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♯

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 citizens chose which households should receive aid, and show that the networks that the model predicts to be more diffusive differentially benefit from community targeting.

JEL Classification Codes: D83, D85

Keywords: Networks, Diffusion of information, Targeting, Development

Date: July 2015.

We thank Emily Breza, Pascaline Dupas, Matthew Elliott, Penny Goldberg, Ben Golub, Matthew O. Jackson, Laura Schechter, Adam Szeidl, Chris Udry, three anonymous referees, 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 Mounu Prem, 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. All views expressed are those of the authors, and do not necessarily reflect the views of the World Bank, the Royal Netherlands Embassy, Mitra Samya, SurveyMeter or the Indonesian Central Bureau of Statistics.

*World Bank. Email: valatas@worldbank.org.

†Massachusetts Institute of Technology, Department of Economics. Email: banerjee@mit.edu.

‡Stanford University, Department of Economics. Email: arungc@stanford.edu.

⋆Harvard University, Kennedy School of Government. Email: rema_hanna@ksg.harvard.edu.

♯Massachusetts Institute of Technology, Department of Economics. Email: bolken@mit.edu.

1. Introduction

Economists are increasingly conscious of the important role played by neighbors and friends. In particular, there is a growing interest in how communities aggregate information: individuals may have information that is useful or interesting to others, but does this information get to those who need it? And, how does the answer to this question vary by the community's social network? The answer to these types of questions can be important for policy design as information spreading is important for technology adoption (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)).

A related trend in developing countries is 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. This is predicated, in part, on the idea that communities have more information, and can more effectively aggregate it, than central governments. The particular example that motivates us here is the role of the community in *targeting* the poor for government assistance programs.[1] The idea behind community targeting is that it is difficult for the central government to effectively use 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 (Alatas et al., 2012). 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 be able to identify the types of villages where the networks are such that information will be aggregated well and therefore these decentralized mechanisms can be used more effectively.

However, despite a number of important theoretical contributions to the question of how information is aggregated within a community – many of which we discuss below – laying out the general relationship between network features and the extent of information sharing is challenging. Networks are very complex objects: they can differ along many dimensions, and how each network characteristic relates to the level of information aggregation can depend both on the network structure and the underlying model of social learning.[2] For example, consider the fact that while more

[1]Examples of community-targeted programs include Bangladesh's Food-For-Education (Galasso and Ravallion, 2005), Albania's Economic Support Safety Net (Alderman and Haque, 2006), and BRAC's Ultra-Poor program (Bandeira et al., 2012).

[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 preceding individuals' 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 asymptotic learning occurs under sequential observational learning in stochastic networks, provided that agents have expanding observations.. Gale and Kariv (2003) move away from the case of irreversible actions taken in sequence, and explore a special case in which a finite set of individuals in a network simultaneously take actions in every period after observing their neighbors' actions from previous periods; Mueller-Frank (2011) extends this setting. Under myopic Bayesian behavior, they provide conditions under which a consensus emerges, making use of the martingale convergence theorem. Mossel et al. (2011) show that in a world with binary uncertainty, if agents in a given network reach consensus asymptotically, then this consensus is the optimal aggregation of all agents' signals. Furthermore, as the number of agents grows, the consensus approaches the truth. That is, whenever there is asymptotic agreement, individuals agree on the truth. Meanwhile, another strand of literature studies various rule of thumb social learning processes. For instance DeGroot (1974), DeMarzo et al. (2003), Golub and Jackson (2010) and Golub and Jackson (2012) look at DeGroot learning on networks and study the speed

connections typically facilitate better communication, having a higher average number of connections (i.e., in the language of network theory, a higher average degree) does not guarantee better information aggregation.[3] To see why, consider an example where there could be a group of people in the community who are all connected to each other, but are entirely disconnected from the rest of the network, making information aggregation inefficient relative to a network where the average number of connections is lower but where everyone is indirectly connected to each other, so there are no isolated people (i.e., low clustering in the language of network theory).[4]

In the above example, the networks differ on both the average number of connections and the clustering of those connections. This suggests that if we want, for example, a general prediction for the effect of number of connections, we should compare networks that have similar clustering patterns, as well as similar patterns for other network features. However, no one measure of clustering summarizes all of the relevant information, just as no one measure of number of links is sufficient (i.e., the variance of the degree distribution 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 a case where the average clustering in the two networks is the same because in the first network everyone outside the one densely connected component is not connected at all. More generally, real networks can differ on so many dimensions that, except in special cases, the theoretical results do not provide clear predictions as to which networks will facilitate better information diffusion.

We take a more rough and ready approach to the problem of predicting the extent of information aggregation based on network characteristics. We exploit unusual data on networks from 631 Indonesian villages 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 (600+) 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 assess whether the villagers are right in their rankings. We can then explore whether network characteristics predict individuals having accurate information about others.

We begin with some reduced form facts from our setting to motivate the way that we model information aggregation below. In particular, we examine the relationship between people's network position and what they know. While these results are purely descriptive and do not address

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 (2007b) study information transmission as a percolation or contagion process in a model in which the network directs the probability of individuals meeting others.

[3]This is made clear, for instance, by Jackson and Rogers (2007b), who require first order stochastic dominance of the degree distribution (which is much stronger than a higher average degree) to ensure greater information diffusion in a meeting model where nodes meet other nodes with probability proportional to their degree.

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

the important and difficult identification issues (in particular, the challenge that being better connected may be correlated with other unobserved household characteristics that may also determine knowledge), the associative patterns are very strong: better connected people are better at ranking others, especially if we measure being better connected by number of connections. Similarly, people 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 in the ranking exercise is also related to network structure (i.e., a person is more likely to say he doesn't know about two other households if he is socially distant from these households). 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. We would ideally have clear theoretical predictions from a learning model about what network characteristics should matter. Unfortunately, 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 obtain them by using what we call "numerical theorizing."

Specifically, we use the *within village* variation 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. These predicted correlations will be our benchmark for examining the empirical cross-village comparisons: 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 observe in the data are qualitatively similar in sign to those predicted by our (numerical) network theory.[5]

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 another household in the village), but this state variable is changing over time as households' wealth evolves. Information about households' wealth flows over the network. Each period, people receive noisy signals about the wealth of others from their neighbors, and then noisily transmit information to their own neighbors, and so on. Agents aggregate the sequence of signals that they have received to develop a guess about the wealth of the target household. In our model, they aggregate this information using a Kalman filter, treating each piece of information they receive over the network as an independent signal.[6] Additionally, inspired by what we observe in the data, the model allows individuals to make a judgement about the quality of their evidence before deciding whether or not to report it.

[5] 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.

