NBER WORKING PAPER SERIES

NETWORK STRUCTURE AND THE AGGREGATION OF INFORMATION: THEORY AND EVIDENCE FROM INDONESIA

Vivi Alatas
Abhijit Banerjee
Arun G. Chandrasekhar
Rema Hanna
Benjamin A. Olken

Working Paper 18351
http://www.nber.org/papers/w18351

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2012

We thank Emily Breza, Pascaline Dupas, Matthew Elliott, Ben Golub, Matthew O. Jackson, Laura Schechter, Adam Szeidl, Chris Udry, seminar participants at Yale, Harvard/MIT Applied Theory, University of Wisconsin- Madison AAE, LSE/UCL, CEU, Berkeley, Princeton CSDP, MIT Political Science, NBER Summer Institute 2013, the Calvó-Armengol Workshop and NEUDC 2010 for helpful comments. We thank Francisco Muñoz, Ritwik Sarkar, Prani Sastiono, Ririn Purnamasari, Hendratno Tuhiman, Matthew Wai-Poi, and Chaeruddin Kodir for outstanding research assistance and thank Mitra Samya, SurveyMeter, and the Indonesian Central Bureau of Statistics for their cooperation implementing the project. Most of all we thank Lina Marliani for her exceptional work leading the field implementation teams. Funding for this project came from a World Bank – Royal Netherlands Embassy trust fund and AusAid. Chandrasekhar is grateful for support from the National Science Foundation GRFP. The views expressed herein are those of the authors and do not necessarily reflect the views of the World Bank, the Royal Netherlands Embassy, the National Bureau of Economic Research, or any other organization.

Network Structure and the Aggregation of Information: Theory and Evidence from Indonesia
Vivi Alatas, Abhijit Banerjee, Arun G. Chandrasekhar, Rema Hanna, and Benjamin A. Olken
NBER Working Paper No. 18351
August 2012, Revised May 2014
JEL No. D83,D85,H23,O12

ABSTRACT

We use unique data from 600 Indonesian communities on what individuals know about the poverty status of others to study how network structure influences information aggregation. We develop a model of semi-Bayesian learning on networks, which we structurally estimate using within-village data. The model generates qualitative predictions about how cross-village patterns of learning relate to different network structures, which we show are borne out in the data. We apply our findings to a community-based targeting program, where villagers chose which households should receive aid, and show that networks the model predicts to be more diffusive differentially benefit from community targeting.

Vivi Alatas
World Bank
Jakarta Stock Exchange Building
Tower 2, 12th & 13th Floor
Jakarta, Indonesia
valatas@worldbank.org

Abhijit Banerjee
Department of Economics, E17-201B
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
banerjee@mit.edu

Arun G. Chandrasekhar
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305
and NBER
arungc@stanford.edu

Rema Hanna
Kennedy School of Government
Harvard University
79 JFK Street
Cambridge, MA 02138
and NBER
Rema_Hanna@hks.harvard.edu

Benjamin A. Olken
Department of Economics, E17-212
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
bolken@mit.edu

An online appendix is available at:
http://www.nber.org/data-appendix/w18351

1. Introduction

Economists are increasingly conscious of the influences of neighbors and friends. In particular, there is a growing interest in how information is aggregated within communities. Many individuals may have information that is useful or interesting to others, but does this information get transmitted to those who need it, either through direct communication between parties or through the observation of their choices? And, how does the answer to this question vary with the nature of the social network within the community? Being able to answer these types of questions is important for policy design: for example, recent evidence suggests that the speed with which new agricultural technologies are adopted depends on who talks to whom about what (e.g., Munshi (2004), Bandiera and Rasul (2006), Duflo et al. (2004), and Conley and Udry (2010)), and social connections have been shown to be important in spreading information about jobs, microfinance, and public health (e.g., Munshi (2003), Bandiera et al. (2009), Banerjee et al. (2013), Kremer and Miguel (2007)).

Similarly, the increasing trend in developing countries towards the decentralization of policy to the local level – e.g., community monitoring of teachers and health professionals or decentralized budgeting of local public goods – is predicated, in part, on the idea that communities have more information – and can more effectively aggregate it – than central governments. For example, decentralization has become increasingly popular for *targeting* the poor for government assistance programs.[1] The idea is that it is costly for the central government to use asset surveys to identify the poorest people within a village, whereas the community may know who they are, simply by virtue of living next to them. In designing these types of community-based targeting systems, it is crucial to understand how information about poverty flows within villages and how it is aggregated through intra-village processes. It is also important to understand in which types of villages the network characteristics are such that information will be aggregated well, and these decentralized mechanisms can be used effectively, and in which types of villages information will not be aggregated well and other mechanisms may be more appropriate.

However, despite several important theoretical contributions to the question of how information is aggregated within a community – many of which we discuss below – laying out a general relationship between a network's characteristics and the extent of information sharing remains challenging due to the complexity of networks. They can differ along many dimensions, and how each individual network characteristic relates to the degree of information aggregation within the network can depend both on the network structure and the underlying model of social learning.[2] To illustrate

[1]The Bangladesh Food-For-Education (Galasso and Ravallion, 2005), Albanian Economic Support safety net (Alderman and Haque, 2006), and BRAC Ultra-Poor program (Bandeira et al., 2012) are examples of community targeted programs.

[2]Due to the difficulty of describing transitional learning dynamics, much of the social learning literature has focused on asymptotic learning. The early literature on observational learning, where agents observe others' actions and attempt to learn the state of the world through these observations, showed how even Bayesian agents may inefficiently herd and ignore their own information (Banerjee, 1992; Bikhchandani et al., 1992). More recently, Acemoglu et al. (2011) show that under sequential observational learning in stochastic networks, provided that agents have expanding observations, asymptotic learning occurs. Gale and Kariv (2003), moving away from the case of irreversible actions taken in sequence, explore a special case in which a finite set of individuals in a network each simultaneously take an action in every period having observed their neighbors' actions in previous periods, which Mueller-Frank (2011) extends. Under myopic Bayesian behavior, they provide conditions under which a consensus emerges, making use of the martingale convergence theorem. Mossel et al. (2011) show in a world with binary uncertainty that with probability tending to one a sequence of growing networks which lead to consensus have consensus on the right state of the world. That

this point, consider the fact that while more connections typically facilitate better communication, having a higher average number of connections (i.e., in the language of network theory, a higher average degree), is not enough to guarantee better information aggregation. This is made clear, for instance, by Jackson and Rogers (2007), who require first order stochastic dominance of the degree distribution (which is much stronger than a higher average degree) to ensure greater diffusion of information in a meeting model where nodes meet other nodes with probability proportional to their degree. To see why, consider the possibility that there could be a group of people in the community who are all connected to each other (leading to a high average degree), but are entirely disconnected from the rest of the network, making information aggregation very inefficient relative to a network where average degree is lower but there is little clustering in any one part of the network.[3]

In this simple example, the networks differ on both degree and clustering patterns. This suggests that if we want, for example, a general prediction for the effect of degree, we might want to only compare networks that have similar clustering patterns, as well as similar patterns for other network features. However, there is no one measure of clustering that summarizes all the relevant information, just as no one measure of degree is sufficient (i.e. the variance of degree matters, as do higher moments). In particular, controlling for the average amount of clustering in the network is not sufficient (see, for instance, Jackson (2010), Watts and Strogatz (1998), among others). In the example above, one can imagine cases where the average clustering in the two networks is the same because everyone outside the one densely connected component in the first network is not connected at all. More generally, real networks can differ on so many dimensions that theoretical results do not provide clear predictions as to which networks will experience better information diffusion except in special cases.

In this paper, we take a more rough and ready approach to the problem of predicting the extent of information aggregation based on network characteristics. We exploit an unusual data set which is in many ways ideal for this purpose. In particular, we have network data from 631 villages in Indonesia that we collected as part of a study on the effectiveness of different targeting methodologies. This data has several key features. First, the very large number of independent networks in our data – which is extremely rare in this literature – makes it possible to make cross-network comparisons. Second, the data contain a natural measure of information aggregation: we know how a sample of villagers rank a set of others in their village in terms of relative economic well-being (i.e., which of the two households is richer). Finally, we have the actual rankings of these households (based either on their true per capita consumption levels or on their subjective assessment of their poverty status), so we can generate the "right" ranking. We use the accuracy

is, not only is there agreement, but that individuals agree while learning the truth. Meanwhile, another strand of literature studies various rule of thumb social learning processes. For instance Golub and Jackson (2010) and Golub and Jackson (2012) look at DeGroot learning on networks and study the speed of convergence in a model in which, in every period, individuals average the beliefs of their network neighbors and communicate their updated beliefs to their neighbors in the following period. In contrast, Jackson and Rogers (2007) study information transmission as a percolation or contagion process in a model in which the network directs the probability of individuals meeting others.

[3]See also the echo-chamber effect discussed in Golub and Jackson (2012) describing how information aggregation may be slower in more segregated networks.

of households in ranking others in their village as our measure of information diffusion and then we ask how this relates to various network characteristics.

We begin with some reduced form facts about information in our setting that will motivate the way we choose to model information aggregation below. In particular, we examine the relationship between people's network position and what they know. This is related to other results in the literature (see for example, Munshi (2004), Bandiera and Rasul (2006), Kremer and Miguel (2007), Duflo et al. (2004), and Conley and Udry (2010)). While these results are purely descriptive and do not address the important and difficult identification issues (Manski, 1993), the patterns are striking in how clear and strong they are: we show that better connected households are better at ranking other households, especially if we measure being better connected by average degree. Similarly households that are socially closer (in terms of path length) to their ranker are more likely to be more accurately ranked. We also document that non-response (i.e., saying that respondent's "don't know" whether person j or k is richer) is an important feature of our data, and is also related to network structure (i.e., a person is more likely to say he doesn't know about j and k if he is socially distant from j and k). Therefore, there is at least prima facie evidence for the importance of network channels for information transmission.

The main focus of our paper, however, is on cross-village comparisons: whether, and if so how, the features of the network as a whole predict how well information gets aggregated. To make these cross-village comparisons, we would ideally have clear theoretical predictions from a diffusion model about what network characteristics should matter. Unfortunately, as explained earlier, there is no available analytical theory rich enough for our setting. Therefore, rather than getting the predictions of what network characteristics matter (and how) from analytical theorems, we get them by using what we call numerical theorizing.

Specifically, we take the following approach: we use the *within village v*ariation in our data to estimate parameters of a model of learning on networks and use that model to simulate information diffusion in every village. We then run regressions in the simulated data to estimate the *cross-village* correlations between network characteristics and the extent of information diffusion. We then use these predicted correlations as our benchmark for examining the empirical cross-village comparisons. This is what we mean by numerical theorizing: by comparing the cross-village regression estimates on actual data to the counterparts generated from simulated data from the model, we can see whether the patterns we pick up in the data are qualitatively similar to those predicted by network theory.[4] For this exercise, we develop a simple, estimable model of learning on networks. The core idea is that individuals are trying to learn a state variable – the economic well-being of others – in the village, but this state variable is changing over time as individuals' wealth evolves. Information about individuals' wealth flows over the network. Individuals who are linked to the information source receive noisy signals each period about others' wealth, and then noisily transmit this information to their neighbors, and so on. Agents aggregate the sequence of signals that they have received to develop a guess as to what the source's current status is. In our model, they aggregate this information using a Kalman filter, treating each piece of information they receive

[4]Simulations have also been used to study other network phenomena that are too complicated to solve analytically. See, for example, Golub and Jackson (Forthcoming), who use simulations to complement an analytic study of a homophily-based link formation model.

over the network as an independent signal. We show that this corresponds exactly to optimal Bayesian learning on simple, directed networks where people receive each bit of information only once. On arbitrary networks, however, where people may receive the same information bit source through many different paths, the Kalman filter assumes the information is independent, and is orders of magnitude easier for the agent to compute than full Bayesian learning which would require the agent to undo all the double-counting.[5] Additionally, based on what we observe in the data, we also allow individuals to make a judgement about the quality of their evidence before deciding whether or not to report it in the model.

After fitting the model via simulated method of moments, we simulate the estimated model for each of the 631 individual networks in our data, generating a predicted information flow for each network. We estimate the cross-village correlations between network characteristics and the extent of information diffusion in the simulated data. Specifically, we regress the predicted information aggregation in the simulated data on a number of commonly used network statistics (size, average degree, average clustering, first eigenvalue of adjacency matrix, link density, and fraction of nodes in giant component) from each network separately as well as jointly, to generate predictions from the model for the relationship between those networks statistics and the extent of information aggregation.[6] We use these predicted correlations as the benchmark for the actual, empirical cross-village comparisons. This is what we mean by numerical theorizing: we test these theoretical predictions by asking whether the empirical cross-network correlation between the observed degree of information aggregation and network characteristics ("the empirical results") are qualitatively similar to the simulated predictions from our theory.

The empirical patterns match up reasonably well with what our theory predicts. In particular, we show that the Jackson and Rogers (2007) result on stochastic dominance of the degree distribution that we described earlier, as well as the Bollobás et al. (2010) result, saying the threshold probability for a percolation process to spread to most of the nodes is related to the first eigenvalue of the adjacency matrix, hold up both in our model and in the actual data.[7] To the best of our knowledge, this is the first "test" of these theories in empirical economics contexts. Moreover, we find that, for the most part, whenever either the predicted (simulated) or the actual (empirical) correlations are significantly different from zero, they have the same sign and for the most part this sign matches what we would have expected based on existing theoretical research.[8] For example, networks with larger first eigenvalues exhibit lower error rates, both in the predicted and actual data.

[5] This behavioral mistake is consistent with data from lab experiments (Chandrasekhar et al., 2012).

[6] The choice of these network characteristics is inspired by important analytical results in the literature on the determinants of information aggregation in networks (even though they cannot be directly applied to our context). For example, Jackson and Rogers (2007), as mentioned above, focus on the effects of first order shifts in the degree distribution, and Bollobás et al. (2010) focus on the role of the first eigenvalue of the adjacency matrix. More generally, though, there are an enormous number of ways of summarizing the properties of the adjacency matrix, so it was impossible to consider all such permutations.

[7] Bollobás et al. (2010) say that for an arbitrary sequence of dense graphs a percolation process with probability p_n is associated with a giant component of nodes being "informed" under this process and the emergence of this component has a sharp threshold at $p_n = \lambda_{1,n}^{-1}$. Thus, less contagious percolation processes are required, the higher the first eigenvalue of the graph. Their results do not directly apply as (i) our information process is more complex (though related) and (ii) our networks are sparse as opposed to dense.

[8] In some instances, the theoretical claims that we have in mind are based on intuitive discussions rather than formal proofs.

