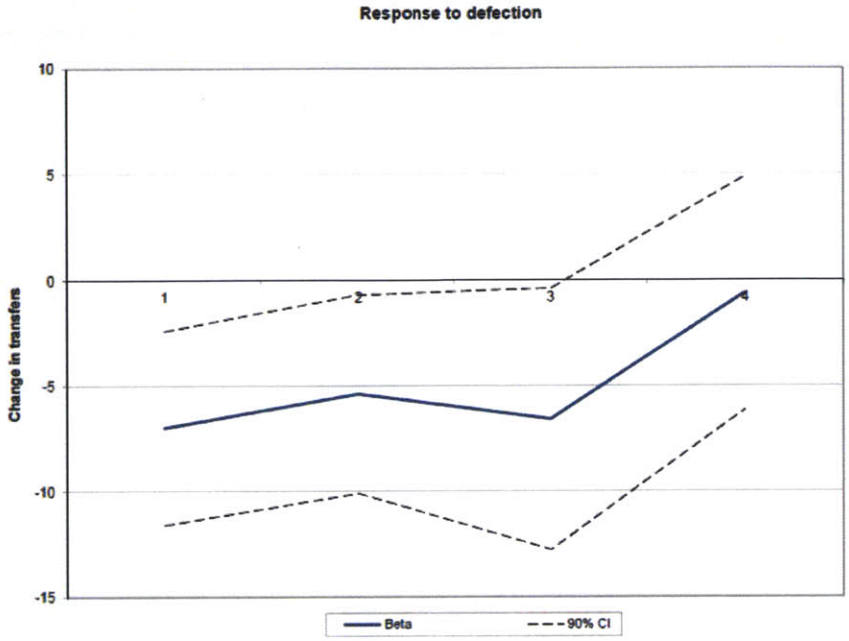


Figure 5b: Response to defection

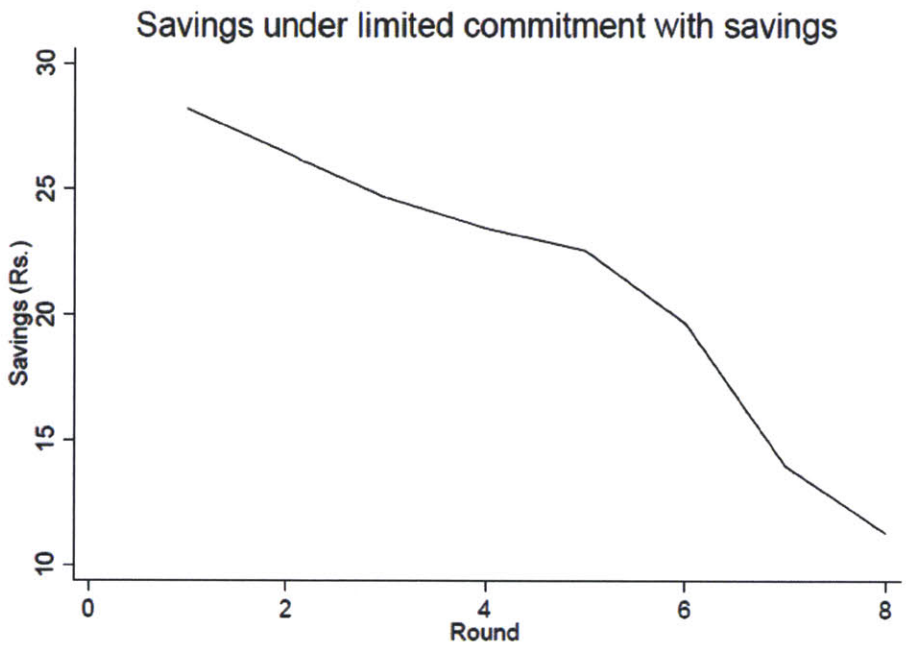


Figure 6: Savings under limited commitment, by round

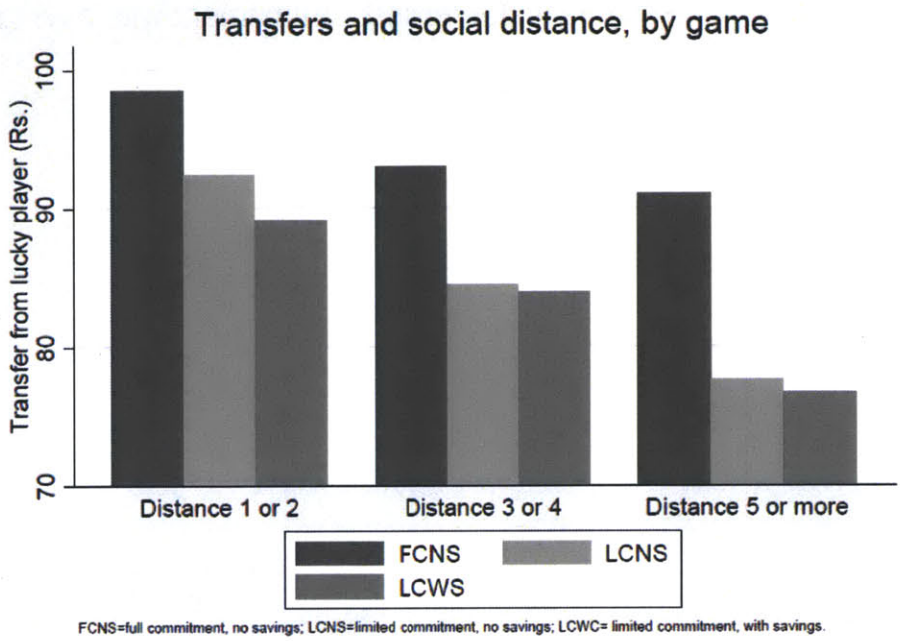


Figure 7: Transfers by social distance and game

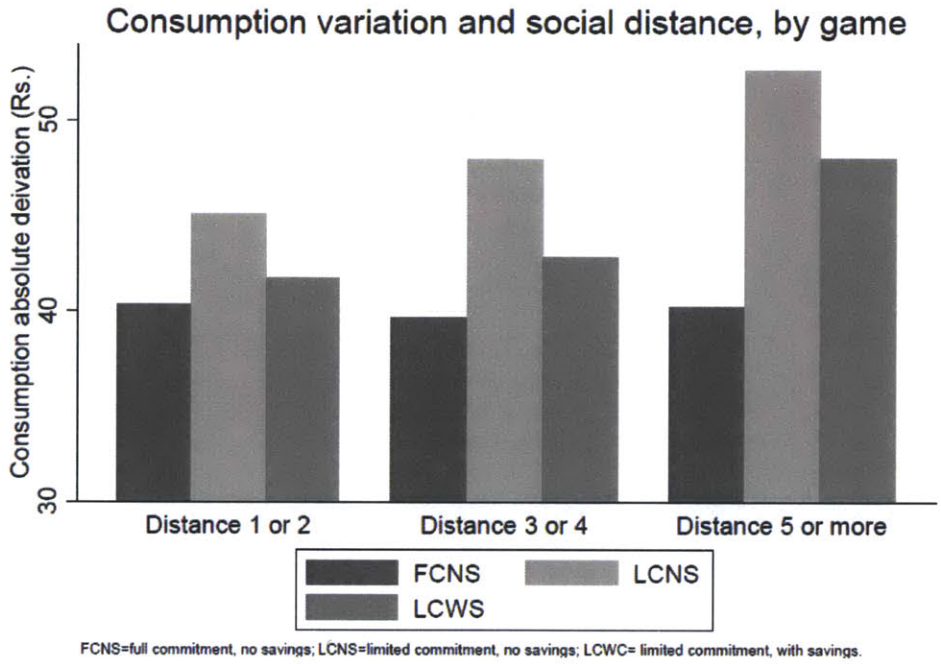


Figure 8: Consumption variability by social distance and game

Table 1a: Summary statistics from survey data

		Mean	N
Roof:	Thatch	0.0107	648
	Title	0.3088	648
	Stone	0.3648	648
	Sheet	0.1776	648
	RCC	0.1008	648
	Other	0.0399	648
Number of Rooms		2.4736	648
Number of Beds		0.9392	648
Has Electricity		0.4183	648
Owner of house		0.8977	648

Notes: Caste variables were only collected for for the random subset of individuals who completed the individual baseline survey. OBC= "other backward caste."

Table 1b: Summary statistics collected in experiment

	Mean	N
Individual-level characteristics		
Male	0.4706	648
Married	0.7324	648
Age	29.9427	648
Education	7.4948	648
Degree	10.1108	622
Eigenvector Centrality	0.0223	622
Pair-level characteristics		
Reachable	0.987	1864
Distance if reachable	3.6337	1840

Notes: See Appendix ?? for definitions of network variables. Reachability defined for pairs with valid network data. Distance defined for reachable pairs.

Table 2: Average transfers, consumption and use of smoothing mechanisms

	(1) Promised Transfers	(2) Realized Transfers	(3) Consumption	(4) Consumption Abs. Dev.	(5) Savings
Panel A: Means and standard deviations					
Full commitment (FCNS)	92.35	92.35	131.04	40.91	-
	(36.31)	(36.31)	(51.97)	(32.05)	
Lim. comm., no savings (LCNS)	92.73	82.73	131.36	49.47	-
	(42.83)	(40.50)	(61.06)	(35.79)	
Lim. comm. with savings (LCWS)	90.79	82.24	128.97	44.65	22.65
	(39.27)	(38.18)	(54.67)	(31.59)	(28.63)
Panel B: Regression-adjusted F-test results					
FCNS=LCNS p-val	0.704	0.000	0.839	0.000	-
FCNS=LCWS p-val	0.049	0.000	0.000	0.001	-
LCNS=LWNS p-val	0.015	0.295	0.000	0.002	-
N	7025	7025	14050	14050	4680

Notes: Standard deviations in parentheses. P-values account for clustering at the village-game level and are regression-adjusted for individual FEs, reachability and distance between partners, surveyor and team effects, and controls for order and round of play (corresponding to the specification in Table 3, below). "Promised transfer" is the amount promised to the unlucky individual (who earned INR 0) by the lucky individual (who earned Rs 250). "Realized transfer" is the amount actually given. "Consumption" is the amount individuals chose to place in their consumption cup. Individuals were paid one randomly chosen consumption value at the end of the game. "Consumption abs. dev." is the deviation of consumption in a a given round from the individual's average consumption throughout the game. "Savings" is the amount individuals chose to place in their savings cup for use in later rounds. * p<.1, ** p<.05, *** p<.01

Table 3: Transfers and consumption smoothing, by treatment and distance

	(1)	(2)	(3)
	Transfers		Consumption
	Unconditional	Conditional	Abs. Dev.
LCNS	-8.99***	-5.61***	8.87***
	[1.56]	[2.05]	[1.35]
LCWS	-11.26***	-6.21***	4.90***
	[21.71]	[1.90]	[1.37]
Reachable	-0.289	-0.323	-6.25
	[6.62]	[11.26]	[4.95]
Reachable	-1.484**	-2.18***	.894*
×Distance	[.665]	[.817]	[.467]
Constant	90.54***	87.99***	52.24***
	[7.15]	[11.14]	[5.54]
LCNS=LCWS			
F-stat	1.111	1.325	10.17
p-value	0.2945	0.2523	0.0019
FCNS mean	92.3512	93.0808	40.912
FCNS std. dev.	36.3129	36.6006	32.0513
N	6369	3845	12752
R^2	0.384	0.448	0.2923
Adjusted R^2	0.312	0.335	0.2533

Notes: Regressions at the individual-game-round level. Regressions include individual-fixed effects, reachability and distance between partners, surveyor and team effects, and controls for order and round of play. Transfer regression includes individuals with high income only. Robust standard errors, clustered at the village by game level, in brackets. Column 1 uses all rounds, column 2 uses rounds where defection has not occurred. Mean of Reachable is 0.987; mean of Reachable × Distance is 3.6337. * p<.1, ** p<.05, *** p<.01

Table 4: Consumption smoothing (absolute deviation of consumption) by in-game income terciles

	(1)	(2)	(3)
		Tercile	
	Lower tercile	Middle tercile	Upper tercile
LCNS	15.53***	4.004**	14.5***
	[3.163]	[1.907]	[2.439]
LCWS	9.968***	4.129**	5.564**
	[3.744]	[1.77]	[2.522]
Reachable	-45.48***	-6.033	16.38*
	[13.34]	[10]	[8.812]
Reach×Distance	0.9573	-0.3695	1.818
	[2.126]	[.7243]	[1.316]
Constant	81.87***	56.41***	27.94***
	[14.28]	[10.92]	[8.873]
LCNS=LCWS			
F-stat	3.255	0.0052	14.3
p-value	0.0743	0.9428	0.00026
FCNS mean	39.7506	40.8573	40.7789
FCNS std. dev.	31.2281	31.8222	31.7478
N	2562	5646	4522
R^2	0.47	0.35	0.37
Adjusted R^2	0.3912	0.2842	0.2993

Notes as in previous table.

Table 5: Transfers and total consumption by endowment

	(1)	(2)	(3)	(4)	(5)	(6)
	FCNS vs. LCNS			LCNS vs. LCWS		
	Cons.	Transfers given	Transfers rec'd	Cons.	Transfers given	Transfers rec'd
Lucky	29.24**	15.37	-13.51	28.88***	-6.64	-2.543
	[13.93]	[15.26]	[16.88]	[10.86]	[15.22]	[14.66]
LCNS	-3.235	-30.41**	-39.04***			
	[15.91]	[14.69]	[9.92]			
LCNS×lucky	-6.334	-6.307	9.613			
	[18.5]	[17.81]	[21.12]			
LCWS				-23.63	-13.06	-4.288
				[14.23]	[12.45]	[13.28]
LCWS×lucky				2.685	16.13	0.0357
				[16.84]	[21.73]	[21.45]
Reachable	4.929	-79.34	51.91	65.38*	19.58	104.7*
	[61.04]	[71.21]	[45.04]	[33.31]	[38.01]	[61.11]
Distance	15.84	-9.396	-0.2837	0.8642	13.43	-42.35***
	[17.91]	[25.92]	[18.18]	[12.01]	[15.75]	[15.71]
Constant	1067***	481.2***	312***	997.7***	254.8***	270.4***
	[58.73]	[57.86]	[38.46]	[48.54]	[30.6]	[57.93]
Reference game	FCNS	FCNS	FCNS	LCNS	LCNS	LCNS
Reference game mean	909.22	319.43	319.07	903.84	284.26	284.18
Std. dev	150.03	134.39	134.28	153.92	137.60	137.57
N	1222	1222	1222	1238	1238	1238
R^2	0.6819	0.671	0.6445	0.7508	0.6966	0.6676
Adjusted R^2	0.3078	0.2839	0.2262	0.4611	0.3439	0.2811

"Lucky" means that player received the INR 60 endowment. Remaining notes as in previous table.

Table 6: Defection rates

LCWS	-0.0207
	[.0133]
Reachable	-.3075**
	[.1205]
Distance	0.002
	[.0093]
Constant	.5078***
	[.1357]
LCNS mean	0.2375
N	4252
R^2	0.4245
Adjusted R^2	0.3183

Notes: Robust standard errors, clustered at the village by game level, in brackets.
p<.1, ** p<.05, *** p<.01

Table 7: Response to defection

	(1)	(2)	(3)	(4)	(5)
RHS variable is transfers from lucky to unlucky player					
Defection	-6.999**				-10.73**
1 Period Ago	[2.805]				[5.075]
Defection		-5.39*			-8.315**
2 Periods Ago		[2.869]			[3.727]
Defection			-6.579*		-6.714
3 Periods Ago			[3.773]		[4.778]
Defection				-0.6261	0.0999
4 Periods Ago				[3.355]	[3.34]
Reachable	11.61	14.94	20.7	17.7	-0.0368
	[8.86]	[15.27]	[17.33]	[20.04]	[18.00]
Reach * Distance	-1.707	-1.887	-1.719	-0.3321	0.1502
	[1.42]	[1.838]	[1.918]	[2.006]	[2.052]
Constant	74.87***	69.94***	63.76***	62.39***	72.08***
	[9.607]	[13.63]	[15.01]	[21.87]	[17.7]
N	1795	1500	1192	884	884
R^2	0.5716	0.6113	0.6729	0.7035	0.714
Adjusted R^2	0.4344	0.4529	0.4873	0.4476	0.4638

Notes as in previous table. Defection is defined as a high-income player transferring less thank promised to her partner.

Table 8: Social distance, consumption smoothing and limited commitment

	(1)	(2)	(3)	(4)
	Transfers		Consumption Dev.	
LCNS	-8.491***	-31.77**	8.786***	33.00***
	[1.426]	[13.94]	[1.227]	[12.34]
Reachable	-8.504	-25.02***	-0.627	17.05**
	[7.655]	[7.705]	[6.386]	[5.99]
Distance	-1.817**	-0.3402	1.128*	-0.2454
	[.9015]	[1.115]	[.6735]	[.8771]
LCNS×Reachable		34.46**		-34.51***
		[15.04]		[12.38]
LCNS×Distance		-2.996*		2.744***
		[1.618]		[1.024]
Constant	102.2***	113.2***	43.73***	31.19**
	[7.878]	[8.014]	[6.753]	[6.180]
FCNS Mean	92.35	92.35	40.91	40.91
FCNS Std. Dev.	36.31	36.31	32.05	32.05
N	4234	4234	8485	8485
R^2	0.45	0.45	0.35	0.36
Adjusted R^2	0.34	0.35	0.30	0.30

Notes: Sample is data for FCNS and LCNS only. Remaing notes as in previous table.

Table 9: Social distance, consumption smoothing and access to savings

	(1)	(2)	(3)	(4)
	Transfers		Consumption Dev.	
LCWS	-1.908	-4.885	-3.961***	-0.3133
	[1.338]	[16.78]	[1.063]	[14.41]
Reachable	6.436	6.291	-16.54***	-14.20
	[6.129]	[14.11]	[6.197]	[13.29]
Distance	-0.7995	-1.159	1.387**	1.339
	[.7912]	[1.147]	[.5984]	[.8598]
LCWS×Reachable		0.6575		-4.631
		[16.85]		[15.55]
LCWS×Distance		0.6497		-0.0823
		[1.289]		[.9407]
Constant	73.64***	75.06***	69.15***	66.97***
	[6.454]	[14.59]	[6.957]	[13.83]
No savings mean	82.73	82.73	49.47	49.47
Std. dev	40.5	40.5	35.79	35.79
N	4252	4252	8507	8507
R^2	0.46	0.46	0.36	0.36
Adjusted R^2	0.36	0.36	0.30	0.30

Notes: Sample is data for LCNS and LCWS only. Remaining notes as in previous table.

Table 10: Savings by distance

Distance	.8311***
	[.3224]
Constant	28.87***
	[2.478]
Distance=1 mean	23.57
Std. dev	24.76
N	4211
R^2	0.22

Notes as in previous table.

Table 11: Defection rates by game and social distance

	(1) FCNS vs. LCNS	(2) LCNS vs. LCPS
High distance	0.0216	-0.0019
	[.0311]	[.0368]
LCNS	.1757***	
	[.0193]	
LCWS		-0.0216
		[.0188]
High distance×LCNS	.1033***	
	[.0273]	
High distance×LCWS		0.0018
		[.0267]
Reachable	-0.0295	-.3094**
	[.0657]	[.1248]
Distance	-.0213*	0.0023
	[.0125]	[.0165]
Constant	0.0376	.5303***
	[.0589]	[.1329]
Reference game	FCNS	LCNS
Reference game mean	0.0000	0.2375
N	4234	4252
R^2	0.4062	0.4245
Adjusted R^2	0.2967	0.3179

Notes as in previous table.

3.B Model setups and proposition proofs

Limited commitment with and without savings

3.B.1 Groups, income, and utility

We consider risk-sharing groups composed of two individuals, $i = 1, 2$. In each period $t = 1, 2, ...,$ individual i receives an income $y^i(s) \geq 0$ of a single good, where s is an i.i.d. state of nature drawn from the set $S = \{1, 2\}$. Income follows the process:

$$y^i(s) = \begin{cases} y & \text{if } i = s \\ 0 & \text{otherwise} \end{cases}.$$

The income process is i.i.d. across time, and perfectly negatively correlated ($\rho = -1$) across individuals. In other words, in each period, one individual will earn positive income y while the other individual will earn no income, with each player equally likely to be lucky. There is no aggregate risk: total group income is y each period.

Individuals have a per-period von Neumann-Morgenstern utility of consumption function $u(c^i)$, where c^i is the consumption of individual i. We assume that $c^i \geq 0$. Individuals are assumed to be risk averse, with $u'(c^i) > 0$, and $u''(c^i) < 0$ for all $c^i > 0$. Individuals are infinitely lived and discount the future with a common discount factor β.[39]

Individuals may enter into risk sharing agreements with their partners. A contract $\tilde{\tau}(\cdot)$ will specify for every date t and for each history of states, $h_t = (s_1, s_2, ..., s_t)$, a transfer $\tilde{\tau}^1(h_t)$ to be made from individual 1 to individual 2, and correspondingly a transfer $\tilde{\tau}^2(h_t)$ to be made from individual 2 to individual 1. For simplicity we denote $\tau^i(h_t) \equiv \tilde{\tau}^i(h_t) - \tilde{\tau}^j(h_t)$, that is, the (positive or negative) net transfer that individual i makes to individual j after history h_t.

Denote $V^i(h_t)$ to be the continuation value of remaining in the insurance agreement, that is, the expected utility of individual i from a contract from period t onwards, discounted to period t, if history $h_t = (h_{t-1}, s_t)$ occurs up to period t and s_t is already known. $V^i(\cdot)$ obeys the recursive relation (Spear and Srivastava 1987):

$$V^i(h_t) = u\left(y^i(h_t) - \tau^i(h_t)\right) + \beta \mathbb{E}_{h_{t+1}|h_t} V^i(h_{t+1}). \tag{3.4}$$

where $\tau^i(h_t)$ follows optimally from (3.8).

[39] In our experiment the $\beta = \frac{5}{6}$, the chance the game will continue after each period, as explained in Section 3.3.

3.B.2 The role of savings

In some of the cases we consider below, individuals have access to a savings technology. The gross return on savings is assumed to be

$$R = \begin{cases} 1 \text{ when saving is available} \\ 0 \text{ otherwise} \end{cases}.$$

That is, when saving is available, one unit of the consumption good saved today delivers one unit in the next period. Savings amounts are restricted to be positive: no borrowing is possible.

In the case that individuals have access to a savings technology, a risk-sharing contract will not only determine net transfers $\tau^1(s_t)$ to be made from individual 1 to individual 2 but also an amount $z^i(s_t)$ that an individual i, for $i = 1, 2$, saves from period t to period $t+1$. For simplicity we then denote as a sharing agreement $(\tau(s_t), z(s_t)) = (\tau^i(s_t), z^i(s_t))$ for $i = 1, 2$.