[6] 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 piece of information through many different paths, the Kalman filter will not typically be Bayes optimal since it assumes that each piece of information is independent when it is not. However, for that reason, it is orders of magnitude easier for the agent to compute than full Bayesian learning which would require the agent to understand and undo all the sources of correlation between his signals. This behavioral mistake is consistent with data from lab experiments (Chandrasekhar et al., 2012).

In addition to being useful to put structure on the data, our model is of independent interest. Learning about a changing parameter is rarely studied in the theoretical literature (see Frongillo et al. (2011) for an example); however, it is a useful description in certain contexts of the world. Beyond capturing how people learn about others' incomes, this kind of knowledge transmission may be important for understanding topics such as matching of individuals to transient labor market opportunities, technology adoption when the benefits of technology evolve over time given the state of the world (e.g., weather, other types of inputs), etc.

After fitting our 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.[7] 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.[8] We use these predicted correlations as a benchmark for the actual, empirical cross-village comparisons. This is what we mean by numerical theorizing: we then ask whether the cross-village results from the simulated model are qualitatively similar in terms of sign to the empirical cross-network correlations between the information aggregation and network characteristics in our data.

The empirical patterns match up reasonably well with the results generated by our model. In most cases, whenever either the simulated or the actual empirical correlations between network characteristics and measures of information aggregation 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.[9] To the best of our knowledge, ours is the first exercise of this sort, looking at descriptive evidence on social learning across many independent networks to see if it is consistent with an underlying theory.

[7]Note, to ensure that our results are not driven by the specific parameter values (the degree of noise as well as the threshold of certainty that a belief needs to cross before an agent is willing to pass information) 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 centered approximately around the estimated values (Appendix G). The basic predictions turn out to be remarkably robust to different parameter values, implying that the patterns that we observe may be portable across varying contexts.

[8]The choice of these network characteristics is inspired by analytical results in the literature on the determinants of information aggregation in networks. The analytic results come from simplified models of diffusion that researchers have used to gain traction on an otherwise difficult problem. For example, Jackson and Rogers (2007b), as mentioned above, focus on the effects of first order shifts in the degree distribution. Bollobás et al. (2010) focus on the role of the first eigenvalue of the adjacency matrix. More generally, there are many ways to summarize the adjacency matrix, so the list of network characteristics that we consider is necessarily incomplete.

[9]For example, we show that if network A has a degree distribution (i.e. the distribution of the number of neighbors) that first order stochastically dominates the degree distribution of network B, then both in the simulations and the empirical analysis network A is more likely to have more information aggregation. This echoes the Jackson and Rogers (2007b) result on stochastic dominance described earlier. Further, we find that the first eigenvalue of the link matrix predicts information flow, which echoes the results in the viral transmission model studied by Bollobás et al. (2010).

However, we also see interesting divergences from what one might have intuitively expected. For example, the effect of higher average number of connections on information aggregation, *controlling for other network characteristics,* is negative both in our simulations and in the empirical results. Though there is a standard intuition that more connections are better, this is not true once one conditions on other dimensions of the network.

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. Our 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 better reflects people's self-assessment of their poverty status in villages that our network model predicts should have better information passing properties.

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. Second, the findings offer insights into policy design problems where governments aim to 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 where we would expect effective information aggregation. 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 them. 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.

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 (Alatas et al. (2012)). Between November 2008 and March 2009, we conducted a randomized experiment to compare the accuracy of three common methods to identify beneficiaries for transfer 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).

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, described below.

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.[10] For each village, we obtained a list of the smallest administrative unit within it, and randomly selected one of these units (henceforth "hamlets") for the experiment. Best thought of as neighborhoods, each hamlet has an elected or appointed administrative head ("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 experiment's baseline survey. SurveyMeter, an independent survey organization, administered the baseline survey in November to December 2008, before the experiment or the social program was announced. For each hamlet, we constructed a census of households and then randomly selected eight households to be surveyed. In addition, we also surveyed the hamlet heads. From this survey, we used information on social networks and on both the perceived and actual income distribution in the hamlet.

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). 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, and then to list all of the relatives of each person named.[11] Second, we asked each respondent to name the social groups within the hamlet that any members of his/her household had participated in, including neighborhood associations, religious groups, school groups, ROSCAs, farmers' associations, etc. This allowed us to relate people through common membership in groups.

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 knowledge and to compare their beliefs to the "truth." To collect data on knowledge, we asked each surveyed 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*). We then collected two measures of "truth." First, we collected a measure of actual per capita expenditure levels at the time of the baseline survey, using the standard 28-question expenditure module from the Indonesian SUSENAS survey. 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)

[10]Note that our results are for the most part robust to just constraining our sample to rural villages, despite the much smaller sample size. See Appendix L.

[11]We can check the quality of this data as follows. If we look at all the surveyed households (about 9 per hamlet) and consider their relatives, we can ask what share of their kin were named by others when others listed these individuals as among the five richest, five poorest, or leaders. This number is 80% in our data.

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 Alatas et al. (2012), 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, 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. For example, if the true rank of person j is 1 and of person k is 7, people who ranked j above k in their own rankings would get credit for a correct answer, regardless of the distance between their ranking of person j and k.[12] The hamlet level error rate is then the mean over the nine households in the hamlet.

2.3.2. *Network Data.* We construct undirected, unweighted network graphs from the familial and social group data for each of the 631 sampled hamlets. This is very unusual data, as most typical studies have closer to five independent networks and thus cannot make cross-network comparison.

To construct each network, we first construct edges between the households that we sampled and those that they identify as their family members. This fills in nine rows and columns in the adjacency matrix. However, while we only sampled nine households per hamlet, our data is considerably richer than that, because as mentioned above, for each surveyed households, we asked them to identify and list the the relatives of the five wealthiest and five poorest households in their hamlet, as well as all the formal and informal hamlet leaders and their respective relatives.

While the households that are named here are non-random, it provides us with complete set of kin for a total of 68.3 percent of households in a median hamlet. Further, by the transitivity of kin, we can connect each pair of these relatives of a given household. In other words, 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). Finally, for our sampled households, we also construct an edge between any two households that are registered as part of the same social group. We then take the union of these graphs.

Note that in addition to having full kin data for all of the surveyed or named households (i.e. 68.3% of the network), we also have partial network data for others who were listed as related to someone who was either surveyed or named as poor, rich or a leader. This is because for any j that we have not sampled, we know if it is connected to any sampled household or any named household. We can conduct a simple back-of-the-envelope calculation to estimate what share of potential links ij we could be missing in our data. Assume for a moment that we have complete data on 68.3% of households uniformly at random. This implies that we miss kin link data for pairs of households ij for only $1/10th$ of potential links, since $(1 - 0.683)^2 \approx 1/10$. Thus, for about 90% of all pairs ij we should know if they are kin or not. This is consistent with the empirical frequency, where we

[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.

can directly compute for which pairs ij do we definitively know if ij are kin or ij are not kin. In the median hamlet the share of missing data on such pairs is 9.5%.[13]

These missing links are unlikely to undermine the credibility of our results. First, while regression analysis on partial samples of network data can generate biases due to non-classical measurement error, Chandrasekhar and Lewis (2012) develops a graph reconstruction technique to deal with this issue. Our results are robust to their correction (described and shown in in Appendix H). Second, we observe that for a subset of the claims that we are interested in, such as the result on first order stochastic dominance of a hamlet's degree distribution, our results are underestimates since the direction of the bias is to attenuate coefficients.[14] Finally, we conducted a series of robustness checks to look at the sensitivity of the results to missing kin data that we discuss in Section 5.3 (in particular, collecting network data for all households in 10 randomly chosen hamlets in our data).