However, we also see interesting divergences from what we might have intuitively expected. For example, the effect of higher average degree on information aggregation, *controlling for other network characteristics,* is negative both in our "numerical theoretical predictions" and in the reduced form empirical results. Though there is a standard intuition that more connections are better, this is not true as a conditional correlation.

To ensure that our results are not driven by the specific parameter values that we estimate in the diffusion model (especially since the bounds on estimates are not very tight), we redo the cross village simulation and regression exercise for a wide interval of parameter values more or less centered around the estimated values. The basic predictions turn out to be remarkably robust to different parameter values, implying that the patterns that we observe may be portable.[9]

Finally, we return to one of our motivating examples and explore whether the characteristics of the network can predict which communities are better at targeting. This data-set comes from an experiment in which villages were randomly assigned to determine eligibility for an anti-poverty program using either community-based targeting, in which a village meeting ranked households from poorest to richest and assigned benefits to the poorest, or using proxy-means tests (PMT), which assign benefits based on a deterministic function of a household's assets. We find that community targeting in villages that our network model predicts should have better information passing properties better reflects people's self-assessment of their poverty status.

Our overall findings are useful for at least two reasons. First, they suggest that the standard intuitions about how networks function may not be so far from the truth, despite the absence of general analytical results behind them, at least if the way we model transmission is broadly correct. For example, networks that have higher first eigenvalues of their adjacency matrices do seem to aggregate information better, and probably for the reasons that we understand from previous theoretical work. Second, the findings highlight the role of social networks in actual community decision making, thus offering insights into policy design problems where governments aim to seek out and harness aggregate local information (e.g., to whom to provide a loan, where local infrastructure should be built) or those that rely on understanding the ways that information spreads within a network (e.g., public health campaigns, agricultural extension programs). They suggest the possibility of using standard network statistics to predict whether in a particular context we would expect effective information aggregation. Of course, this points to a need for further work to think about which network characteristics could be sufficient for these purposes and how to cost-effectively collect these network statistics. We provide some guidance on this type of future work below.

The paper is organized as follows. Section 2 describes the data. Section 3 presents reduced form evidence at the individual level and Section 4 establishes the framework and describes the predictions of the numerical model. Section 5 describes our main empirical results. Section 6 makes the connection with targeting. Section 7 concludes.

[9]See Appendix G.

2. Context and Data

2.1. **Context.** This study stems from a broader data collection effort that was designed to study the efficacy of different targeting methodologies in Indonesia. Between November 2008 and March 2009, we conducted a randomized evaluation to compare the accuracy of three common methods to identify beneficiaries for targeted social programs: proxy-means testing (PMT), wherein one collects asset and demographic information on everybody in the census and uses the data to predict consumption; a community targeting approach, wherein decisions on beneficiaries are made in a communal meeting; and a methodology that combined both community and PMT methods (Hybrid). A detailed description and findings from this study are described in Alatas et al. (2012).

In this paper, we utilize the detailed data that we collected on social networks, as well as data on individuals' reports about the relative incomes of other villagers. We first discuss the sample construction. In Section 2.3, we provide a detailed description of the survey data collected, describe the construction of the network, and then briefly discuss the design of the targeting experiment. Finally, we report key sample statistics in Section 3.1.

2.2. **Sample Description.** The initial sample consists of 640 hamlets spread across three Indonesian provinces: North Sumatra, South Sulawesi, and Central Java. The provinces were chosen to be broadly representative of Indonesia's diverse geography and ethnic makeup, with one province located on each of the three most populous islands (Sumatra, Sulawesi, and Java). Within these three provinces, we randomly selected a total of 640 villages, stratifying the sample to consist of approximately 30 percent urban and 70 percent rural locations. For each village, we obtained a list of the smallest administrative unit within it (a *dusun* in North Sumatra and a *Rukun Tetangga* (RT) in South Sulawesi and Central Java), and randomly selected one of these units (henceforth "hamlets") for the experiment. The hamlets are best thought of as neighborhoods. Each hamlet has an elected or appointed administrative head, whom we refer to as the hamlet head, and contains an average of 54 households. We make use of 631 hamlets that have network data available.

2.3. **Data.**

2.3.1. *Data Collection.* We primarily use data that was collected as part of the baseline survey for the experiment. SurveyMeter, an independent survey organization, administered the baseline survey in the field in November to December 2008, before any mention of the experiment or the social program were made to villages. For each randomly selected hamlet in the village, we constructed a census of households and then randomly selected eight households to be surveyed. In addition, we always surveyed the hamlet head to obtain the "leadership" perspective. From this survey, we used information on social networks and on both the perceived and actual income distribution within the village.

To construct the social networks (discussed in Section 2.3.2), we used two forms of social connections data. First, we used a series of data on familial relationships within each hamlet. Specifically, we asked each of the surveyed households to name all other households in the hamlet to whom they were related (either through blood or marriage).[10] We then asked the respondent to name the formal and informal leaders, the five poorest households in the hamlet, and five richest households in the hamlet, along with all of the relatives of each person named.[11] Second, we asked each respondent to name the social groups that each household member participated in within the hamlet, and prompted them with various types of groups to ensure a complete list. The social groups included, but were not limited to, neighborhood associations, religious groups, school groups, ROSCAs, farmers' associations, etc.

In this study, we are concerned with how accurately information about households' economic status diffuses within a hamlet. Thus, we needed to construct a measure of each household's beliefs, and needed to compare it to a measure of the "true" distribution within the hamlet. To collect data on knowledge, we conducted a poverty ranking exercise where we asked each household to rank the other eight households that were interviewed from their hamlet from the "most well-off" (*paling mampu*) and to the "poorest" (*paling miskin*). Note that this was done in the baseline survey, before any of the targeting treatments were implemented or even discussed in the village, so individual responses should not be affected by the subsequent targeting experiment.

We then collected two measures of the "true" distribution of economic well-being of households. First, we collected a measure of actual per capita expenditure levels at the time of the baseline survey, using the standard 28-question Indonesian SUSENAS expenditure module. Second, we asked households to self-assess their own poverty status. Specifically, each household was asked "Please imagine a six-step ladder where on the bottom (the first step) stand the poorest people and on the highest step (the sixth step) stand the richest people. On which step are you today?" Each respondent responded with a number from 1 to 6. In previous work, we show that when asked to assess the poverty status of others, Indonesian households use a concept that may more closely correspond to the self-assessed welfare metric than to objective per capita consumption (Alatas et al., 2012), which is why we include both in this study. We then construct an error rate for each household's knowledge by computing the fraction of times that the surveyed household makes an error in the (8 choose 2) comparisons in the poverty ranking exercise, where the right answer is either per capita consumption or the household self-assessment.[12] Note that we construct a village level error rate analogously.

2.3.2. *Network Data.* The networks utilized in this paper are undirected, unweighted graphs that are constructed from the familial and social group data in a way we now describe. Specifically, we first construct edges between the households that we sample and those that they identify as their family members. Second, we consider each household that was named as one of the poorest or

[10]On the forms, there was room to list up to 10 households in the village. If households were related to more than 10 households in the hamlet, the enumerator could add additional related households to the survey. On average, households reported that they were related to about 3.1 households in the hamlet.

[11]The quality of our kinship data can be seen as follows. Of the 1,658 households named as family by survey respondents who self-identify as formal elites, 1,303 of them (79%), were also identified as elite relatives by at least two other households in our survey. This cross-validation exercise suggests that measurement error in kinship is relatively small. See Alatas et al. (2012, 2013) for further details.

[12]If a respondent was unable to rank a household during the poverty ranking exercise (i.e. since he or she did not know members from the household or anything about their income level), we assigned this as an "error," i.e., they were unable to correctly rank the households.

richest, or as a leader by any household we surveyed, and then construct an edge between the named household and all of their named relatives. Moreover, we construct an edge between each pair of these relatives (i.e., if household i is named as being in the same extended family as household j, and household j is separately named (potentially by another respondent) as being in the same extended family as household k, we construct edge (i, k) in addition to (i, j) and (j, k). Third, we construct an edge between any two households who are registered as part of the same social group. Finally, we take the union of these graphs.

Two specific facts are worth mentioning. First, the data consists of a set of subgraphs of the target graphs that we are ultimately interested in. As noted in Chandrasekhar and Lewis (2012), regression analysis on partial samples of network data can show biases due to non-classical measurement error.[13] However, on average, we have complete family data on 65 percent of households in each hamlet. In addition, for a number of key quantities and specifications, for instance the first order stochastic dominance of a village's degree distribution against another's, our results are conservative as the bias will generate attenuated coefficients.[14] Second, our data is unique in terms of the sheer number of networks we have at our disposal. Typical papers have very few graphs in their sample (closer to 5 than 50). Having a sample with over 600 networks puts us in a unique position to shed light on questions about how cross-network variation in social structure is associated with the outcome of a diffusion process.

2.3.3. *Aggregation of Data in Community Based Targeting.* Whether to decentralize "targeting" – the selection of beneficiaries to social programs aimed towards the poor – to local communities has become a key policy question in recent years as household income is challenging and costly to measure. The data used in the paper was collected prior to an experiment in which we compared community targeting with nationally-imposed, data driven approaches. Specifically, in each hamlet, the Central Statistics Bureau (BPS) and Mitra Samya, an Indonesian NGO, implemented an unconditional cash transfer program, where a fixed number of households would receive a one-time, Rp. 30,000 (about \$3) cash transfer. The amount of the transfer is equal to about 10 percent of the median beneficiary's monthly per capita consumption, or a little more than one day's wage for an average laborer. Each hamlet was randomly allocated to one of three main targeting treatments: PMT, Community or Hybrid. In the PMT treatment, program beneficiaries were determined through a regression-based formula that mapped easily observable household characteristics into a single index. In the community treatment, the hamlet residents determine the list of beneficiaries through a poverty-ranking exercise at a public meeting. In the hybrid treatment, the community ranking procedure was done first, followed by a subsequent PMT verification. Additional details of these three procedures can be found in Appendix C and in Alatas et al. (2012).

[13] Most of the bias correction solutions discussed in Chandrasekhar and Lewis (2012) are not applicable as they rely on missing-at-random data. In addition, the estimates in our structural model described in Section 4 are generated by fitting a diffusion process taking place on sub-graphs of the true underlying network which then, in turn, are likely to affect the relationship between the network regressors and the simulated outcomes. We discuss below how this affects our qualitative predictions.

[14] At least in the univariate case, note that, conditional on sign-consistency, any *standardized* effect has to decrease even with non-classical measurement error. Following the Cauchy-Schwarz inequality it is easy to show that $\beta_0 \cdot \sigma_x \geq \operatorname{plim} \widehat{\beta} \sigma_{\bar{x}} = \beta_0 \frac{\operatorname{cov}(x, \bar{x})}{\sigma_{\bar{x}}}$ as $\sigma_x \sigma_{\bar{x}} \geq \operatorname{cov}(x_i, \bar{x}_i)$ where $\widehat{\beta}$ is the estimated regression coefficient, β_0 is the true value, x is the true regressor, and $\bar{x}$ is the mismeasured regressor.

Using intuitions from network theory on information aggregation, we can then test whether the network characteristics that are typically associated with a better informed population also predict where community-based targeting does better. Following Alatas et al. (2012), we create two metrics to assess the degree to which these methods correctly assign benefits to poor households. First, we compute the rank correlation between the results of the targeting experiment (the "targeting rank list") and per capita consumption. Second, we compute the rank correlation of the targeting experiment with respondents' self-assessment of poverty, as reported in the baseline survey. To assess the degree to which different network structures affect the targeting outcomes, we can examine whether the difference in these rank correlations between community / hybrid treatments (which use community information) and the PMT treatment (which does not) is greater in villages with network structures that should lead to better information transmission.

3. Sample Statistics and Information at the Household Level

In this section, we provide prima facie evidence of information diffusion through the network to help motivate the diffusion model we develop in section 4. We first provide sample statistics to help describe the knowledge environment. Next, we explore how a household's place in the network is correlated with their ability to rank others within the hamlet (section 3.1.1). Finally, we test whether households are better at ranking those who are more connected to them (section 3.1.2). Note that these are descriptive regressions, not causal estimates. This section establishes that there is substantial micro-heterogeneity that matters for information. Individuals who are more central in the network are better able to rank others, and individuals are better able rank those to whom they are more closely connected. We use these stylized facts in the next section to motivate our model.

3.1. **Sample Statistics.** Table 1 reports descriptive statistics for the primary network and outcome variables (Appendix A provides definitions of each network variable). Panel A provides the statistics for the hamlet level variables, while Panel B provides corresponding household level statistics. We report variable means in Column 1 and standard deviations in Column 2.

The sampled hamlets tend to be small, with an average of 53 households (Panel A). The number of connections per household, called a household's *degree*, averages 8.18. Villages exhibit significant *clustering,* with a mean of 0.42; this means that about 42 percent of pairs of an individual's contacts are also linked to each other. The average *path length* is about 2, which suggests that two randomly chosen households will be separated by one household in between, conditional on being in the same component. The networks have an average *fraction of nodes in the giant component* of only 0.51, which means that about half of the households are interconnected to each other through some chain of connections.[15]

[15] It is likely that the true underlying network is in fact fully connected, and the fact that this number differs greatly from 1 comes from the sampling of the graph. Note that more dense graphs will exhibit a higher fraction of nodes in the giant component under sampling. There is considerable variation in the fraction of households in the giant component, with a standard deviation of 0.24, which implies that there is significant heterogeneity in the sparsity of the underlying true graphs. As discussed in Footnote 27, despite the sampling problem the correlations of the data are still in line with those predicted from the model.

Households struggle with making the comparisons of the households' economic status. The mean average error rate at the village level based on consumption is 0.52, while the mean error rate based on the self-assessment is about 0.46. However, there is substantial heterogeneity in the error rate across villages – the standard deviation for both variables is about 0.2, which means that in the very best villages the error rate is as little as 0.1.[16] These levels need to be interpreted carefully, however, as part of what we are calling error may be due to measurement error in consumption Alatas et al. (2013).

Many households refuse to make certain comparisons, claiming that they do not know: the rate of reporting do not know is 0.19. This suggests that the appropriate model should account for this aspect of reality. Note, that even when reporting, the individual error rate is still high: 0.36 and 0.27 for consumption and self-assessment, respectively with standard deviations of 0.48 and 0.45.

Panel B provides corresponding sample statistics at the household level. Most notable is the fact that the average clustering coefficient in our networks is 0.64. This differs from the aggregated data in Panel A because we have more information about sampled individuals than we have about the rest, which is natural because everything we know about non-sampled individuals comes from reports from the sampled group.