For the case that individuals have access to a savings technology $V^i(\cdot)$ is denoted as

$$V^i\left(h_t, z^i(h_{t-1})\right) = u\left(z^i(h_{t-1}) + y^i(h_t) - \tau^i(h_t) - z^i(h_t)\right) + \beta \mathbb{E}_{h_{t+1}|h_t} V^i\left(h_{t+1}, z^i(h_t)\right) \quad (3.5)$$

where $\tau^i(h_t), z^i(h_t)$ follow optimally from (3.17) on page 143.

3.B.3 Autarky

Thus far we have assumed that individuals can make transfers with other individuals. However, individuals may choose not to make such transfers. In particular, they might initially promise to make certain transfers, but later change their minds. To characterize the payoffs to an individual who reneges on promises to his or her partner, we assume that if either party reneges upon the contract, both individuals consume autarky levels thereafter. The grim trigger or "autarky forever after defection" case is used for expositional clarity and because it supports the most on-equilibrium risk-sharing. In our experimental setup, players are free to choose any post-defection response. The qualitative properties of the equilibrium do not depend on the grim trigger assumption, as argued by Ligon et al. (2002).

If individuals have access to a savings technology they can smooth consumption only intertemporally. Without a savings technology, an individual in autarky will simply live "hand to mouth," consuming his or her income in each period. By choosing not to make transfers with others, an individual gives up the benefits of interpersonal consumption smoothing: the option to receive transfers from others when unlucky, in exchange for making transfers to others when lucky is lost forever. When individuals are risk-averse, such interpersonal insurance will be welfare-enhancing, and giving it up is a cost of choosing autarky instead. (We discuss below why individuals might make this choice.) There may also be other costs of choosing autarky, which we consider next.

3.B.3.1 Autarky without savings

If individuals do not have access to a savings technology, then after the violation of a contract both individuals consume their income in every period. Denote $V^i_{A,NS}(s_t)$ to be the expected utility of autarky for an individual i, who has reneged against individual j in period t, after observing s_t:

$$V^i_{A,NS}(h_t) = u\left(y^i(s_t)\right) + \beta \mathbb{E}_{h_{t+1}} V^i_{A.NS}(h_{t+1}) \tag{3.6}$$

There is no maximization because, in autarky with no savings, i simply consumes her income each period.

3.B.3.2 Autarky with savings

If individuals have access to a savings technology, and its use cannot be barred from those who have defected, then after the violation of a contract individuals are not constrained to consume their income period-by-period as they can make use of the storage technology. After the violation of a contract, both individuals keep any savings they have.

We denote $V^i_{A,S}(h_t, z^1_{t-1})$ to be the expected utility of autarky for an individual i in period t with savings z^1_{t-1}, after observing s_t:

$$V^i_{A,S}\left(h_t, z^1_{t-1}\right) = \max_{z^i(h_t)} u\left(z^i_{t-1} + y^i(s_t) - z^i_t(h_t)\right) + \beta \mathbb{E}_{h_{t+1}} V^i_{A.S}\left(h_{t+1}, z^1_t\right) \tag{3.7}$$

Unlike the no-savings case, i has a choice variable, namely $z^i(h_t)$, the amount of savings that will be carried into the next period.

3.B.4 Risk-sharing with no commitment, no savings

We now set up the problem characterizing the set of constrained efficient risk-sharing contracts for the case where there is no access to savings. As a risk-sharing contract can be seen as a non-cooperative equilibrium of a repeated game, and reversion to autarky is the most severe subgame-perfect punishment, this assumption allows us to characterize the most efficient set of non-cooperative subgame-perfect equilibria (Abreu 1988).

The set of efficient risk-sharing contracts for the no commitment, no savings case solves the following dynamic programming problem[40]:

$$\begin{aligned} V^1\left(V^2_t(s_t)\right) \quad = \quad & \max_{\substack{\tau^1(s_t), \\ \left\{V^2_{t+1}(s_{t+1})\right\}_{s \in S}}} \left\{u\left(y^1(s_t) - \tau^1(s_t)\right) + \beta \mathbb{E}_{s_{t+1}} V^1\left(V^2_{t+1}(s_{t+1})\right)\right\} \\ & \text{s.t.} \end{aligned} \tag{3.8}$$

[40] This will also be the set of decentralizable equilibrium allocations since the conditions of the 2nd welfare theorem are satisfied.

$$\lambda \quad : \quad u\left(y^2\left(s_t\right)+\tau_t^1\left(s_t\right)\right)+\beta \mathbb{E}_{s_{t+1}} V_{t+1}^2\left(s_{t+1}\right) \geq V_t^2\left(s_t\right), \forall\, s_t \in S \tag{3.9}$$

$$\beta\phi_t \quad : \quad V_{t+1}^2\left(s_{t+1}\right) \geq V_{A,NS}^2\left(s_{t+1}\right)-f\left(\gamma(2,1)\right), \forall\, s_{t+1} \in S \tag{3.10}$$

$$\beta\mu_t \quad : \quad V^1\left(V_{t+1}^2\left(s_{t+1}\right)\right) \geq V_{A,NS}^1\left(s_{t+1}\right)-f\left(\gamma(1,2)\right), \forall\, s_{t+1} \in S \tag{3.11}$$

$$\psi_1 \quad : \quad y^1\left(s_t\right)-\tau_t^1\left(s_t\right) \geq 0, \forall\, s_t \in S \tag{3.12}$$

$$\psi_2 \quad : \quad y^2\left(s_t\right)+\tau_t^1\left(s_t\right) \geq 0, \forall\, s_t \in S \tag{3.13}$$

where $V_{A,NS}^i\left(s_t\right)$ is as in (3.6). We have written $\tau^1\left(s_t\right)$ and $V_t^2\left(s_t\right)$ instead of $\tau^1\left(h_t\right)$ and $V_t^2\left(h_t\right)$ because, due to the recursive nature of the problem, all previous history of the efficient risk-sharing contract is encoded in s_t. This recursivity also allows us to write $V^1\left(V_t^2\left(s_t\right)\right)$ instead of $V_t^1\left(V_t^2\left(s_t\right)\right)$, because player 1's value function will be the same whenever an amount $V_t^2\left(s_t\right)$ is promised to player 2 (Ligon et al. 2002).

Due to the strict concavity of $u\left(c^i\right)$, it follows that $V_t^i\left(\cdot\right)$ is also strictly concave for $i=1,2$. The set of constraints is convex (this follows from the concavity of $u\left(\cdot\right)$ and the linearity in $V^i\left(\cdot\right)$). Consequently, the problem is concave, and the first-order conditions are both necessary and sufficient.

The first-order conditions for this problem are the following:

$$\tau_t^1\left(s_t\right) \quad : \quad \frac{u'\left(y^1\left(s_t\right)-\tau_t^1\left(s_t\right)\right)}{u'\left(y^2\left(s_t\right)+\tau_t^1\left(s_t\right)\right)}=\lambda-\frac{\psi_1-\psi_2}{u'\left(y^2\left(s_t\right)-\tau^1\left(s_t\right)\right)}, \forall\, s_t \in S, \tag{3.14}$$

$$V_t^2 \quad : \quad -V^{1\prime}\left(V_t^2\left(s_t\right)\right)=\frac{\lambda+\phi_t}{\left(1+\mu_t\right)}, \forall\, s_t \in S. \tag{3.15}$$

Further, the envelope condition is given by

$$V^{1\prime}\left(V_t^2\left(s_t\right)\right)=-\lambda, \forall\, s_t \in S. \tag{3.16}$$

The terms $f\left(\gamma(i,j)\right)$ do not enter directly into the first-order conditions since they enter the problem additively, but their presence will affect the likelihood of the continuation constraints (3.10) and (3.11) binding, as discussed below.

Ligon et al. (2002) note that a constrained efficient risk-sharing contract can be characterized in terms of the evolution of λ, the multiplier on individual 2's promise-keeping constraint, which from (3.16) measures the rate at which individual 1's expected utility can be traded off against that of individual 2, once the current state is known. (This is a measure of individual 2's relative importance or bargaining power within the relationship.) Once the state of nature for the following period s_{t+1} is known, the new value for λ is determined by equation (3.15). Furthermore, λ completely determines the current transfers (3.14) once the state of nature s_t has been realized.

The intuition for this result is the following. For simplicity assume that the non-negativity

constraints never bind,[41] and hence that $\psi_1 = \psi_2 = 0$. Then, we can rewrite (3.14) as

$$\lambda = \frac{u'\left(y^1\left(s_t\right) - \tau\left(s_t\right)\right)}{u'\left(y^2\left(s_t\right) - \tau\left(s_t\right)\right)}$$

Proof. *[Proof of Proposition 3.4.1]* The first-best risk sharing contract keeps the ratio of individuals' marginal utilities constant across states and over time. Then, if (3.10) and (3.11) never bind, λ never updates, and hence full insurance can be achieved. Then individuals each consume a constant share of the per-period endowment y where the share is given by the initial value of λ, λ_0. However, if either (3.10) or (3.11) ever bind, λ is no longer constant and full insurance is no longer achievable. Because the only player who may be constrained is the player with the high income realization, who would be required to make a transfer to the other under full insurance, binding continuation constraints will cause consumption to be positively correlated with income (Coate and Ravallion 1993). Moreover, when consumption is positively correlated with income, transfers are lower than under full insurance. □

3.B.5 No commitment, with savings

As before, if either party reneges upon the contract, both individuals consume autarky levels thereafter. However, now after the violation of a contract, individuals are not constrained to consume their income period-by-period as now they can make use of the storage technology. After the violation of a contract, both individuals keep any savings they have.

The set of efficient risk-sharing contracts for the no commitment case with savings solves the following dynamic programming problem:

$$V^1\left(V_t^2\left(s_t, z_{t\text{-}1}^2\right), z_{t\text{-}1}^1\right) = \tag{3.17}$$
$$\max_{\substack{\tau_t^1(s_t), z_t(s_t) \in \mathbb{R}^+, \\ V_t^2\left(s_{t+1}, z_t^2\right)}} \left\{u\left(z_{t-1}^1 + y^1\left(s_t\right) - \tau_t^1\left(s_t\right) - z_t^1\left(s_t\right)\right) + \beta \mathbb{E}_{s_{t+1}} V^1\left(V_t^2\left(s_{t+1}, z_t^2\right), z_t^1\right)\right\}$$
s.t.

$$\begin{aligned}
\lambda \quad &: \quad u\left(z_{t-1}^2 + y^2\left(s_t\right) + \tau_t^1\left(s_t\right) - z_t^2\left(s_t\right)\right) + \beta \mathbb{E}_{s_{t+1}} V_t^2\left(s_{t+1}, z_t^2\right) \qquad (3.18)\\
&\geq \quad V_t^2\left(s_t, z_{t-1}^2\right), \forall\ s_t \in S \\
\beta\phi_t \quad &: \quad V_t^2\left(s_t, z_{t-1}^2\right) \geq V_{A,S}^2\left(s_t, z_{t-1}^2\right) - f\left(\gamma(2,1)\right), \forall\ s_t \in S \qquad (3.19)\\
\beta\mu_t \quad &: \quad V_t^1\left(V_t^2\left(s_t, z_{t-1}^2\right)\right) \geq V_{A,S}^1\left(s_t, z_{t-1}^2\right) - f\left(\gamma(1,2)\right), \forall\ s_t \in S \qquad (3.20)\\
\psi_{1t} \quad &: \quad z_{t-1}^1 + y^1\left(s_t\right) - \tau_t^1\left(s_t\right) - z_t^1\left(s_t\right) \geq 0, \forall\ s_t \in S \qquad (3.21)\\
\psi_{2t} \quad &: \quad z_{t-1}^2 + y^2\left(s_t\right) + \tau_t^1\left(s_t\right) - z_t^2\left(s_t\right) \geq 0, \forall\ s_t \in S \qquad (3.22)
\end{aligned}$$

where as before the problem is characterized recursively, and $V_{A,S}^i\left(s_t\right)$ is as in (3.7). Note that now the constraint set is non-convex due to (3.19) and (3.20) and consequently the problem may not

[41] This is automatic when $\lim_{c \to 0} u(c) = -\infty$, as is the case for CRRA utility with relative risk aversion > 1.

be concave. To avoid such issues, lotteries can be used to convexify the problem, as in Ligon et al. (2000).

Proof. [*Proof of Proposition 3.4.3*] As noted by Ligon et at. (2000), savings access tightens participation constraints when $V^i_{A,S}(s_t) > V^i_{A,NS}(s_t)$: when autarky with savings is preferable to autarky without savings ("hand to mouth"). If participation constraints are more likely to bind, the correlation between consumption and income is increased, and transfers fall. □

Proof. [*Proof of Proposition 3.4.2*] Since the value of the endowment is known when the sharing agreement is made, $S = 1$ (there is only one possible state for each partner: the one that was realized). Therefore, for the high-endowment partner (say, individual 1), the promise-keeping plus participation constraints require that individual 1's promised total consumption increases by INR 30. □

3.B.5.1 The role of social networks

Proof. [*Proof of Proposition 3.4.4*] Ceteris paribus, participation constraints are less likely to bind when partners are socially close, and hence that transfers fall more under limited commitment when social distance is greater. To see this, assume that after some history, i is just indifferent between reneging and staying in the insurance agreement with j when i is lucky (when income is y), for a given promised transfer $\tau^i_t(y)$, promised utility $V^i_t(y)$, and penalty, $f(\gamma(i,j))$, meaning that i's participation constraint binds when i's income is y. Now, decrease the social distance between i and j, holding the promised transfer and promised utility fixed. Since i was just indifferent between reneging and staying at the lower penalty, when the penalty increases, i will no longer be tempted to renege. Thus, denoting as ϕ_{it} the Lagrange multiplier on i's time t participation constraint, and taking expectations over the possible states of nature at t:

$$\frac{\partial \mathbb{E}_{t-1}\phi_{it}}{\partial f(\gamma(i,j))} < 0.$$

and similarly for i's partner, j. The expected magnitude of the multiplier on the promise-keeping constraint is lower the greater the penalty for reneging, i.e., the lower the pair's social distance.

Manipulating the first-order conditions on the limited commitment no-savings problem (3.11), (3.10) and (3.16) yields the following relationship between i and j's marginal utilities, as a function of i's relative bargaining power λ_{it} :

$$\lambda_{it} = \frac{u'(y_{jt} + \tau^j_t)}{u'(y_{it} + \tau^i_t)}$$

and the following updating rule for the multiplier on i's time t promise-keeping constraint (Ligon et al. 2002):

$$\lambda_{i,t+1} = \lambda_{it}\left[\frac{1+\phi_{i,t+1}}{1+\phi_{j,t+1}}\right]$$

This yields the following expression for the ratio of i and j's time $t+1$ marginal utility:

$$\frac{u'(y_{j,t+1} - \tau^i_{t+1})}{u'(y_{i,t+1} + \tau^i_{t+1})} = \frac{u'(y_{jt} + \tau^j_t)}{u'(y_{it} + \tau^i_t)}\left[\frac{1+\phi_{i,t+1}}{1+\phi_{j,t+1}}\right] \tag{3.23}$$

Therefore, the more often i or j have binding participation constraints (i.e., a positive ϕ_{it} or ϕ_{jt}), and the more binding they are (larger positive values of ϕ_{it} or ϕ_{jt}), the more each player's consumption $c_{it} = y_{i,t+1} - \tau^i_{t+1}$ is expected to vary. Thus, when participation constraints are more binding, less interpersonal insurance is possible. This implies that players will on average transfer less to each other under limited commitment when they are more socially distant. □

Proof. [*Proof of Proposition 3.4.5*] Proposition 3.4.4 implies that, under limited commitment, consumption is more strongly correlated with contemporaneous income when social distance is greater. Hence, consumption smoothing is

worse under limited commitment when social distance is greater. □

Proof. [*Proof of Proposition 3.4.6*]

From Ligon et al. (2000)'s equation (14), the motive to save arises from the expectation that, without savings, expected marginal rates of substitution would differ across dates. By our Proposition 3.4.5, with savings, consumption smoothing is worse, i.e. expected marginal rates of substitution differ more, the more socially distant the pair. Therefore distant pairs have the greatest incentive to save. □

Proof. [*Proof of Proposition 3.4.7*] We assume that the value of reneging on a particular promise contains an additive, mean-zero, i.i.d. error term, v, unforecastable by the individual. In the no savings case:

$$V^i_{A,NS}(s_t) = u\left(y^i(s_t)\right) + \beta \mathbb{E}_{h_{t+1}} V^i_{A,NS}(s_{t+1}) + v^i_t$$

and similarly in the case with savings. The probability of defection when y_H is realized is then the probability that v exceeds the surplus the lucky individual had anticipated when receiving y_H and making the promised transfer τ. In the no-savings case:

$$\Pr(defect) = \Pr\left(v^i_t > u\left(y_H - \tau^j_t(y_H)\right) + \beta \mathbb{E}_{s_{t+1}} V^i_{t+1}(s_{t+1}) - V^i_A(y_H)\right)$$

By Proposition 3.4.4, the surplus i obtains from not reneging is, ceteris paribus, decreasing in the social distance between i and j. Therefore, the probability of defection is increasing in the social distance between i and j. □

Proof. [*Proof of Lemma 3.3.1*] If players 1 and 2 fully insure their idiosyncratic risk ($\alpha = 1$), and player 1 has a Pareto weight/bargaining power factor of λ, 1 transfers an amount

$$\tau^1_{FI} = (1-\lambda)\,250$$

to 2 when 1 is lucky, and 2 transfers an amount

$$\tau^2_{FI} = \lambda 250$$

to 1 when 2 is lucky. Since each player is lucky 50% of the time on average, average transfers will be

$$.5\tau^A_{FI} + .5\tau^2_{FI} = .5\,(\lambda + 1 - \lambda)\,250 = 125$$

regardless of λ. Similarly, if players 1 and 2 insure, on average, fraction α of their idiosyncratic risk, $\tau^1_\alpha = \alpha\,(1-\lambda)\,250$ and $\tau^2_\alpha = \alpha\lambda 250$, and average transfers will be

$$.5\tau^1_\alpha + .5\tau^2_\alpha = \alpha 125$$

Even if transfers change over the course of the game in response to binding participation constraints, as we expect to happen in a limited commitment setting, average transfers will be $\alpha 125$, where α is the fraction of risk that is insured, averaging across rounds. Note that the independence of average transfers and bargaining weights relies on the fact that the income process is independent of bargaining weights. This holds in our setting because each player has a 50% chance of being lucky or unlucky in each round. However, in non-experimental data, bargaining weights would typically be correlated with the individuals' income processes, and it would not be possible to map average transfers into the degree of insurance without knowledge of bargaining weights. □

3.C Network Statistics

Here we introduce basic social network terminology used in the paper. A graph or network $G := (V; E)$ consists of a set of vertices, V, and edges,E. We assume that the graph is undirected *undirected* and *unweighted*; households i and j are either connected or not and this relationship is symmetric. The graph is represented by its adjacency matrix $A := A(G)$, where $A_{ij} = \mathbf{1}\{ij \in E\}$ indicates whether i and j share an edge.

The graph admits a natural metric of distance between nodes. The *social distance* between i and j is given by the geodesic, or the shortest path

$$\gamma(i,j) := \min_{k \in \mathbb{N}} \left[A^k\right]_{ij} > 0$$

which is finite if there exists a path between nodes i and j through the graph. We define *reachability* of i and j as

$$\rho(i,j) := \mathbf{1}\{\gamma(i,j) < \infty\}$$

which indicates whether such a path exists. If i and j are in the same connected component of the network, then they are certainly reachable. We caution that it is essential to control for reachability when studying sampled networks as individuals with few links, who are distant from most other households, may appear in the sampled data having only close ties where the social distance is finite. Encoded in this manner, sign switching may occur if reachability is not controlled for (Chandrasekhar and Lewis 2010).

The social network data of Banerjee et al. (2011) contains data from 12 dimensions of relationships: (1) visitors who come to the household, (2) households that a person visits, (3) relatives, (4) non-related friends, (5) those who provide medical advice, (6) those with whom one goes to temple, (7) those from whom one borrows material goods, (8) those to whom one lends material goods, (9) those from whom one borrows money, (10) those to whom one lends money, (11) those to whom a person gives advice, and (12) those from whom one receives advice. Instead of working with the multigraph or constructing an ad hoc weighted graph, we take the approach of Banerjee et al. (2011) and consider the union network, $G^{all} = (V, \cup_{r=1}^{R} E^r)$, where

$$A_{ij}^{all} := 1\left\{\sum_{r=1}^{R} A_{ij}^r > 0\right\},$$

where r indexes the dimension of the relationship. The network studied in this paper is precisely G^{all} constructed in the above manner.

Chapter 4

Testing Models of Social Learning on Networks: Evidence from a Lab Experiment in the Field

4.1 Introduction

Understanding the mechanisms of social learning is very important, particularly in developing nations. Due to the lack of formal institutions, missing markets, and information aggregating mechanisms, agents in developing economies often rely on social connections for information and opportunities. Social learning is often characterized by word-of-mouth communication. Given that social learning is a fundamental component of many economic processes, before employing models of social learning to make policy recommendations we must first understand which models best describe features of empirical social learning. The mechanics of the learning process are of policy interest. If the features of social learning are better described by certain models, those models should be the environment in which the relevant economic outcomes are studied. The two leading classes of social learning models employed in the network literature are Bayesian learning and DeGroot rules of thumb models. The models often exhibit distinct behavior. Individuals employing rules of thumb often double-count information and may not reach consensus in the long run.

In this paper, we will address the question of whether Bayesian learning or DeGroot rules of thumb models do a better job of describing empirical learning processes on networks. To study this question we conduct a unique lab experiment in the field across 19 villages in rural Karnataka, India. We ran our experiments in villages so that we could study the relevant population of interest, namely those who could be potentially targeted by policy that depends greatly on social learning. Our approach was to design simple networks that provide statistical power to distinguish between the different learning models. We then conducted a lab experiment in the field using these networks to address the proposed question.

We created networks of seven individuals and gave each individual a map of the entire graph so that the full informational structure was comprehended. The underlying state of the world was either one or zero with equal probability. At $t = 0$ each individual received an independent identically distributed (iid) signal about the underlying state of the world and were informed that signals were correct with probability $5/7$. After receiving the signal each individual privately made a guess about the state of the world. These guesses were communicated to each individual's network neighbors at the start of the first period, $t = 1$.

Thereafter, in any given period, each individual knew the guesses of all of her network neighbors from all past periods. Using this information, she made a guess about the state of the world, which in turn was communicated to each of her network neighbors at the beginning of the following period. Every individual was paid for her guess in a randomly chosen round from the set of rounds that she played that day over the course of all the experiments. Consequently, participants had strong incentives to make their best guess in each round.[1]

The work most closely related to ours is Mobius et al. (2012), who study how information decays as it spreads through a network. They test between DeGroot models and a streams model that they develop in which individuals "tag" information by describing where it comes from. Their experiment uses Facebook network data in conjunction with a field experiment and finds evidence in favor of the streams model. In our experiment, we shut down this ability for individuals to "tag" information to be able to compare the Bayesian model to DeGroot alternatives.