It is important to note that all our analysis assumes that the individuals who are part of the networks are fully described by their observables (including their network position). However, in practice, the networks – and individual network positions – are likely to be endogenously determined. For example, more central individuals are likely to be different on unobservable dimensions compared to less central individuals. These unobserved characteristics may in turn be correlated with what they know about others, and this may be a part of the reason why central people turn out to know more. Our approach in this paper is thus a descriptive one: we do not have random variation in network structure that would make network position uncorrelated with all possible unobservables. As a result, the claims we make here are non-causal – we only ask whether the data can be rationalized by a natural model of network interaction, and if that can teach us something about which communities are likely to know more about their wealth distribution. On the other hand, it is also worth noting that most network data sets are subject to the same limitation without having the great advantage of having over 600 independent networks (typical studies have closer to 5) to work with as well as a measure of information for each of them, which makes our data ideal for carrying out the kind of cross network comparisons we are interested in.

2.3.3. *Aggregation of Data in Community-Based Targeting.* Whether to assign the responsibility for "targeting" – the selection of beneficiaries to social programs aimed towards the poor – to local communities has become an important policy question with increasing recognition of the challenges of accurately measuring household income. The data used in the paper was collected prior to an experiment in which we compared community targeting with the status quo in Indonesia, which

[13]Moreover, the data coverage is even better for the relevant parts of the network: if we ask household i to rank household j vs k, since we also randomly sampled j and k and know their complete kin networks, we would know for sure if j and k were connected at distance 1 or 2 (since we know all of j's connections and all of k's connections, our data would tell us if j and k were connected of distance 2 or less). So if a relevant link is missing, we can infer that they are at least distance 3 or more.

[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 provided the measurement error is uncorrelated with the structural error in the regression of interest. This covers the case when the value of the measurement error is correlated with the value of x itself, which is likely to be true in a network setting but assumed away under classical measurement error. Following the Cauchy-Schwarz inequality it is easy to show that $\beta_0 \cdot \sigma_x \geq \text{plim}\, \widehat{\beta}\sigma_{\bar{x}} = \beta_0 \frac{\text{cov}(x,\bar{x})}{\sigma_{\bar{x}}}$ as $\sigma_x \sigma_{\bar{x}} \geq \text{cov}(x_i, \bar{x}_i)$ 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. Note that if the argument holds in the univariate case, it also holds for the multivariate case where covariates other than the covariate of interest are not measured with error, by the Frisch-Waugh theorem.

is to use data collected by the central statistical system. 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: Proxy Means Test (PMT), Community or Hybrid. In the PMT treatment, program beneficiaries were determined through a regression-based formula that mapped easily observable household characteristics collected by the statistical system into a single index. In the community treatment, the hamlet residents determined 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).

Using intuitions from network theory on information aggregation, we can look at 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 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 hamlets with network structures that should lead to better information transmission.

3. Sample Statistics and Information at the Household Level

In this section, we establish stylized facts to motivate our diffusion model (section 4). We first provide sample statistics to describe the knowledge environment. Next, we explore how a household's network position is correlated with their ability to rank others within the hamlet (section 3.1.1). Finally, we look at whether households are better at ranking those who are more connected to them (section 3.1.2). Note that these are descriptive regressions and not causal estimates.

3.1. **Sample Statistics.** Table 1 reports descriptive statistics (Appendix A provides more detailed 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 the variable means in Column 1 and standard deviations in Column 2.

The sampled hamlets are small, with an average of 53 households (Panel A). The largest has 263 households, the smallest has 11, and the inter-quartile range is 25-64. The number of connections per household, called *degree* in the network literature, averages 8.18. Networks 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, suggesting that two randomly chosen households will be separated by one household in between, conditional on there being a path that connects the two households. 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]

Households struggle with making the comparisons of the households' economic status. The mean hamlet error rate based on consumption is 0.52, while that based on the self-assessment is about 0.46. However, there is substantial heterogeneity in the error rate across hamlets – the standard deviation for both variables is about 0.2, which means that in the very best hamlets 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 are likely due to errors in our measure of consumption (Alatas et al., 2013). Under classical measurement error in the outcome variable, regression coefficients will be unaffected, so this is not a problem per se for the paper, but worth keeping in mind for interpreting the levels of these variables.

Many households refuse to make certain comparisons: the rate of reporting "do not know" is 0.19. This suggests that the appropriate model should account for this aspect of reality. Even when reporting, the individual error rate is still high: 0.36 and 0.27 for consumption and self-assessment.

Panel B provides corresponding sample statistics at the household level. It is worth noting that these networks exhibit a high clustering coefficient; the average clustering coefficient is 0.64.

3.1.1. *Network Position of those Ranking Others.* We first ask whether more central individuals have a lower error rate in ranking other households. In Table 2 and 3, 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 categorizes incorrectly) or its rate of not knowing at least one of the households in the ranked pair, W_{ir} are i's network characteristics, X_{ir} are demographic characteristics, $\mu_r$is a hamlet fixed effect, and ϵ_{ir} is the error term.

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 eigenvector centrality (Columns 3 and 7), which is a measure of the node's importance, defined recursively, to be proportional to the sum of the importance levels of her neighbors. Detailed 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 (μ_r) and Columns 5-8 add hamlet fixed effects in order to sweep out any cross-network average differences and focus just on within-network differences in position. Since network position may be correlated with

[15] It is possible that the underlying network is completely connected. The fact that the share of nodes in the giant component is less than one may be due to the fact that we have sampled the network. If the true network was more dense, then a random sample from it is more likely to be completely connected in the sense that it is more likely that the researcher observes a path between any two nodes. If the true network was sparse, even if there was a giant component, sampling the network could make these paths break in the observed graph thereby reducing the share of nodes in the giant component.

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

other household characteristics, we also explore whether the results are sensitive to controlling for ranker demographic characteristics, 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, is from an ethnic minority, is from a religious minority, and whether the respondent is female. Table 2 reports the results with no covariates (i.e. constraining δ to be zero) and Table 3 reports results with covariates.