3.1.1. *Network Position of those Ranking Others.* We begin by asking whether individuals that are more central within the network have a lower error rate in ranking other households in the hamlet. Specifically, we estimate:

$$Error_{ir} = \beta_0 + \beta_1' W_{ir} + X_{ir}'\delta + \mu_r + \epsilon_{ir} \tag{3.1}$$

where i is the household doing the ranking, r is a hamlet, $Error_{ir}$ is household i's error rate in ranking (the share of the $\binom{8}{2}$ comparisons that i correctly categorizes), W_{ir} are i's network characteristics, ϵ_{ir} is the error term, and μ_r is a hamlet fixed effect in order to estimate the effect of the household's characteristics within the network conditional on others within the network. Hamlet fixed effect sweeps out any cross-network average differences and focuses just on within-network differences in position. Since network position may be correlated with other household characteristics, we also explore whether the results are sensitive to controlling for demographic characteristics for household i, X_{ir}; these include log consumption, years of education of the respondent, and dummy variables that indicate whether the household is a formal or informal leader within the village, whether the household is from an ethnic minority, whether the household is from a religious minority, and whether the respondent is female. Table 2A reports the results with no covariates (i.e. constraining δ to be zero) and Table 2B reports them when we include a full set of covariates (X_{ir}).[17]

We consider several network characteristics: degree (Columns 1 and 5), which is the number of links to other households; the clustering coefficient (Columns 2 and 6), which is the fraction of a household's neighbors that are themselves neighbors; and the eigenvector centrality (Columns 3 and

[16]The 5th percentile for these variables are 0.254 and 0.138, respectively.

[17]The remainder of the tables present results conditional on covariates, unless otherwise noted, though we include appendix versions without covariates.

7), where eigenvector centrality is a measure of the node's importance defined, recursively, to be proportional to the sum of her neighbors' importances. Formal definitions are included in Appendix A. In Columns 4 and 8, we estimate the effect of each of these three network characteristics, conditional on one another. Columns 1-4 do not include hamlet fixed effects and Columns 5-8 add hamlet fixed effects.

In Panel C, we replicate the analysis in Panels A-B, but change our outcome variable. Instead of using the error rate, we use the share of "don't know"s reported by household i about the ranks of the other households.

Overall, households that are more connected within the network have an easier time ranking other households. Using consumption as the measure of the truth (Panel A of Table 2A), the bivariate regressions (Columns 1-3) show that households that have a higher number of links with other households in the network (degree), that have more interwoven social neighborhoods (clustering), and households that are a more important node in the network (eigenvector centrality) are less likely to make errors in ranking others. Conditional on each other, we find that a one standard deviation increase in average degree is associated with a 5.4pp drop in the household's error rate and similarly a one standard deviation increase in the clustering coefficient is associated with a 1.4pp decrease (Column 4). Including hamlet fixed effects, degree (Column 5) and eigenvector centrality (Column 7) continue to predict a household's error rate (both at the 1 percent level), but clustering is no longer significant. When all three measures are included in Column 8, while magnitudes remain similar to the bivariate cases (and smaller than the regressions without hamlet fixed effects), we no longer are able to detect a statistically significant relationship. Similarly, as Panel B illustrates, households that are more connected also have an easier time ranking other households as compared against their self-assessment. The coefficient estimates of all models are similar across Panels A and B, both in terms of sign and magnitude. In Column 8, we find that a one standard deviation increase in degree corresponds to roughly a 1.4pp decrease in the error rate (significant at the 5 percent level). It is worth noting that the inclusion of hamlet fixed effects systematically leads to a decline in the coefficient magnitude – a fact borne out in the simulations as well (see Appendix Tables E.2A and E.2B). This suggests that network-level effects may be important for information aggregation, a subject we explore in much more detail below.

The results are robust to two key changes in specification. First, including the control variables in Table 2B does not alter the findings, suggesting that the results are not driven by observable household demographic characteristics. For example, the coefficient estimates in Column 4, Panel A, imply that a one standard deviation increase in degree is associated with a 1.3pp decline in the consumption error rate; the analogous impacts from Table 2A (with no covariates X_{ir}) was 1.4pp. Again, the patterns in the data look similar using self-assessment (Panel B) as the measure of the truth rather than consumption (Panel A). Second, we also explore these relationships excluding the cases where individuals claim they do not know (Appendix Tables D.2A and D.2B).[18] The results are similar to the main specification, implying that even when individuals decide to venture a guess, they are still more likely to get it right if they are more connected within the network.

[18]Note that in the main specifications, if i has declared that she does not know the wealth of j and refuses to rank j, then in all comparisons involving j, we treat it as an error.

Finally, we study the relationship between willingness to report another's wealth and network characteristics (Panel C). Is a more central individual more likely to receive information and therefore less likely to declare that she doesn't know the answer? A one standard deviation increase in the degree of an individual is associated with a 6.5 or 1.6pp decrease in the likelihood of reporting "don't know" (without or with hamlet fixed effects, respectively) in the bivariate regressions. Recalling that the mean of "don't know" is 0.19, this indicates a large effect size. A similar result is true for the eigenvector centrality. We see that the results are robust to including demographic covariates (Table 2B, Panel C).

3.1.2. *Connections Between Ranker and Rankee.* The preceding analysis explored how one's place in the network affected the accuracy of the ranking. We now test whether the ranker is more accurate when he or she is more connected to the households that he or she is ranking. Specifically, in Table 3, we address whether a household i does a better job of ranking nodes j versus k if the pair is closer to i. To measure distance on the network, we use the shortest path length. The distance between i and j is denoted $d(i,j)$. Because of sampling, many nodes cannot be connected by any path and therefore have infinite distance between them. Thus, we simply use the average inverse distance between (i,j) and (i,k): $\frac{1}{2}\left(\frac{1}{d(i,j)}+\frac{1}{d(i,k)}\right)$ which now scales to a measure of closeness in $[0,1]$.[19] Specifically, we estimate:

$$Error_{ijkr} = \beta_0 + \beta_1' W_{ijkr} + X_{ijkr}'\delta + \mu_r + \nu_i + \epsilon_{ijkr} \tag{3.2}$$

where $Error_{ijkr} = \mathbf{1}\{i \text{ ranks } j \text{ versus } k \text{ incorrectly}\}$ (which is done for all $j<k,\ j\neq i,\ k\neq i$), W_{ijkr} is the average network characteristics of the households that are being ranked (j and k), and X_{ijkr} are the covariates. In some specifications we include hamlet fixed effects μ_r, to isolate the within-hamlet differences, and ranker fixed effects ν_i, to isolate the differences among the people a given individual is connected to and to control flexibly for all aspects of the person doing the ranking.

In Column 1, we show the basic correlations between the error rate and average inverse distance from i to j and k, conditional on demographic controls. In Column 2, we introduce additional network characteristics (average degree, average clustering coefficient and average eigenvector centrality, where the average is across the two people being ranked). In Columns 3 and 4, we include hamlet fixed effects and ranker fixed effects, respectively. All standard errors are clustered by village.

Average inverse distance between them tends to be highly predictive of the accuracy in the ranking. Using consumption as the measure of truth (Panel A), if both j and k are at distance 1 from i as compared to each being distance 3 from i, then household i is 6 to 13 percentage points less likely to rank them incorrectly, and if the average distance of the ranked pair increases by one standard deviation, then there is a resulting increase of 1 to 1.5 percentage points in the probability that household i ranks them incorrectly. These results are generally robust to including hamlet fixed effects (Columns 3-4). However, we lose considerable power with ranker fixed effects (Column 4), and the effect is no longer significant at conventional levels, although the sign and magnitudes

[19] If there is no path from node i to j, the distance is by convention infinite.

of the coefficients are generally similar to Column 3. Using self-assessment as the truth (Panel B), the average reachability and distance predict the error of the ranked pairs with physical controls and hamlet fixed effects (Column 3). Again when controlling for ranker fixed effects (Column 4), it is no longer significant.

We look at how the distance of i from nodes j and k that are being ranked influences i's propensity to declare "don't know" in Panel C. Again we find that if the average distance between the ranker and the rankees is 1 as opposed to 3, then the ranker is anywhere from 5 to 13pp less likely to declare a don't know in the assessment of one of the ranked parties. Finally, it is once again true that even among individuals who offer a view, the ranking is more likely to be correct when the ranker and rankee are more closely connected (Appendix Table D.3).

Taken together, the evidence shows that even when looking within a network and even after conditioning on a large vector of demographic observable characteristics, (i) more central nodes are more likely to offer a guess; (ii) more central nodes are less likely to make a mistake in their guess; (iii) nodes are more likely to offer a guess when they are closer to the nodes they are ranking; (iv) and nodes are more likely to make mistakes in their assessments, the farther the distance to the nodes they are ranking. These findings motivate our model.

4. Model, Estimation, and Simulation Results

4.1. **Model.** In the preceding analysis, we empirically described how an individual's network status related to their knowledge of others' relative economic well-being. In this section, we aim to build a parsimonious model that relates network characteristics to information diffusion, capturing the key features of the environment that we discussed above:

(1) Individuals who are more socially proximate to those they are ranking are more likely to correctly rank them.
(2) More central individuals in the network are more likely to correctly rank others.
(3) Individuals often report that they don't know, implying that their posteriors may be too imprecise to be worth reporting.
(4) When individuals claim that they know, they are still often wrong. In other words, being willing to speak does not necessarily mean that they know they got a perfect signal of the truth.
(5) Individuals further away from those being ranked are more likely to say that they don't know.

A natural model for capturing these attributes is one where individuals learn about the wealth of other members of their community through communication on a social network. We capture the "don't knows" by assuming that individuals make some judgment about quality of their evidence that they receive before deciding whether to report it or not.

Specifically, each individual j has a wealth, $w_{j,t}$ that evolves stochastically over time. Each period, j transmits a noisy signal about his current wealth to everyone that he is connected to. Each person i in the network also passes, with noise, some information they received about j in the previous period, to everyone that i is connected to. Person i then receives signals about j from everyone he is connected to, and i updates his beliefs accordingly, and so on. This means that the further an individual i is from j, the noisier his information about j will be, both because it will have passed through more steps en route from j to i and hence have a lower signal-to-noise ratio, and because it will be older and so does not incorporate more recent changes to j's wealth level.

The two key issues here are what part of j's information gets passed on and how different pieces of information get aggregated. To see why it may not make sense to require that all of the information is passed on, note that people typically receive information in a given period from multiple pathways, not all of which would be up-to-date. While we do not explicitly model the cost of communication, it is easy to see why someone may be reluctant to go through that whole list. Instead, we assume that people only pass on the most up-to-date information they receive. Moreover, we assume that for any given person j in the network, everyone in the network knows the distance from all their neighbors to j (as measured by the shortest path through the network) and passes on just the report that came from the person closest to j. Under the assumption that both the rule for passing and the fact that everyone knows the shortest distance to any other network member are common knowledge within the network, members can always know identify the latest information that they have and this is what they pass on. Intuitively, one can think of this as "gossip" – people are only excited to pass on the latest tidbit of information.

We also assume that people don't find it worth their while to pass on really stale information. If their information is sufficiently outdated, people skip passing it on (i.e., they would say they "don't know" anything about j if asked and otherwise say nothing).

In terms of aggregation of information, assuming that people are fully Bayesian in this context may be somewhat unrealistic. The full Bayesian aggregation rule requires people to properly weight all the various alternative pathways through which the information could have reached them, taking into account the fact that different pieces of information may have come from the same ultimate sources (i.e., passed through many of the same nodes before they diverged and followed different paths) and therefore may be subject to correlated errors. And, it is not enough to do this for just the current signals – since signals are noisy, a Bayesian would want to take account of all signals, past and present, and correctly average them. To give a sense of scale to this computation, note that enumerating all such paths is #P-complete and a random graph with n nodes and edges with probability p_n has an expected number of paths between nodes 1 and n given by $(n-2)!p_n^{n-1}e\,(1+o\,(1))$, which is potentially an enormous number (Roberts and Kroese, 2007). For example, in our data, with an average of 52 nodes and $p_n = 0.1$, there would be in expectation 82,674,076,879,277 paths between individuals i and j. Why would anyone go through such a difficult exercise in order to answer a surveyor's question?[20]

We therefore adopt the following approach. The decision-maker treats all of the signals that he receives as if they were independent (conditional on the truth) and applies Bayes' rule to them correctly (under that potentially incorrect assumption). Since the weight given to each signal then only depends on its precision, which in turn depends on the distance to the source, our previous assumptions about the knowledge of distance and the passing of only the latest information are

[20]To be clear we are not saying that Bayesian learning on network model always requires doing all these calculations. For instance, if individuals always pass on their their entire information sets, the computations would be simpler–the cost is that they would have to keep track of and communicate a much larger object.

sufficient to allow the decision-maker to compute the weights. With normal distributions for the evolution of wealth and noise, the decision-maker's aggregation rule is a Kalman filter.

This set of assumptions considerably simplifies the decision-maker's problem. Instead of keeping track of an exponential number of paths (i.e., 82 trillion paths for the typical node in our data), the average node receives just $p_n n$ – in our villages the average degree is 8 – signals in each period, each of which has a precision given by its distance from the source. It is also, arguably, more realistic; failure to properly account for the correlation between signals appears to be one of the more consistent ways in which people deviate from the fully Bayesian behavior in experimental settings (Chandrasekhar et al., 2012). Indeed this is one of the arguments used in favor of a DeGroot model, in which agents simply take an average of their neighbors' opinions, over the full Bayesian model (DeGroot, 1974; DeMarzo et al., 2003; Golub and Jackson, 2012). DeGroot learning is a simple model of weighted averaging. Individuals have a belief about the state of the world. They look at their neighbors' beliefs from the previous period, they average the opinions, and form a new opinion which is then passed into all the neighbors so that the process continues.[21] Indeed, one interpretation of our model is essentially an extension/refinement of a time-varying DeGroot model where we micro-found the weights. Typical DeGroot models embody fixed weights that, while they may vary by node, do not vary by source. In our model, by contrast, the distances from sources give the appropriate weights for an individuals' neighbors and then the Kalman filter tells us the right way to combine historical data into a posterior belief.

In the next sub-section 4.2, we outline the formal setup of the model. In sub-section 4.3, we then discuss the model's properties, including how it differs from a fully-optimizing Bayesian model.

4.2. **Model Setup.** n individuals are arranged in an unweighted graph $G = (V, E)$ consisting of a set of vertices V and edges E. If $ij \in E$, then i is linked to j, and if $ij \notin E$, then i is not linked to j. Let N_i denote the neighborhood of node i, with $j \in N_i$ meaning that $ij \in E$.