We are able to analyze the data at two levels: the network level and the individual level. Network level analysis considers the entire network and sequence of actions as a single observation. That is, theory predicts a path of actions under a model of social learning, for each individual in each period given a network. At the network level, we address a question about how well social learning behaves; the observational unit in this approach is the social network itself. Meanwhile individual level analysis considers the action of an individual, given a history, as the observational unit.

Our core results are as follows. First, at the network level, we find evidence that a DeGroot rule of thumb model better explains the data than the Bayesian learning model. This is not to say, however, that the social learning process does not resemble the data generated by Baysian learning. In fact, the Bayesian learning model explains 62% of the actions taken by individuals while the best DeGroot rule of thumb explains over 76% of the actions taken by individuals.[2]

Second, at the individual level, we find that a DeGroot rule of thumb model of learning performs significantly better than Bayesian learning in explaining the actions of an individual given a history

[1]While it could be the case that players were extremely sophisticated and engaged in experimentation in early rounds, anecdotal evidence from participants suggests that this is not the case. In addition, the theoretical and experimental literature uses this assumption.

[2]When we say a model explains $x\%$ of the actions, we are interested in $x := \frac{y-50}{50}$, where y is the percent of actions predicted correctly. This is the right normalization since we could always explain half the actions by flipping a coin.

of play. In fact this model explains nearly 87% of the actions taken by individuals given a history.

Third, we make a methodological contribution by identifying key problems in estimating models of Bayesian learning on networks with trembles or quantal response equilibrium (QRE), the standard technique in the literature. We demonstrate that networks that are small enough to avoid computational constraints are not large enough to tease out the differences between DeGroot and Bayesian learning with trembles. Meanwhile those that are large enough to separate the models become computationally infeasible to study using trembles or QRE.

We also establish several supplementary results which may also be of independent methodological interest. First, we identify a key asymmetry in the social learning on networks literature. Usually this literature has focused on asymptotic learning, whether individuals end up learning the state of the world as time goes to infinity. Models of Bayesian learning have focused on action models; namely, individuals observe the actions taken by their neighbors and update their beliefs accordingly. Under some regularity conditions, this model of learning has nice asymptotic properties such as consensus formation. Meanwhile, the literature emphasizing DeGroot learning, e.g. (Golub and Jackson 2010a), has focused on a communication model wherein each individual communicates her belief to her neighbors every period. This literature, too, has found that asymptotic learning happens with minimal regularity conditions. However, we point out that such models allow for myopic individuals (those following rules of thumb) to transmit too much information relative to the Bayesian learning models. By considering the fair comparison, an action model of DeGroot learning, we highlight an example of networks which satisfy the properties to have asymptotic learning under both Bayesian (action) and DeGroot (communication) learning models, but fail to have asymptotic learning with DeGroot (action) learning models.

Second, we establish an algorithm to simulate Bayesian learning on networks which is computationally tight in the sense that asymptotically there can be no faster algorithm. Namely, the algorithm is $O(T)$ were T is the number of rounds played. This algorithm is actually asymptotically efficient given our setup.[3]

Third, we provide a discussion of why such questions of social learning may be better addressed in a structural manner and, in particular, using a lab experiment. Researchers may be tempted to use the intuitions of Bayesian learning and DeGroot learning to test for correlations in regression analysis of social learning data. Two natural examples are (a) regressions that check whether individuals overweight the same information if they receive it through multiple channels and (b) regressions that explore whether individuals treat old information that cycles back to them as if it was new, additional information. Under these frames, the null hypothesis is the Bayesian model, since in these specifications the Bayesian models suggest that the relevant parameters ought to be zero. A rejection of zero in the coefficients of these regressions provides evidence in the direction of the DeGroot rules of thumb. To test the effectiveness of regression-based analysis,

[3]An algorithm is $O(T)$ if the number of computations as a function of T, $f(T)$, is such that $\frac{f(T)}{T} \to M$ for some constant M. In particular, this is true if $f(T) = MT$, as it is in our algorithm

we simulate learning data from a Bayesian model using the aforementioned algorithm as well as from DeGroot models. We show that such regression-based approaches do not work, namely, the data generated even by Bayesian learning models do not conform to the intuition motivating the regressions. We maintain that, in turn, the researcher ought to proceed by a structural analysis. Given the computational constraints for structural estimation of learning models in large networks, this suggests that separating models of social learning are best addressed in a lab setting.

There is little empirical evidence comparing Bayesian learning with rules of thumb learning in non-experimental contexts. Without experimental data it is difficult to control priors of agents in the network and the signal quality. Moreover, even in field experiments separating between Bayesian and DeGroot models may be difficult. First, structural approaches are computationally infeasible even with moderately sized networks (10 nodes), as it will become clear below in our discussion of computational complexity. Second, reduced form tests may not suffice for separating between these models. Third, empirical network data may not be precisely measured, affecting the conclusions of a researcher who is trying to select between these models of learning. There may be problems with estimating a structural model on a sampled network and the survey-obtained social network may not be precisely the communication channels used in practice, both of which would induce biases ((Chandrasekhar and Lewis 2010)). Since network-based estimation (which is inherently structural even when using reduced form regressions) is often sensitive to misspecification of the network, it is difficult to cleanly identify which model best describes the data in a non-laboratory context. Fourth, we are unable to know exactly what information is being transmitted in empirical data. Without knowing whether the information transmitted in this context is beliefs, actions or something else all together, one may mistakenly select the wrong model because of not properly specifying the information that is communicated.

Meanwhile, we believe that for our purposes, conducting a lab experiment outside the field of interest is insufficient because we desire to describe the social learning process for our population of interest. We are precisely interested in studying the social behavior of rural populations in a developing country as this is the relevant population in the aforementioned literature.

Related Literature

(Acemoglu, Dahleh, Lobel, and Ozdaglar 2010) and (Jackson 2008b) provide extensive reviews of the social learning on networks literature. The literature is partitioned by whether the learning is Bayesian or myopic (following some rule of thumb). On top of this, the literature layers a myriad of questions such as whether individuals learn from the communication of exact signals (or beliefs or payoffs of other agents) or by observing others' actions, whether the information arrives once or enters over time, whether the interaction is simultaneous or sequential, etc.

Bayesian Learning

The benchmark models of Bayesian learning come from (Banerjee 1992) and (Bikhchandani, Hirshleifer, and Welch 1992). They examine sequential decisions by Bayesian agents who observe past actions. These papers point out that consensuses are formed and thereafter agents end up making the same decision, which may in fact be the wrong decision. (Smith and Sorensen 2000) and (Celen and Kariv 2005) look at Bayesian-rational sequential decision making and explore the conditions under which asymptotic learning is attainable. (Acemoglu, Dahleh, Lobel, and Ozdaglar 2010) extend the framework to consider a social network environment in which individuals have stochastic neighborhoods. Their main result is that asymptotic learning occurs even with bounded beliefs for stochastic topologies such that there is an infinitely growing subset of agents who are probabilistically "well informed" (i.e. with some probability observe the entire history of actions) with respect to whom the rest of the agents have expanding observations. (Gale and Kariv 2003) study an environment that is closest to ours. They only focus on Bayesian learning and extend the learning model to a finite social network with multiple periods. At time t each agent makes a decision given her information set, which includes the history of actions of each of their neighbors in the network. Via the martingale convergence theorem, they point out that connected networks with Bayesian agents will yield uniform actions in finite time with probability one.

DeGroot Learning

(Ellison and Fudenberg 1993; Ellison and Fudenberg 1995) are benchmarks for the rule of thumb learning models. (Bala and Goyal 1998) investigate an extension of the two-armed bandit problem to the network context with boundedly rational agents and demonstrate that there will be payoff equalization and uniformity of action in the limit via an imitation principle.

In a network context, (DeGroot 1974) provides the most influential non-Bayesian framework. Agents observe signals just once and communicate with each other and update their beliefs via a weighted and possibly directed trust matrix. (Golub and Jackson 2010a) characterize the asymptotic learning for a sequence of growing networks. They argue that crowds are wise, provided that there are not agents that are too influential. (DeMarzo, Vayanos, and Zwiebel 2003) also consider a DeGroot style model and show that as agents fail to account for the repetition of information propagating through the network, persuasion bias may be a serious concern affecting the long run process of social opinion formation. Moreover, they show that even multidimensional beliefs converge to a single line prior to obtaining a consensus, thereby demonstrating how a multidimensional learning process can be characterized by a uni-dimensional convergence. (Chatterjee and Seneta 1977), (Berger 1981), (Friedkin and Johnsen 1997), and (Krackhardt 1987) are among other papers that examine variations on the DeGroot models.

The main contribution of this paper is to bridge the experimental literature with common myopic models of learning, motivated by the DeGroot model tradition, in a framework that allows

us to study the Bayesian and DeGroot-like features of empirical social learning processes. We fill the gap by conducting experiments designed to precisely aid us in model selection procedures and establish regularities of the learning process on networks.

Learning Experiments

The literature on social learning experiments begins with (Anderson and Holt 1997), (Hung and Plott 2001), and (Kubler and Weizsacker 2004). Explicit network structure are considered in a series of papers by (Gale and Kariv 2003) , (Choi, Gale, and Kariv 2005; Choi, Gale, and Kariv 2009), and Çelen, Kariv, and Schotter (2010).

(Choi, Gale, and Kariv 2005; Choi, Gale, and Kariv 2009) make a seminal contribution to the empirical literature of social learning by testing the predictions derived by (Gale and Kariv 2003) in a laboratory experiment. They are able to show that features of the Bayesian social learning model fit the data well for networks of three individuals. However, they do not allow for statistical power under the DeGroot alternatives. In extremely simple networks, such as the ones studied in their paper, there are few (if any) differences in the predicted individual learning behavior by the Bayesian and the rule of thumb learning models.

As mentioned above, Mobius et al. (2012) have conducted a related a field experiment using the Facebook network of Harvard undergraduates to distinguish between streams models versus DeGroot models of social learning. This paper as complementary to ours. We address whether individuals observing only actions of their neighbors behave in a Bayesian or DeGroot manner and they are interested in whether myopic individuals, who can communicate more information by tagging, appear to behave in a streams manner or a DeGroot manner.

Structure of the Paper

The rest of the paper is organized as follows. Section 4.2 develops the theoretical framework. Section 4.3 contains the experimental setup. Section 4.4 describes the structural estimation procedure and the main results of the estimation. Section 4.5 presents the discussion of the difficulties of reduced form approaches. Section 4.6 concludes.

4.2 Framework

4.2.1 Notation

Let $G = (V, E)$ be a graph with a set V of vertices and E of edges and put $n = |V|$ as the number of vertices. We denote by $A = A(G)$ the adjacency matrix of G and assume that the network is an undirected, unweighted graph, with $A_{ij} = \mathbf{1}\{ij \in E\}$. Let $N_i = \{j \in V : ij \in E\} \cup \{i\}$ the set of neighbours of i (i.e. the individuals who are connected to i) plus himself. Individuals in the

network are attempting to learn about the underlying state of the world, $\theta \in \Theta = \{0, 1\}$. Time is discrete with an infinite horizon, so $t \in \mathbb{N}$.

At $t = 0$, and only at $t = 0$, agents receive iid signals $s_i|\theta$, with $\mathrm{P}(s_i = \theta|\theta) = p$ and $\mathrm{P}(s_i = 1 - \theta|\theta) = 1 - p$. The signal correctly reflects the state of the world with probability p. In every subsequent period, the agent takes action $a_{i,t} \in \{0, 1\}$ which is her best guess of the underlying state of the world. Figure 4-3 provides a graphical illustration of the timeline.

In addition, we denote by W the set of all possible combinations of signals among agents, which we will refer to as "worlds". Therefore $s \in S$ is an element $s = (s_1, ..., s_n)$ with $s_i \in \{0, 1\}$. Note that $|W| = 2^n$. We will use $d_i = \sum_j A_{ij}$ to refer to the vector of degrees for $i \in \{1, ..., n\}$ and ξ for the eigenvector corresponding to the maximal eigenvalue of A.

4.2.2 Bayesian Learning

In our analysis we consider a model of Bayesian learning with incomplete information. Individuals will have common priors over the relevant state spaces (described below) and update according to Bayes' rule in each period. We formalize the model in Appendix 4.C. Each agent is drawn from a population which has π share Bayesian agents and $1 - \pi$ share DeGroot agents and this fact is common knowledge, as is the structure of the entire network. However, there is incomplete information about the types of the other agents in the network, and the Bayesian individuals will attempt to learn about the types of the other agents in the network along the path while attempting to learn about the underlying state of the world. The incomplete information setup is a useful step beyond the fully Bayesian environment, restricting $\pi = 1$. For instance, if an individual believes that her neighbor does not act in a Bayesian manner, she will process the information about observed decisions accordingly; as outside observers, the econometricians might think that she is not acting as a Bayesian. This is a very common problem when testing Bayesian learning, because we need to make very strong assumptions about common knowledge. A model in which there is incomplete information about how other players behave attempts to address this issue while only minimally adding parameters to be estimated in an already complicated system.

4.2.3 DeGroot Learning

We begin with a classical model of rule of thumb learning on networks and discuss three specific and natural parametrizations. (Jackson 2008c) contains an excellent review of DeGroot learning models. In our experiment, we consider DeGroot action models as opposed to communication models. In action models individuals observe the historical actions of their network neighbors, while in communication models individuals will be able to communicate their beliefs to their neighbors.

We are interested in action models for several reasons. First, the models of Bayesian learning on networks are action models, so it is the appropriate comparison. Second, it is extremely difficult to get reliable, measurable, and believable data of beliefs in a communication model for a lab experiment conducted in the field in rural villages. Third, as it is difficult to control and map into

data exactly what is (or is not) communicated by various agents in a more general communication model, we are able to focus on the mechanics of the learning process by restricting communication to observable actions. Fourth, this also fits with the motivating literature wherein individuals may observe the actions, such as technology or microfinance adoption decisions, of their neighbors.

Let $T = T(A)$ be a weighted matrix which parametrizes the weight person i gives to the action of person j. We study three natural parametrizations of the DeGroot model. The first is *uniform weighting* wherein each individual weights each of her neighbors exactly the same. The weight matrix $T^u(A)$ is given by

$$T_{ij}^u = \frac{A_{ij}}{d_i + 1} \text{ and } T_{ii}^u = \frac{1}{d_i + 1}$$

meaning that each individual puts $(d_i+1)^{-1}$ weight on each of her d_i neighbors as well as on herself.

The second model we consider is *degree weighting*. Each individual weights her neighbors by their relative popularity, given by degree. $T^d(A)$ is given by

$$T_{ij}^d = \frac{d_j}{\sum_{j \in N_i} d_j + d_i} \text{ and } T_{ii}^d = \frac{d_i}{\sum_{j \in N_i} d_j + d_i}.$$

The third model is *eigenvector weighting*. An individual places weight on her neighbor proportional to the neighbor's relative importance, given by eigenvector centrality. $T^e(A)$ is given by

$$T_{ij}^e = \frac{\xi_j}{\sum_{j \in N_i} \xi_j + \xi_i} \text{ and } T_{ii}^e = \frac{\xi_i}{\sum_{j \in N_i} \xi_j + \xi_i}$$

where ξ is the eigenvector corresponding to the maximal eigenvalue of A. This is motivated by the idea that an individual may put greater weight on more information-central neighbors, which eigenvector centrality captures.

While a very natural parametrization of learning, the DeGroot model misses strategic and inferential features of learning. Behavior is as follows. At time $t = 0$, individuals receive signals $s = (s_1, s_2, ..., s_n)$, and accordingly, actions $a_{i,0} = \mathbf{1}\{s_i = 1\}$ are taken. Let $\mathbf{a}_0 = (a_{1,0}, a_{2,0}, ..., a_{n,0})$ At the beginning of $t = 1$, individual i observe $a_t^{(j)}$ for all $j \in N_i$ and aggregates information to form beliefs to $b_1 = T\mathbf{a}_0$. In turn, actions are chosen $\mathbf{a_1}=\mathbf{1}\{b_1 > 1/2\}$. Now consider time $t = k$ with action profile $\mathbf{a}_k$. Then beliefs for stage $k+1$ are formed $b_{k+1} = T\mathbf{a_k}$ and accordingly actions are chosen $a_{k+1}=\mathbf{1}\{b_{k+1} > 1/2\}$. In turn, we have if the limit exists,

$$\begin{aligned} a_\infty &= \lim_{k \to \infty} \mathbf{1}\{Ta_{k+1} > 1/2\} \\ &= \lim_{k \to \infty} \mathbf{1}\{T \cdot \mathbf{1}\{Ta_k > 1/2\} > 1/2\}, \; a_k = \mathbf{1}\{Ta_{k-1} > 1/2\}. \end{aligned}$$

While we cannot easily analyze this using the theory of linear operators (due to nested indicator functions), we will discuss its implications in section 4.2.5.

4.2.4 A Fair Comparison

We argue that there is an important gap in the literature of learning on networks. Models of Bayesian learning in the literature have largely been action based; individuals observe the actions taken by their neighbors and accordingly update their beliefs and in this manner asymptotic consensus is achieved under certain regularity conditions ((Gale and Kariv 2003)). In a sequential learning environment, provided no finite set of individuals is asymptotically excessively influential, asymptotic learning occurs (Acemoglu et al.). However, the literature on DeGroot learning, e.g. Golub and Jackson (2010), has focused on models of communication to obtain asymptotic learning results under similar regularity conditions. So long as there is no finite set of agents that are too central in a particular sense, the network is wise in the sense that asymptotic learning occurs. Notice that the wisdom of the DeGroot learning hinges on the fact that an extensive amount of information is passed along such a model relative to the action model of Bayesian learning. For a parallel, in Bayesian learning if we introduced a communication model, then the filtering problem would be somewhat simpler since an agent would know her neighbors' posteriors exactly.

We argue that there is an asymmetry in the literature; the right abstraction to think about social learning ought to be parallel across the competing Bayesian and DeGroot models. For motivation we examine a key example of a sequence of networks which satisfy the properties to have asymptotic learning under both Bayesian (action) and DeGroot (communication) learning models, but fail to have asymptotic learning with DeGroot (action) learning models.

4.2.5 An Illustrative Example: Concentric Social Quilts

We present a simple setup which yields asymptotic learning under communication DeGroot models and consensus under action Bayesian models, but fails asymptotic learning and violate consensus with action DeGroot models. Namely, a number of nodes will become "stuck" in an information trap in a local neighborhood of the network. This demonstrates a wedge between DeGroot and Bayesian learning and therefore motivates our learning experiment.

The motivation for the graph structure comes from (Jackson, Barraquer, and Tan 2010) who study network architecture that arise as equilibria in favor exchange games. They show that these networks will be social quilts; a social quilt is a patchwork of substructures (e.g., triangles) pasted together in specific ways. We take a very simple example of this style of a network. From the applied perspective, the intuition is that if graphs are constructed as equilibria of risk-sharing or favor-exchange games, then they may have such quilt-like substructures. While these quilt-like structures enable network members to maintain favor-exchange relationships in equilibrium through local punishment of misbehavior, the same networks are also the surface on which information passes among members. We caution that it may be the case that information does not transmit efficiently through social quilts.

To illustrate this we define a concentric social quilt (CSQ) as a graph that consists of triangles

quilted together around a central triangle such that every triangle (in the interior of the structure) is connected to exactly three triangles in the following way. Consider a sequence of CSQs which can be constructed following a recursive process as the number of nodes goes to infinity. We index this sequence by $r \in \mathbb{N}$.

1. Take CSQ_{r-1} and let T_r be the set of terminal nodes of CSQ_{r-1}.

2. To each terminal node $i \in T_r$, attach a triangle with two new nodes added.

Figure 4-1 shows such a network and the model is developed in detail in Appendix 4.E.

DEFINITION 4.2.1. *We say that node $i \in V_r$ is* stuck *if there exists a $t_i \in \mathbb{N}$ such that for all $t \geq t_i$, $a_{i,t} = 1 - \theta$.*

A node is stuck if the node for all but finitely many periods takes the same (wrong) action. Figure 4-2 provides two examples of nodes that get stuck despite the majority of nodes in the network receiving the right signal.

Panel A of Figure 4-2 illustrates the problem. Assume that for some subtree of the CQR, which connects to the rest of the network through the top-most node, we have the initial signal endowment shown. Any information from the rest of the graph will come via top-most node, which we will call parent node. To get a lower bound on the number of nodes that get stuck in the wrong action, we can simply assume that the parent node of the subtree always chooses the right action for all rounds. However, even in this case the nodes in the lower right triangle act in the same (wrong) manner for all but finitely many periods. As the sequence of networks grow, $r \to \infty$, there will be a non-vanishing fraction of subtrees with this initial configuration of signals. These subtrees will have at least 3/7 nodes which become stuck. This example has demonstrated the following result.

PROPOSITION 4.2.1 (Wedge in Learning Models). *For a sequence of concentric social quilts with iid signals with probability p, with probability approaching one*

1. *under the DeGroot communication model with uniform weighting the network is wise,*[4]

2. *and under the DeGroot action model with uniform weighting a non-vanishing fraction of nodes get stuck.*

Proof. All proofs are contained in Appendix 4.F. □

That asymptotic learning occurs with the DeGroot communication model follows from Corollary 1 of (Golub and Jackson 2010a), and the result for the DeGroot action model is apparent from the previous example. To illustrate the severity of Proposition 4.2.1, in Table 1 we show lower

[4]Let $\mu = \theta p + (1-p)(1-\theta)$ and T_n is a sequence of convergent row-normalized matrices. As defined in (Golub and Jackson 2010a), the sequence is wise if $\text{plim}_{n\to\infty} \sup_{i \leq n} \left|\lim_{t\to\infty} T_n^t s_n - \mu\right| = 0$. In our context wisdom corresponds to asymptotic learning since in the limit a share of nodes that have belief μ goes to one and therefore the nodes can distinguish $\mu > 0$ or $\mu < 0$, as p is known.

bounds on the expected fraction of nodes that are stuck. Even with high quality signals ($p = 0.7$) at least 16% of nodes become stuck and do not asymptotically learn. In particular, recall that the benchmark for mistakes is 50%, since a node can always randomly guess. Therefore, relative to the expected fraction of nodes that should have learned, at least 25% actually get stuck with the wrong information.

In addition to motivating the study of DeGroot action models in our experiment, this example is of independent interest as it illustrates a wedge in the learning on networks literature between the usual DeGroot and Bayesian models. It also raises the question about whether certain network structures are better for social learning, given that asymptotic learning may not occur due to this stuck property. We conjecture that graphs with sufficiently good expansion properties will generate asymptotic learning even with action DeGroot models.

4.3 Experiment

4.3.1 Setting

Our experiment was conducted at 19 villages in Karnataka, India for a total of 95 experimental sessions for each of three chosen networks. The villages range from 1.5 to 3.5 hours' drive from Bangalore. A village setting was chosen because social learning through networks such as by word-of-mouth communication is of the utmost importance in rural environments; information about new technology ((Conley and Udry 2010)), microfinance ((Banerjee, Chandrasekhar, Duflo, and Jackson 2011)), among other things propagates through the social network.