Overall, being a more connected household is associated with a lower error rate in ranking other households. Using consumption as the measure of the truth (Panel A of Table 2), 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 with hamlet fixed effects, magnitudes remain similar to the bivariate cases with fixed effects, but we are no longer 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 in terms of their self-assessment. The coefficient estimates of all models are similar across Panels A and B, both in terms of sign and magnitude. 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 that we discuss below as well (see Appendix Tables E.1 and E.2). This suggests that network-level effects may be important for information aggregation, a subject we explore in much more detail below.

In Panel C, 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 6.3 or 1.5pp decrease in the likelihood of reporting "don't know" (without or with hamlet fixed effects, respectively) in the bivariate regressions. Recall that the mean of "don't know" is 0.19, which indicates that these effects are large. A similar result is true for eigenvector centrality.

The results are robust to two key changes in specification. First, including the control variables in Table 3 does not alter the findings, suggesting that the results are not driven by these observable household demographic characteristics. Nonetheless, from now on, we always include the demographic control variables unless otherwise specified.[17] Second, we also explore these relationships excluding the cases where individuals claim they do not know (Appendix Tables D.2). The results

[17] For regressions that study within-hamlet variation, we present tables both with and without demographic control variables. When we look at across hamlet regressions, versions without are in Appendix F.

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.

3.1.2. *Connections Between Ranker and Rankee.* The preceding analysis explored how one's network position affected her accuracy in ranking others. In Table 4, we now explore whether the ranker is more accurate when he is more connected to the households that he is ranking; i.e., does household i do 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 $d(i,j)$. Many nodes cannot be connected by any path; by convention,the distance between them is infinite. We 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 scales to a measure of closeness in $[0,1]$. Specifically, we estimate:

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

where $Error_{ijkr} = \mathbf{1}\{i \text{ ranks } j \text{ versus } k \text{ incorrectly}\}$, W_{jkr} is the average network characteristic of the households being ranked (j and k), and X_{ijkr} are the covariates. The sample is all i, j and k in hamlet r such that $j < k$, $j \neq i$, $k \neq i$. In Column 1 of Table 4, we show the basic correlations between the error rate and average inverse distance from i to j and k, conditional on the same set of demographic covariates as above (log consumption, education, etc) for both ranker i and the average for rankees j and k. 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 (μ_r) and ranker fixed effects (ν_i,), respectively. All standard errors are clustered by hamlet.

Average inverse distance is highly predictive of ranking accuracy. 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 1.5 to 3.9 percentage points less likely to rank 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), although the sign and magnitudes of the coefficients are generally similar to Column 3. Using self-assessment as the truth (Panel B), the average reachability and inverse distance predicts the error of the ranked pairs with demographic controls and hamlet fixed effects (Column 3). Again, when controlling for ranker fixed effects (Column 4), the effect of average inverse distance is no longer significant.

In panel C, 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." Again we find that if the ranker is at distance 1 to each of the rankees, as opposed to distance 3 then the ranker is anywhere from 1.9 to 4.9pp less likely to declare a don't know in the assessment of one of the ranked parties.[18]

3.1.3. *Summary of Results thus far and Outline of Subsequent Approach.* In short, we find that, both with and without conditioning on observable demographic characteristics, (i) more central households are more likely to rank other households rather than say they don't know; (ii) more central households are less likely to guess incorrectly; (iii) households are more likely to guess rather

[18]In panels A and B, non-response is always coded as error. Even when we drop households that are not ranked, the ranking is more likely to be correct when the ranker and rankee are more closely connected (Appendix Table D.3).

than say they don't know when they are closer in the network to the people that they are ranking; and (iv) households are less likely to make ranking mistakes the smaller the distance in the network to the people they are ranking.[19]

We use this description of the environment to motivate a novel (though straightforward) quasi-Bayesian model of social learning (Section 4). Since we, the researchers, ask a household i to assess the wealth $w_{j,t}$ of some household j in period t, the model deals with characterizing i's estimates of $w_{j,t}$, given i's history of observations. We assume that household wealth can change over time. Agents are trying to learn about this, but it takes time before they hear about shocks to distant household's wealth, since this information needs to travels through the network. Moreover, every time an individual transmits information to her neighbor, a little bit of noise gets added (communication is noisy). As a result, if i and j are close in the network, then i will learn newer (therefore more predictive) information about j's wealth more quickly and with less noise. The model individuals use to aggregate this information is exactly Bayesian for certain special classes of networks, but simpler and less computationally demanding for others. The deviations from Bayesian learning of this model are consistent with evidence from laboratory experiments (Chandrasekhar et al., 2012). We then take the model to the data and estimate structural parameters of the model using moments obtained using within-village variation.

In Section 5, we simulate the learning process on our networks in order to generate predictions about the relationship between the network structure in a hamlet and the average error-rate in predicting wealth. We then estimate these relationships in our actual data and observe whether the actual empirical results appear qualitatively similar to the theoretical predictions and are thus potentially consistent with the model. It is worth noting that fitting the model using within-village variation does not automatically imply that the model would be successful in explaining the cross-village variation. This is due to the complexity of the relationship between individual level information transmission and its overall aggregation through the network, which is what our model is meant to help with.

Finally, we explore whether the networks that we predict to spread information better do better in a real policy settings where aggregate information is required (Section 6).

4. Model, Estimation, and Simulation Results

4.1. **Model Overview.** We build a parsimonious model that relates network characteristics to information diffusion, capturing the key features of the environment 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.

[19]Note that this evidence suggests that a story of social learning is plausible. However, it is also possible that alternative stories may explain these patterns. For example, it is possible that more central individuals are more likely to know people and learn about them directly (from talking or observing them). In this case, they would be learning individually about other individuals, but not necessarily passing along information to others. They would just be more likely to meet others and the network would be describing a meeting process.

(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 that they received 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 others through communication on a social network. We assume that individuals receive information from others and make some judgment about the quality of that information before deciding whether to report it. We outline the key aspects of the model here; the next section writes down the model formally.

More 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.[20] 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 also receives signals about j from anyone he is connected to who has such a signal, and 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 because it will have passed through more steps en route and acquired noise at each stage, and also because the information it is based on is older and therefore 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 be passed on, note that people typically receive information in a given period from multiple pathways, some of which is outdated. 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 (or if there many such people, the average of their reports). 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 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 do not find it worth their while to pass on stale information. If their information is sufficiently outdated, people do not pass it on (i.e., they do not pass the information on to other households, and they would say they "don't know" anything about j if asked in a survey).

[20]One may wonder why people get into conversations about each others' wealth. The primary motive may be as simple as a desire to gossip. Some reflection on conversations one engages in surely illustrates that individuals talk about others' purchases and so on. They may also be interested in their status relative to those of their peers, in which case it is possible that people may try to hide their consumption from others. However, that should reduce the advantage of central people, since they are the ones who are most likely to spread that information. This is the subject of Banerjee et al. (2012), who study information diffusion in a rival setting.