We model people's wealth as an evolving stochastic process. Specifically, every individual j has wealth that evolves according to an AR(1) process:

$$w_{j,t} = \rho w_{j,t-1} + c + \epsilon_{j,t},$$

with $\epsilon_{j,t} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$. Individuals $i \in V\backslash\{j\}$ want to guess $w_{j,t}$ when surveyed at period t, given an information set $\mathcal{F}^j_{i,t-1}$ that is informed by social learning. Every individual has a prior which is a normal distribution given by the invariant distribution: $\mathcal{N}\left(\frac{c}{1-\rho}.\frac{\sigma_\epsilon^2}{1-\rho^2}\right)$.

Individuals communicate with each others as follows. At period t:

- *Signals from the source:* j sends signal s^k_{t-1} to neighbor $k \in N_j$ where

$$s^k_{t-1} = w_{j,t-1} + u^k_{j,t-1}.$$

  N_j is the set of j's immediate neighbors (i.e. those households connected directly to j), and $u^k_{j,t-1} \sim \mathcal{N}\left(0, \sigma_u^2\right)$ is an iid noise term.

[21] In the canonical DeGroot model, each agent receives a noisy signal about the state of the world in period 0 and then in subsequent period, each individual generates beliefs by taking an average of her neighbors and her previous opinion. Our setting involves time-varying states, given by individuals' wealths, and that only individuals in the source' neighborhood receive signals before the social learning process commences.

- *Signals from arbitrary nodes:* For each pair i and j, define the closest neighbor $k^{\star}(i,j)$ as $\text{argmin}_{k\in N_i} d(k,j)$, where again $d(k,j)$ is the shortest path length between the nodes k and j in the network. That is, $k^{\star}$ is the neighbor of i closest to j. If i is sufficiently close to the source j that he deems the information that he got from her closest neighbor to be of high enough quality, i.e. if $d(k(i,j),j) \leq \tau$, then each period, $i \in V\setminus\{j\}$ sends signal $s^{i,l}_{t-d(j,i)}$ to $l \in N_i$ where:
$$s^{i,l}_{t-d(j,i)} = s^{k^{\star},i}_{t-1-d(k^{\star}(i,j),i)} + u^{i,l}_{t-d(j,i)},$$
with $u^{i,l}_{t-d(j,i)} \sim \mathcal{N}(0,\sigma^2_u)$. In other words, if he is close enough to the source, every period, i passes on to each of his neighbors the most "up-to-date" piece of information that i received about j. If i has two closest neighbors $k^{\star}, k'$ that are equally close to j, we assume he passes on the average of the two signals. If j is too far from i, i.e. if $d(k(i,j),j) > \tau$, no information is passed.
- *Forming a posterior:* $i \in V\setminus\{j\}$ forms a posterior about $w_{j,t}$ by using a Kalman filter on her historical data. This is a vector
$$\mathbf{s}^{i,t} = \left(s^{i,t}_1, ..., s^{i,t}_{t-d(j,i)}\right).$$
A Kalman filter uses the entire history of (noisy) signals to help predict the future distribution of the state variable of interest. Essentially, each signal provides information about the current value $w_{j,t}$ since the entire observed history is measured with noise. Because agent i possibly receives signals from her neighbors at varying distances from the source, the information she has about j's past wealths at various periods can vary over time. We discuss this in greater detail below and in Appendix B.

The model, though it may seem daunting, is actually very quite simple. Each individual has some signals about how wealthy j was in each period in the past. When i receives some incremental information about j's wealth in any period, she updates it using a standard Bayesian updating rule treating signals as independent, but weighting the information optimally according to precision, which depends only on distance from the source and then combining them to make an optimal prediction about j's wealth today.

Figure 1 illustrates the model using simulations. We consider a network of 20 nodes arranged in a directed line, where all nodes are attempting to track node 1's wealth. Panel (A) shows the predictions of 1's wealths by other nodes, over time. Nodes that are closer to the source are better-able to estimate the current period wealth. Panel (B) depicts the posterior variance for several nodes. In Panel (C) we show the correlation of a node's estimate of 1's wealth with the true value, by distance to the source. Panel (D) shows that for the chosen parameters, only 4 nodes speak and, as node 5's posterior variance is above the threshold, nodes 5-20 do not speak in the learning process.

4.3. **Discussion of properties and assumptions.** We adopt the independence assumption and Kalman filter because it exactly replicates full Bayesian learning under the assumption that the different signals that each decision-maker receives are statistically independent, conditional on the truth, yet is dramatically computationally simpler on more general networks. The following result makes the equivalence with Bayesian learning clear:

Proposition. *For any directed graph where the source is the root and every node is connected to the source only through independent paths, every agent is fully Bayesian under our above model.*

Proof. It is clear that for any node with $d(1,j) > \tau$, since node j receives no signals, the node retains her prior, which is the correct Bayesian computation. For the remainder of the proof, consider $d(1,j) \leq \tau$.

First consider the case of a directed tree with the source, node 1, being the root. Let j be a node with $d(1,j) = k \leq \tau$. Note there is exactly one path (a line from 1 to j). Then, j receives a k period lagged signal about $w_{1,t}$ that has been disturbed by the equivalent of noise distributed $\mathcal{N}(0, k\sigma_u^2)$. Thus, the problem can be recast as an agent j making a prediction about state $w_{1,t}$ given signals $s_0^j, ..., s_{t-k}^j$. In such a linear system with normal disturbances, the Bayesian belief about a state given a history is given by the Kalman filter (Kalman, 1960; Masreliez and Martin, 1977). Note that this is exactly the computation which is done in our model.

For a case where j has L independent paths from node 1, with $k_l = d_l(j,1) \leq \tau$ for $l \in \{1, ..., L\}$, the computation is as follows. Let $k = \min_l \{k_l\}$. In period t, an individual has $s_0^{j,t}, ..., s_{t-k}^{j,t}$, where $s_\tau^{j,t}$ are computed using the period τ signal along each independent path. Again this generates a sequence of Kalman filters, indexed by, t. That is, the Bayesian prediction of $w_{1,t}$ given the signal sequence $s_0^{j,t}, ..., s_{t-k}^{j,t}$ is given by a Kalman filter and prediction. By definition, this is exactly the computation that our agents do in the model. $\square$

The set of networks covered by this includes direct lines, more generally directed trees, as well as other configurations. For instance, see the networks in Figures 2a, 2b and 2c. In particular Figure 2c depicts a graph with arbitrarily long but independent paths that lead from the source to other nodes; even in this case, our model is equal to the full Bayesian model.

To highlight where our model deviates from the full Bayesian case, consider Figure 2d. We see that a signal from A passes through B and whatever transmission error takes place there is therefore propagated through all n subsequent paths before arriving at C. Under our model, C processes the information as if she is in the graph depicted in Figure 2b. This comes from the (incorrect) assumed independence of the paths where she only accounts for the vintage of the information.

In sum, the case for our simplifications from the full Bayesian model is that it (i) requires very limited knowledge of the network structure, (ii) requires limited amount of communication, (iii) allows for confidence and self-censoring, and (iv) coincides exactly with the Bayesian model for a class of network structures. Additionally, the deviations from Bayesian learning in this model are familiar in the social learning literature: agents do not properly account for double-counting, just as in DeGroot classes of models.

4.4. **Empirical implementation.** Here we briefly outline the estimation procedure (further details are provided in Appendix B). We use data from the Indonesian Family Life Survey to estimate ρ, the AR(1) coefficient on wealth, uniformly across the hamlets. From our survey data we can estimate c. We also estimate σ_ϵ^2 from our survey data, namely since we observe $\mathrm{var}[w]$ and then know that by assumption on the model $\mathrm{var}[w] = \frac{\sigma_\epsilon^2}{1-\rho^2}$.

Given these parameters, we use the simulated method of moments to estimate the key model parameters: σ_u^2 (the noise term for passing information) and τ (the threshold distance beyond which people say "don't know" rather than speak). We use the following within-village moments:

(1) The correlation of whether i ranks j vs k correctly with $\frac{1}{d(i,j)+d(i,k)}$.
(2) The correlation of the eigenvector centrality of i with how many don't knows i reports.

Our implementation imposes the additional assumption that the rule that people use to decide whether to pass on a signal is the same as the rule they use to decide to report to us.[22]

The parameter values from the estimation are shown in Table 4. For ease of interpretation, we present a normalization of the first parameter, $\alpha := \frac{\sigma_u^2}{\sigma_\epsilon^2}$. We estimate it as $\widehat{\alpha} = 0.2$. This means that the transmission error is a fifth of the size of the structural error. However the standard errors are such that even 0.4 would be a reasonable estimate of the transmission error to structural error ratio. We also find that $\widehat{\tau} = 4$. This means that a node on the same component as a source will tend to have heard some information about the source, since there are likely to be paths of distance less than 4 to the source (the average path length is 2.02). However, the standard errors are such that anywhere from 3 to 6 would be reasonable parameter estimates.

Given the estimated parameters, we generate simulations from the model as described in Appendix B. We generate a 50 samples of draws of the wealth-learning process and then ask whether our motivating observations – that more central individuals know more and that individuals know less about others the further they are – are borne out in the simulations. Specifically, we rerun the same regressions as in Tables 2A, 2B, and 3 using the simulated data from the model; the results are provided in Panel C of Appendix Tables E.2A, E.2B and E.3.of each respective table.[23] By and large, the results confirm our intuition. Households that have a higher degree are associated with lower error rates, households that have higher clustering are associated with lower error rates, and households that are more eigenvector central are associated with lower error rates (Panel C of Tables E.2A/B). We find that being in the same component as those who an individual is ranking reduces the error rate while being several steps farther away increases the error rate (Panel C of Table E.3).

4.5. **Simulation Results at the Network Level: Numerical Propositions.** A key question we wish to ask of the model is how network-level characteristics affect information diffusion across the network. We start from the important analytical result in Jackson and Rogers (2007) showing that if network I's degree distribution and neighbor degree distribution first-order stochastic

[22]This assumption makes sense in the environment of our model since we would expect that them to be at least as willing to speak when we ask them as they are when they are actually volunteering information. Moreover, the decision to pass on information depends on their latest signal's quality; the decision to answer our question should depend on the quality of their overall information, which is higher. On the other hand, someone (i) who is further away from the source (j) than $k^\star(i,j)+1$ gets no signals and has nothing to pass on. Therefore, the only choice is whether to set the cutoff for reporting to the survey at $k^\star(i,j)$ or at $k^\star(i,j)+1$. We set it at $k^\star(i,j)$ on the grounds that it does not make a huge difference and its perhaps somewhat less confusing.

[23]It could be the case that projecting a complex diffusion process into a specific linear regression specification may itself generate unintuitive coefficient estimates. However, as our method compares the signs of those generated by simulations from the model and the real data, if the model is a good description of the information transmission process, the unintuitive projections should be similar across both the simulations and the real data. Thus, it is the case that comparing two regressions – one with a simulated outcome variable and another with an empirical one – turns out to be a reasonable test of whether the real-life process is similar to the model process.

dominates network J's degree distribution and neighbor degree distribution, respectively, then in steady state of a mean-field approximation to the matching process described above, network I should have a higher equilibrium information rate than network J.[24]

This result, however, unfortunately cannot be directly applied to our context for at least two reasons. First, their model uses a mean-field approximation to a matching process, which itself tries to approximate a contagion process, to gain analytic tractability. However, we are precisely interested in the cases where the mean-field approximation may not be apt, i.e., where we do not believe that everyone's local neighborhood essentially contains the same average information as the global average. The approximation does not work well when, for instance, nodes vary systematically in the proportion of neighbors who have information, which is likely to be true in our case (this is presumably why the network position matters for accuracy of the ranking). Second, to rank two households, each node needs to have two pieces of information, whereas there is only one thing to learn in Jackson and Rogers (2007).

We therefore use the numerical simulations of our model to examine whether we should expect the equivalent result to hold in our context. The simulations are described in detail in Appendix B. As discussed in Appendix B, we generate $\overline{Error}_{ijkr}^{SIM}$ – the average error rate from our simulations of i ranking j versus k in hamlet r – via the aforementioned simulation process. By averaging over pairs j,k, we construct individual level simulated error rates $Error_{ir}^{SIM}$, and then we construct hamlet level error rates ($Error_r^{SIM}$ for hamlet r) by averaging over the individual level error rates.

Our main outcome variable of interest is whether $Error_I^{SIM} > Error_J^{SIM}$ for hamlets I and J. We regress this variable on whether I stochastically dominates J or vice versa:

$$\mathbf{1}\left\{Error_I^{SIM} > Error_J^{SIM}\right\} = \beta_0 + \beta_1 \mathbf{1}\left\{I \succ J\right\} + \beta_2 \mathbf{1}\left\{J \succ I\right\} + X_{IJ}'\delta + \epsilon_{IJ}. \tag{4.1}$$

We include fixed effects for geographically clustered groups of hamlets, and specify two-way clustered standard errors, for hamlet I and hamlet J.[25] The results, which are reported in Table 5, suggest that the Jackson and Rogers (2007) pattern holds in our context. Because stochastic dominance is a partial ordering, the omitted category in Columns 1 and 3 is the non-comparable groups of villages. In Columns 2 and 4 we focus only on comparable village pairings, in which case we only include a dummy $\mathbf{1}\left\{I \succ_{FOSD} J\right\}$. We find that if I dominates J (instead of vice versa), there is a 24pp decrease in the probability that I has a larger error rate than J – a large effect relative to a mean of 0.5 (by construction).

We can also apply the same methodology to examine the predictions of the model regarding the role of other fundamental network characteristics. Again, analytical propositions about the impact of these networks on information diffusion have been hard to derive in general, which is why we simulate them in our structure. We choose six standard measures used in various related, but otherwise different, models – network size, average degree, average clustering, first eigenvalue of adjacency matrix, link density, and fraction of nodes in giant component – and simulate how they

[24]The neighbor degree distribution is the empirical cdf of the number of links a neighbor has, taken over all neighbors as we count over all nodes. Stochastic dominance was determined at the decile level. If the distribution function for the degree of hamlet I was weakly lower than J at all deciles (and was strict for at least one), then we say that I dominates J.

[25]Specifically, we include fixed effects for the stratification group from the Alatas et al. (2012) experiment. This a subdistrict (*kecamatan*), or in some cases when we had few subdistricts, a group of 2 or 3 subdistricts.

affect diffusion within our estimated model, described in detail in Appendix B. As discussed above, we generate $\overline{Error}_{ijkr}$ via the aforementioned simulation process and we then construct hamlet level error rates by averaging over the individual level error rates $Error_{ir}^{SIM}$.