4.3.2 Overall Game Structure

In each village, individuals played the social learning game three times, each time with a different network structure. The three networks (see Figures 4-4) were played with a random order in each village. Every network consisted of seven individuals and each participant was shown the entire network structure as well as her own location in the network.

At the beginning of each game, every individual was shown two identical bags, one with three yellow balls and one blue ball and the other which has three blue balls and one yellow ball. One of the two bags was chosen at random to represent the state of the world and the goal of the game was that the participant had to independently guess whether the blue bag or the yellow bag had been selected. Since there was an equal probability that either bag could be chosen, we induced priors of $1/2$. As the selected bag contained three balls reflecting the state of the world, individuals anticipated receiving independent signals that were correct with probability $5/7$.

After receiving their signals in round zero, all individuals simultaneously and independently made their best guesses about the underlying state of the world (which bag had been selected). The game continued to the next round randomly and on average lasted 6 rounds. If the game

continued to the second round, at the beginning of the second round each individual was shown the round one guesses of the other individuals in her neighborhood, N_i. Agents updated their beliefs about the state of the world and then again made their best guesses about the state of the world. The game again continued to the following round randomly. This process repeated until the game came to a close. Notice that after the time zero set of signals, no more signals were drawn during the course of the game. Individuals would only observe the historical decisions of their neighbors and update their own beliefs accordingly.

Individuals were paid for a randomly chosen round from a randomly chosen game and therefore faced non-trivial incentives to submit a guess which reflected their belief about the underlying state of the world. Figure 4-3 depicts the timing.

4.3.3 Network Choice

We selected networks specifically so that we could separate between various DeGroot and Bayesian models considered in the paper. The previous experimental literature on Bayesian learning on networks ((Choi, Gale, and Kariv 2005; Choi, Gale, and Kariv 2009)) make use of several interesting and parsimonious three-person networks. However, we are unable to borrow these networks for our study as they were not designed for the purpose of separating between DeGroot and Bayesian learning. In fact, the networks utilized in (Choi, Gale, and Kariv 2005; Choi, Gale, and Kariv 2009) lack power to pit Bayesian learning against the DeGroot alternatives posited above. Panel A of Table 2 shows the fraction of observations that differ across complete information Bayesian learning and the DeGroot alternatives for each of the three networks used in (Choi, Gale, and Kariv 2005) and (Choi, Gale, and Kariv 2009). In two of the networks, there are no differences between the equilibrium paths of Bayesian learning and each of the DeGroot alternatives and in the third network the differences are on the order of 15% of the observations.

Given our goal of separating between Bayesian and DeGroot alternatives, we move to an environment with seven agents as opposed to three agents, so that we obtain more power to distinguish between these models while still maintaining computational tractability.[5]

We considered all connected, undirected networks with seven nodes. Next, we established a model selection criterion function. This criterion function depended on power to detect a DeGroot alternative against a complete information Bayesian null, using our pilot data to generate an estimate of the noise, as well as a divergence function. The divergence function measures the share of node-time observations for which the Bayesian model (with $\pi = 1$) and a DeGroot model pick different actions,

$$D(G) := \frac{1}{nT} \sum_{w \in W} \sum_{t=1}^{T} \sum_{i=1}^{n} \left| a_{i,t}^{B}(w \mid G) - a_{i,t}^{m}(w \mid G) \right| \cdot \mathrm{P}(w \mid \theta = 1),$$

where $a_{i,t}^{B}(w|G)$ is the action predicted under the Bayesian model and $a_{i,t}^{m}(w|G)$ is the action

[5]Moving to eight agents, for instance, would be exponentially more difficult for our structural estimation.

predicted under DeGroot with m-weighting, where m is uniform, degree, or eigenvector weighting. Figure 4-5 depicts the Pareto frontier between power and divergence and shows one of the networks that we have selected. The procedure yields the following networks shown in Figure 4-4.

4.4 Testing the Theory

In order to test how well a model m fits the data in village r, we will use a divergence function, defined as

$$D(m,r) := \frac{1}{n(T_r-1)} \sum_{i=1}^{n} \sum_{t=2}^{T_r} D^m_{i,t,r}$$

where $D^m_{i,t,r} = |a^{obs}_{i,t,r} - a^m_{i,t,r}|$. This computes the share of actions taken by players that are not predicted by the model m.[6] To examine how poorly model m predicts behavior over the entirety of the data set, we define the *divergence function* as

$$\mathbf{D}(m) := \frac{1}{R} \sum_{r=1}^{R} \frac{1}{(T_r-1)n} \sum_{i=1}^{n} \sum_{t=2}^{T_r} D^m_{i,t,r}.$$

Model selection will be based on the minimization of this divergence metric, which can also be thought of as the fraction of actions not explained by model m.

The divergence is the deviation of the observed data from the theory. We may define the action prescribed by theory in one of two ways: network level, which looks at the social learning process as the unit of observation, and individual level, in which the unit of observation is the individual.

When studying *network level divergence*, we consider the entire learning process as a single observation. Theory predicts a path of actions under the true model for each individual in each period given a network and a set of initial signals. This equilibrium path that model m predicts is gives the theoretical action, $a^m_{i,t,v}$. When using this approach, we try to assess how the social learning process as a whole is explained by a model. This method maintains that the predicted action under m is not path-dependent and is fully determined by the network structure and the set of initial signals.

When we consider the individual level divergence, the observational unit is the individual. The action prescribed by theory is conditional on the information set available to i at $t-1$: $a^m_{i,t,v}$ is the action predicted for agent i at time t in treatment r given information set $I_{i,r,t}$.

4.4.1 Learning at the Network Level

We begin by treating the observational unit as the social network itself and take the whole path of actions $\{a_{i,r}\}_{r \leq R}$ as a single observation.

[6] Since all models and all empirical data have a fixed first stage (given by the signal endowment), the first round should not enter into a divergence metric. In turn, we restrict attention to $t \geq 2$.

In our incomplete information Bayesian learning model, we consider three natural candidates for π (the probability that an agent is Bayesian as opposed to DeGroot). First, we can consider the parameter that minimizes the network level divergence: the best-fitting value of π, given by $\widehat{\pi} = \text{argmin } \mathbf{D}(m, \pi)$. Second, we can consider the limit as $\pi \to 1$. This is the limiting case in which all agents are almost certain that everyone is Bayesian. Third, we can consider the limit as $\pi \to 0$.

We find that the optimal $\widehat{\pi}$ is a corner solution. In our dataset, the network level divergence is invariant to π.[7] In particular, while the beliefs change with π , the threshold for choosing an alternative action is never crossed. In turn, the optimal value corresponds to $\pi \to 0$ or $\pi \to 1$ as a function of which ever has the lower divergence. Thus, we are relegated to a situation identical to that of the complete information environment, wherein we can consider the two cases of all agents being Bayesian and all agents being DeGroot.

Table 3 presents the network level divergence for each of the three DeGroot models as well as the complete information Bayesian model across all the networks, uniform weighting fails to explain 12% of the data, degree weighting fails to explain 14% of the data, complete information Bayesian learning fails to explain 19% of the data, and eigenvector centrality fails to explain 27% of the data. This suggests that the degree and uniform DeGroot models as well as the Bayesian learning models each explain between 60 to 80% of the observations.

Uniform weighting dominates each of the other models in terms of fit. Meanwhile, eigenvector centrality weighting performs uniformly worse. The data shows that these models outperform the eigenvector weighting model considerably: they explain approximately 84% of all agent-round observations, while the latter explains less than 67% of these observations.

Next, we study the incomplete information Bayesian learning model with DeGroot alternatives. We estimate the optimal $\hat{\pi}$. Figures 4-6, 4-7, 4-8 show the divergence against π and our optimal estimate is given by minimizing the expected diveregence. As the optimal $\hat{\pi}$ are corner solutions, we need only compare the complete information Bayesian learning model to each of the DeGroot alternatives. To be able to statistically test the difference of fits across these different models, we conduct a non-nested hypothesis test in a GMM setting (e.g. Rivers and Vuong, 2002). Specifically, we perform a nonparametric bootstrap at the village-game level wherein we draw, with replacement, a 22 village-game blocks of observations, and compute the network level divergence. We then create the appropriate test statistic, which is a normalized difference of the divergence functions from the two competing models.

Our key hypothesis of interest is a one-sided test with the null of Bayesian learning against the alternative of the DeGroot model. Table 4 contains the results of the inference procedure. First, when we look across all topologies, we find evidence to reject the Bayesian model in favor of the uniform weighting alternative and the degree weighting alternative.Finally, we note that eigenvector centrality is dominated. When looking across all networks, it is summarily rejected in favor of any

[7]Details are provided in a supplementary appendix, available upon request from the authors.

alternative model. Looking at the bootstrap for a given topology, we find that only in the case of network 1 against Bayesian learning and network 3 against uniform weighting does eigenvector weighting remain statistically indistinguishable.

Ultimately, the bootstrap provides strong evidence that the null of Bayesian learning is rejected in favor of the degree weighted DeGroot alternative, suggestive evidence that Bayesian learning underperforms relative to the uniform weighted DeGroot alternative, and support for the claim that eigenvector weighting explains the data quite poorly.

4.4.2 Learning at the Individual Level

Having looked at the network level divergence, we turn our attention to individual level divergence. While this does not purely address the mechanics of the social learning process as a whole, it does allow us to look at individual learning patterns. Understanding the mechanics of the individual behavior may help us microfound the social learning process.[8]

4.4.2.1 DeGroot Models

We begin by calculating the individual level divergence for the DeGroot models.[9] Panel A of Table 5 contains the results of the exercise. First, uniform weighting systematically outperforms eigenvector weighting (by a large margin) and degree weighting (by a smaller margin). It is worth noting how well the DeGroot models perform in terms of predicted individual behavior. Across all three networks, the uniform weighting model explains approximately 88.9% of all individual observations. Degree and eigenvector centrality weighting models predict 86% and 75% of all individual observations, respectively.

Furthermore, we look at the individual level divergence village by village. Panel B of Table 5 shows the number of villages for which each DeGroot model was the best fitting. For network 1, uniform weighting is the best fitting (17 villages), for network 2 both uniform and degree weighting are tied (11 villages each), and for network 3 uniform weighting is by far the best fitting model (21 villages).

4.4.2.2 Bayesian Learning with Incomplete Information

We now turn our attention to the Bayesian learning model. Unlike the myopic models, when considering the empirical divergence and the subsequent predicted action $a^m_{i,t,v}$, we need to consider the whole path of observed actions for all agents.

A potential problem arises: since our model of Bayesian learning implies that actions taken by individuals are deterministic functions of the underlying environment, this implies that the support

[8]It may be the case that agents themselves are do not each behave in according to a particular model while the aggregate social group may best be described by such a model

[9]When an agent faces a tie, they stay with their previous action. We considered a random tie-breaking alternative as well, which does not change the results much.

of the set of potential paths that individuals could have observed is rather limited. Therefore, there is a possibility that empirically, Bayesian agents may arrive to an information set that has zero probability of occurrence. This is problematic for identification, since the Bayesian learning model is mute when agents have to condition their inference on zero probability events; any observed action from then on would be admissible for a Bayesian learning agent. Table 6 shows that zero probability information sets are hit quite frequently.

With any DeGroot alternative, 100% of the treatments in networks 1 and 2 have at least one agent hitting a zero-probability information set. Moreover, at least 62% of players in these networks have hit a zero-probability information set at some point. Though one may be tempted to interpret the lack of support itself as evidence against a Bayesian model, this is a delicate issue requiring a more careful treatment.

To compute the divergence across all observations, we need to make the support of possible paths extend over our entire data set. The usual way to deal with this problem is to introduce disturbances. In the following subsection we explore the possibility of estimating a trembling hand or quantal response equilibrium (QRE) style version of Bayesian learning in which we introduce the possibility of making mistakes by all agents. In such a model, individuals can make mistakes with some probabilities, and Bayesian agents, knowing the distribution of these disturbances, integrate over this possibility when updating.

4.4.2.3 Bayesian Learning with Disturbances and Complexity Problems

To simplify exposition, we restrict attention to the case of a complete information Bayesian model where each agent is Bayesian. In this environment, each agent makes a mistake with probability ε and chooses the opposite action that a Bayesian agent would chose. This guarantees full support; any agent can take any action given any history with positive probability.[10] Moreover, one could potentially describe what a Bayesian learning model prescribes by taking the limit $\varepsilon \to 0$.

Introducing disturbances comes at great computational cost in an environment where agents learn on networks. The only sufficient statistic for the information set that each agent sees is the information set itself, as there is no deterministic function between signal endowments and information sets. This means that through time, the relevant state space (the histories each agents could have seen) grows exponentially. We show that this makes the problem intractable for any practical purpose.

First, we note that the algorithm that we use to simulate the Bayesian learning model without trembles is computationally "tight" in the sense that, asymptotically, there is no faster algorithm.[11] Because any algorithm would have to take order T steps to print output for each of the T periods,

[10] (Haile, Hortaçsu, and Kosenok 2008) show that QRE imposes no falsifiable restrictions and can rationalize any distribution of behavior in normal form games. Relating this intuition to our context, one may be able to pick a distribution of ε such that it rationalizes the incomplete information Bayesian model as describing the data well.

[11] Our environment consists of finite graphs where n does not grow in T.

an algorithm that is $O(T)$ is asymptotically tight.

PROPOSITION 4.4.1. *The algorithm for computing Bayesian learning with no disturbances is $O(T)$. Moreover, it is asymptotically tight; i.e. any algorithm implementing Bayesian learning running time must be at least $O(T)$*

Specifically, the algorithm is $\Theta(n4^nT)$.[12] Notice that if n was growing this algorithm would be exponential time. Second, we show that the extension of this algorithm to an environment with disturbances is computationally intractable.

PROPOSITION 4.4.2. *Implementing the Bayesian learning algorithm with disturbances has computational time complexity of $\Theta(4^{nT})$.*

To see how computationally intractable is this algorithm, take as an example our experimental design. Assume that the original code takes one second to run. With $n = 6$ and $T = 5$ the computational time is on the order of 6.9175×10^{18} seconds, which is 8.0064×10^{13} days or 2.1935×10^{11} years. To get some perspective, let us compare the number of calculations with this very simplistic algorithm using the (Choi, Gale, and Kariv 2005; Choi, Gale, and Kariv 2009) enviroment in which $n = 3$. In this setup, the expected time would be 1.2288×10^5 seconds or 1.42 days which is entirely reasonable for structural econometrics.

In the above exercise, we used the most natural algorithm and one that was efficient for the case without disturbances; an objection may be made that there could perhaps be a more efficient algorithm. The decision problem we are interested in is determining whether an agent i in time period t given a history always picks the same action under a proposed algorithm as under the Bayesian model with trembles. We conjecture that the problem is NP-hard, which we are investigating in ongoing work. This means that the computational problem is at least as hard as NP-complete problems.[13] Whether there may or may not be polynomial time solutions for NP-hard problems is open; if P $\neq$ NP, then none would exist. The computer science literature studying Bayesian learning networks shows that obtaining the probabilities is NP-hard (Cooper, 1990) in any given network of events. In this context the networks are networks of events. Translating our framework into this setup involves constructing a network of belief states for each individual in the network and each time period, so a node in the Bayesian learning network would be a pair (i, t), so the size of it would be NT. Our ongoing work seeks to extend their argument to our decision problem which involves checking that the action taking by each person in each time period is identical when comparing a proposed algorithm with the true Bayesian learning model. The intuition is that the learning network is growing linearly in the number of periods and individuals and therefore for any algorithm there can be some action sequence such that to be able to decide whether individual i at time t, given the history, needs to decide whether to guess 0 or 1, one needs all the probabilities.

[12]Recall that we say $f_1(n) \in \Theta(f_2(n))$ if f_1 is asymptotically bounded above and below by f_2, upto a constant scale. Formally, if $\exists c_1, c_2 > 0, \underline{n}$ such that $\forall n > \underline{n}$, $c_1 \cdot |f_2(n)| < |f_1(n)| < c_2 \cdot |f_2(n)|$.

[13]A problem is said to be NP-complete if (a) it is NP which is to say that a given solution can be verified in polynomial time and (b) it is NP-hard so that any NP problem can be converted to this problem in polynomial time.

Based on Cooper (1990), which applies to a broader class of networks (and therefore will have weakly worse complexity), we conjecture that the argument for our sub-class of networks will also be NP-hard.

4.4.2.4 Results

We have argued that estimating the Bayesian model with trembles has computational complexity constraints. In turn, we now turn to studying which model best fits the data, taking these constraints into account. We are interested in studying which model best fits the data. We look at the deviation of each agent's action, given the history that the agent has observed at that time, from the predicted action by the model for that agent given the history. The formalities are developed in Appendix B.

To compute the individual level divergence for the model with incomplete information, we need an estimate of π. Again we use the filter described in Appendix C to compute the probability that an agent i in village v is a Bayesian learner. Recall that we have three natural candidates for π : the best-fitting value, given by $\hat{\pi} = \arg\min \mathbf{D}(m, \pi)$, the limit as $\pi \to 1$, and the limit as $\pi \to 0$.

Since guaranteeing full support in this model by reintroducing trembles induces computational problems, we make the following arguments regarding the relative performance of the Bayesian model. First, the fact that we repeatedly observe agents facing zero probability events, even when there is positive probability that agents may be behaving in another manner, may be taken as prima facie evidence supporting the idea that this model of Bayesian learning with incomplete information on networks fails to explain the experimental data.

Second, one could make the objection that the considered incomplete information Bayesian model is not sufficiently rich to capture the characteristics of the data and that, perhaps, one needs a more nuanced model. This could indeed be the case, but as demonstrated in Proposition 4.4.2, it would be computationally infeasible to estimate a model generating full support.

Third, it might be the case that we have the right model of incomplete information Bayesian model but we lack a theory of what individuals do once they hit zero probability events. This implies that we assume the existence of a correct set of beliefs when encountering zero probability events that rationalizes individuals' actions. If this is the case we may take two different approaches. First, we could be agnostic about the correct off equilibrium beliefs. Second, we could consider the case of a richer Bayesian model that rationalizes the actions taken after an agent hits a zero probability event and precisely matches the supposed off equilibrium behavior. Such a model, of course, has the degree-of-freedom problem.

We begin with the first approach, by being agnostic about the off-equilibrium behavior and instead restrict attention to observations for which agents were in an information set that had a positive probability of occurrence. This is the only feasible comparison we can do given our assumption and agnosticism about the off-equilibrium beliefs. In this subset of observations, we can calculate the individual level divergence, since Bayes' rules applies and the Bayesian learning models

gives us a concrete prediction. Of course, we have not eliminated observations at random, but rather we have eliminated those that were not in accordance to the Bayesian learning model equilibrium (i.e. those that happened with zero probability). This is an admittedly ad hoc approach, requiring the assumption that the DeGroot model does not perform sufficiently worse off-equilibrium (where the Bayesian model in principle could rationalize anything). We will return to this issue below.

Under such an assumption, if it turns out that even in this subset of observations, Bayesian performs worse than the alternative myopic models considered, then this would be further evidence that, at the individual level, the Bayesian learning model would not seem to fit the experimental data well. Based on this idea, we present the calculation of the individual level divergence measure for observations that were in the support of the model. Table 7 contains the results of this exercise. It shows that across all networks, as well as for each network, both degree and uniform weighting DeGroot models have a lower divergence than the Bayesian model. Specifically, it suggests that a model with a limiting $\pi \to 0$ dominates the corresponding model with $\pi \to 1$.

As in the case of the network level analysis, the individual level divergence does not vary with π , as long as $\pi \in (0,1)$. That is, by varying the ex-ante probability of any given agent being Bayesian, and as long as it is strictly different from 0 or 1 (as we have assumed throughout), a Bayesian agent would choose the same action on identical information sets. Upon inspection, this is not because there is no actual effect on the inference of each agent on the true state of the world θ, but instead because the change in inference is not sufficiently strong that it changes a guess from 0 to 1 (or vice versa). Of course, there are equilibrium considerations here. Even if, in principle, varying the parameter π affects the decision, if no one chooses different actions for different π's, then there cannot be any difference in the inference made by agents on the actions of others.

To be able to perform inference on the null hypothesis of the Bayesian learning model against the alternative DeGroot models, again we perform a nonparametric bootstrap. Our main hypothesis of interest is a one-sided test of the Bayesian learning null against the DeGroot alternatives. Table 8 shows us that both the complete and incomplete information Bayesian models are strongly rejected against the alternatives of uniform and degree weighting DeGroot models. Moreover, as in the case of the network level analysis, we find that eigenvector centrality is uniformly rejected against all alternative models. Third, and perhaps more puzzling, at the individual level we find that uniform weighting is a better description of the behavior than degree weighting; the result holds as significant at a 10% one-sided test. While the evidence is somewhat weak, it still overturns the finding in the network level case wherein degree weighting performed better than uniform weighting. This suggests that while the entire process may be best described by degree weighting, conditional on an information set, an individual's action may be best predicted by uniform weighting. That being said, the divergence for each of these models is quite low, suggesting that preferential weighting is a strong fit as well.

We now return to the second approach. If we assume that indeed we have the right model of incomplete information Bayesian model but we are simply missing the correct off equilibrium

behavior, we could consider the case that a richer Bayesian model could be the one that rationalizes the actions taken after an agent hits a zero probability event and precisely matches the supposed off equilibrium behavior. Notice that even if, for short T, the Bayesian model might be underperforming, with probability 1 this will be the opposite in the long run, inverting the present results. This follows because, if we consider the Bayesian model as rationalizing anything off-equilibrium, once we are off equilibrium, as $t \to \infty$, Bayesian would never be penalized while DeGroot will be penalized infinitely often (assuming the behavior does not identically match the DeGroot model for all but finitely many periods).

To summarize this section's results, first we have presented evidence that the considered model of Bayesian learning result arrives at zero probability information sets extremely often. This can be taken as evidence against these particular models. Second, we provide computational theory that shows that models with trembles, which would smooth out the zero probability information set problem, are of little practical use to structurally evaluate empirical data. In turn, methodologically, structural approaches must restrict themselves to models which allow for zero probability information sets. Third, we take a pass at the data by ignoring the off-equilibrium information sets. This evidence suggests that, when restricting the analysis to positive probability information sets, the divergence minimizing models have in the limit no Bayesian agents. Finally, we point out that this approach, while ad hoc, may be inappropriate for a model of incomplete information behavior wherein the off-equilibrium behavior is well-matched. However, assuming the researcher is interested in incomplete information models (because of the computational infeasibility of QRE models), the argument in favor of the ad hoc approach is rejected only if the researcher believes in an untestable model which performs well off equilibrium (since we know on equilibrium it performs poorly). But such a model is unlikely to be the empirically relevant object.