In terms of aggregation of information, assuming that people are fully Bayesian in this context may be somewhat unrealistic. Full Bayesian aggregation 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 source (and have 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 accounts for all signals, past and present, and correctly averages 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). 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?[21]

We therefore adopt the following approach. The decision-maker treats the signals that he receives as if they were independent (conditional on the truth) and applies Bayes' rule, under the potentially incorrect assumption about independence. Since the weight given to each signal 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 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 vastly 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 $p_n n$ signals in each period, each of which has a precision given by its distance from the source. To get a sense of the magnitude of this number, note that in our data the average degree is 8. The independence assumption 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 laboratory experimental settings (Chandrasekhar et al., 2012) as well as in more recent field experiments (Bai et al., 2014).[22]

[21] Note that we are not saying that Bayesian learning on a network always requires doing all these calculations. For example, if individuals always pass on 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 and fast-growing object. An alternative possibility is suggested by recent work by Mossel and Tamuz (2013), who study a context where all agents receive signals, and show that the decision-maker can compute the Bayesian beliefs using an algorithm that is polynomial in n, the number of nodes in the network. However, this computation requires that everyone knows the entire graph, which is not particularly realistic. It remains to be seen in what way this result extends to settings where the graph is not known. Moreover, even if it turns out that the required computation is easier than we think, it may well be harder than what people want to undertake – based on both lab experimental evidence (see Chandrasekhar et al. (2012)) and field evidence (Bai et al. (2014)).

[22] Indeed this is one of the arguments routinely 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 simply weighted averaging with exogenously given weights. Individuals start with a belief about the state of the world. They then look at their neighbors' beliefs from the previous period, they average the opinions using fixed weights and form a new opinion which is then passed into all the neighbors so that the process continues. One interpretation of our model is as an extension/refinement of a DeGroot model where we micro-found the time-varying weights.

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, possibly directed, 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 $ij \in E$. The model applies to directed graphs, where information flows along (directed) edges. In our application we consider undirected graphs, namely information always flows both ways.

We model people's wealth as an evolving stochastic process in discrete time. 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 $\epsilon_{j,t} \sim \mathcal{N}\left(0, \sigma_\epsilon^2\right)$ that are independent across j and t. All households know the fundamental parameters ρ, c, and σ_ϵ^2, and this is common knowledge.

In what follows, we fix a given node j about whose wealth the remainder of the nodes are learning. Individuals $i \in V \backslash \{j\}$ have beliefs over $w_{j,t}$ that are informed by social learning. At period t, given the entire history of information that i has ever received from her neighbors, i has beliefs about $w_{j,t}$ given her information set. At $t = 0$, every individual has a prior, which is a normal distribution given by the invariant distribution: $\mathcal{N}\left(\frac{c}{1-\rho}, \frac{\sigma_\epsilon^2}{1-\rho^2}\right)$.

The model will have a transmission error at every step when an individual speaks to another individual. For instance, when l communicates with i in period t, l may be passing information about $w_{j,r}$ for some $r < t$. This communication is disturbed by some $u_r^{l \mapsto i}$. We will assume that every $u_r^{l \mapsto i}$ is independently and identically distributed according to $\mathcal{N}\left(0, \sigma_u^2\right)$ which again is known to all agents.

To preview the remainder of the setup, recall that we have fixed j and everyone learns about j's wealth, which evolves over time. In every period t, for every pair of agents l and i that are linked, l sends at most one piece of information about j's wealth to i. This information is a noisy signal about j's wealth at time $r < t$, $w_{j,r}$. This corresponds to the period that is the newest piece of information that l has about j, and therefore this implies that $r = t - d(l, j)$ since it takes that many steps for information to come from j to l. Because we will assume that individuals only pass on information if they are certain enough about it, an immediate result is that it is equivalent to write the model such that agents only pass on information if they are close enough to the source, since the degree to which information is distorted is exactly proportional to the distance it has traveled.

We now formally define the communication protocol. At period t we look at what node i receives from others and we consider her updating problem:

- *Signals from the source j:* Every period, the source j generates a signal about her $t-1$ wealth that she transmits to each of her neighbors, $i \in N_j$.

$$S_{t-1}^{j \mapsto i} = w_{j,t-1} + u_{t-1}^{j \mapsto i}.$$

- *Signals from an arbitrary node l to i:* Every period, a node l noisily transmits the most recent piece of gossip she has heard about j's wealth to each of her neighbors. The noise is independent across transmissions.
  - Let $k^* := k^*(l,j)$ be the neighbor of l that is closest to j.[23] The signal that l received from k^* the previous period is what will then be passed on.
  - Passing only occurs if l is sure enough about the quality of this information. An immediate consequence of this assumption is that we can write that there exists some threshold τ such that if $d(k^*,j) \leq \tau$, then l passes information to each of her neighbors. If $d(k^*,j) > \tau$, then no information is passed.
  - When l passes information, it is

$$S_{t-d(l,j)}^{l \mapsto i} = S_{t-1-d(k^*,j)}^{k^* \mapsto l} + u_{t-d(l,j)}^{l \mapsto i}.$$

    Notice that $S_r^{l \mapsto i}$ denotes the information about j's wealth at time r (i..e, $w_{j,r}$) that l passes on to i at time t. In the above $r = t - d(l,j)$, since it takes $d(l,j)$ periods for the information to come from the source j to node l.
- *Forming a posterior:* $i \in V \setminus \{j\}$ forms a posterior about $w_{j,t}$ by using a Kalman filter on her historical data which is all information that has ever been passed to her from her neighbors at any period in the past. This is a vector

$$\mathbf{s}^{i,t} = \left(s_1^{i,t}, ..., s_{t-d(j,i)}^{i,t}\right)$$

  where the signals that i has about $w_{j,r}$ at period t, denoted $s_r^{i,t}$, can be constructed from the signals that i has received in various periods from her neighbors when they transmitted period r information to i, $\left\{S_r^{l \mapsto i} : \ l \in N_i, \ r \leq t - d(l,j)\right\}$. This is simply $s_r^{i,t} = \sum_{l \in N_i} \omega_{l,r,t,i} \cdot S_r^{l \mapsto i}$, where the weights are the appropriate precision-based weights, defined below.

  A Kalman filter uses the entire history of (noisy) signals $\mathbf{s}^{i,t}$ to help predict $w_{j,t}$. Essentially, each signal provides information about the current value $w_{j,t}$ since the entire observed history is measured with noise. Notice that because agent i may be receiving signals from her neighbors at varying distances from the source, the information she has about j's wealth at some given past period r can vary over time.[24] We discuss this in greater detail below and in Appendix B.