Given these simulation-based hamlet level error rates, we estimate:

$$Error_r^{SIM} = \beta_0 + W_r'\beta_1 + X_r'\delta + \epsilon_r \tag{4.2}$$

where $Error_r^{SIM}$ is the average error rate in hamlet r from the simulations and W_r is a vector of graph level statistics including average degree, average clustering, the number of households in the hamlet, first eigenvalue, link density, and fraction of nodes in giant component. Together with the set of hamlet-level covariates X_r, we include many potentially correlated network variables in the specification of the regression model. It is not ex ante obvious that the conditional correlations of network features with the outcome variables will behave the same as the unconditional correlations, and so this is also where our numerical simulations can guide us.

As shown in Table 6, when the network characteristics W are included one by one, most of the network statistics of interest have significant effects on the error rate and they all go in the "intuitive" direction: there are lower error rates in villages where the average degree is higher, clustering is higher, the first eigenvalue of adjacency matrix is bigger, the link density is higher, and there are more households that are in the giant component.[26] The inclusion of hamlet level covariates make no difference (see Appendix F, Table F.6). When we jointly estimate the relationship of all of these network variables with the error rate, we observe some counterintuitive patterns (Column 7). In particular, while most of the effects remain significant, average degree and average clustering now have the "wrong" sign. This could mean either that the actual partial correlation of these two variables with the error rate in the types of networks we have in our data is actually positive in our model once we condition on the other network statistics; or it is the case that even with more than 600 hamlets, we do not have enough independent variation to properly estimate these effects separately when included together in the same regression (the first eigenvalue is has a correlation of 0.88 with average degree in our data).[27] In other words, it is either just very hard to separately identify the effects of these very interconnected variables, or some of our intuition is off when we are considering the variables conditional on others.

5. Cross hamlet comparisons

We are now ready to explore how network-level characteristics affect diffusion across the network in the actual data, and to compare how the actual diffusion patterns across networks compare to what the model predicts. We begin by exploring empirically whetherJackson and Rogers (2007)'s result on stochastic dominance extends to our environment and then more generally examine other claims about the role of other fundamental network characteristics.

[26] We do not interpret the effect of hamlet size because it is difficult to do so in our framework.

[27] A natural worry is that average degree, number of households, and link density (which amounts to average degree over number of households) may be generating too much collinearity. However, conditional on the other covariates in column 7, omitting link density makes no difference to the "wrong" sign that degree takes on in the the regression. It appears, instead, that conditioning on the first eigenvalue and clustering leaves average degree to not matter in an obvious way. A table documenting this is available upon request.

5.1. **Stochastic Dominance Results.** We begin by testing the Jackson and Rogers (2007) prediction about first-order stochastic dominance of the degree distribution in our data. To our knowledge, the claim about stochastic dominance has not been empirically tested before due to data limitations. In order to do so, a large sample of independent networks alongside data on information diffusion is necessary. The data collected for this study provides a plausibly large enough sample (631 hamlets) to do so. Note also that in addition to being interesting in its own right, focusing on stochastic dominance has a major advantage in our context. Working with a sampled graph, rather than the full network, may result in biases that could lead us to end up with estimates of the effects of network characteristics that are biased to the point of having the wrong sign. An advantage of working with FOSD is that while there may be attenuation bias in our estimates, we would not expect a sign reversal (Chandrasekhar and Lewis, 2012).[28] As such, our results would provide a lower bound of the predictive capabilities of the network.

We estimate a regression of whether the error rate of the hamlet I exceeds the error rate of hamlet J ($\mathbf{1}\{Error_I > Error_J\}$) on dummy variables that indicate whether hamlet I stochastically dominates hamlet J ($\mathbf{1}\{I \succ J\}$) and vice versa ($\mathbf{1}\{J \succ I\}$):

$$\mathbf{1}\{Error_I > Error_J\} = \beta_0 + \beta_1 \cdot \mathbf{1}\{I \succ J\} + \beta_2 \cdot \mathbf{1}\{J \succ I\} + X'_{IJ}\delta + \epsilon_{IJ}. \tag{5.1}$$

The omitted category is when hamlet I's and hamlet J's degree distribution are not comparable. We can also estimate regressions where we drop hamlets that are not comparable:

$$\mathbf{1}\{Error_I > Error_J\} = \beta_0 + \beta_1 \cdot \mathbf{1}\{I \succ J\} + X'_{IJ}\delta + \epsilon_{IJ}. \tag{5.2}$$

These are the same empirical specifications we examined in the simulated data in Table 5.

Panels A and B of Table 7 presents the results of these regressions on the actual empirical data on error rates from our survey. Column 1 presents the results from estimating equation (5.1), while Column 2 presents the results from estimating equation (5.2). For both models, as above, we include stratification group fixed effects, estimate with OLS, and specify two-way clustered standard errors, for hamlet I and hamlet J. Once again, we compute error rates with consumption as the measure of truth (Panel A) and with self-assessment as the measure of truth (Panel B).

The results validate the implications of the model that are provided in Table 5 and discussed above: if a hamlet's degree distribution first-order stochastic dominates another hamlet's distribution, it will have lower error rates in ranking the income distribution of the hamlet (for both measures of truth). Specifically, as Panel B, Column 2 shows, if hamlet I dominates J, then I has on average a 17pp lower error rate than J (significant at the 1 percent level). Columns 3 and 4 repeat the exercises of Columns 1 and 2, respectively, adding in demographic controls to which the results are robust and the coefficients remain stable.

5.2. **General Cross-Hamlet Results.** We now present the general hamlet level regression. Our theoretical benchmarks are the numerical simulations described above and presented in Table 6. We thus present equivalent reduced form analysis in Panel A and B of Table 8. In Columns 1 to

[28] Sign-switching would be possible only when over half of the categorizations of I dominating J become flipped due to sampling, which is very unlikely to happen.

6, we present the univariate regressions, while in Column 7 we present the multivariate regression. Once again, we consider error rates based on consumption (Panel A) and the self-assessment metric (Panel B).[29]

The results look very similar whether we use the consumption or the self-assessment metric. The univariate regressions tend to have the expected sign and they match up quite closely with our numerical predictions: whenever both the simulated and actual coefficients are significant (which is most of the time), they always have the same sign. For instance, an increase in the average degree of the hamlet is associated with a lower error rate (Column 1), an increase in the average clustering coefficient is associated with a lower error rate (Column 2), and an increase in the number of households is not associated with the error rate (Column 3, Panel A) or associated with a higher error rate (Column 3, Panel B). In addition, as seen in Column 4, a higher first eigenvalue of the adjacency matrix is associated with a considerable reduction of the error rate (a one standard deviation increase is associated with a 4pp drop in error rate). Column 5 shows that a higher fraction of nodes being in the giant component is associated with an extremely lowered error rate. As expected, Column 6 shows that a higher density of links corresponds to a lower error rate.

Including all network variables in the regression model (Column 7), we once again find a good match between the *actual* and *simulated* results. Strikingly, higher average degree appears to be a positive and significant predictor of error rate (higher degree means more errors) across both our reduced form and simulated results in this column.

The first eigenvalue of the adjacency matrix and the fraction of nodes in the giant component both come negative (and significant) in Panels A and B, exactly as our simulations would have had us expect and confirming intuitions that come from Bollobás et al. (2010), among others. The one exception is clustering which comes in with the "right" sign in the data, but was positive in the simulations.[30]

6. Application: Targeting

In this section, we test whether network characteristics predict the quality of real-world decisions that rely on communal information. We examine the targeting experiment discussed in Section 2.3.3. In particular, we test whether community-based targeting is relatively more effective than proxy-means testing (PMT) at identifying the poor in networks that we expect to be better at diffusing information about poverty. If communities efficiently aggregate information, we would expect that this would be the case, since community-based targeting utilizes local information and the findings thus far have shown that better networked communities hold more accurate information. However, if there are distortions in the processes through which information is aggregated, this may not necessarily be the case.

We estimate regressions of the form:

$$y_r = \alpha + \beta_C \mathbf{1}\{r \in C\} \cdot \rho_r + \beta_H \mathbf{1}\{r \in H\} \cdot \rho_r + \tau_c \mathbf{1}\{r \in C\} + \tau_h \mathbf{1}\{r \in H\} + \gamma \rho_r + \epsilon_r, \tag{6.1}$$

[29]See Appendix F, Table F.8 for comparison without covariates.

[30]A natural worry is that this may be due to sampled network data. The true process takes place on an unobserved network which we have sampled and fit a process on that takes the sampled network data as given data. Simulation results, available upon request, show that by generating the data under the model, sampling the network data, and then running analogous regressions, we are unable to overturn this feature.

where y_r is the rank correlation between the poverty assessments generated by the program and the benchmark of true poverty (either based on per capita consumption or based on the self-assessment), $\mathbf{1}\{r \in C\}$ and $\mathbf{1}\{r \in H\}$ are dummies for the experimental assignment of hamlet r to either the community or the hybrid treatment (the omitted category is PMT), and ρ_r is a measure (discussed below) of how diffusive a network is. We are mostly interested in β_C, and to a lesser extent, β_H. Given that higher ρ_r indicates that a network is better at spreading information, we expect that $\beta_C > 0$ – in other words, we expect community-based targeting to perform better relative to a proxy-means test when networks are more diffusive.

We take two approaches to computing ρ_r. In Table 9, we first compute ρ_r by using principal-components to aggregate the six measures of network diffusiveness from Table 6: average degree, clustering, first eigenvalue, number of households, link density, and fraction of nodes in the giant component. We then take the first principal component vector corresponding to the data matrix of these six network attributes and define $\rho_r = \sum_{k=1}^{6} v_k W_{k,r}$, where v_k are the entries of the principal component vector and $\{W_{k,r}\}$ are the six network features for hamlet r. For ease of interpretation, we normalize the regressor by percentile in the sample.

Network diffusiveness as measured in this way appears to predict whether communities are more effective than a PMT at classifying individuals based on their self-assessed poverty, but not based on consumption (Table 9). Panel A shows that β_C and β_H are not distinguishable from zero when we take y_r to be the rank correlation using consumption data, i.e. we do not observe that community targeting is more accurate in more diffusive communities relative to the PMT (Columns 2-5 of Panel A of Table 9). However, when we take y_r to be the rank correlation using self-assessment data, we find positive and significant estimates of β_C and β_H (Columns 2-5 of Panel B). Conditional on community targeting, going from the 25th to 75th percentile in diffusiveness corresponds to a 0.115 increase in the rank correlation of the targeting outcome with the self-assessment benchmark (which has a mean of 0.4) relative to the PMT (Column 4, Panel B). Not surprisingly, when we pool the treatments, the relationship persists. The fact that ρ_r only matters for the effectiveness of community targeting when assessed using self-assessment is consistent with the experimental findings in Alatas et al. (2012). That paper also showed that in general, community meetings increased the rank correlation with self-assessment, but not with per capita consumption relative to the traditional approach of using a PMT for targeting. The results here show that the impact of the community treatments on improving the correlation of targeting outcomes with self-assessed poverty status is considerably stronger in villages with more diffusive network characteristics.

A second approach is to use the model and simulations from Section 4 to compute ρ_r. Specifically, we use the one minus the simulated error rate for a hamlet, $1 - Error_r^{SIM}$, as a measure of its diffusiveness since, by definition, networks that are better at spreading information should exhibit lower error rates. Table 10 then replicates the exercises in Table 9, but now uses the the percentiles of $1 - Error_r^{SIM}$ as a measure of diffusiveness of the network. Again for ease of interpretation we normalize ρ_r by percentile in the sample. We find that community targeting differentially works better when a hamlet has lower error rates when measured using self-assessment. Going from the 25th to the 75th percentile of ρ_r, conditional on community targeting, corresponds to a 0.11 increase in the rank correlation of the targeting outcome with the self-assessment benchmark, relative to the PMT (Column 4, Panel B).[31]

Taken together, the findings show that the network structure and our learning model model not only accurately predict how information spreads, but are also useful in understanding how real decisions are made using that information. This clearly points to a need for further work to think about which network characteristics are the most useful for these purposes and how to cost-effectively obtain relevant network data (since the data-collection process may be expensive). There are several options available to researchers and policymakers. First, they can ask a simple question of prediction: is it the case that given a vector of observables from a standard data source (e.g., a census), policymakers can predict which networks are organized in a manner that encourages diffusion? These are likely to be the communities where community-based targeting would work as opposed to using a proxy-means test. This approach would work particularly well in an environment where policymakers get multiple rounds of data from the same distribution. Second, they could pursue an avenue along the lines of work by Banerjee et al. (2014) – making use of the fact that individuals in the network may have knowledge about the features of the network structure. Banerjee et al. (2014) show that if asked to name the person who would be best to initially inform in order to spread information, individuals name a small set of villagers who turn out to be eigenvector central in the network. Along these lines, one could imagine other simple questions that could be added to a standard survey with the goal of extracting knowledge of network organization from network members themselves. Finally, one could explore whether relevant network data – membership in social groups and/or kinship information – can be obtained directly from a village or sub-village head. This would also be considerably cheaper than surveying many members of the community and could be of great policy value.

7. Conclusion

We estimate a simple model of how information about poverty status is transmitted within the network, and then use the estimated model to predict the relationship between a village's network characteristics and how information on poverty status is aggregated within the village. We then compare our predictions with empirical evidence from a unique data-set of 631 villages, where we have both detailed social network data and measures of how accurately households can describe the poverty status of other households. The empirical results match up nicely with the model predictions: the characteristics that predict better information aggregation in the model also do so in the data. For example, we provide evidence supporting the Jackson and Rogers (2007) claim that if a network's degree distribution first-order stochastic dominates another's distribution, it will have overall lower error rates in ranking the income distribution of the hamlet.

[31] We note that in Panel B of both Tables 9 and 10, a more diffusive network is correlated with worse targeting under PMT when measured by the correlation with self-assessment. In fact, we can show that the covariance between consumption based wealth ranking and self-assessment based wealth ranking decreases as we look at more diffusive hamlets. Therefore, it seems that high ρ_r hamlets makes the self-assessment based notions of poverty harder to detect by conventional means. However, it seems that the community does know more about who is poor by this criterion; as the community also puts weight on this criterion, the community pulls the outcome closer to the self-assessment metric.

We then show that the network characteristics can help predict where policies that rely on information diffusion are likely to be effective: for example, we show that community-based targeting appears more effective than a more traditional, data-driven approach in areas where networks are more diffusive. The results are encouraging because they suggest the possibility of using standard network statistics to predict whether in a particular context we would expect effective information aggregation, or conversely, whether some outside intervention will be needed to supplement information flows through the network. Moreover the results give us some confidence that we are not very far off in using very simple social learning models to study communications in networks.