A natural exercise may be to project the intuitions of these models into a regression framework. Such a reduced form approach would allow us to avoid computational constraints and see whether the patterns in regression provide suggestive evidence of DeGroot learning as opposed to Bayesian learning. We discuss the prospect of this in the following section.

4.5 Reduced Form

In this section, we discuss two reduced form approaches to study the data. Our motivation for this exercise is twofold. First, given the computational limits of the structural approach, we are interested in seeing whether reduced form patterns of Bayesian learning (as opposed to DeGroot learning) may be obtained from the data. Second, since larger networks, such as those found in empirical data sets, do not lend themselves to structural approaches for computational reasons, it is worth looking into the effectiveness of reduced form approaches to address these questions.

The central intuition we focus on has to do with double counting information. Under any of the aforementioned Bayesian models, Bayesian agents should not double-count information. However,

DeGroot agents do double-count information, and it is on this intuition that we build the exercise.

We provide two examples of regressions which researchers may run. The first set of regressions explores whether individuals overweight the same information if they receive it through multiple channels. The second set of regressions explore whether individuals treat old information that cycles back to them as if it is new, additional information. The null in these regressions is Bayesian model, since one would assume that the relevant parameters ought to be zero. Thus, a rejection of a zero may provide evidence in the direction of the DeGroot rules of thumb. The empirical data shows that both these reduced form analyses seem to provide support in favor of the DeGroot alternatives. However, because we are able to simulate out the data under the null, we show that these intuitions are wrong. Specifically, when we simulate social learning data under the Bayesian null, the coefficients are not as one may have expected.

4.5.1 Multiplicity

We define a variable which is a dummy for whether individual i makes a guess of 1 in the final period T, $y_i := 1\{a_{i,T} = 1\}$. As before, d_i is the degree of individual i and N_i is the set of (direct) neighbors $N_i = \{j \in V : ij \in E\}$. Note that $d_i = |N_i|$. Moreover, N_{2i} is the set of second-neighbors of person i; that is, $j \in N_{2i}$ means that there is at least one path of length two between i and j, but no path of length one. Finally, we define N_{2i}^l to be the set of second neighbors to whom she has exactly l paths.

The first regression we run is of the form

$$y_i = \beta_0 + \beta_1 s_i + \beta_2 \mathbb{E}_{N_i}[s_j | j \in N_i] + \sum_l \beta_{3l} \mathbb{E}_{N_{2i}^l}[s_j | j \in N_{2i}^l] + \epsilon_i. \tag{4.1}$$

This is a regression of whether or not individual i ultimately makes a guess of 1 on whether the individual's signal is 1 (s_i) the share of ones ($\mathbb{E}_{N_i}[s_j | j \in N_i]$) in individual i's neighborhood, and the share of ones given to each subset of second neighbors to whom i has exactly l paths ($\mathbb{E}_{N_{2i}^l}[s_j | j \in N_{2i}^l]$).

The interpretation is as follows. β_2 measures the impact of her neighborhood receiving a greater share of ones on an individual's guess. We expect $\beta_2 > 0$. Moreover, β_{3l} measures the impact of the subset of her second-neighborhood with multiplicity l. The intuition is that as the signals of individuals with greater multiplicity ought not be double-counted under a Bayesian frame, $\beta_{3l+1} > \beta_{3l}$ would be evidence of overweighting redundant information that has arrived via multiple channels, while $\beta_{3l+1} = \beta_{3l}$ would provide evidence in favor of the Bayesian hypothesis.

Table 9 provides the simulation and empirical results. Looking at the empirical results, as expected, an individual's own signal being one increases the probability of the final guess being one by 38pp, in a linear probability model framework. Moreover, if the share of individuals in one's neighborhood with signals of one increase by 10%, an agent has a 4.6pp increase in the probability of her final guess being one. While this seems to be consistent with the intuition that agents engage

in double-counting, the simulation exercise shows that these patterns cannot be interpreted in that manner.

Given the learning model, the network structure, and signal endowment, we simulated out the learning path and then ran the relevant regressions. We present results when simulating the learning process from the complete information Bayesian model (every agent is Bayesian) as well as each of the three DeGroot alternatives. The table shows that the Bayesian null does not have coefficients that are near identical across multiplicities 1, 2 and 3. Furthermore, the increasing correlation with indirect friends of higher multiplicities is also not uniformly found across the DeGroot models. Ultimately, the regressions suggest that the linear projection of this learning process is complex and may depend crucially on the network structure and set of initial signals.

4.5.2 Historical Information

Another reduced form that one may look at is whether individuals re-incorporate historical information that they have previously observed. Consider an individual at period 3. They have observed both their own signals and the signals of their direct neighbors (insofar as the first period guesses of their neighbors will be identical to their signals). In period three, therefore, a Bayesian individual's guess should not re-incorporate this information. Instead, it should only update using information about second-neighbors and the like, about whom they have yet to receive information.

To examine this formally, we perform the following regression. We regress the period three guess of individual i on her own signal (s_i) and the average signal of her neigbhorhood ($\mathbb{E}_{N_i}[s_j|j \in N_i]$) which she would have seen in period three. We also include as regressors the average signal of second neighbors ($\mathbb{E}_{N_{2i}}[s_k|k \in N_{2i}]$) which should be new information in period three. Lastly, we include the average signal of direct neighbors whose signals can cycle back via a path of length two back to individual i. Of course, we also include the agent herself in this set. (Formally, we use $\mathbb{E}_{C_i}[s_j|j \in C_i]$, where $C_i = \{j \in V - \{i\} : A_{ij}^2 A_{ij} > 0\} \cup \{i\}$.) The regression is as follows.

$$y_i = \alpha_0 + \alpha_1 s_i + \alpha_2 \mathbb{E}_{N_i}[s_j] + \alpha_3 \mathbb{E}_{N_{2i}}[s_k|k \in N_{2i}] + \alpha_4 \mathbb{E}_{C_i}[s_j|j \in C_i] + \epsilon_i. \tag{4.2}$$

We test the hypothesis of whether $\alpha_4 = 0$, which is our Bayesian null. Notice that $\alpha_4 > 0$ provides evidence that individuals reincorporate information that they already knew as it cycles through the network.

Table 10 presents the simulation and empirical results. The empirical results show that $\alpha_4 > 0$ and $\alpha_3 > 0$. While this seems to provide suggestive evidence for the intuition that DeGroot weighting reincorporates old information, the simulation results provide evidence that for our environment $\alpha_3 < 0$ for some of these models and $\alpha_4 > 0$ even for the Bayesian model.

4.5.3 Reflection on Reduced Forms

Taken together, Tables 9 and 10 have shown that natural reduced form approaches to test between these models may be misguided without first checking whether the patterns by the learning processes actually match the intuitions. We are able to study the reduced form projections of the Bayesian model using our simulation algorithm and find evidence that, when projected onto a regression for these networks with this environment, the Bayesian data suggests that the coefficients can deviate greatly from our intuitions. This, we argue, provides a strong motivation for the structural approach to studying the models.

4.6 Conclusions

In this paper we have investigated whether social learning patterns on small networks resemble those which would emerge if agents behaved in a Bayesian manner or if they are better modeled with DeGroot alternatives which are myopic and more simplistic models. To do so, we developed a simple experiment on networks that were designed to distinguish between these models, large enough to give us power on this dimension, but small enough to ensure that simulating a Bayesian learning on networks model was not computationally intractable. Given our experimental data, we were able to study the social learning process as a whole by taking the network as the unit of observation and study the behavior of an individual, which addresses whether an agent acts in a Bayesian manner.

At the network level we find evidence that the degree weighting DeGroot model best explains the data. The Bayesian learning null is rejected in favor of this alternative model. However, we maintain that Bayesian learning did an adequate job of describing the data process, explaining (beyond pure random guessing) 62% of the actions taken as compared to 76% by the DeGroot alternative.

At the individual level we find that uniform weighting DeGroot performs the best and both uniform and degree weighting DeGroot models outperform the Bayesian model. However, we show that the Bayesian model encounters the problem that many individuals come across zero probability information set. First, this provides suggestive evidence of the lack of fit of this incomplete information Bayesian model. Second, we demonstrate that introducing disturbances to smooth out the distribution cannot be a solution in this environment. The computational complexity of the problem is damaging to the very approach of applying QRE or trembles to the Bayesian learning on networks environment. As such, we recommend that researchers focus on computationally tractable models which will be easier to falsify.

We also show that reduced form approaches may be problematic. We provide two natural examples of regressions which build on intuitions separating DeGroot and Bayesian learning patterns. Equipped with our Bayesian learning algorithm, we simulate learning data from the Bayesian model as well as from DeGroot models and show that the reduced form regression outcomes do not

conform to the intuitions.

Ultimately, the findings suggest that agents and the learning process as a whole may better be thought of as coming form DeGroot action models where individuals myopically weight their neighbors' actions when updating their own beliefs. This may imply that social learning processes empirically may be sub-optimal, with information often getting stuck in pockets of the network. Having constructed an example of a network which satisfies asymptotic learning for DeGroot communication models, but where asymptotic learning fails for DeGroot action models, we argue that in action-learning environments DeGroot processes may be more damaging to the wisdom of society than previously anticipated.

4.A Figures and Tables

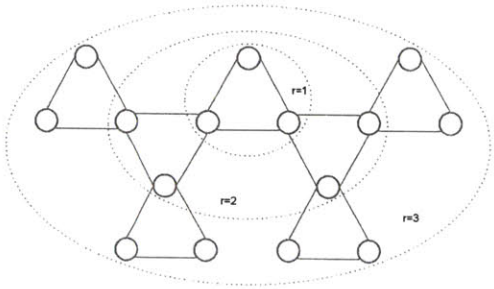
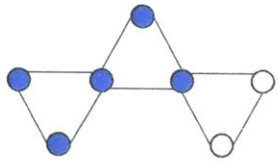
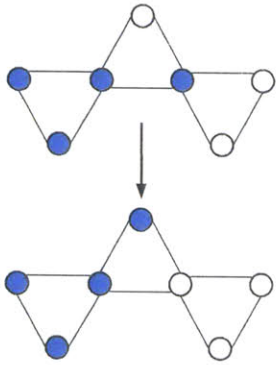


Figure 4-1: A concentric social quilt.

Panel A:





Panel B:

Figure 4-2: In Panel A two nodes are stuck for all periods $t \in \mathbb{N}$, even though 5 of the 7 nodes have received the true signal. In Panel B in the first period 4 nodes receive the true signal, and after one node switches, 3 are asymptotically stuck.

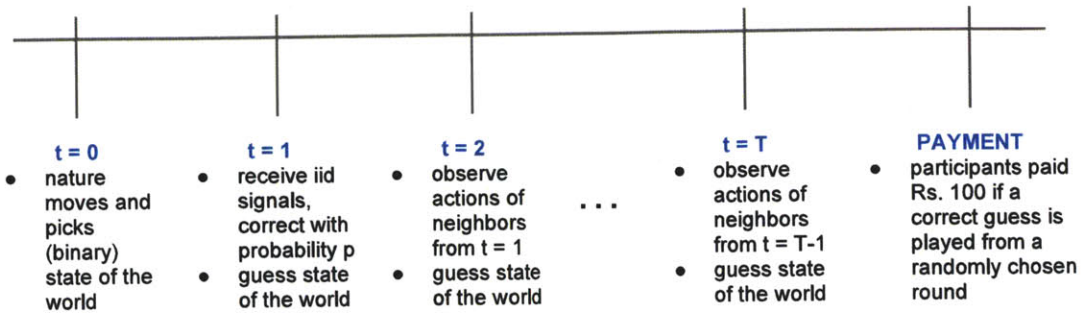
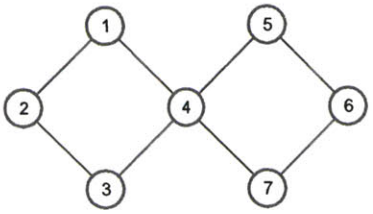
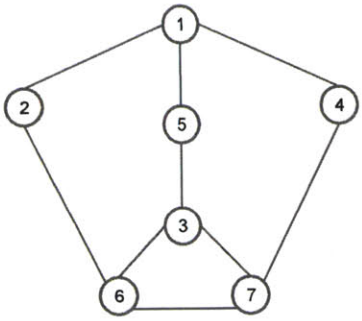


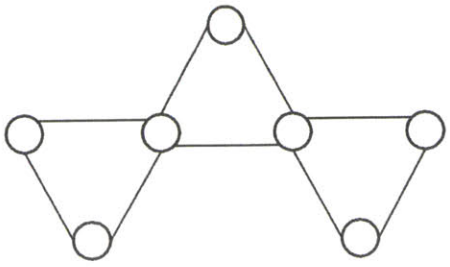
Figure 4-3: Timeline



Panel A: Network 1



Panel B: Network 2



Panel C: Network 3

Figure 4-4: Network structures chosen for the experiment.

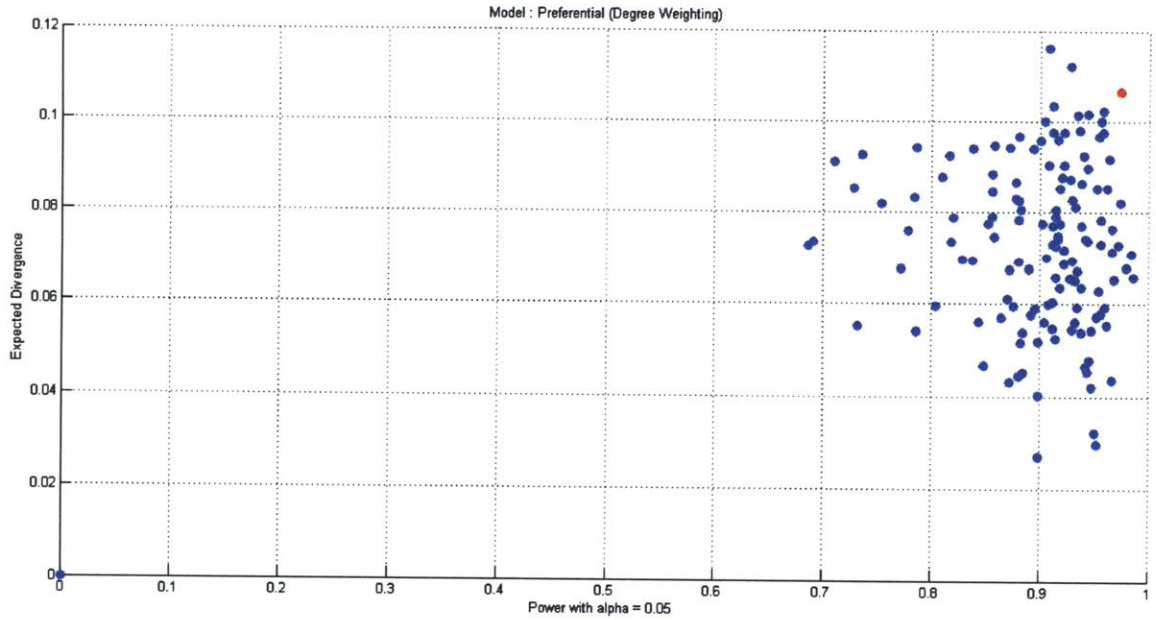


Figure 4-5: Divergence vs power frontier for network selection

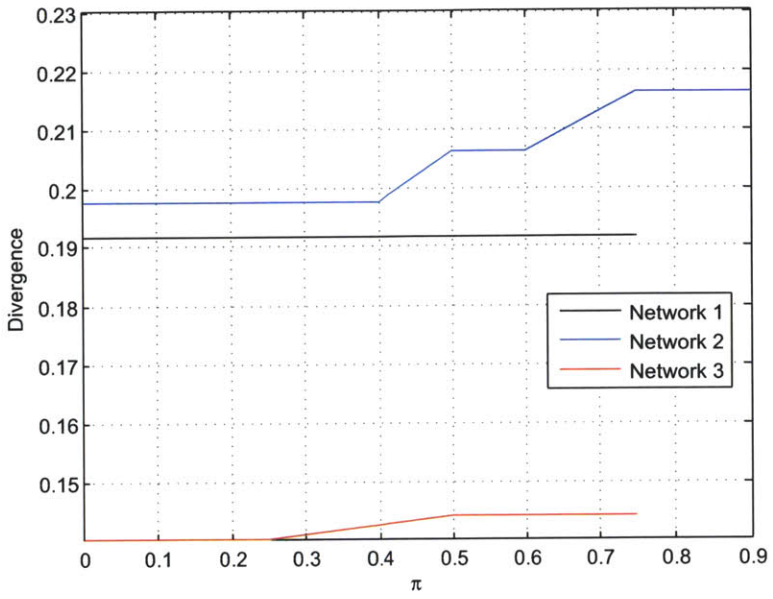


Figure 4-6: Divergence with Uniform Weighting

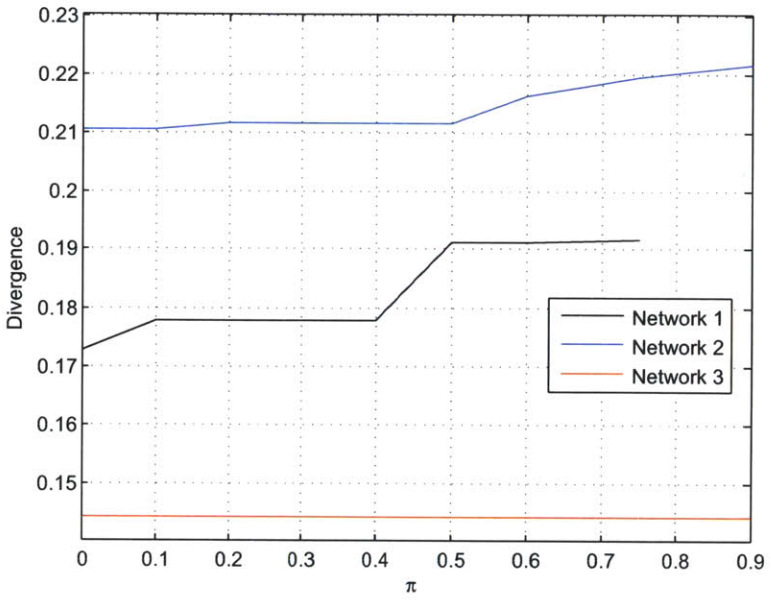


Figure 4-7: Divergence with Degree Weighting

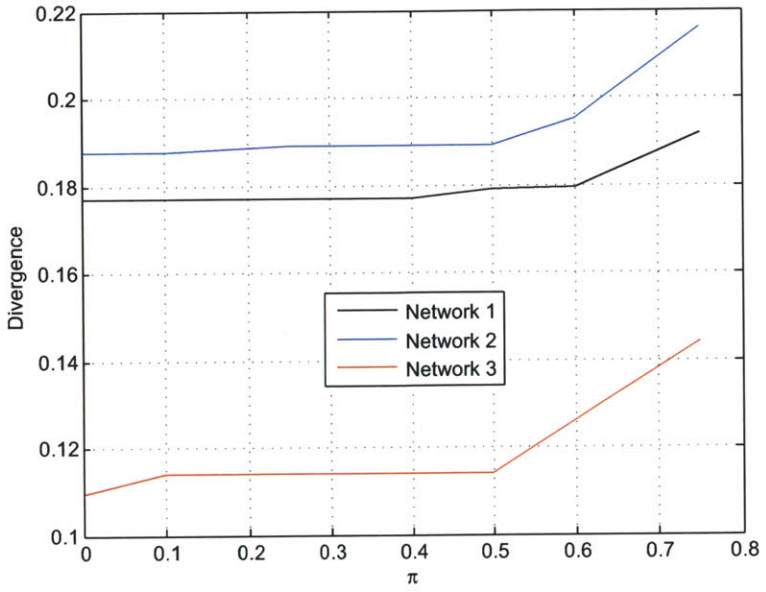


Figure 4-8: Divergence with Eigenvector Weighting

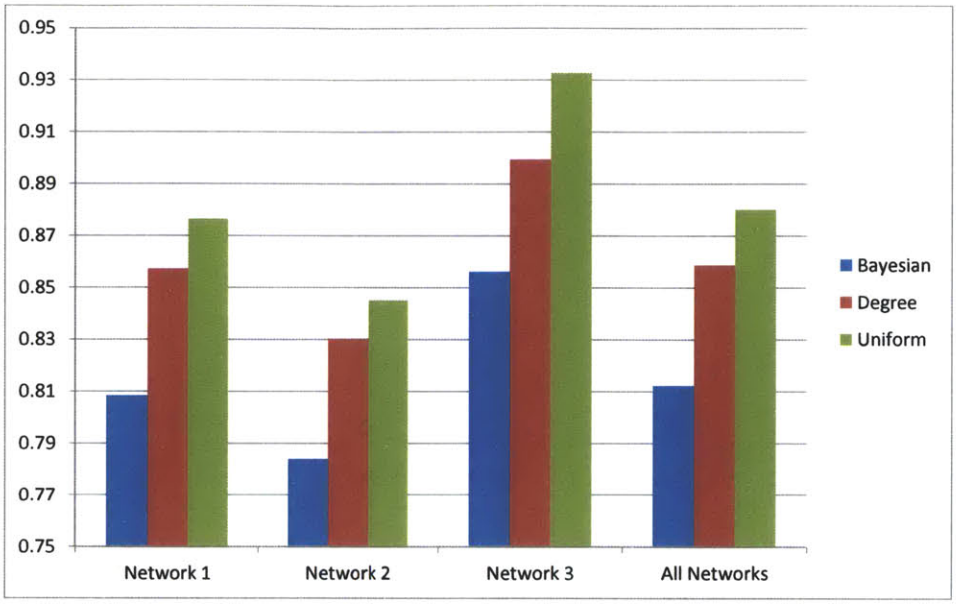


Figure 4-9: Fraction of actions explained at network level

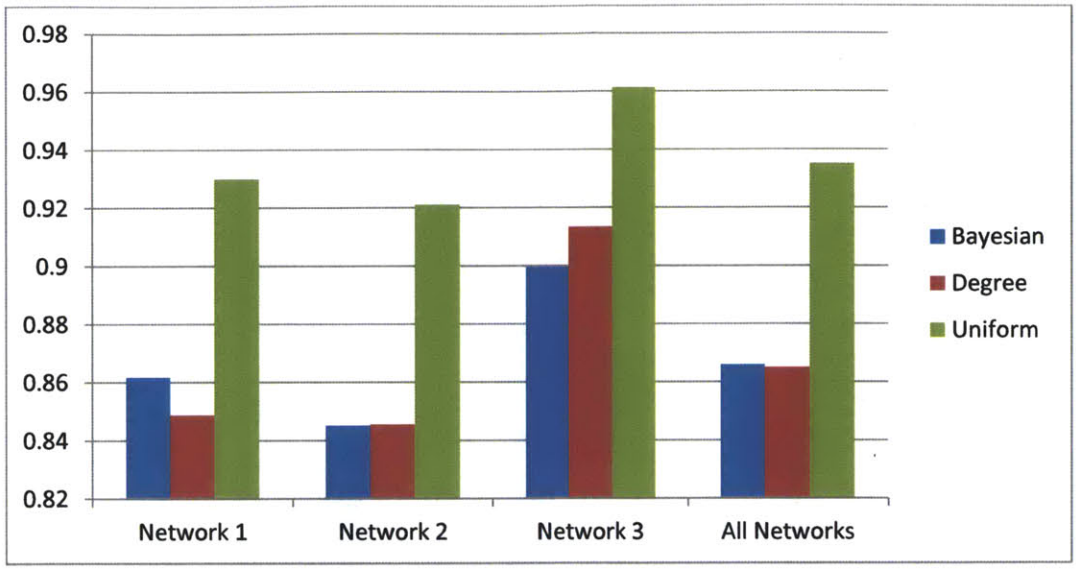


Figure 4-10: Fraction of actions explained at individual level

Table 1: Asymptotic Lack of Learning

	Fraction Stuck		*Extent of Learning*	
Signal Prob.	(1)	(2)	(3)	(4)
0.51	41.82%	55.06%	7.18%	14.64%
0.55	35.69%	48.80%	9.31%	20.69%
0.6	28.40%	41.40%	11.60%	29.00%
0.65	21.74%	33.80%	13.26%	37.89%
0.7	15.85%	27.14%	14.15%	47.17%
0.75	10.86%	21.30%	14.14%	56.56%
0.8	6.82%	16.42%	13.18%	65.90%
0.9	1.62%	9.59%	8.38%	83.80%

Note: Column (1) presents a lower bound on the fraction of nodes who get stuck and column (2) presents an upper bound. Column (3) presents an upper bound on the fraction of nodes that are right in the limit less p, the fraction of agents who receive the right signal in the beginning. Column (4) presents an upper bound on the fraction of agents who began with the wrong initial signal who ended up with the right guess in the limit.