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

$$s_r^{i,t} \sim \mathcal{N}\left(w_{j,r}, \sigma_{r,t,i}^2\right)$$

  where i's tth period set of signals about $w_{j,r}$ can only come from neighbors that are close enough to j. This is because only neighbors of i that are within $t-r-1$ steps of j can reveal an estimate of $w_{j,r}$ to i by period t. Every time the signal is transferred across individuals, it is disturbed by a shock with variance σ_u^2, leading to a variance of $\sigma_u^2 \cdot d(l,j)$.

[23] For presentation purposes we assume this is unique. If it is not unique, and there are two or more such closest signals, then we assume that l passes the average.

[24] That is, $s_r^{i,t}$ need not be equal to $s_r^{i,t-1}$ since at period t individual i could have received a signal from some other neighbor at a further distance about $w_{j,r}$, which now updates $s_r^{i,t-1}$ to $s_r^{i,t}$.

In this case, we can compute i's period t variance of its signal about $w_{j,r}$ as

$$\sigma^2_{r,t,i} = \sum_{l \in N_i} \omega^2_{l,r,t,i} \cdot \sigma^2_u d\left(l,j\right),$$

where $\omega_{l,r,t,i} = \frac{\mathbf{1}\{t-r \geq d(l,j)+1\}/\left[\sigma^2_{u,} d(l,j)\right]}{\sum_{k \in N_i} \mathbf{1}\{t-r \geq d(k,j)+1\}/\left[\sigma^2_u d(k,j)\right]}$ is the weight that i puts on l's estimate of $w_{j,r}$ in period t.

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

This model is actually much simpler than it might seem. 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 on a directed line, where all nodes are attempting to track node 1's wealth. Panel (A) shows the predictions of 1's wealth 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 it is dramatically computationally simpler on more general networks. The following result makes the equivalence with Bayesian learning precise:

Proposition 4.1. *For any directed graph where the source j is the root and every i node is connected to the source only through independent paths, i's learning process about j is fully Bayesian under our above model.*

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

First consider the case of a directed tree with the source, node j, being the root. Let i be a node with $d\left(i,j\right) = q \leq \tau$. Note there is exactly one path from j to i. It is useful to denote $j = 1$, the first node, and then label nodes in sequence $1, ..., n$, where node $i-1$ communicates to node i. Then a generic node i receives a q period lagged signal about $w_{j,t}$, $S^{i-1 \mapsto i}_{t-d(i-1,j)}$ in the previous notation, that has been disturbed by the equivalent of noise distributed $\mathcal{N}\left(0, q\sigma^2_u\right)$. Thus, the problem can be recast as an agent i making a prediction about state $w_{j,t}$ given a history of signals $s^{i,t}_0, ..., s^{i,t}_{t-q}$, where in this case $s^{i,t}_\kappa = S^{i-1 \mapsto i}_\kappa$ for any period $\kappa \leq t - d\left(i,j\right)$. 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 i has L independent paths from node j, with $q_l = d_l(i,j) \leq \tau$ for $l \in \{1,...,L\}$, the computation is as follows. Let $q = \min_l \{q_l\}$. In period t, an individual has information $s_0^{i,t}, ..., s_{t-q}^{i,t}$, where $s_\kappa^{i,t}$ are computed using the period κ signal along each independent path. Again this generates a sequence of Kalman filters, indexed by t. That is, the Bayesian prediction of $w_{j,t}$ given the signal sequence $s_0^{i,t}, ..., s_{t-q}^{i,t}$ is given by a Kalman filter and prediction. By definition, this is exactly the computation that our agents do in the model. □

The set of networks covered by Proposition 4.1 includes direct lines, more generally directed trees, as well as other configurations. For instance, see the networks in Figures 2a, 2b and 2c. of these, Figure 2c depicts a graph with arbitrarily long but independent paths that lead from the source to other nodes.

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. **Model Estimation.** 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.[25] From our survey data, we estimate c. We also estimate σ_ϵ^2 from our survey data. Note that the AR(1) model implies the relationship $\text{var}[w] = \frac{\sigma_\epsilon^2}{1-\rho^2}$. We use data on wealth to estimate $\text{var}[w]$, and then estimate σ_ϵ^2 using the previous relationship and the estimate for ρ.

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 to the source, beyond which people stop transmitting information about the source). 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.

[25]We use the 1993-1997 and 2000-2007 periods to estimate ρ, avoiding the 1997-2000 period where ρ was likely much lower due to the Asian Financial Crisis. Doing so does not substantially affect the main conclusions of the exercise.

Our estimation of the model 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.[26]

The parameter values from the estimation are shown in Table 5. For ease of interpretation, we present a normalization of the first parameter, $\alpha := \frac{\sigma_u^2}{\sigma_\epsilon^2}$. We estimate it as $\widehat{\alpha} = 0.397$. This means that the transmission error is two-fifths of the size of the structural wealth shocks. However, the standard errors are such that $\alpha = 0.5$ would be a reasonable estimate of the transmission error to structural shock ratio. We also find that $\widehat{\tau} = 4$. This means that a node connected to 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, conditional on being connected, 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. We generate 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 2, 3, and 4 using the simulated data from the model; the results are provided in Panel C of Appendix Tables E.1, E.2 and E.3 of each respective table. 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.1 and E.2). We find that inverse distance is correlated with a reduction in 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 analytical result in Jackson and Rogers (2007b) showing that if network I's degree distribution and neighbor degree distribution first-order stochastic 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.[27]

Jackson and Rogers (2007b) was the first result to note that under some regularity conditions, networks that are more diffusive in the sense of first order stochastic dominance of the distribution of agents' links should have more information diffused in the equilibrium. In more layman's terms: "if we look at two networks A and B, which has more diffusion and can we tell based on the

[26] This assumption makes sense in the environment of our model since we would expect the respondents 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 this likely does not make any significant difference; it is also simpler to assume that households use the same rule when passing information as when they respond to the survey.

[27] 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.

distribution of links in the network (the degree distribution)?" This is an important and sensible question because it asks if the basic trait involved in learning – the distribution of how many links one has to their learning partners – will tell us something about whether a community has more diffusion than another. Jackson and Rogers (2007b) noted that this was a particularly difficult question to study on a fixed network, but by moving to a random matching model, they were able to simplify the analytics to be able to generate a suggestive answer.

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 all local neighborhoods essentially contain the same average information as the global average. The approximation does not work well when, for example, 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 piece of evidence to learn in Jackson and Rogers (2007b).

We therefore use the numerical simulations of our model to examine whether we should expect the equivalent result to hold in our context (see Appendix B for details). 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 $\overline{Error}_{ir}^{SIM}$, and then we construct hamlet level error rates ($\overline{Error}_{r}^{SIM}$ for hamlet r) by averaging over the individual level error rates.