References

Acemoglu, D., M. Dahleh, I. Lobel, and A. Ozdaglar (2011): "Bayesian learning in social networks," *The Review of Economic Studies*, 78, 1201–1236.

Alatas, V., A. Banerjee, R. Hanna, B. Olken, R. Purnamasari, and M. Wai-Poi (2013): "Self-Targeting: Evidence from a Field Experiment in Indonesia," Tech. rep., MIT Working Paper.

Alatas, V., A. Banerjee, R. Hanna, B. Olken, and J. Tobias (2012): "Targeting the Poor: Evidence from a Field Experiment in Indon," *American Economic Review*, 102.

Alderman, H. and T. Haque (2006): "Countercyclical safety nets for the poor and vulnerable," *Food Policy*, 31, 372–383.

Bandeira, O., R. Burgess, S. Gulesci, I. Rasul, and M. Sulaiman (2012): "Can Entrepreneurship Programs Transform the Lives of the Poor?"," Tech. rep., LSE.

Bandiera, O., I. Barankay, and I. Rasul (2009): "Social connections and incentives in the workplace: Evidence from personnel data," *Econometrica*, 77, 1047–1094.

Bandiera, O. and I. Rasul (2006): "Social networks and technology adoption in northern mozambique*," *The Economic Journal*, 116, 869–902.

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 (2014): "Gossip: Identifying central individuals in a social network," .

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013): "The diffusion of microfinance," *Science*, 341.

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.

Bollobás, B., C. Borgs, J. Chayes, and O. Riordan (2010): "Percolation on dense graph sequences," *The Annals of Probability*, 38, 150–183.

Chandrasekhar, A., H. Larreguy, and J. Xandri (2012): "Testing models of social learning on networks: Evidence from a framed field experiment," .

Chandrasekhar, A. and R. Lewis (2012): "Econometrics of sampled networks," MIT working paper.

Conley, T. and C. Udry (2010): "Learning about a new technology: Pineapple in Ghana," *The American Economic Review*, 100, 35–69.

DeGroot, M. (1974): "Reaching a consensus," *Journal of the American Statistical Association*, 69, 118–121.

DeMarzo, P., D. Vayanos, and J. Zwiebel (2003): "Persuasion Bias, Social Influence, and Unidimensional Opinions*," *Quarterly journal of economics*, 118, 909–968.

Duflo, E., M. Kremer, and J. Robinson (2004): "Understanding technology adoption: Fertilizer in Western Kenya, preliminary results from field experiments," *Unpublished manuscript, Massachusetts Institute of Technology.*

Galasso, E. and M. Ravallion (2005): "Decentralized targeting of an antipoverty program," *Journal of Public Economics*, 89, 705–727.

Gale, D. and S. Kariv (2003): "Bayesian learning in social networks," *Games and Economic Behavior*, 45, 329–346.

Golub, B. and M. Jackson (2010): "Naive Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2, 112–149.

——— (2012): "How homophily affects learning and diffusion in networks," *Quarterly Journal of Economics*.

——— (Forthcoming): "Does Homophily Predict Consensus Times? Testing a Model of Network Structure via a Dynamic Process," *Review of Network Economics*.

Jackson, M. and B. Rogers (2007): "Relating network structure to diffusion properties through stochastic dominance," *The BE Journal of Theoretical Economics*, 7, 1–13.

Jackson, M. O. (2010): *Social and Economic Network*, Princeton.

Kalman, R. E. (1960): "A new approach to linear filtering and prediction problems," *Journal of basic Engineering*, 82, 35–45.

Kremer, M. and E. Miguel (2007): "The Illusion of Sustainability*," *The Quarterly Journal of Economics*, 122, 1007–1065.

Manski, C. (1993): "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 60, 531–542.

Masreliez, C. and R. Martin (1977): "Robust Bayesian estimation for the linear model and robustifying the Kalman filter," *Automatic Control, IEEE Transactions on*, 22, 361–371.

Mossel, E., A. Sly, and O. Tamuz (2011): "From agreement to asymptotic learning," *Arxiv preprint arXiv:1105.4765*.

Mueller-Frank, M. (2011): "A general framework for rational learning in social networks," .

Munshi, K. (2003): "Networks in the Modern Economy: Mexican Migrants in the US Labor Market*," *Quarterly Journal of Economics*, 118, 549–599.

——— (2004): "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution," *Journal of Development Economics*, 73, 185–213.

Roberts, B. and D. P. Kroese (2007): "Estimating the Number of st Paths in a Graph." *J. Graph Algorithms Appl.*, 11, 195–214.

Watts, D. and S. Strogatz (1998): "Collective dynamics of small-world networks," *Nature*, 393, 440–442.

Figures

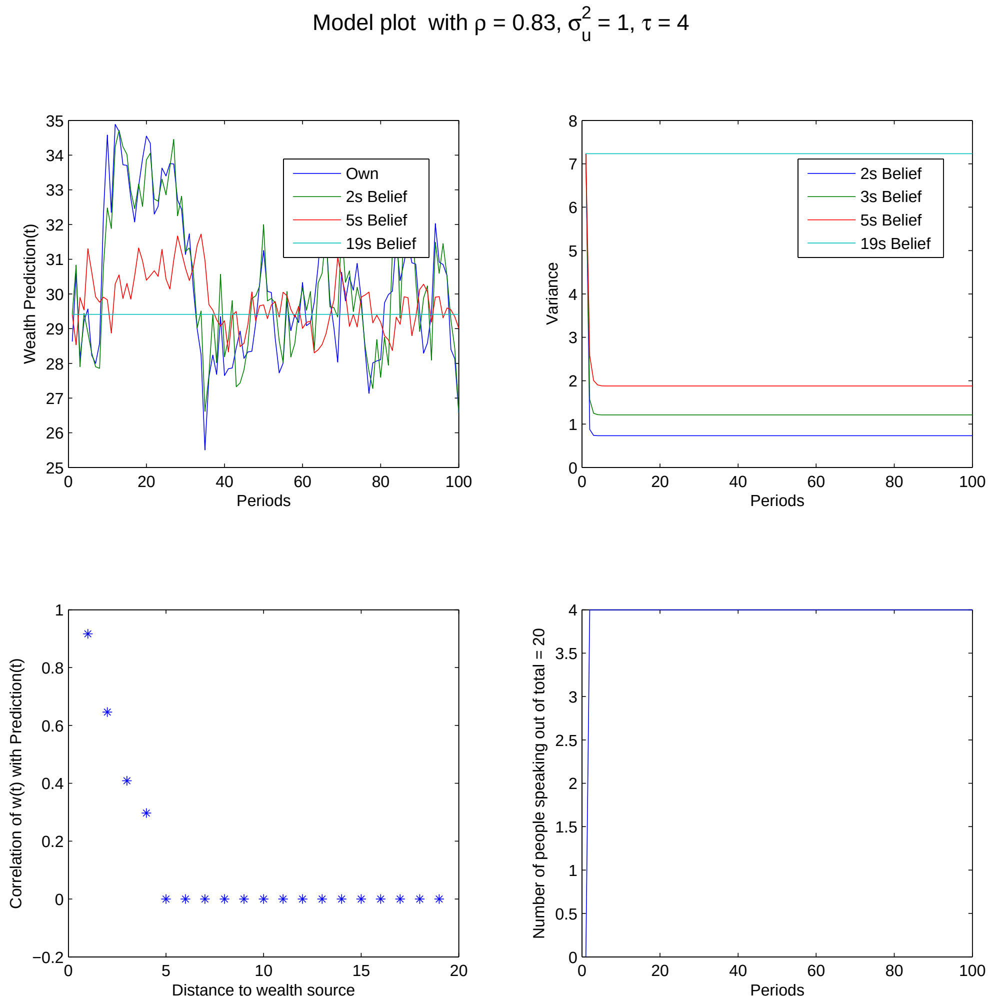
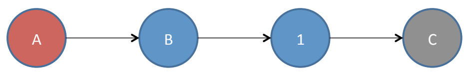
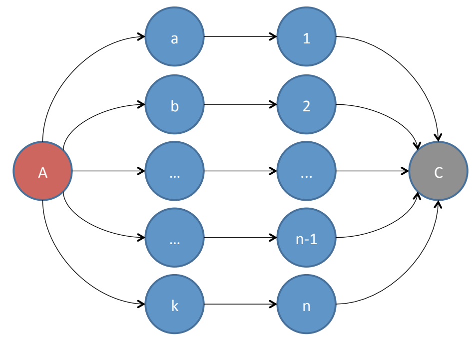


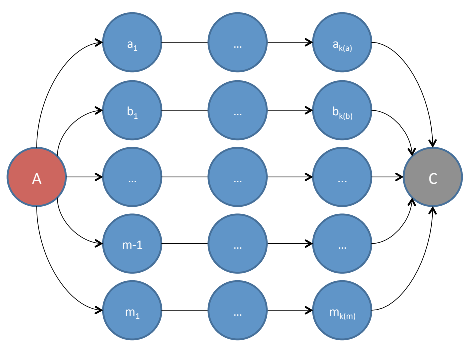
Figure 1. Simulations for a directed line with $n = 20$ nodes where individuals are learning about node 1's wealth and parameters are $\rho = 0.83$, $\sigma_u^2 = 1$, and $\tau = 4$. (A) shows the predictions $\hat{w}_{t,i}^j$ (posterior mean) by agents i. (B) depicts the posterior variance. (C) shows the correlation of $\hat{w}_{t,i}^j$ with w_t^j by distance $d(j,i)$. All individuals beyond the cutoff distance to not speak have zero correlation mechanically. (D) shows the number of individuals speaking per period.



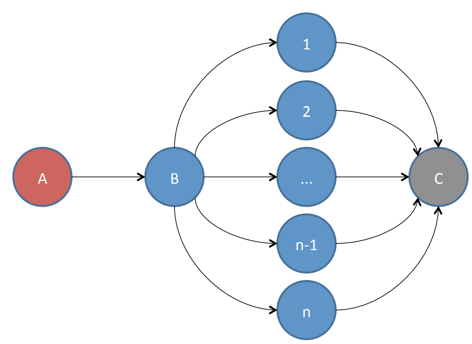
(A) A directed line.



(B) Numerous transmission errors subsequently distorted through independent paths.



(C) Independent paths of arbitrary length.



(D) Single transmission error subsequently distorted through numerous paths.

Figure 2. Various network configurations.

Tables

Table 1: Descriptive Statistics

	Mean	Standard Deviation
	(1)	(2)
Panel A: Village level		
Number of households	53.04	27.31
Average degree	8.18	3.81
Variance of degree distribution	16.34	13.62
Average clustering coefficient	0.42	0.18
Fraction of nodes in giant component	0.51	0.24
Average path length	2.02	0.50
First eigenvalue	8.57	3.13
Inequality	1.02	0.39
Link Density	0.10	0.11
Error rate (consumption)	0.52	0.19
Error rate (self-assessment)	0.46	0.22
Share don't knows	0.19	0.22
Error rate given report (consumption)	0.36	0.48
Error rate given report (self-assessment)	0.27	0.45
Panel B: Household level		
Degree	8.35	4.91
Clustering coefficient	0.64	0.30
Eigenvector centrality	0.23	0.14
Error rate (consumption)	0.52	0.23
Error rate (self-assessment)	0.45	0.26

Notes: Panel A provides sample statistics on the network characteristics of the 631 villages in the sample. It also provides information on the average level of competency in the village in assessing the poverty level of other members of the village. Panel B provides equivalent sample statistics for the 5,633 households in the sample.

Table 2A: The Correlation between Household Network Characteristics and the Error Rate in Ranking Income Status of Households

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Consumption Metric, Error Rate				
Degree	-0.0104***			-0.0109***	-0.00281***			-0.00182
	(0.00110)			(0.00139)	(0.000712)			(0.00116)
Clustering		-0.0590***		-0.0474***		-0.0118		-0.00859
		(0.0142)		(0.0134)		(0.00889)		(0.00986)
Eigenvector Centrality			-0.164***	0.0619			-0.0847***	-0.0409
			(0.0372)	(0.0484)			(0.0232)	(0.0380)
R-squared	0.049	0.006	0.009	0.052	0.667	0.666	0.667	0.668
				Panel B: Self-Assessment Metric, Error Rate				
Degree	-0.0130***			-0.0141***	-0.00386***			-0.00276**
	(0.00124)			(0.00160)	(0.000712)			(0.00122)
Clustering		-0.0568***		-0.0459***		-0.00283		0.000172
		(0.0158)		(0.0145)		(0.00999)		(0.0108)
Eigenvector Centrality			-0.174***	0.107**			-0.103***	-0.0439
			(0.0415)	(0.0545)			(0.0247)	(0.0408)
R-squared	0.061	0.004	0.008	0.065	0.674	0.672	0.673	0.674
				Panel C: Share of Don't Knows				
Degree	-0.0132***			-0.0146***	-0.00333***			-0.00169
	(0.00122)			(0.00156)	(0.000677)			(0.00115)
Clustering		-0.0378**		-0.0365**		0.00269		0.00577
		(0.0172)		(0.0159)		(0.0109)		(0.0119)
Eigenvector Centrality			-0.156***	0.118**			-0.102***	-0.0662
			(0.0422)	(0.0542)			(0.0256)	(0.0411)
R-squared	0.068	0.002	0.007	0.072	0.722	0.720	0.722	0.722
Hamlet Fixed Effect	No	No	No	No	Yes	Yes	Yes	Yes

Notes: This table provides estimates of the correlation between a household's network characteristics and its ability to accurately rank the poverty status of other members of the village. The sample comprises 5,633 households. The mean of the dependent variable in Panel A (a household's error rate in ranking others in the village based on consumption) is 0.50, while the mean of the dependent variable in Panel B (a household's error rate in ranking others in the village based on a household's own self-assessment of poverty status) is 0.46. The mean of the dependent variable in Panel C (what fraction of others does a household report "don't know" about) is 0.19. Standard errors are clustered by village and are listed in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2B: The Correlation between Household Network Characteristics and the Error Rate in Ranking Income Status of Households, Controlling for Household Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				Panel A: Consumption Metric, Error Rate				
Degree	-0.00854***			-0.00875***	-0.00215***			-0.00127
	(0.00105)			(0.00133)	(0.000697)			(0.00112)
Clustering		-0.0553***		-0.0449***		-0.0116		-0.00860
		(0.0135)		(0.0131)		(0.00877)		(0.00974)
Eigenvector Centrality			-0.145***	0.0382			-0.0684***	-0.0363
			(0.0358)	(0.0472)			(0.0230)	(0.0372)
R-squared	0.070	0.045	0.047	0.073	0.671	0.670	0.671	0.671
				Panel B: Self-Assessment Metric, Error Rate				
Degree	-0.0102***			-0.0109***	-0.00302***			-0.00209*
	(0.00119)			(0.00154)	(0.000697)			(0.00118)
Clustering		-0.0517***		-0.0423***		-0.00279		-5.26e-05
		(0.0146)		(0.0141)		(0.00972)		(0.0106)
Eigenvector Centrality			-0.148***	0.0709			-0.0819***	-0.0376
			(0.0394)	(0.0529)			(0.0243)	(0.0398)
R-squared	0.098	0.066	0.069	0.100	0.679	0.677	0.678	0.679
				Panel C: Share of Don't Knows				
Degree	-0.0108***			-0.0117***	-0.00267***			-0.00110
	(0.00114)			(0.00148)	(0.000670)			(0.00113)
Clustering		-0.0378**		-0.0343**		0.00182		0.00544
		(0.0163)		(0.0156)		(0.0107)		(0.0117)
Eigenvector Centrality			-0.144***	0.0768			-0.0854***	-0.0634
			(0.0394)	(0.0520)			(0.0254)	(0.0403)
R-squared	0.049	0.006	0.009	0.052	0.726	0.724	0.726	0.726
Hamlet Fixed Effect	No	No	No	No	Yes	Yes	Yes	Yes