Table 2: Fraction of Observations that Differ with the Bayesian Model

Panel A: Networks from Choi et al. (2005, 2009)

	Total Divergence			Divergence in Final Period		
Network	*Uniform*	*Degree*	*Eigenvector*	*Uniform*	*Degree*	*Eigenvector*
1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
3	9.37%	21.87%	8.98%	12.67%	18.67%	7.67%

Panel B: Networks Selected in This Paper

	Total Divergence			Divergence in Final Period		
Network	*Uniform*	*Degree*	*Eigenvector*	*Uniform*	*Degree*	*Eigenvector*
1	6.89%	10.64%	6.89%	9.13%	13.89%	9.13%
2	9.91%	10.00%	10.48%	15.08%	13.66%	12.10%
3	8.75%	9.67%	9.67%	12.24%	13.18%	13.18%

Notes: Fraction of observations for which the complete information Bayesian model differs with the DeGroot alternative. In Panel A, network 1 is the "complete network", network 2 is the "directed network", and network 3 is the "incomplete network" of Choi et al. (2005, 2009).

Table 3: Network Level Divergence

Network	*Total Obs*	*Bayesian*	*Uniform*	*Degree*	*Eigenvector*
All Networks	9,205	0.1878	0.1198	0.1413	0.2703
1	3,045	0.1917	0.1236	0.1428	0.2229
2	3,031	0.2161	0.1548	0.1698	0.3026
3	3,129	0.1440	0.0673	0.1006	0.2909

Notes: Network level divergence for the complete information Bayesian model, uniform DeGroot weighting, degree DeGroot weighting, and eigenvector DeGroot weighting.

Table 4: Significance Tests for Network Level Divergence

H_0	H_a	All Networks	Network 1	Network 2	Network 3
Bayesian	Degree	0.0001	0.0041	0.0019	0.0133
Bayesian	Uniform	0.0001	0.0001	0.0001	0.0001
Bayesian	Eigenvector	0.9999	0.9397	0.9999	0.9999
Degree	Uniform	0.0006	0.0615	0.0696	0.0001
Degree	Eigenvector	0.9999	0.9999	0.9999	0.9999
Uniform	Eigenvector	0.9999	0.9999	0.9999	0.9999

Notes: The test statistic is the normalized difference in the divergence functions of the null and the alternative model. We show the probability that the test statistic is less than 0, estimated via bootstrap with replacement.

Table 5a: Individual Divergence for DeGroot Models

Panel A: Divergence Across Networks

Network	*Observations*	Divergence		
		Uniform	*Degree*	*Eigenvector*
All	*9,205*	*0.0648*	*0.135*	*0.1083*
1	3,045	*0.0699*	*0.1513*	*0.0961*
2	3,031	*0.0788*	*0.1544*	*0.1562*
3	3,129	*0.0386*	*0.0866*	*0.0598*

Panel B: Share of Best-Fitting Sessions

H_0	H_1	Informative	H0 beats H1	Best model
Uniform	*Degree*	*0.65*	*0.8225*	*Uniform*
Uniform	*Eigenvector*	*0.4538*	*0.7712*	*Uniform*
Degree	*Eigenvector*	*0.3577*	*0.172*	*Eigenvector*

Notes: Panel A shows the individual level divergence across the three networks for each of the three DeGroot models.
In Panel B, the first column indicates the share of informative sessions over all sesssions, and the second column indicates the share of informative sessions where the H0 model beats the H1 model.

Table 5b: Lack of Bayesian Individual Behavior

H_1	Observations	Share Bayesian
Degree	46	17.39%
Uniform	98	17.35%
Eigenvector	98	17.35%

Note: "Observations" are the number of cases where there are discrepancies between the parent node action (which from t>3 Bayesian prescribes peripheral nodes should follow) and the action that the H1 model prescribes peripheral nodes. "Share Bayesian" is the share of observations where peripheral nodes indeed follow the parent node action.

Table 6: Zero Probability Information Sets Reached

Panel A: Complete Information Model

Network	% Individuals	% Treatments	% Observations
1	98.95%	98.95%	32.35%
2	100.00%	100.00%	33.18%
3	100.00%	100.00%	36.96%

Panel B: Incomplete Information Model

Degree Weighting Alternative

Network	% Individuals	% Treatments	% Observations
1	98.95%	98.95%	23.46%
2	100.00%	100.00%	19.78%
3	100.00%	100.00%	23.02%

Uniform Weighting Alternative

Network	% Individuals	% Treatments	% Observations
1	98.95%	98.95%	22.40%
2	100.00%	100.00%	29.29%
3	100.00%	100.00%	26.30%

Eigenvector Weighting Alternative

Network	% Individuals	% Treatments	% Observations
1	98.95%	98.95%	23.28%
2	100.00%	100.00%	25.91%
3	66.67%	100.00%	20.29%

Notes: Panel A presents results for the complete information Bayesian model. Panel B presents results for the incomplete information Bayesian model against DeGroot alternatives. % Individuals refers to the fraction of individuals who reach a zero probability information set. % Treatments refers to the fraction of treatments (network x village) that reaches a zero probability information set. % Observations refers to the fraction of individual x time units that reach a zero probability information set.

Table 7: Individual Level Divergence For Bayesian Model

Alternative	*Network*	*No. Obs.*	*At optimal* π
Degree	1	1858	0.173
	2	1898	0.210
	3	1358	0.144
Uniform	1	1847	0.192
	2	1675	0.198
	3	1300	0.140
Eigenvector Centrality	1	1822	0.177
	2	1750	0.188
	3	1383	0.109

Note: We present the individual level divergence for the Bayesian model on information sets in the support. No. of observations denotes the number of triples (individual, village, treatment) that were taken at non-zero probability information sets. Divergence is calculated conditional on all agents being Bayesian, with other potential types being Alternative. Optimal $\pi = 0$ for all networks and alternatives.

Table 8: Significance Tests for Individual Level Divergence

H_0	H_a	All Networks	Network 1	Network 2	Network 3
Bayesian	Degree	0.6438	0.794	0.357	0.0768
Bayesian	Uniform	0.0001	0.0001	0.0001	0.0001
Bayesian	Eigenvector	0.0001	0.0001	0.4267	0.0001
Degree	Uniform	0.0001	0.0001	0.0001	0.0001
Degree	Eigenvector	0.0001	0.0001	0.9999	0.0001
Uniform	Eigenvector	0.9999	0.9992	0.9999	0.9999

Notes: The test statistic is the normalized difference in the divergence functions of the null and the alternative. We show the probability that the test statistic is less than 0, estimated via bootstrap with replacement.

Table 9: Weight on indirect neighbors according to the number of multiple direct neighbors

	Data		Bayesian		Degree		Uniform	
	All	Restricted	All	Restricted	All	Restricted	All	Restricted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Signal	.4974***	.5351***	.3965***	.4429***	.5725***	.6204***	.6017***	.677***
	[.0218]	[.0279]	[.017]	[.0223]	[.0169]	[.0211]	[.0159]	[.0201]
Direct	.7735***	.7485***	.775***	.8048***	.9125***	.8913***	.8954***	.9078***
	[.0311]	[.0438]	[.0323]	[.0384]	[.0335]	[.0425]	[.0381]	[.0475]
One Way	.3157***	.2609***	.7056***	.8189***	0.0461	-0.0614	.1306***	.077*
	[.0405]	[.0542]	[.0363]	[.0486]	[.035]	[.0421]	[.0368]	[.0447]
Two Ways	.2177***	.1751***	.3582***	.4096***	.0797*	0.024	.1985***	.1841***
	[.0355]	[.0424]	[.0337]	[.0338]	[.0421]	[.0449]	[.0459]	[.0519]
TwoWays > OneWay								
T-statistic	-1.97	-1.557	-8.653	-8.136	0.7675	1.925	1.481	2.215
Probability	0.9741	0.9386	1	1	0.2224	0.0287	0.071	0.0146
Observations	1750	1271	1750	1271	1750	1271	1750	1271
R -squared	0.4724	0.373	0.5322	0.4632	0.588	0.5453	0.5974	0.5571

Note: Ouctome variable is action in round 3. "Direct" is the average signal of direct neighbors, "One Way" is the average signal of indirect neighbors only thorugh one direct neighbor, and "Two Ways is the average signal of indirect neighbors thorugh two direct neighbors.

Robust standard errors, clustered at the village by game level, in brackets. * p<.1, ** p<.05, *** p<.01 Odd columns are regressions with all data. Even columns are regressions restricting to treatments that are informative for the comparisons Bayesian - Degree and Bayesian - Uniform.

Table 10: Weight on indirect neighbors according to whether they provide new information

	Data		Bayesian		Degree		Uniform	
	All	Restricted	All	Restricted	All	Restricted	All	Restricted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Signal	.4284***	.4573***	.4729***	.5264***	.4507***	.482***	.5694***	.6468***
	[.0347]	[.0425]	[.0286]	[.0356]	[.032]	[.0359]	[.0307]	[.0357]
Direct	.7508***	.7011***	.8095***	.8007***	.8436***	.8167***	.7576***	.7275***
	[.0363]	[.0471]	[.0285]	[.0346]	[.0402]	[.0489]	[.0391]	[.0492]
Indirect New	.2455***	.1803***	.3768***	.3489***	.1558***	.0883**	.2548***	.2063***
	[.0331]	[.0377]	[.0261]	[.0317]	[.0333]	[.0384]	[.0283]	[.0338]
Indirect Repeated	.1715***	.1669***	.1029***	.0801*	.2463***	.2619***	.1724***	.1554***
	[.0416]	[.048]	[.0354]	[.0427]	[.0398]	[.0447]	[.0417]	[.0492]
Observations	1587	1250	1587	1250	1587	1250	1587	1250
R -squared	0.4628	0.3958	0.4953	0.4135	0.5819	0.5475	0.5945	0.5687

Note: Ouctome variable is action in round 3. "Direct" is the average signal of direct neighbors, "Indirect New" is the average signal of indirect neighbors that provide new information, and "Indirect Repeated" is the average signal of indirect neighbors that do not provide new information.

Robust standard errors, clustered at the village by game level, in brackets. * p<.1, ** p<.05, *** p<.01 Odd columns are regressions with all data. Even columns are regressions restricting to treatments that are informative for the comparisons Bayesian - Degree and Bayesian - Uniform.

Table 11: Weight on neighbors according to distance.

	Data		Bayesian		Degree		Uniform	
	All	Restricted	All	Restricted	All	Restricted	All	Restricted
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Signal	.5131***	.5535***	.4118***	.4457***	.5907***	.6599***	.6489***	.7337***
	[.0235]	[.0303]	[.0219]	[.0266]	[.0209]	[.0259]	[.0197]	[.022]
Direct	.7902***	.7847***	.9039***	.957***	.9347***	.9666***	.9118***	.9379***
	[.0345]	[.0537]	[.0367]	[.0486]	[.0369]	[.0508]	[.0407]	[.0544]
Distance 2	.3261***	.2872***	.6484***	.7207***	.1645***	.1392**	.2369***	.2264***
	[.0398]	[.0573]	[.0394]	[.0518]	[.045]	[.0548]	[.0428]	[.0523]
Distance 3	.1042**	0.0764	.3517***	.3947***	.1125***	0.0673	.1609***	.1315***
	[.043]	[.0587]	[.0428]	[.0533]	[.0357]	[.045]	[.0335]	[.0425]
Distance 4	0.0583	0.0266	.15**	.1778**	0.0295	-0.0278	0.0498	0.0081
	[.0656]	[.0777]	[.0594]	[.0733]	[.0581]	[.0668]	[.05]	[.0562]
Test Distance 2 > Distance 3								
T-statistic	5.021	4.289	6.807	7.036	0.9166	1.132	1.319	1.453
Probability	1.20E-06	2.30E-05	4.60E-10	2.00E-10	0.1808	0.1304	0.0951	0.0749
Test Distance 3 > Distance 4								
T-statistic	0.8282	0.8034	3.893	3.686	1.433	1.43	1.809	1.758
Probability	0.2048	0.2119	9.20E-05	2.00E-04	0.0776	0.0782	0.0368	0.0411
Observations	1330	1001	1330	1001	1330	1001	1330	1001
R -squared	0.4822	0.3903	0.5706	0.5078	0.6279	0.5833	0.6442	0.6133

Note: Ouctome variable is action in last round. "Direct" is the average signal of direct neighbors, "Distance #" is the average signal of indirect neighbors with distance #.

Robust standard errors, clustered at the village by game level, in brackets. * p<.1, ** p<.05, *** p<.01 Odd columns are regressions with all data. Even columns are regressions restricting to treatments that are informative for the comparisons Bayesian - Degree and Bayesian - Uniform.

4.B Complete Information Bayesian Algorithm

4.B.1 Setup

To begin with, suppose that all agents learn about the state of the world using Bayes' rule. Assume that this fact is common knowledge for all agents. We will extend this framework below, but it is the most pedagogical model to help establish notation. It is crucial to keep in mind that the signal endowment w is a sufficient statistic over the actions that agents take, since θ is never observed. In turn, the inference that agents need to do from the other agents' play is concerned with the signal endowment. To proceed we must establish some notation.

We define $\mu_0^{\theta}(w)$ as the probability of signal endowment w when the true state of the world is θ. Based on the previous assumptions, we have that:

$$\mu_0^{\theta=1}(w) := \Pr(w \mid \theta = 1) = p^{\#\{i:w_i=1\}}(1-p)^{N-\#\{i:w_i=1\}} = p^{\left(\sum_{i=1}^{n} w_i\right)}(1-p)^{n\left(1-\sum_{i=1}^{n} w_i\right)} \quad (4.3)$$

$$\mu_0^{\theta=0}(w) := \Pr(w \mid \theta = 1) = p^{\left(\sum_{i=1}^{i=N} \frac{w_i}{n}\right)}(1-p)^{\left(N-\sum_{i=1}^{i=N} \frac{w_i}{n}\right)} \quad (4.4)$$

Following the same reasoning, define $\mu_{i,t}^{\theta}(w, \overline{w})$ as the belief probability distribution that agent i has at period t of the game over states $w \in W$, given that the true signal endowment is $\overline{w}$. Observe that for different signal endowments, the information sets that each agent will observe are clearly going to be different. In turn, the belief over signal endowments that each agent has at each state depends on what was the true signal endowment. Of course, this measure will be constant over signal endowments that generate the same information sets.

Also, define $p_{i,t}(\overline{w})$ as the belief probability that agent i has at stage t of the game puts on the event $\theta = 1$, if the information set he observes comes from the true signal endowment being $\overline{w}$. Moreover, put $\mathbf{a}_{i,t}(\overline{w}) \in \{0,1\}$ as the action that agent i takes at stage t of the game, if the information set reached comes from the true signal endowment $\overline{w}$. The objects that will be iteratively found will be $\mu_{i,t}^{\theta}(w, \overline{w})$, $\mathbf{a}_{i,t}(\overline{w})$ and $p_{i,t}^{\theta}(\overline{w})$.

To start the iterations, we need to find these objects at $t = 1$. These are given by

$$p_{i,1}^{\theta}(\overline{w}) = \begin{cases} p \text{ if } \overline{w}_i = s_i = 1 \\ 1-p \text{ if } \overline{w}_i = s_i = 0 \end{cases} \quad (4.5)$$

$$\mathbf{a}_{i,1}(\overline{w}) = s_i \quad (4.6)$$

$$\mu_{i,1}^{\theta=1}(w, \overline{w}) = \begin{cases} 0 \ \forall w \in W : \ w_i \neq s_i \\ \frac{\mu_0^{\theta=1}(w)}{\Pr(s_i \mid \theta=1)} \text{ if } w_i = s_i \end{cases} \quad (4.7)$$

$$\mu_{i,1}^{\theta=0}(w, \overline{w}) = \begin{cases} 0 \ \forall w \in W : \ w_i \neq s_i \\ \frac{\mu_0^{\theta=0}(w)}{\Pr(s_i \mid \theta=0)} \text{ if } w_i = s_i \end{cases} \quad (4.8)$$

To model what is exactly what each agent observes, in order to see how beliefs are updated over the

course of the game, we need to introduce some network notation. In particular, $N(i) := N_i \cup \{i\}$, the set of neighbors of agent i, including i herself. Next, define

$$\mathbf{a}_t^{(i)}(\overline{w}) := \left(\underbrace{\mathbf{a}_{j_1,t-1}(\overline{w}), \mathbf{a}_{j_2,t-1}(\overline{w}), ..., \mathbf{a}_{j_{d(i)},t-1}(\overline{w})}_{\text{actions taken by neighbors}}, \underbrace{\mathbf{a}_{i,t-1}(\overline{w})}_{\text{own past action}} \right) \tag{4.9}$$

to be the action profile that agent i sees at the beginning of state t, when the true state of the world is $\overline{w}$. If we just write $a_t^{(i)}$ (without bold letters) we refer to a particular observed action profile.

4.B.2 Time $t+1$ iteration

At time t, we have

$$\begin{aligned}
\mu_{i,t}^{(\theta=1)}(w,\overline{w}) &= \Pr(w \mid \theta = 1, I_{i,t}(\overline{w})) \\
\mu_{i,t}^{(\theta=0)}(w,\overline{w}) &= \Pr(w \mid \theta = 0, I_{i,t}(\overline{w})) \\
p_{i,t}^{(\theta=1)}(\overline{w}) &= \Pr(\theta = 1 \mid I_{i,t}(\overline{w})) \\
p_{i,t}^{(\theta=0)}(\overline{w}) &= \Pr(\theta = 0 \mid I_{i,t}(\overline{w}))
\end{aligned}$$

where $I_{i,t}(\overline{w})$ is the information set agent i is at stage t of the game, given that the true type endowment is $\overline{w}$. Of course, agent i does not know $\overline{w}$, but only the information set $I_{i,t}$. Suppose now the agents receive new information, namely, $w \in I_{t+1}(\overline{w}) \subseteq I_t(\overline{w})$. Then

$$\mu_{i,t+1}^{(\theta=1)}(w,\overline{w}) := \Pr(w \mid \theta = 1, I_{i,t+1}(\overline{w})) = \begin{cases} 0 \text{ if } w \notin I_{i,t+1}(\overline{w}) \\ \frac{\mu_{i,t}^{(\theta=1)}(w,\overline{w})}{\sum_{w' \in I_{t+1}} \mu_{i,t}^{(\theta=1)}(w,\overline{w})} \text{ if } w \in I_{i,t+1}(\overline{w}) \end{cases}. \tag{4.10}$$

Likewise,

$$\mu_{i,t+1}^{(\theta=0)}(w,\overline{w}) := \Pr(w \mid \theta = 0, I_{i,t+1}(\overline{w})) = \begin{cases} 0 \text{ if } w \notin I_{i,t+1}(\overline{w}) \\ \frac{\Pr(w|\theta=0, I_t(\overline{w}))}{\sum_{w' \in I_{t+1}} \Pr(w'|\theta=0, I_{i,t}(\overline{w}))} \text{ if } w \in I_{i,t+1}(\overline{w}) \end{cases}. \tag{4.11}$$

Based on the new probability distribution over signal endowments, we can get the probability over θ as

$$\begin{aligned}
p_{i,t+1}^{(\theta=1)}(\overline{w}) &= \Pr(\theta = 1 \mid I_{i,t+1}(\overline{w})) = \Pr(\theta = 1 \mid I_{i,t}(\overline{w}) \cap I_{i,t+1}(\overline{w})) \\
&= \frac{p_{i,t}^{(\theta=1)}(\overline{w}) \sum_{w \in I_{i,t+1}(\overline{w})} \mu_{i,t}^{(\theta=1)}(w,\overline{w})}{p_{i,t}^{(\theta=1)}(\overline{w}) \sum_{w \in I_{i,t+1}} \mu_{i,t}^{(\theta=1)}(w,\overline{w}) + \left(1 - p_{i,t}^{(\theta=1)}(\overline{w})\right) \sum_{w \in I_{i,t+1}} \mu_{i,t}^{(\theta=0)}(w,\overline{w})}
\end{aligned} \tag{4.12}$$

Therefore, we need to compute the relevant information sets. Let $\mathbf{a}_{i,t}(w)$ be the action that agent i takes at time t if configuration of signals is w. Then we can consider the set of worlds that have

positive probability at time t, given by

$$W_{t+1}^{(i)}\left(\mathbf{a}_t^{(i)}\right) = \{w \in W : \mathbf{a}_{j,t}(w) = a_{j,t} \text{ for all } j \in N(i)\} \tag{4.13}$$

and hence,

$$I_{i,t+1}(\overline{w}) = W_{t+1}^{(i)}\left(\mathbf{a}_t^{(i)}(\overline{w})\right) \tag{4.14}$$

In deriving the information set, we were able to eliminate the path of actions observed by agent i so far by realizing that the actions taken by an agent are deterministic functions of what they observe. Thus, once we have conditioned on the signal endowment w, the actions observed $a_t^{(i)}(w)$ are completely determined; conditioning on them does not change the probability calculation. However, in environments where we allow for random actions, as opposed to deterministic actions, this fact is no longer true. While it is a perfectly sensible and reasonable model, it complicates things computationally in an astronomical manner. We will return to this problem below.