Our main outcome variable of interest is a dummy equal to one if $\overline{Error}_{I}^{SIM} > \overline{Error}_{J}^{SIM}$, and zero otherwise. We regress this variable on whether I stochastically dominates J or vice versa:

$$\mathbf{1}\left\{\overline{Error}_{I}^{SIM} > \overline{Error}_{J}^{SIM}\right\} = \beta_0 + \beta_1 \mathbf{1}\left\{I \succ_{FOSD} J\right\} + \beta_2 \mathbf{1}\left\{J \succ_{FOSD} I\right\} + X_{IJ}'\delta + \epsilon_{IJ}. \quad (4.1)$$

We include fixed effects for geographically clustered groups of hamlets, hamlet-level control variables, and specify two-way clustered standard errors, for hamlet I and hamlet J.[28] The results, which are reported in Table 6, suggest that the Jackson and Rogers (2007b) pattern holds in our context. Since stochastic dominance is a partial ordering, the omitted category in Columns 1 and 3 is the non-comparable groups of hamlets. In Columns 2 and 4 we focus only on comparable hamlet 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 25pp 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. 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 affect diffusion within our estimated model.

[28]Specifically, we include fixed effects for the stratification group from the Alatas et al. (2012) experiment.

These measures are described at length in Jackson (2008). The average degree is an obvious and basic measure for a diffusion process, since it captures the average number of links. Similarly, the density of the links, which is the average degree scaled by the size of the network, captures the probability that a randomly chosen node is linked to another randomly chosen node in the network. Basic intuition suggests that higher linking rates may correspond to higher learning probabilities, an idea which is articulated more formally in Bollobás et al. (2010).[29]

Another important feature that could be relevant for learning is the correlation of links. A basic way to capture this is using the average clustering in the network, which measures the share of a nodes' neighbors that are themselves linked (Jackson, 2008; Jackson et al., 2010). More correlated links can re-enforce beliefs and present a divergence between rule-of-thumb and Bayesian learning (DeMarzo et al., 2003; Gale and Kariv, 2003; Golub and Jackson, 2012; Chandrasekhar et al., 2012), since Bayesian agents will have to undo correlation in signals that emerge through clustering.

Moreover, because information flows along paths, a natural measure to include is the average of path lengths in the network (Albert and Barabasi, 2002; Jackson, 2008; Golub and Jackson, 2012).

Finally, we include the fraction of nodes in the giant component. A well-understood empirical regularity is that there exists a path between many (if not most) pairs of nodes and therefore most nodes are part of a very large component called the giant component (Albert and Barabasi, 2002; Jackson and Rogers, 2007a; Jackson, 2008; Bollobás et al., 2010). Mechanically, in a learning-on-networks model, if two nodes are not part of the same component there cannot be any direct or indirect exchange of information, since there is no path of information from one node to the other.

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 $\overline{Error}_{ir}^{SIM}$.

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

$$\overline{Error}_{r}^{SIM} = \beta_0 + W_r'\beta_1 + X_r'\delta + \epsilon_r \tag{4.2}$$

where $\overline{Error}_{r}^{SIM}$ 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 variable will behave the same as the unconditional correlations, and so this is also where our numerical simulations can guide us.

As shown in Table 7, 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 hamlets where the average degree is higher,

[29]Bollobás et al. (2010) build on these intuitions formally for a specific class of models called percolation models. Let $\lambda_1(G)$ be the maximal/first eigenvalue corresponding to the adjacency matrix of G. The idea here is that every link is activated, independently, with probability q. Then a random node receives a piece of information that is transmitted through the network along activated links. They show that if the transmission probability q is high enough, specifically $q \geq 1/\lambda_1(G)$, then almost all nodes will become informed. This is because $\lambda_1(G)$ is a general notion of density, weighting both direct links and indirect paths, so we hypothesize that this should be positively associated with learning.

clustering is higher, the first eigenvalue of adjacency matrix is larger, the link density is higher, and there are more households in the giant component. The inclusion of hamlet level covariates make no difference (see Appendix F, Table F.1).

When we jointly estimate the relationship of all of these network variables with the error rate, we observe some counter-intuitive patterns (Column 7). In particular, while most of the effects remain significant, average degree and average clustering now have the "wrong" sign. This could either mean that the actual partial correlation of these two variables with the error rate in the types of networks we examine 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 has a correlation of 0.88 with average degree in our data).[30] A proposed explanation goes as follows. Holding the eigenvalue fixed, raising the average degree involves removing central links at the expense of adding less central links. It could be the case a priori that the marginal link added is less valuable than the one removed in this thought experiment.[31] The more general take-away is that partial correlations conditional on other network statistics are complicated.

5. Cross hamlet comparisons

We now explore how network-level characteristics are related to diffusion through the network in the actual data, and compare how the actual diffusion patterns across networks compare to the model predictions. We begin by exploring empirically whether Jackson and Rogers (2007b)'s result on stochastic dominance extends to our environment, and then more generally examine the role of other fundamental network characteristics.

5.1. **Stochastic Dominance Results.** The first cross-network comparison we carry out is based on the Jackson and Rogers (2007b) prediction about first-order stochastic dominance of the degree distribution being related to better aggregation of information. To our knowledge, this prediction has not been empirically documented before due to data limitations. In order to do so, one needs a large sample of independent networks combined with data on information diffusion, which we have here given data from 631 hamlets.[32]

In Table 8, we estimate the same specifications as in Table 6, but now in the actual data. Specifically, we estimate a regression of whether the error rate of the hamlet I exceeds the error

[30] 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.

[31] We thank a referee for pointing this out.

[32] 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 (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). As such, our results would provide a lower bound of the predictive capabilities of the network.

rate of hamlet J ($\mathbf{1}\left\{\overline{Error}_I > \overline{Error}_J\right\}$) on dummy variables that indicate whether hamlet I stochastically dominates hamlet J ($\mathbf{1}\{I \succ J\}$) and vice versa ($\mathbf{1}\{J \succ I\}$):

$$\mathbf{1}\left\{\overline{Error}_I > \overline{Error}_J\right\} = \beta_0 + \beta_1 \cdot \mathbf{1}\{I \succ J\} + \beta_2 \cdot \mathbf{1}\{J \succ I\} + X'_{IJ}\delta + \epsilon_{IJ}. \tag{5.1}$$

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:

$$\mathbf{1}\left\{\overline{Error}_I > \overline{Error}_J\right\} = \beta_0 + \beta_1 \cdot \mathbf{1}\{I \succ J\} + X'_{IJ}\delta + \epsilon_{IJ}. \tag{5.2}$$

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. We compute error rates with consumption as the measure of truth (Panel A) and with self-assessment as the measure of truth (Panel B). Columns 3 and 4 report results from the first two columns adding socio-demographic controls.