Notes: This table provides estimates of the correlation between a household's network characteristics and its ability to accurately rank the poverty status of other members of the village, controlling for the household's characteristics including leadership status, consumption, education, minority status, religion, respondent gender. The sample comprises 5,630 households for Panels A and B, and 5,406 for Panel C. The mean of the dependent variable in Panel A (a household's error rate in ranking others in the village based on consumption) is 0.50, while the mean of the dependent variable in Panel B (a household's error rate in ranking others in the village based on a household's own self-assessment of poverty status) is 0.46. The mean of the dependent variable in Panel C (what fraction of others does a household report "don't know" about) is 0.19. Standard errors are clustered by village and are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: The Correlation Between Inaccuracy in Ranking a Pair of Households in a Village and the Average Distance to Rankees

	(1)	(2)	(3)	(4)
	Panel A: Consumption Metric, Error Rate			
Inverse of the Distance	-0.0594***	-0.0401***	-0.0226***	-0.0128
	(0.00836)	(0.00823)	(0.00569)	(0.0125)
Average Degree		-0.00508***	0.00267	0.00266
		(0.00174)	(0.00315)	(0.00321)
Average Clustering Coefficient		-0.00643	0.0282	0.0286
		(0.0252)	(0.0272)	(0.0275)
Average Eigenvector Centrality		0.0668	-0.0735	-0.104
		(0.0668)	(0.0912)	(0.0945)
R-squared	0.007	0.012	0.136	0.202
	Panel B: Self-Assessment Metric, Error Rate			
Inverse of the Distance	-0.0680***	-0.0403***	-0.0224***	-0.00351
	(0.00938)	(0.00904)	(0.00607)	(0.0134)
Average Degree		-0.00620***	0.000102	-0.000553
		(0.00193)	(0.00336)	(0.00345)
Average Clustering Coefficient		-0.0404	0.00400	0.00403
		(0.0270)	(0.0301)	(0.0301)
Average Eigenvector Centrality		0.126*	0.0532	0.0130
		(0.0746)	(0.103)	(0.106)
R-squared	0.033	0.035	0.183	0.264
	Panel C: Share of Don't Knows			
Inverse of the Distance	-0.0771***	-0.0451***	-0.0294***	-0.00434
	(0.00946)	(0.00985)	(0.00718)	(0.0130)
Average Degree		-0.00968***	-0.00258	-0.00295
		(0.00210)	(0.00306)	(0.00306)
Average Clustering Coefficient		-0.0393	-0.0167	-0.0196
		(0.0302)	(0.0286)	(0.0284)
R-squared	0.011	0.012	0.136	0.202
Physical Controls	No	Yes	Yes	Yes
Hamlet FE	No	No	Yes	Yes
Ranker FE	No	No	No	Yes

Notes: This table provides an estimate of the correlation between the accuracy in ranking a pair of households in a village and the characteristics of the households that are being ranked. In Panel A, the dependent variable is a dummy variable for whether person i ranks person j versus person k incorrectly based on using consumption as the metric of truth (the sample mean is 0.497). In Panel B, the self-assessment variable is the metric of truth (the sample mean is 0.464). The sample is comprised of 106,305 ranked pairs in Panel A and 105,292 in Panel B. In Panel C, the dependent variable is a dummy variable for whether person i does not know person j or person k. Physical covariates are as in Table 2B, averaged for nodes j and k. Standard errors are clustered by village and are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Structural Parameters

α	0.2
	(0.1598)
τ	4
	(1.0693)

Notes: Standard errors computed using 1000 simulations of Bayesian bootstrap, as described in Appendix B. The bootstrap weights every network by a mean-normalized exponential random variable, which is equivalent to drawing 631 hamlets with replacement when computing the objective function.

Table 5: Numerical Predictions on Stochastic Dominance

	(1)	(2)	(3)	(4)
I fosd J	-0.121***	-0.245***	-0.120***	-0.236***
	(0.0163)	(0.0245)	(0.0166)	(0.0250)
J fosd I	0.122***		0.125***	
	(0.0170)		(0.0172)	
Observations	199,396	147,460	199,396	147,460
Non-Comparable	Yes	No	Yes	No
Physical Controls	No	No	Yes	Yes
Stratification Group FE	Yes	Yes	Yes	Yes

Notes: In these regressions, the outcome variable is a dummy for whether the error rate of village *I* exceeds the error rate of village *J*. When included, physical controls are differences between the standard controls for villages *I* and *J*. The controls include consumption, education, PMT score, agricultural share, education of household head and RT head, rural/urban and inequality. Results for error rates using simulated data, as described in Appendix B. Standard errors in parentheses, two-way clustered at I and J. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Numerical Predictions on Correlation between Village Network Characteristics and Village-Level Error Rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Average Degree	-0.0301***						0.0765***
	(0.00272)						(0.00789)
Average Clustering		-0.360***					0.383***
		(0.0431)					(0.0738)
Number of Households			0.000899***				0.000293
			(0.000251)				(0.000288)
$\lambda_1(A)$				-0.0286***			-0.0523***
				(0.00236)			(0.00461)
Fraction of Nodes in Giant Component					-0.442***		-0.785***
					(0.0268)		(0.0503)
Link Density						-0.505***	-0.500***
						(0.0745)	(0.0882)
R-squared	0.198	0.135	0.030	0.267	0.297	0.118	0.512

Notes: This table provides village network characteristics and the error rate in ranking others in the village. Columns 1-6 show the univariate regressions, while column 7 provides the multvariate regressions. Physical covariates include consumption, education, PMT score, agricultural share, education of household head and RT head, urban dummy, stratification group FE, and inequality. The sample comprises 631 villages. Results for error rates using simulated data, as described in Appendix B. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Empirical Results on Stochastic Dominance

	(1)	(2)	(3)	(4)
		Panel A: Consumption Metric		
I fosd J	-0.0935***	-0.136***	-0.0875***	-0.119***
	(0.0193)	(0.0298)	(0.0191)	(0.0281)
J fosd I	0.0465**		0.0474***	
	(0.0184)		(0.0178)	
Observations	200,028	148,090	200,028	148,090
		Panel B: Self-Assessment Metric		
I fosd J	-0.100***	-0.170***	-0.0756***	-0.123***
	(0.0177)	(0.0264)	(0.0180)	(0.0260)
J fosd I	0.0730***		0.0587***	
	(0.0168)		(0.0167)	
Observations	200,028	148,090	200,028	148,090
Non-Comparable	Yes	No	Yes	No
Physical Controls	No	No	Yes	Yes
Stratification Group FE	Yes	Yes	Yes	Yes

Notes: In these regressions, the outcome variable is a dummy for whether the error rate of village *I* exceeds the error rate of village *J*. When included, physical controls are differences between the standard controls for villages *I* and *J* as in Table 5. Panel A presents results for error rates using the consumption metric. Panel B presents results for error rates using the self-assessment metric. Standard errors in parentheses, two-way clustered at I and J. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Empirical Results on Correlation between Village Network Characteristics and Village-Level Error Rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: Consumption Metric				
Average Degree	-0.0139***						0.0347***
	(0.00293)						(0.0115)
Average Clustering		-0.284***					-0.301***
		(0.0456)					(0.0967)
Number of Households			0.000854***				0.000665*
			(0.000294)				(0.000391)
$\lambda_1(A)$				-0.0118***			-0.0218***
				(0.00222)			(0.00583)
Fraction of Nodes in Giant Component					-0.208***		-0.153**
					(0.0342)		(0.0704)
Link Density						-0.209***	0.0858
						(0.0769)	(0.138)
R-squared	0.122	0.144	0.108	0.125	0.143	0.106	0.176
			Panel B: Self-Assessment Metric				
Average Degree	-0.0177***						0.0280**
	(0.00310)						(0.0124)
Average Clustering		-0.366***					-0.385***
		(0.0500)					(0.107)
Number of Households			0.00119***				0.000688*
			(0.000325)				(0.000415)
$\lambda_1(A)$				-0.0124***			-0.0167**
				(0.00253)			(0.00652)
Fraction of Nodes in Giant Component					-0.257***		-0.139*
					(0.0370)		(0.0766)
Link Density						-0.309***	0.123
						(0.0793)	(0.149)
R-squared	0.182	0.212	0.167	0.173	0.205	0.166	0.231

Notes: This table provides village network characteristics and the error rate in ranking others in the village. Columns 1-6 show the univariate regressions, while column 7 provides the multvariate regressions. Physical covariates include consumption, education, PMT score, agricultural share, education of household head and RT head, urban dummy, stratification group FE, and inequality. The sample comprises 631 villages. Panel A presents results for error rates using the consumption metric. Panel B presents results for error rates using the self-assessment metric. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Rank Correlation on Targeting Type Interacted with Diffusiveness (Principal Component)

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Rank Correlation (Consumption)			
Community x Diffusiveness		0.110	0.0831	0.0747	0.100	
		(0.112)	(0.117)	(0.119)	(0.102)	
Hybrid x Diffusiveness		0.0126	-0.0252	-0.0355		
		(0.114)	(0.116)	(0.118)		
Community	-0.0588*	-0.109*	-0.0967	-0.0965	-0.0817	
	(0.0319)	(0.0617)	(0.0647)	(0.0661)	(0.0596)	
Hybrid	-0.0614*	-0.0649	-0.0464	-0.0351		
	(0.0327)	(0.0592)	(0.0609)	(0.0627)		
Diffusiveness		-0.0993	-0.0787	-0.0763	-0.105	-0.0758
		(0.0793)	(0.0953)	(0.0958)	(0.0769)	(0.0955)
Community or Hybrid						-0.0642
						(0.0543)
Community or Hybrid x Diffusiveness						0.0163
						(0.103)
R-squared	0.014	0.018	0.088	0.096	0.092	0.094
			Panel B: Rank Correlation (Self-Assessment)			
Community x Diffusiveness		0.187	0.224*	0.230*	0.114	
		(0.116)	(0.123)	(0.124)	(0.106)	
Hybrid x Diffusiveness		0.213*	0.232*	0.199		
		(0.118)	(0.122)	(0.122)		
Community	0.108***	0.0261	0.00938	0.00671	0.0135	
	(0.0321)	(0.0650)	(0.0678)	(0.0691)	(0.0605)	
Hybrid	0.0839**	-0.0128	-0.0230	-0.00365		
	(0.0331)	(0.0640)	(0.0666)	(0.0679)		
Diffusiveness		-0.248***	-0.284***	-0.281***	-0.162**	-0.279***
		(0.0833)	(0.102)	(0.105)	(0.0809)	(0.104)
Community or Hybrid						0.00193
						(0.0588)
Community or Hybrid x Diffusiveness						0.213**
						(0.107)
R-squared	0.033	0.049	0.115	0.132	0.115	0.131
Keca FE	No	No	No	Yes	Yes	Yes
Demographic Covariates	No	No	No	Yes	Yes	Yes

Notes: The outcome variable is the rank correlation. Panel A presents rank correlation using the consumption metric. Panel B presents rank correlation using the self-assessment metric. Diffusiveness is the percentile of the predicted value based on the first principal component vector of the covariance matrix of the network characteristics described in Table 4. Physical covariates include consumption, education, PMT score, agricultural share, education of household head and RT head, urban dummy, stratification group FE, and inequality. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 10: Rank Correlation on Targeting Type Interacted with Diffusiveness (1-Simulated Error Rate)

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Rank Correlation (Consumption)			
Community x Diffusiveness		-0.0831	-0.0842	-0.0984	-0.0976	
		(0.117)	(0.117)	(0.121)	(0.124)	
Hybrid x Diffusiveness		-0.0618	-0.0643	-0.0854		
		(0.113)	(0.113)	(0.122)		
Community	-0.0588*	-0.0207	-0.0167	-0.0142	-0.0107	
	(0.0319)	(0.0633)	(0.0632)	(0.0656)	(0.0661)	
Hybrid	-0.0614*	-0.0330	-0.0288	-0.0131		
	(0.0327)	(0.0657)	(0.0661)	(0.0739)		
Diffusiveness		-0.0364	-0.0108	0.0398	0.0553	0.0393
		(0.0756)	(0.0784)	(0.0948)	(0.107)	(0.0945)
Community or Hybrid						-0.0144
						(0.0583)
Community or Hybrid x Diffusiveness						-0.0899
						(0.102)
R-squared	0.014	0.014	0.017	0.095	0.151	0.094
			Panel B: Rank Correlation (Self-Assessment)			
Community x Diffusiveness		0.249**	0.247**	0.225*	0.209*	
		(0.112)	(0.112)	(0.118)	(0.120)	
Hybrid x Diffusiveness		0.246**	0.243**	0.227*		
		(0.110)	(0.112)	(0.117)		
Community	0.108***	-0.0170	-0.00894	0.00149	0.00686	
	(0.0321)	(0.0675)	(0.0669)	(0.0706)	(0.0720)	
Hybrid	0.0839**	-0.0450	-0.0372	-0.0290		
	(0.0331)	(0.0678)	(0.0681)	(0.0736)		
Diffusiveness		-0.205***	-0.151*	-0.147	-0.145	-0.144
		(0.0790)	(0.0819)	(0.101)	(0.112)	(0.101)
Community or Hybrid						-0.0109
						(0.0624)
Community or Hybrid x Diffusiveness						0.220**
						(0.102)
R-squared	0.033	0.029	0.043	0.127	0.161	0.125
Keca FE	No	No	No	Yes	Yes	Yes
Demographic Covariates	No	No	No	Yes	Yes	Yes

Notes: The outcome variable is the rank correlation. Panel A presents rank correlation using the consumption metric. Panel B presents rank correlation using the self-assessment metric. Diffusiveness is the percentile of 1 - simulated error rate, as described in Appendix B. Simulated error rate is the expected predicted value of the error rate in a hamlet under the estimated parameters of the diffusion model. Physical covariates include consumption, education, PMT score, agricultural share, education of household head and RT head, urban dummy, stratification group FE, and inequality. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix A. Network definitions

In this section, we provide basic definitions and interpretations for the different network characteristics that we consider. At the household level, we study:

- Degree: the number of links that a household has. This is a measure of how well connected a node is in the graph.
- Clustering coefficient: the fraction of a household's neighbors that are themselves neighbors. This is a measure of how interwoven a household's neighborhood is.
- Eigenvector centrality: recursively defined notion of importance. A household's importance is defined to be proportional to the sum of its neighbors' importances. It corresponds to the i^{th} entry of the eigenvector corresponding to the maximal eigenvalue of the adjacency matrix. This is a measure of how important a node is, in the sense of information flow. We take the eigenvector normalized with $\|\cdot\|_2 = 1$.
- Reachability and distance: we say two households i and j are reachable if there exists a path through the network which connects them. The distance is the length of the shortest such path.