4.B.3 Actions

The algorithm described so far gives us how a Bayesian agent i would update her beliefs if she knows $\mathbf{a}_{j,t}(\overline{w})$ for all j and has prior beliefs on signal endowments and states of the world given by

$$\left\{\mu_{i,t}^{(\theta=1)}(w,\overline{w}), \mu_{i,t}^{(\theta=0)}(w,\overline{w}), p_{i,t}^{(\theta=1)}(\overline{w})\right\}. \tag{4.15}$$

If agent i is Bayesian, then the decision at $t+1$ is given by (when there are no ties)

$$\mathbf{a}_{i,t+1}(\overline{w}) = \begin{cases} 1 \text{ if } p_{i,t+1}^{(\theta=1)}(\overline{w}) > \frac{1}{2} \\ 0 \text{ if } p_{i,t+1}^{(\theta=1)}(\overline{w}) < \frac{1}{2} \\ a \in \{0,1\} \text{ if } p_{i,t+1}^{(\theta=1)}(\overline{w}) = \frac{1}{2} \end{cases} \tag{4.16}$$

Note that when $p_{i,t+1}^{(\theta=1)}(\overline{w}) = \frac{1}{2}$, we need to use some tie breaking rule. We will use the "past action" rule. That is, when faced with a tie, an individual will play the action she played in the previous round, $p_{i,t+1}^{(\theta=1)}(\overline{w}) = \frac{1}{2} \implies \mathbf{a}_{i,t+1}(\overline{w}) = \mathbf{a}_{i,t}(\overline{w})$. Of course, one could think of many other tie breaking rules, including random tie breaking rule, where the agent plays each action with the same probability. However, as we will see, this such a model will be computationally intractable in our framework.

Observe that the above framework extends to situations where some agents play ad hoc decision rules via DeGroot learning. Suppose that each agent i may be of some type $\eta \in H := \{\eta_1, \eta_2, ..., \eta_K\}$. For example, take the type space to be

$$H = \{\text{Bayesian, Eigenvector, Naive, Preferential}\} \tag{4.17}$$

so each agent may be either a Bayesian agent or a DeGroot agent who constructs simple linear

indexes from the past actions taken by neighbors. In particular, suppose that agent i has type η_i = Preferential. In world $\overline{w}$ and time $t+1$, she observes actions $\mathbf{a}_t^{(i)}$. Based on this, she defines the following index:

$$Preferential_{i,t}(\overline{w}) := \sum_{j \in N(i)} \mathbf{a}_{j,t}(\overline{w}) T_{ij}^{Pref} \tag{4.18}$$

Therefore, the corresponding action rule is

$$\mathbf{a}_{i,t+1}^{P}(\overline{w}) = \begin{cases} 1 \text{ if } Preferential_{i,t}(\overline{w}) > \frac{1}{2} \\ 0 \text{ if } Preferential_{i,t}(\overline{w}) < \frac{1}{2} \\ a \in \{0,1\} \text{ if } Preferential_{i,t}(\overline{w}) = \frac{1}{2} \end{cases}. \tag{4.19}$$

Similarly, we can construct $\mathbf{a}_{i,t+1}^{N}(\overline{w})$, $\mathbf{a}_{i,t+1}^{E}(\overline{w})$ using T^{Naive} and T^{Eig}, respectively.

As long as *agents' types are common knowledge*, the algorithm described so far can handle heterogeneity in agents' types without changing the nature of the Bayesian updating.

4.B.4 Algorithm To Find Action Rules

Step 1: Initiate the algorithm with $\mu_{i,1}^{\theta=1}(w,\overline{w})$ and $\mu_1^{\theta=0}(w,\overline{w})$ given by (4.3) and (4.4), action rule $\mathbf{a}_{i,1}(\overline{w})$ as in (4.6) and $p_{i,1}(\overline{w})$ as in (4.5).

Step 2: At period $t+1$, start with $\left\{\mu_{i,t}^{\theta=1}(w,\overline{w}), \mu_{i,t}^{\theta=0}(w,\overline{w}), p_{i,t}(\overline{w}), \mathbf{a}_{i,t}(\overline{w})\right\}$. Derive $\mu_{i,t+1}^{\theta=1}(w)$ and $\mu_{i,t+1}^{\theta=0}(w)$ using using equations (4.11) and (4.10).
Obtain $p_{i,t+1}(\overline{w})$ from (4.12), and then derive the action that each agent takes, $\mathbf{a}_{i,t+1}(\overline{w})$ depending on the agent's type.

Step 3: Repeat Step 2 until $t = T$.

4.C Incomplete Information Bayesian Algorithm

We only include the modification from the above. The only difference now is how to start the algorithm: see that now

$$\mu_0^{\theta=1}(w) = \mu_0^{\theta=1}(s,\eta) = \left(p^{\sum_{i=1}^n s_i}(1-p)^{n-\sum_{i=1}^n s_i}\right)\prod_{i=1}^n \pi_i^{\eta_i}(1-\pi_i)^{1-\eta_i} \tag{4.20}$$

$$\mu_0^{\theta=0}(w) = \mu_0^{\theta=1}(s,\eta) = \left((1-p)^{\sum_{i=1}^n s_i} p^{n-\sum_{i=1}^n s_i}\right)\prod_{i=1}^n \pi_i^{\eta_i}(1-\pi_i)^{1-\eta_i} \tag{4.21}$$

See that the assumption of independent types is immaterial to the description of the algorithm: we could substitute in principle the term $\prod_{i=1}^n \pi_i^{\eta_i}(1-\pi_i)^{1-\eta_i}$ for some function $F(\eta) = \Pr(\eta)$.

After each agent sees their "signal" (s_i, η_i), we can calculate the derived measures over worlds

as

$$\begin{aligned}\mu_{i,1}^{\theta=1}(w,\overline{w}) &= \Pr\left((s,\eta),(\overline{s_i},\overline{\eta_i}) \mid (s_i,\eta_i),\theta=1\right) \\ &= \frac{\Pr(w \mid \theta=1)\Pr\left((s_i,\eta_i) \mid w,\theta=1\right)}{\sum_{z\in W}\Pr(z)\Pr\left((s_i,\eta_i) \mid z,\theta=1\right)} = \begin{cases} 0 \text{ if } s_i \neq \overline{s_i}, \text{ or } \eta_i' \neq \overline{\eta_i} \\ \frac{\mu_0^{\theta=1}(w)}{\sum_{z\in W}\mu_0^{\theta=1}(z)\Pr(s_i,\eta_i|z,\theta=1)} \text{ otherwise} \end{cases}\end{aligned}$$

and similarly for $\mu_{i,1}^{\theta=0}(w,\overline{w})$. For the other needed objects for the algorithm, we use:

$$\mathbf{a}_{i,1}(\overline{w}) = \begin{cases} 1 \text{ if } \overline{s_i}=1 \\ 0 \text{ otherwise} \end{cases}$$

$$p_{i,1}(\overline{w}) = \begin{cases} p \text{ if } \overline{s_i}=1 \\ 1-p \text{ if } \overline{s_i}=0 \end{cases}$$

For the rest of the algorithm, the action rule will depend on the type. Let $\mathbf{a}_{i,t}^{B}(\overline{w})$ denote the probability with which agent i at stage t plays $a=1$ in world $\overline{w}$ if acts as a Bayesian, and $\mathbf{a}_{i,t}^{M}(\overline{w})$ be the analogous if the agent act as a $M-$weighter (and then, does not depend on which state of the world she's in, which only affect the learning that she has done so far). Then, the action profile at time t is

$$\mathbf{a}_{i,t}(\overline{w}) = \begin{cases} \mathbf{a}_{i,t}^{B}(\overline{w}) \text{ if } \overline{\eta_i}=1 \\ \mathbf{a}_{i,t}^{M}(\overline{w}) \text{ if } \overline{\eta_i}=0 \end{cases}$$

4.D Filtering

Define

$$F_0(\eta) = \Pr(\eta \mid I_0,\hat{\pi})$$

with I_0 the information set of the statistician at $t=0$. Since we are controlling the experiment, we know that $\{s,\theta\}=I_0$. Call s^* and θ^* the chosen values by the experimentalist. For example, if type endowments are independent of both θ and s and the location on the network

$$F_0(\eta) = \prod_{i=1}^{n}\pi_i^{\eta_i}(1-\pi_i)^{1-\eta_i} \tag{4.22}$$

Now, suppose that we have calculated at to time $t-1$:

$$F_{t-1}(\eta) = \Pr(\eta \mid I_{t-1},\hat{\pi})$$

Define

$$\mathcal{A}_{i,t-1}^{*}(\eta) = \Pr(a_{i,t}=1 \mid I_{t-1},(s^*,\eta),\hat{\pi}) = \begin{cases} 1-\varepsilon \text{ if } A_{i,t-1}(\eta,s^*)=1 \\ \varepsilon \text{ if } A_{i,t-1}(\eta,s^*)=1 \end{cases} \tag{4.23}$$

that means, is the probability distribution that the statistician has over actions if she knew the true type endowment. The probability ε is to ensure that as statisticians, we put positive weight on every history, and later we take $\varepsilon \to 0$. This will not be a problem empirically, since all histories have positive probability empirically.

See that $\mathcal{A}^*_{i,1}(\eta) = \mathbf{1}(s_i = s^*_i)$ for all $\eta \in H$ for any (reasonable) model that for any agent, if they see only their signal, they choose their own signal. Let $\mathbf{a}^*_t$ be the $N \times 1$ action vector observed by the experimenter at time t.

This function allows us to get the conditional probabilities over actions of all agents as:

$$\Pr(\mathbf{a}_t \mid I_{t-1}, (s^*, \eta), \hat{\pi}) = \left(\prod_{j:a_j=1} \Pr(a_{j,t} = 1 \mid I_{t-1}, (s^*, \eta), \hat{\pi})\right)\left(\prod_{j:a_j=0} (1 - \Pr(a_{j,t-1} = 1 \mid I_{t-1}, (s^*, \eta)), \hat{\pi})\right) =$$

$$\left(\prod_{j:a_j=1} \mathcal{A}^*_{i,t-1}(\eta)\right)\left(\prod_{j:a_j=0} \left(1 - \mathcal{A}^*_{i,t-1}(\eta)\right)\right) \tag{4.24}$$

Using in (4.24) that agents randomize over actions independently. Then, after observing action vector $a^*_t \in \{0,1\}^n$, the statistician updates her beliefs over types as

$$F_t(\eta) = \Pr(\eta \mid I_{t-1}, a^*) = \frac{\Pr(\eta \mid I_{t-1}, \hat{\pi}) \Pr(\mathbf{a}^*_t \mid I_{t-1}, \eta, \hat{\pi})}{\sum_{\tilde{\eta}} \Pr(\tilde{\eta} \mid I_{t-1}) \Pr(\mathbf{a}^*_t \mid I_{t-1}, \tilde{\eta}, \hat{\pi})}$$

$$= \frac{F_t(\eta)\left(\prod_{j:a^*_j=1} \mathcal{A}^*_{i,t-1}(\eta)\right)\left(\prod_{j:a^*_j=0} \left(1 - \mathcal{A}^*_{i,t-1}(\eta)\right)\right)}{\sum_{\eta \in H} F_t(\tilde{\eta})\left(\prod_{j:a^*_j=1} \mathcal{A}^*_{i,t-1}(\tilde{\eta})\right)\left(\prod_{j:a^*_j=0} \left(1 - \mathcal{A}^*_{i,t-1}(\tilde{\eta})\right)\right)} \tag{4.25}$$

To finish up the algorithm, we need to calculate $\mathcal{A}^*_{i,t}(\eta)$, but this comes directly from the algorithm described above.

The algorithm then to get the distribution of η conditional on the whole set of information is:

Step 1: Initiate algorithm with $F_0(\eta)$ as in (4.22)and an action function $\mathcal{A}^*$ as described above. Moreover, introduce information about s^* (only thing that we actually need)

Step t < T : Taking $\mathcal{A}^*_{i,t-1}(\eta)$ as given, run learning code as in the previous section and calculate $\mathcal{A}^*_{i,t}(\eta)$ as in (4.23)

Step t = T : Once $\mathcal{A}^*_{i,T-1}(\eta)$ is calculated, calculate likelihood over type endowments as

$$F_T(\eta) = \frac{F_{T-1}(\eta)\left(\prod_{j:a^*_{Tj}=1} \mathcal{A}^*_{i,T-1}(\eta)\right)\left(\prod_{j:a^*_T=0} \left(1 - \mathcal{A}^*_{i,T-1}(\eta)\right)\right)}{\sum_{\tilde{\eta} \in H} F_{T-1}(\tilde{\eta})\left(\prod_{j:a^*_{Tj}=1} \mathcal{A}^*_{i,T-1}(\tilde{\eta})\right)\left(\prod_{j:a^*_T=0} \left(1 - \mathcal{A}^*_{i,T-1}(\tilde{\eta})\right)\right)} \tag{4.26}$$

Last Step : Get the probability of being Bayesian of agent i as

$$\Pi_{i,t,v}(\pi) \equiv \Pr(\eta_i = 1) = \sum_{\tilde{\eta} \in H : \tilde{\eta}_i = 1} F_T(\tilde{\eta}) \tag{4.27}$$

4.E Stuck Nodes and Concentric Social Quilts

4.E.1 Stuck Nodes

Given an undirected graph $G = (V, E)$ and a subset of nodes $v \subseteq V$ we define $G_v = (v, E_v)$ as the induced subgraph for subset v, where $E_v = \{(ij) \in E : \{i, j\} \subseteq v\}$. Given a subgraph G_v, let $d_i(G_v)$ be the degree of node i in subgraph G_v. Let $a_{it} \in \{0, 1\}$ be the action that node $i \in V$ takes at round $t \in \mathbb{N}$, which we will assume follows the uniform DeGroot action model; i.e. $a_{i,t} = \mathbf{1}\left\{\frac{1}{d_i+1}\sum_{j \in N_i} a_{j,t-1} > \frac{1}{2}\right\}$. We allow for any tie-breaking rule when $\frac{1}{d_i+1}\sum_{j \in N_i} a_{j,t-1} = \frac{1}{2}$.

LEMMA 4.E.1. *Take a subset of individuals $v \subseteq V$ such that there exists $h \in \mathbb{N}$ with*

$$h \leq d_i(G_v) \leq d_i < 2h + 1 \text{ for all } i \in v$$

If agents behave according to the uniform weighting DeGroot action model, and at some $T \in \mathbb{N}$ we have $a_{i,T} = a \in \{0, 1\}$ for all $i \in v$, then $a_{i,t} = a$ for all $t \geq T$.

Proof. The proof is by induction: without loss of generality, suppose $a_{i,T} = 1$ for all $i \in v$. Of course, for $t = T$ the result is trivially true. Suppose now that $a_{i,t} = 1$ for all $i \in v$ and $t \geq T$, and we need to show that $a_{i,t+1} = 1$ too. Let $I_{i,t+1} = \frac{1}{d_i+1}\sum_{j \in N_i} a_{j,t}$ be the index of uniform weighting. We then now that $I_{i,t} \geq \frac{1}{2}$ for all nodes in v, and it suffices to show that this implies $I_{i,t+1} \geq \frac{1}{2}$. Observe,

$$\begin{aligned} I_{i,t+1} &= \frac{\sum_{j \in N_i} a_{j,t}}{d_i + 1} = \frac{\sum_{j \in v \cap N_i} \underbrace{a_{j,t}}_{=1} + \sum_{j \in N_i - v} a_{j,t}}{d_i + 1} \underset{(i)}{\geq} \frac{h + 1 + \sum_{j \in N_i - v} a_{j,t}}{d_i + 1} \\ &\geq \frac{h+1}{d_i+1} \underset{(ii)}{>} \frac{1}{2}. \end{aligned}$$

We have used in (i) that $d_i(G_v) \geq h$ and $a_{j,t} = 1$ for all $j \in v$. Inequality (ii) comes from the fact that

$$\frac{h+1}{d_i+1} > \frac{1}{2} \iff d_i < 2h + 1.$$

Therefore, we have that $I_{i,t+1} > \frac{1}{2}$ for any $i \in v$, implying then that $a_{i,t+1} = 0$, as we wanted to show. □

This lemma says that whenever we find a subset of nodes v such that each node has more connections to nodes in v than it has outside v, then whenever they reach consensus, they would remain there forever. We present an useful corollary of Lemma 4.E.1, which we will use when studying the family CSQ_r.

COROLLARY 4.E.1 (Regular Subgraphs). *Take a family of nodes $v \in V$ such that there exists $k \in \mathbb{N}$ such that*

1. *G_v is a k-regular graph*
2. *$d_i < 2k + 1$ for all $i \in v$.*

Then, if agents behave according to the uniform weighting DeGroot action model, and at some $T \in \mathbb{N}$ *we have* $a_{i,T} = a \in \{0,1\}$ *for all* $i \in v$*, then* $a_{i,t} = a$ *for all* $t \geq T$*.*

Proof. Simply take $h = k$ and apply Lemma 4.E.1. □

See that any triangle in CSQ_r is a $2-$ regularsubgraph, and that each node in it has $d_i = 4 < 2 * 2 + 1 = 5$, so we apply Corollary 4.E.1 with $k = 2$. So, whenever a triangle achieves concensus, it remains there forever.

4.E.2 Concentric Social Quilts: Preliminaries

Let's now focus on Concentric Social Quilts. Define $S_r = \{i \in V_r : i \text{ gets stuck}\}$ and let $N_r = \#(V_r)$, the number of nodes in CSQ_r. Our object of interest is the random variable

$$\mathcal{F}_r(r) = \text{Fraction of nodes in } CSQ_r \text{ that gets stuck} \equiv \frac{\#S_r}{N_r}$$

which is a random variable. Our objective is to get an asymptotic bound on $\mathcal{F}_r$. Since we do not yet know whether $\mathcal{F}_r$ has a limit for almost every realization, we define $\underline{\mathcal{F}}$ and $\overline{\mathcal{F}}$as

$$\underline{\mathcal{F}} = \liminf_{r\to\infty} \mathcal{F}_r \text{ and } \overline{\mathcal{F}} = \limsup_{r\to\infty} \mathcal{F}_r \tag{4.28}$$

which is well defined for all realizations of the sequence $\mathcal{F}(r)$ and so it is a well defined random variable. Namely, we want to get the tightest asymptotic lower and upper bounds for the fraction of stuck nodes. Our objective is to get a number $F \in [0,1]$ such that $\underline{\mathcal{F}} \geq \underline{F}$ and $\overline{\mathcal{F}} \leq \overline{F}$ almost surely; i.e. $\mathrm{P}\left\{\underline{F} \leq \underline{\mathcal{F}} \leq \overline{\mathcal{F}} \leq \overline{F}\right\} = 1$.

Any CSQ_r has an initial parent triangle $P = (i_0, i_1, i_2)$. For any node $i \in V_r$ we define the distance to parent $d(i,P) = \min\{d(i,i_0), d(i,i_1), d(i,i_2)\}$ where $d(i,j)$is the minimum number of links we have to go through to connect node i with node j. Likewise, given a triangle $T = (i_s, i_k, i_j)$, we define the distance between triangle T and the parent triangle P as

$$d(T,P) = \max_{i\in T}\{d(i,P)\}$$

DEFINITION 4.E.1 (Ring). *Given a graph* $CSQ_r = (V_r, E_r)$ *and* $s \in \{1,...,r\}$ *we define* R_s*, the level* s *ring of* CSQ_r *as the subgraph* $R_s = (V_r^s, E_r^s)$ *of nodes that lie in triangles with distance to parent* $d(T,P) = s-1$*.*

Intuitively, a ring is just the collections of triangles that lie in the $s-th$ level of the concentric social quilt as seen in Figure 4-11. Note that for all s, R_s is a graph that consists of disconnected triangles.

Define

$$OR_r(k) := \bigcup_{s=0}^{s=k} R_{r-s}$$

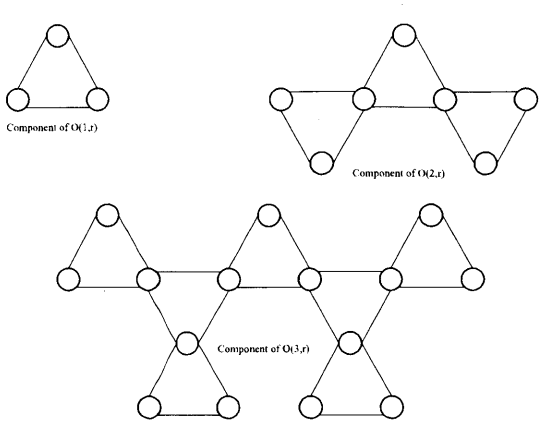


Figure 4-11: Outer Rings.

as the subgraph formed by the outer rings from $r-k$ to r. This subgraph is also disconnected, with a lot of components, which now are no longer triangles, but rather "trees of triangles" as pictured in Figure 4-11.

Let $C \subset OR_r(k)$ be a component (a subgraph as shown in Figure 4-11). The last level of nodes in every component correspond to terminal nodes. The most important property of these components is that **the only connection between each component C and the rest of the graph is the parent node of the component C**, denoted by i_C . This will be the key property of these components, which we will try to explore.

Define

$$\Psi_r(k) = \frac{\#\{OR_r(k) \cap S_r\}}{\#\{OR_r(k)\}}$$

to be the fraction of stuck nodes in $OR_r(k)$

$$\underline{\Psi}(k) := \liminf_{r\to\infty} \frac{\#\{OR_r(k) \cap S_r\}}{\#\{OR_r(k)\}}$$

and

$$\overline{\Psi}(k) := \limsup_{r\to\infty} \frac{\#\{OR_r(k) \cap S_r\}}{\#\{OR_r(k)\}}$$

which is also a well defined random variable. These are the tightest asymptotic lower and upper bounds on the fraction of nodes stuck in the last k rings. That is, a lower bound on the fraction of nodes in $OR_r(k)$ that get stuck.

LEMMA 4.E.2. *For all* $k \in \mathbb{N}$, $\lim_{r\to\infty} \frac{\#\{OR_r(k)\}}{N_r} = \frac{2^{k+1}-1}{2^{k+1}}$.

Proof. Let L_r = number of terminal triangles in ring r. Of course, we have that $L_r = T_{r-1}$. Because of how CSQ_r grows, we have the following recursion for L_r :

$$L_{r+1} = 2L_r \text{ and } L_2 = 3.$$

It can be easily shown that

$$L_r = 3 * 2^{r-2}. \tag{4.29}$$

We also need to calculate $N_r = \#(V_r)$. Again, because of how CSQ_r is generated, we have the following recursion for N_r :

$$N_{r+1} - N_r = 2L_{r+1}$$

and it can be also easily shown that

$$N_r = 3(2^r - 1) \tag{4.30}$$

Finally, let n_k be the number of nodes in a component $C \subset OR_r(k)$. It is also easy to show that

$$n_k = 2^{k+1} - 1$$

Now, we can state the result. Observe that

$$\begin{aligned} \frac{\#\{O_r(k)\}}{N_r} &= \frac{\overbrace{n_k}^{\text{nodes per component}} \times \overbrace{L_{r-k+1}}^{\text{number of components}}}{N_r} = \left(2^{k+1} - 1\right) \frac{3 \times 2^{r-k+1-2}}{3(2^r - 1)} \\ &= \frac{2^{k+1} - 1}{2^{k+1}} \left(\frac{2^r}{2^r - 1}\right) \longrightarrow_{r \longrightarrow \infty} \frac{2^{k+1} - 1}{2^{k+1}} \end{aligned}$$

as we wanted to show. □

The following proposition is the key to understand how to get bounds on $\underline{\mathcal{F}}$ and $\overline{\mathcal{F}}$ by getting bounds on $\underline{\Psi}(k)$ and $\underline{\Psi}(k)$

PROPOSITION 4.E.1. *Suppose there exist functions $\overline{\psi}, \underline{\psi} : \mathbb{N} \to [0,1]$ such that for all k we have*

$$\underline{\psi}(k) \leq \underline{\Psi}(k) \leq \overline{\Psi}(k) \leq \overline{\psi}(k) \text{ almost surely.}$$

Then, for all $k \in \mathbb{N}$ almost surely,

$$\underline{\mathcal{F}} \geq \frac{2^{k+1} - 1}{2^{k+1}} \underline{\psi}(k) \tag{4.31}$$

and

$$\overline{\mathcal{F}} \leq 1 - \left(\frac{2^{k+1} - 1}{2^{k+1}}\right) \left[1 - \overline{\psi}(k)\right]. \tag{4.32}$$

Proof. Lets focus only on inequality 4.31, since 4.32 follows the same reasoning. See that

$$\begin{aligned} \mathcal{F}(r) = & \frac{\#\{O_r(k)\}}{N_r} \left(\frac{\#\{O_r(k) \cap S_r\}}{\#\{O_r(k)\}}\right) + \frac{\#\{S_r - O_r(k)\}}{N_r} \\ \geq & \frac{\#\{O_r(k)\}}{N_r} \left(\frac{\#\{O_r(k) \cap S_r\}}{\#\{O_r(k)\}}\right) \end{aligned}$$

so, for all realizations,

$$\begin{aligned} \underline{\mathcal{F}} &= \liminf_{r \to \infty} \mathcal{F}(r) \geq \left(\lim_{r \longrightarrow \infty} \frac{\#\{O_r(k)\}}{N_r}\right) \left(\liminf_{r \to \infty} \left(\frac{\#\{O_r(k) \cap S_r\}}{\#\{O_r(k)\}}\right)\right) \\ &= \left(\lim_{r \longrightarrow \infty} \frac{\#\{O_r(k)\}}{N_r}\right) \underline{\Psi}(k) = \frac{2^{k+1} - 1}{2^{k+1}} \underline{\Psi}(k). \end{aligned}$$

This, together with the fact that $\underline{\Psi}(k) \geq \underline{\psi}(k)$ almost surely, finishes the proof. □

Note that this proposition is true for any learning model (Bayesian or DeGroot). The learning model plays a role when calculating the bounds $\underline{\psi}$and $\overline{\psi}$. See that condition 4.31 and 4.32 are bounds on $\underline{\mathcal{F}}$ and $\overline{\mathcal{F}}$, which do not depend on k: therefore, these are bounds for all k: the higher k, the tighter the bound we get.

4.E.3 Bounding stuck nodes in the Uniform Weighting model

To normalize, we will assume that the true state of nature is $\theta = 1$, which implies that as $r \to \infty$ the fraction of nodes with true signals is $p > \frac{1}{2}$. The idea is pretty simple: take a component $C = (V_C, E_C) \subset OR(k, r)$. As we mentioned before, the only connection between C and the rest of the graph is through the parent node i_C (as seen in Figure 4-11). Let $W_C = \{0,1\}^{n_k}$ be the set of signal endowments for nodes in C. We will try to find a lower bound $\underline{\psi}_k(w)$ for each signal endowment realization in C such that, when signal endowment is w, the fraction of stuck nodes in C is larger that $\underline{\psi_k}(w)$fraction of stuck nodes in C if endowment is $w \geq \underline{\psi}_k(w)$. If we can find such $\underline{\psi}_k(w)$, then we can use a law of large numbers to argue that

$$\underline{\Psi}(k) \geq \underline{\psi}(k) \equiv \mathbb{E}_{w \in W_C}\left\{\underline{\psi}_k(w)\right\} \text{ almost surely}$$

because the realizations of w in each component C is independent of each other. Likewise, if we can find a function $\overline{\psi}_k(w)$ to bound from above the fraction of stuck nodes, and then

$$\overline{\Psi}(k) \geq \overline{\psi}(k) \equiv \mathbb{E}_{w \in W_C}\left\{\overline{\psi}_k(w)\right\} \text{ almost surely}$$

Imagine first that the signal endowment of the upper triangle in C is $(0,0,0)$. Then, using Lemma 4.E.1we know that the upper triangle of C will get stuck from period $t = 1$ on, and we can get the expected value of stuck nodes in C from there on. See that the fraction of nodes that get stuck in this component is only a function of the realization of $w \in W_C$, which is independent of the realization of the signal endowment of other components on $OR_r(k)$

When the signal endowment of the upper triangle in C is different from $(0,0,0)$, we make use of the other property we knew from C : that the only connection to the rest of the graph is through i_C,the uppermost node in C. Therefore, **a way of getting a lower bound on the number of nodes that get it wrong, is assuming that from round $t = 2$ on, node i_C knows the truth, and plays $a_{i_C,t} = 1$ for all $t \geq 2$.** Intuitively, we are making the graph to have the biggest effect possible in convincing nodes in C that actually, $\theta = 1$, which can only do by making $a_{i_C,t} = 1$ for all rounds other than $t \geq 2$. Once we have that, we can simulate the learning model on C, and calculate $\underline{\psi_k}(w)$ and $\overline{\psi}_k(w)$in this way.

There are two ways of calculating $\mathbb{E}_{w \in W_C}\left\{\underline{\psi}_k(w)\right\}$:

1. **Doing it explicitly:** This can be done for $k = 2$ and $k = 3$, because $\#\{W_C\} = 128$. The bound when $k = 3$ is the one we present in this paper.

2. **Monte-Carlo:** Of course, as k goes bigger, it is computationally unfeasible to calculate the expected value of $\underline{\psi}_k(w)$ explicitly, since

$$\#\{W_C\} = 2^{2^{k+1}-1} = O(\exp(\exp(k)))$$

which grows super-exponentially. However, we can simulate random draws of $w \in W_C$ and get an estimate for $\mathbb{E}_{w \in W_C}\left\{\underline{\psi}_k(w)\right\}$ using law of large numbers.

The above method will also work for different learning models on the CSQ_r family.

4.F Proofs

Proof of Proposition 4.2.1. The first part follows from (Golub and Jackson 2010a), since every node has degree 2 or 4, Corollary 1 applies. Namely, $\max_{1 \leq i \leq n} \frac{d_i}{\sum d_i} \to 0$ along our sequence and therefore the social learning process is wise. The second part follows from Lemma 4.E.2. □

Proof of Proposition 4.4.2. Let $f(n,T)$ be the amount of calculations that should be done for a network of size n and played for T rounds.

$$\begin{aligned} f(n,T) &= \sum_{t=1}^{T} n2^{nt+1}(1+2^{nt}) = 2n\left(\sum_{t=1}^{T} 2^{nt} + \sum_{t=1}^{T} 4^{nt}\right) \\ &= 2n\left[\frac{2^{n(T+1)}-2^n}{2^n-1} + \frac{4^{n(T+1)}-4^n}{4^n-1}\right] = \Theta\left(n4^{nT}\right) \end{aligned}$$

Meanwhile, the amount of calculations for the original model (with no trembles) is simply given by

$$g(n,t) = \sum_{t=1}^{t=T} n(2^{2n} + 2^{n+1}) = nT(4^n + 2^{n+1}) = \Theta(nT4^n)$$

Thus, the complexity ratio between the model with trembles and the model with no trembles is

$$\frac{2n\left[\frac{2^{n(T+1)}-2^n}{2^n-1} + \frac{4^{n(T+1)}-4^n}{4^n-1}\right]}{nT(4^n+2^{n+1})} = \Theta\left(\frac{1}{T}4^{n(T-1)}\right)$$

which completes the proof. □

References

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica 56*, 383–396.

Acemoglu, D., M. A. Dahleh, I. Lobel, and A. Ozdaglar (2010). Bayesian learning in social networks.

Alatas, V., A. Banerjee, A. G. Chandrasekhar, B. Olken, and R. Hanna (2011). Identifying the poor: Social networks and the aggregation of information. *MIT Working Paper*.

Ambrus, A., M. Mobius, and A. Szeidl (2010, January). Consumption risk-sharing in social networks. Harvard University working paper.

Anderson, L. and C. Holt (1997). Information cascades in the laboratory. *The American Economic Review 87*(5), 847–862.

Andrews, D. (1992). Generic uniform convergence. *Econometric Theory 8*(2), 241–257.

Andrews, D. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica: Journal of the Econometric Society 62*(1), 43–72.

Angelucci, M., G. De Giorgi, M. Rangel, and I. Rasul (2009). Family networks and school enrolment: Evidence from a randomized social experiment. *Journal of Public Economics*.

Angelucci, M. and G. DeGiorgi (2009). Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? *American Economic Review 99*, 486–508.

Aral, S. and D. Walker (Forthcoming). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science*.

Bala, V. and S. Goyal (1998). Learning from neighbours. *Review of Economic Studies 65*(3), 595–621.

Banerjee, A. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 797–817.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. Jackson (2011). The diffusion of microfinance. MIT working paper.

Barabasi, A. and R. Albert (1999). Emergence of scaling in random networks. *Science 286*(5439), 509.

Barr, A., M. Dekker, and M. Fafchamps (2008, April). Risk sharing relations and enforcement mechanisms. Centre for the Study of African Economies Series WPS/2008-14.

Barr, A. and G. Genicot (2008). Risk sharing, commitment, and information: An experimental analysis. *Journal of the European Economic Association 6*(6), 1151–1185.

Berger, R. (1981). A necessary and sufficient condition for reaching a consensus using DeGroot's method. *Journal of the American Statistical Association 76*(374), 415–418.

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

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy 100*(5).

Bloch, F., G. Genicot, and D. Ray (2008). Informal insurance in social networks. *Journal of Economic Theory 143*(1), 36–58.

Blume, L., W. Brock, S. Durlauf, and Y. Ioannides (2011). Identification of social interactions. *Handbook of Social Economics 1*, 855–966.

Bollobás, B., C. Borgs, J. Chayes, and O. Riordan (2010). Percolation on dense graph sequences. *The Annals of Probability 38*(1), 150–183.

Bramoulle, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

Burgess, R. and R. Pande (2005). Do rural banks matter? Evidence from the Indian social banking experiment. *American Economic Review 95*, 780–795.

Celen, B. and S. Kariv (2005). An experimental test of observational learning under imperfect information. *Economic Theory 26*(3), 677–699.

Chandrasekhar, A. and R. Lewis (2010). Econometrics of sampled networks. MIT working paper.

Chandrasekhar, A. G. and M. O. Jackson (2011). Tractable and consistent exponential random graph models. *Mimeo*.

Charness, G. and G. Genicot. An experimental test of risk-sharing without commitment. Economic Journal, forthcoming.

Chatterjee, S., P. Diaconis, and A. Sly (2010). Random graphs with a given degree sequence. *arXiv Preprint: 1005.1136*.

Chatterjee, S. and E. Seneta (1977). Towards consensus: Some convergence theorems on repeated averaging. *Journal of Applied Probability*, 89–97.

Chernozhukov, V., I. Fernandez-Val, and B. Melly (2009). Inference on counterfactual distributions.

Choi, S., D. Gale, and S. Kariv (2005). Behavioral aspects of learning in social networks: An experimental study. *Advances in Applied Microeconomics: A Research Annual 13*, 25–61.

Choi, S., D. Gale, and S. Kariv (2009). Social Learning in Networks: A Quantal Response Equilibrium Analysis of Experimental Data. *WP, May*.

Christakis, N., J. Fowler, G. Imbens, and K. Kalyanaraman (2010). An Empirical Model for Strategic Network Formation. *NBER Working Paper*.

Chung, F. (1997). *Spectral Graph Theory*. American Mathematical Society.

Coate, S. and M. Ravallion (1993). Reciprocity without commitment: Characterization and performance of informal insurance arrangements. *Journal of Development Economics 40*, 1–24.

Conley, T. and C. Udry (2010). Learning about a new technology: Pineapple in ghana. *The American Economic Review 100*(1), 35–69.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics 2*(2), 241–275.

De Jong, R. and T. Woutersen (2010). Dynamic time series binary choice. *Econometric Theory 1*(1), 1–30.

Deaton, A. (1991). Saving and liquidity constraints. *Econometrica 59*, 1221–1248.

DeGroot, M. (1974). Reaching a consensus. *Journal of the American Statistical Association 69*(345), 118–121.

DeMarzo, P., D. Vayanos, and J. Zwiebel (2003). Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics 118*(3), 909–968.

Diaconis, P. and D. Freedman (1984). Partial exchangeability and sufficiency. *Statistics: Applications and New Directions*, 205–236.

Dudley, R. (2002). *Real analysis and probability*, Volume 74. Cambridge Univ Pr.

Duflo, E., R. Glennerster, and M. Kremer (2007). Using randomization in development economics research: A toolkit. *Handbook of Development Economics 4*, 3895–3962.

Dupas, P. and J. Robinson (2009, January). Savings constraints and microenterprise development: Evidence from a field experiment in Kenya. NBER Working Paper 14693.

Ellison, G. and D. Fudenberg (1993). Rules of thumb for social learning. *Journal of Political Economy 101*(4), 612–643.

Ellison, G. and D. Fudenberg (1995). Word-of-mouth communication and social learning. *The Quarterly Journal of Economics*, 93–125.

Fafchamps, M. and S. Lund (2003). Risk-sharing networks in rural Phillipines. *Journal of Development Economics 71*, 261–287.

Feigenberg, B., E. M. Field, and R. Pande (2010, May). Building social capital through microfinance. NBER Working Paper No. 16018.

Fischer, G. (2010, February). Contract structure, risk sharing, and investment choice. London School of Economics working paper.

Frank, O. (1980). Sampling and inference in a population graph. *International Statistical Review/Revue Internationale de Statistique*, 33–41.

Frank, O. (1981). A survey of statistical methods for graph analysis. *Sociological methodology 12*, 110–155.

Friedkin, N. and E. Johnsen (1997). Social positions in influence networks. *Social Networks 19*(3), 209–222.

Gale, D. and S. Kariv (2003). Bayesian learning in social networks. *Games and Economic Behavior 45*(2), 329–346.

Ghani, A., C. Donnelly, and G. Garnett (1998). Sampling biases and missing data in explorations of sexual partner networks for the spread of sexually transmitted diseases. *Statistics in Medicine 17*(18), 2079–2097.

Ghatak, M. and T. Guinnane (1999). The economics of lending with joint liability: theory and practicel. *Journal of development economics 60*(1), 195–228.

Goeree, J., M. McConnell, T. Mitchell, T. Tromp, and L. Yariv (2010). The 1/d law of giving. *American Economic Journal: Microeconomics 2*(1), 183–203.

Goldsmith-Pinkham, P. and G. Imbens (2011). Determinants of social networks: Large sample theory. *Harvard Working Paper*.

Golub, B. and M. Jackson (2009). How homophily affects learning and diffusion in networks. *Working Paper*.

Golub, B. and M. Jackson (2010a). Naive Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics 2*(1), 112–149.

Golub, B. and M. Jackson (2010b). Using selection bias to explain the observed structure of internet diffusions. *Proceedings of the National Academy of Sciences 107*(24), 10833.

Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology 78*(6), 1360–1380.

Hahn, J. and G. Kuersteiner (2004). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory 1*(1), 1–40.

Hahn, J. and W. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica 72*(4), 1295–1319.

Haile, P., A. Hortaçsu, and G. Kosenok (2008). On the empirical content of quantal response equilibrium. *The American Economic Review 98*(1), 180–200.

Hall, P. and J. Horowitz (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica: Journal of the Econometric Society*, 891–916.

Handcock, M. and K. Gile (2010). Modeling social networks from sampled data. *The Annals of Applied Statistics 4*(1), 5–25.

Hayashi, F., J. Altonji, and L. Kotlikoff (1996). Risk-sharing between and within families. *Econometrica 64*, 261–294.

Hochberg, Y., A. Ljungqvist, and Y. Lu (2007). Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance 62*(1), 251–301.

Huisman, M. (2009). Imputation of missing network data: Some simple procedures. *Journal of Social Structure 10*(1).

Hung, A. and C. Plott (2001). Information cascades: Replication and an extension to majority rule and conformity-rewarding institutions. *American Economic Review 91*(5), 1508–1520.

Iyer, R. and M. Puri (forthcoming). Understanding bank runs: the importance of depositor-bank relationships and networks. *American Economic Review*.

Jackson, M. (2008a). Average distance, diameter, and clustering in social networks with homophily. *Internet and Network Economics*, 4–11.

Jackson, M. (2008b). *Social and economic networks*. Princeton University Press.

Jackson, M. (2009a). Networks and economic behavior. *Annu. Rev. Econ 1*, 489–513.

Jackson, M. (2009b). An overview of social networks and economic applications. *The Handbook of Social Economics., ed. Jess Benhabib, Alberto Bisin and Matthew O. Jackson. Elsevier Press. forthcoming.*

Jackson, M., T. Barraquer, and X. Tan (2010). Social Capital and Social Quilts: Network Patterns of Favor Exchange.

Jackson, M. and B. Rogers (2007). Relating network structure to diffusion properties through stochastic dominance. *The BE Journal of Theoretical Economics 7*(1), 1–13.

Jackson, M. and L. Yariv (2007). Diffusion of behavior and equilibrium properties in network games. *American Economic Review 97*(2), 92–98.

Jackson, M. O. (2008c). *Social and Economic Networks*. Princeton University Press.

Jackson, M. O., T. Rodriguez-Barraquer, and X. Tan (Forthcoming). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*.

Jenish, N. and I. Prucha (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics 150*(1), 86–98.

Kaboski, J. P. and R. M. Townsend (2009, July). A structural evaluation of a large-scale quasi-experimental microfinance initiative. MIT working paper.

Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl (2009). Trust and social collateral. *Quarterly Journal Of Economics 124*, 1307–1361.

Kinnan, C. and R. Townsend (2011). Kinship and financial networks, formal financial access and risk reduction. *Northwestern University and MIT Working Paper*.

Kleibergen, F. (2007). Generalizing weak instrument robust iv statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics 139*(1), 181–216.

Kocherlakota, N. R. (1996). Implications of efficient risk sharing without commitment. *Review of Economic Studies 63*, 595–609.

Kolotilin, A. (2011). Estimation of a scale-free network formation model. *MIT Working Paper*.

Kossinets, G. (2006). Effects of missing data in social networks. *Social Networks 28*(3), 247–268.

Krackhardt, D. (1987). Cognitive social structures. *Social Networks 9*(2), 109–134.

Kremer, M. and E. Miguel (2007). The Illusion of Sustainability. *The Quarterly Journal of Economics 122*(3), 1007–1065.

Kubler, D. and G. Weizsacker (2004). Limited depth of reasoning and failure of cascade formation in the laboratory. *Review of Economic Studies 71*(2), 425–441.

Lahiri, S. N. (1992). Edgeworth correction by 'moving block' bootstrap for stationary and nonstationary data. In R. Lepage and L. Billard (Eds.), *Exploring the Limits of Bootstrap*, Volume 270, pp. 183–214. Wiley-Interscience.

Leider, S., M. M. Möbius, T. Rosenblat, and Q.-A. Do (2009). Directed altruism and enforced reciprocity in social networks. *Quarterly Journal of Economics 124*(4), 1815–1851.

Liben-Nowell, D. and J. Kleinberg (2008). Tracing information flow on a global scale using internet chain-letter data. *Proceedings of the National Academy of Sciences 105*(12), 4633.

Ligon, E., J. P. Thomas, and T. Worrall (2000). Mutual insurance, individual savings, and limited commitment. *Review of Economic Dynamics 3*, 216–246.

Ligon, E., J. P. Thomas, and T. Worrall (2002). Informal insurance arrangements with limited commitment: Theory and evidence from village economies. *Review of Economic Studies 69*, 209–244.

Lin, C., Z. Lin, and C. Lu (1996). *Limit theory for mixing dependent random variables*, Volume 378. Springer.

Macours, K. (2003). Comparing a direct with an indirect approach to collecting household level data: Who tells the truth about what? *Washington, DC, United States: Johns Hopkins University. Mimeographed document.*

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*(3), 531–542.

Möbius, M. and A. Szeidl (2006). Trust and cooperation in social networks. *Harvard University and UC Berkeley Working Paper*.

Newman, M., S. Strogatz, and D. Watts (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E 64*(2), 26118.

Platteau, J.-P. (2000). *Institutions, social norms, and economic development*, Chapter Egalitarian norms and economic development, pp. 189–240. Harwood Academic.

Rothenberg, R. (1995). Commentary: sampling in social networks. *Connections 18*(1), 104–110.

Rubin, D. (1976). Inference and missing data. *Biometrika 63*(3), 581.

Santos, P. and C. Barrett (2008). What do we learn about social networks when we only sample individuals? not much. *Working Paper*.

Smith, L. and P. Sorensen (2000). Pathological outcomes of observational learning. *Econometrica 68*(2), 371–398.

Snijders, T. (1992). Estimation on the basis of snowball samples: How to weight? *Bulletin de méthodologie sociologique 36*(1), 59.

Spear, S. E. and S. Srivastava (1987). On repeated moral hazard with discounting. *Review of Economic Studies 54*, 599–617.

Stephan, F. (1945). The expected value and variance of the reciprocal and other negative powers of a positive bernoullian variate. *The Annals of Mathematical Statistics 16*(1), 50–61.

Stumpf, M., C. Wiuf, and R. May (2005). Subnets of scale-free networks are not scale-free: sampling properties of networks. *Proceedings of the National Academy of Sciences of the United States of America 102*(12), 4221.

Takasaki, Y., T. Barham Oliver, and L. Bradford (2000). Rapid rural appraisal in humid tropical forests: an asset possession-based approach and validation methods for wealth assessment among forest peasant households. *World Development 28*(11), 1961–1977.

Thompson, S. (2006). Adaptive web sampling. *Biometrics 62*(4), 1224–1234.

Townsend, R. M. (1994). Risk and insurance in village India. *Econometrica 62*, 539–591.

Townsend, R. M. (2007). Townsend thai project household annual resurvey, 1998-2006. [Computer file]. 2nd Release. Chicago, IL: Social Sciences Computing Services, the University of Chicago [Producer & distributor].

Van der Vaart, A. and J. Wellner (1996). *Weak convergence and empirical processes*. Springer Verlag.

Watts, D. and S. Strogatz (1998). Collective dynamics of small-world networks. *Nature 393*(6684), 440–442.