At the hamlet level, we consider:

- Average degree: the mean number of links that a household has in the hamlet. A network with higher average degree has more edges on which to transmit information.
- Average clustering: the mean clustering coefficient of households in the hamlet. This measures how interwoven the network is.
- Average path length: the mean length of the shortest path between any two households in the hamlet. Shorter average path length means information has to travel less (on average) to get from household i to household j.
- First eigenvalue: the maximum eigenvalue of the adjacency matrix. This is a measure of how diffusive the network is. A higher first eigenvalue tends to mean that information is generally more transmitted.
- Fraction of nodes in the giant component: the share of nodes in the graph that are in the largest connected component. Typically, realistic graphs have a giant component with almost all nodes in it. Thus, the measure should be approaching one. For a network that is sampled, this number can be significantly lower. In particular, networks which were tenuously or sparsely connected, to begin with, may "shatter" under sampling and therefore the giant component may no longer be giant after sampling. In turn, it becomes a useful measure of how interwoven the underlying network is.
- Link density: the average share of connections (out of potential connections) that a household has. This measure looks at the rate of edge formation in a graph.

Appendix B. Kalman Filter, Estimation and Simulation

This section develops the formal algorithm for the model and discusses estimation.

B.1. **Model.**

Setup. Without loss of generality, fix node j about whose wealth the remainder of the nodes are learning. Wealth follows an AR(1) process given by

$$w_{j,t} = c + \rho w_{j,t-1} + \epsilon_{j,t}.$$

Individuals $i \in V \backslash \{j\}$ want to guess $w_{j,t}$ when surveyed at period t, given an information set $\mathcal{F}^{j}_{i,t-1}$ that is informed from social learning. Individuals communicate with each others as follows. At period t:

- In every period t, every neighbor of j, $i \in N_j$ receives an iid signal

  $$s^{i,t}_{t-1} = w_{j,t-1} + u^{i,t}_{j,t-1}.$$

  That is a mean zero normally disturbed signal of j's previous period wealth with

  $$u^{i,t}_{j,t-1} \sim \mathcal{N}\left(0, \sigma^2_u\right).$$

- In every period t, a generic node $i \neq j$ with $d(i,j) \leq \tau$ in the graph transmits the newest piece of information it receives to each of its neighbors l. Let $k^{\star} := \mathrm{argmin}_{k \in N_i} d(k,j)$ be i's closest neighbor to j. Then i passes on this newest piece of information – an estimate of $w_{j,t-1-d(j,k^{\star})}$ – to $l \in N_i$:

  $$s^{i,l}_{t-d(j,i)} = s^{k^{\star},i}_{t-1-d(j,k^{\star})} + u^{i,l}_{t-d(j,i)}.$$

  Here again $u^{i,l}_{t-d(j,i)} \sim \mathcal{N}\left(0, \sigma^2_u\right)$. In other words, if he is close enough to the source, every period, i passes on to each of his neighbors the most "up-to-date" piece of information that i received about j. If i has two closest neighbors k^{*}, k' that are equally close to j, we assume he passes on the average of the two signals. If j is too far from i, i.e. if $d(k(i,j), j) > \tau$, no information is passed.

The above protocol defines a signal generation process. Thus, in every period t, a generic node i in the graph has received a vector of signals

$$\mathbf{s}^{i,t} := \left(s^{i,t}_{1}, ..., s^{i,t}_{t-d(i,j)}\right).$$

The signal vector $\mathbf{s}^{i,t}$ is double-indexed since it can have time-varying elements.

- The signal vector is treated as a collection of independent draws (conditional on the wealth sequence) with

  $$s^{i,t}_{r} \sim \mathcal{N}\left(w_r, \sigma^2_{r,t,i}\right)$$

  where is i's t'th period set of signals about time period r, where $r \leq t - d(i,j)$. Moreover, i's period t variance about the signal for period r is given by

  $$\sigma^2_{r,t,i} = \frac{1}{\sum_{k \in N_i} \frac{1}{d(k,j)\sigma^2_u}}.$$

- Given $\mathbf{s}^{i,t}$, node i applies the Kalman filter to obtain the posterior mean and variance.

Kalman Filter. The Kalman filter is as follows. In what follows, we reserve τ to index time (and describe the process only for nodes that are speaking). At period t a node i makes the following computation. She treats the system as the tth period of a linear Gauss-Markov dynamical system with

- state equation is given by
$$w_{j,\tau} = c + \rho w_{j,\tau-1} + \epsilon_{j,\tau},\ \ \tau = 1, ..., t+1.$$
- measurement equation given by
$$s_{\tau}^{i,t} = w_{j,\tau} + v_{\tau}^{i,t},$$
where $v_{\tau}^{i,t} \sim \mathcal{N}\left(0, \sigma_{\tau,t,i}^2\right)$.

Then the computation of the Kalman filter is entirely standard given the vector $\mathbf{s}^{i,t}$ of measurements and knowledge of parameters $c, \rho, \sigma_\epsilon^2, \sigma_u^2$ and $d(k,j)\ \ \forall k \in N_i$. The crucial equations are how to do a time update given prior information and how to incorporate the new measurements to correct the system:

- Time update equations:
$$\begin{aligned} \hat{w}_\tau^- &= \rho \hat{w}_{\tau-1} + c \\ P_\tau^- &= \rho^2 P_{\tau-1} + \sigma_\epsilon^2. \end{aligned}$$
- Measurement update equations:
$$\begin{aligned} K_\tau &= \frac{P_\tau^-}{P_\tau^- + \sigma_{\tau,t,i}^2}. \\ \hat{w}_\tau &= \hat{w}_\tau^- + K_\tau\left(s_\tau^{i,t} - \hat{w}_\tau^-\right). \\ P_\tau &= \left(1 - K_\tau\right) P_\tau^-. \end{aligned}$$

The initialization is at the mean of the invariant distribution $w_0 = \frac{c}{1-\rho}$ and the variance $P_0 = \frac{\sigma_\epsilon^2}{1-\rho^2}$.

B.2. **Estimation.** Before conducting our simulated method of moments, we first estimate some preliminary parameters.

(1) Autocorrelation parameter (ρ): We use data from Indonesia Family Life Survey. We construct a panel data for 1993, 1997, 2000, and 2007. The sample used contains only those households that were surveyed in all the years. We use real total expenditures as our variable of interest.[32] Given that the gap between the years is long and variable, we use the mean gap to compute an approximate yearly ρ. The mean gap is 4 years so we obtain ρ using $(\rho)^4 = \hat{\rho}_{Panel}$ and it's distribution is derived using the delta method. We estimate $\hat{\rho} = 0.53$ and because of the size of the panel, the parameter is tightly estimated (standard error 0.01); thus, we view it as super-consistent relative to the structural parameters in our model.

[32]Expenditures were converted in real terms using the CPI published by the Central Bank.

(2) Variance of the error term (σ_ϵ^2): We obtain this variable using the stationary variance of the consumption process $\sigma_w^2 = \frac{\sigma_\epsilon^2}{(1-\rho^2)}$. Again, given the size of the data set this can be viewed as super-consistent.

B.3. **Simulated Method of Moments.** Equipped with a collection of over 600 graphs, ρ, and σ_ϵ^2, we estimate our model via simulated method of moments. The two parameters we are interested in are (α, τ) where $\alpha := \frac{\sigma_u^2}{\sigma_\epsilon^2}$ and τ is the maximal distance away from the source for an individual to be confident enough to pass information to her neighbors.

Our approach is a grid-based simulated method of moments which allows us to conduct inference on a large simulation quite easily (Banerjee et al., 2013). We let Θ be the parameter space and Ξ be a grid on Θ, which we describe below. We put $\psi(\cdot)$ as the moment function and let $z_r = (y_r, x_r)$ denote the empirical data for network r with a vector of wealth ranking decisions for each surveyed individual, y_r, as well as data, x_r, which includes expenditure data and the graph G_r.

Define $m_{emp,r} := \psi(z_r)$ as the empirical moment for village r and $m_{sim,r}(s,\theta) := \psi(z_r^s(\theta)) = \psi(y_r^s(\theta), x_r)$ as the sth simulated moment for village r at parameter value θ. Finally, put B as the number of bootstraps and S as the number of simulations used to construct the simulated moment. This nests the case with $B = 1$ in which we just find the minimizer of the objective function.

(1) Pick lattice $\Xi \subset \Theta$. For $\xi \in \Xi$ on the grid:
  (a) For each network $r \in [R]$, compute
$$d(r,\xi) := \frac{1}{S} \sum_{s \in [S]} m_{sim,r}(s,\theta) - m_{emp,r}.$$
  (b) For each $b \in [B]$, compute
$$D(b,\xi) := \frac{1}{R} \sum_{r \in [R]} \omega_r^b \cdot d(r,\xi)$$
  where $\omega_r^b = e_{br}/\bar{e}_r$, with e_{br} iid exp(1) random variables and $\bar{e}_r = \frac{1}{R}\sum e_{br}$ if we are conducting bootstrap, and $\omega_r^b = 1$ if we are just finding the minimizer.
  (c) Find $\xi^{\star b} = \text{argmin } Q^{\star b}(\xi)$, with $Q^{\star b}(\xi) = D(b,\xi)' D(b,\xi)$.[33]

(2) Obtain $\{\xi^{\star b}\}_{b \in B}$.

(3) For conservative inference on $\widehat{\theta}_j$, the j^{th} component, consider the $1-\alpha/2$ and $\alpha/2$ quantiles of the $\xi_j^{\star b}$ marginal empirical distribution.

In all simulations we use $B = 10000$, $S = 50$. We set $\Xi = [0.1 : 0.033 : 0.85] \times \{1, ..., 7\}$.

B.4. **Simulations for Regressions.** To generate our synthetic data we fix our parameters $(\widehat{\alpha}, \widehat{\tau})$ and generate 50 draws. We then compute
$$\overline{Error}_{ijk}^{SIM} = \sum_s Error_{ijk}^s / 50.$$
This allows us to aggregate the errors to any level we need. For instance by integrating over all the triples in our sample, we can compute $\overline{Error}_r^{SIM}$, the simulated error rate for village r. We then conduct our regression analysis with these simulated outcomes.

[33]Because we are just identified we do not need to weight the moments.

Appendix C. Details on Poverty Targeting Procedures

This appendix briefly describes the poverty targeting procedures used to allocate the transfer program to households. Additional details can be found in Alatas et al. (2012).

- PMT Treatment: the government created formulas that mapped 49 easily observable household characteristics into a single index using regression techniques.[34] Government enumerators collected these indicators from all households in the PMT hamlets by conducting a door-to-door survey. These data were then used to calculate a computer-generated predicted consumption score for each household using a district-specific PMT formula. A list of beneficiaries was generated by selecting the pre-determined number of households with the lowest scores in each hamlet, based on quotes determined by a geographic targeting procedure.
- Community Treatment: To start, a local facilitator visited each hamlet to publicize the program and invite individuals to a community meeting.[35] At the meeting, the facilitator first explained the program. Next, he or she displayed the list of all households in the hamlet (which came from the baseline survey). The facilitator then spent about 15 minutes having the community brainstorm a list of characteristics that differentiate the poor from the wealthy households in their community. The facilitator then proceeded with the ranking exercise using a set of randomly-ordered index cards that displayed the names of each household in the neighborhood. He or she hung a string from wall to wall, with one end labeled as "most well-off" (paling mampu) and the other side labeled as "poorest" (paling miskin). Then, he or she started by holding up the first two name cards from the randomly-ordered stack and asking the community, "Which of these two households is better off?" Based on the community's response, he or she attached the cards along the string, with the poorer household placed closer to the "poorest" end. Next, the facilitator displayed the third card and asked how this household ranked relative to the first two households. The activity continued with each card being positioned relative to the already-ranked households one-by-one until complete. Before the final ranking was recorded, the facilitator read the ranking aloud so adjustments could be made if necessary. After all meetings were complete, the facilitators were provided with "beneficiary quotas" for each hamlet based on the geographic targeting procedure. Households ranked below the quota were deemed eligible.
- Hybrid Treatment: This method combines the community ranking procedure with a subsequent PMT verification. The ranking exercise, described above, was implemented first.

[34]The chosen indicators encompassed the household's home attributes (wall type, roof type, etc), assets (TV, motorbike, etc), household composition, and household head's education and occupation. The formulas were derived using pre-existing survey data: specifically, the government estimated the relationship between the variables of interest and household per capita consumption. While the same indicators were considered across regions, the government estimated district-specific formulas due to the perceived high variance in the best predictors of poverty across regions

[35]On average, 45 percent of households attended the meeting. Note, however, that we only invited the full community in half of the community treatment hamlets. In the other half (randomly selected), only local elites were invited, so that we can test whether elites are more likely to capture the community process when they have control over the process.

However, there was one key difference: at the start of these meetings, the facilitator announced that the lowest-ranked households would be independently checked by the government enumerators before the beneficiary list was finalized. After the community meetings were complete, the government enumerators indeed visited the lowest-ranked households to collect the data needed to calculate their PMT score. The number of households to be visited was computed by multiplying the "beneficiary quotas" by 150 percent. Households were ranked by their PMT score, and those below the village quota became beneficiaries of the program. Thus, it was possible that some households could become beneficiaries even if they were ranked as slightly wealthier than the beneficiary quota cutoff line on the community list. Conversely, some relatively poor-ranked households on the community list might become ineligible.