The results validate the model's implications that are provided in Table 6: 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). In Columns 3 and 4 we add socio-demographic controls, including a measure of hamlet-level inequality; 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 benchmark is given by the numerical simulations from Table 7. We present analogous reduced form analysis in Table 9. In Columns 1 to 6, we present the univariate regressions, while in Column 7 we present the multivariate regression.[33]

The results look very similar whether we use the consumption or the self-assessment metric. The univariate regressions 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 associated with the error rate (Column 3). In addition, as seen in Column 4, Panel A, 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 1.9pp 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 in terms of sign. Strikingly, higher average degree appears to be a positive and significant predictor of error rate (higher degree means more errors)

[33]See Appendix F, Table F.2 for the version without covariates.

across both our reduced form and simulated results in this column (significant for consumption and in the simulations).

The first eigenvalue of the adjacency matrix and the fraction of nodes in the giant component both come out negative (significantly in Panel A), exactly as our simulations would have had us expect and confirming intuitions 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.[34,35]

5.3. **Sampled Networks and Robustness of our Results.** Since our network data is sampled rather than based on a census of the hamlet, there is some potential for bias. As discussed above, because we asked households to name a series of other households (rich, poor, and leaders) and all of their relatives, we have complete kin-ship data (that means the entire row of the adjacency matrix) on 68.3% of households in a median hamlet, which corresponds to knowing about 90% of the potential kin links. We have less information on the network of social interactions, but note that among our surveyed households 57% percent of their social links are kin. However, we now use a number of techniques to explore the robustness of our results to the sampling strategy.

First, using techniques developed in Chandrasekhar and Lewis (2012), we estimate a model of link formation based on the observed part of the network and use it to predict what we would find if we had the missing data. Specifically, we estimate a model of network formation using the randomly sampled component of the data and then use the estimated model to integrate over the missing link data (both kin links and social links). A detailed description and the results from this exercise are presented in Appendix H; the key findings from the paper remain intact when we apply this correction.

Second, we returned to the field in 2015 and collected new (complete) kinship data in 10 hamlets. First, we augment our old network data with this list of new relationships. We then re-estimate the within-village regressions using this augmented data and show that the results look qualitatively similar to our original within-village regressions on this sample of 10 hamlets. Next, we conduct a similar exercise, but in this case we use the augmented data only for nodes that appear in our original data. That is, we take the 2015 data, but erase all links that we would not have observed had we used our original 2007 sampling scheme. This holds the data fixed, but just varies the sampling scheme. The results are very similar, confirming that our particular sampling of nodes is not driving the results. These results are presented in Appendix I.

Third, in Appendix J we explore what would happen to our results if we had even less network data. To investigate this, we conduct two exercises. First, we drop 25% of links uniformly at random. We then carry out the exact same exercises as in the paper and the results, reported in Appendix J.1, remain quite similar. Second, we drop two of the eight randomly sampled households

[34]A natural worry is that this may be due to sampled network data. The true process takes place on an unobserved network; we sampled from this network and fit a process that takes the sampled network data as if it was the full network. 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.

[35]Another proposed explanation could be that this teaches us a divergence of theory from reality. Under the model, with high clustering, many good signals that are received are not passed since it is not of the most recent vintage. However, less information is lost this way when clustering is low, holding average degree fixed. Thus, the sign switch can be consistent with individuals sharing more than just the latest signal available. We thank a referee for this comment.

at random and then erase the corresponding links. We also drop all the information these surveyed households provided: the kin of the 5 poorest, 5 richest, and elites that they gave us in response to our survey. Again we carry out the exact same exercises as in the paper and the results, reported in Appendix J.2, remain quite similar. This suggests that our results are not driven by the fact that we sampled a relatively small number of households.

Finally, we re-run all our analysis only for small hamlets, where our sampled households comprise a greater share of the network; again, as seen in Appendix K, we find similar results to our main tables.

The combination of these four exercises strongly suggests that our results are likely to be robust to the fact that our network data is sampled.

6. Application: Targeting

In this section, we investigate 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 check whether community-based targeting, where a subset of community members allocate funds to poor households, 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, just because the community members have more information in certain communities does not necessarily mean that this will translate into more accurate targeting decisions.

We estimate regressions of the form:

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

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, which is the pure community-driven targeting treatment, and, to a lesser extent, β_H (since in the hybrid, the community's information is partially verified by the PMT). 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 10, to compute ρ_r we first use a principal-components approach to aggregate the six measures of network diffusiveness from Table 7: 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 $\{W_{k,r}\}_k$ 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 proxy means test at classifying individuals based on self-assessed poverty, but not based on consumption (Table 10). 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 10). 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 0.113 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, in Column 6, 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 hamlets 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 average simulated correct ranking rate for a hamlet, $1 - \overline{Error}_r^{SIM}$, as a measure of its diffusiveness since, by definition, networks that are better at spreading information should exhibit lower error rates. Table 11 then replicates the exercises in Table 10, but now uses the the percentiles of $1 - \overline{Error}_r^{SIM}$ 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 0.13 increase in the rank correlation of the targeting outcome with the self-assessment benchmark, relative to the PMT (Column 4, Panel B).[36]

Taken together, the findings show that the network structure and our learning 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

[36]We note that in Panel B of both Tables 10 and 11, 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.

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 and they have similarly signed relationship in both. For example, we provide evidence supporting the Jackson and Rogers (2007b) 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.

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 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 Indonesia," *American Economic Review*, 102.

Albert, R. and A. Barabasi (2002): "Statistical mechanics of complex networks," *Reviews of modern physics*, 74, 47–97.

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

Bai, J., M. Golosov, N. Qian, and Y. Kai (2014): "Understanding the Influence of Government Controlled Media: Evidence from Air Pollution in China (Preliminary)," .

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., E. Breza, A. G. Chandrasekhar, E. Duflo, and M. Jackson (2012): "Come play with me: Experimental evidence of information diffusion about rival goods," *Work in Progress*.

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.*

Frongillo, R. M., G. Schoenebeck, and O. Tamuz (2011): "Social Learning in a Changing World," *arXiv preprint arXiv:1109.5482*.

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 (2007a): "Meeting strangers and friends of friends: How random are social networks?" *The American economic review*, 97, 890–915.

——— (2007b): "Relating network structure to diffusion properties through stochastic dominance," *The BE Journal of Theoretical Economics*, 7, 1–13.

Jackson, M. O. (2008): *Social and Economic Networks*, Princeton University Press.

——— (2010): *Social and Economic Network*, Princeton.

Jackson, M. O., T. Rodriguez-Barraquer, and X. Tan (2010): "Social Capital and Social Quilts: Network Patterns of Favor Exchange'," *Unpublished Manuscript*.

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.

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*.

Mossel, E. and O. Tamuz (2013): "Efficient bayesian learning in social networks with gaussian estimators," *arXiv preprint arXiv:1002.0747*.

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.

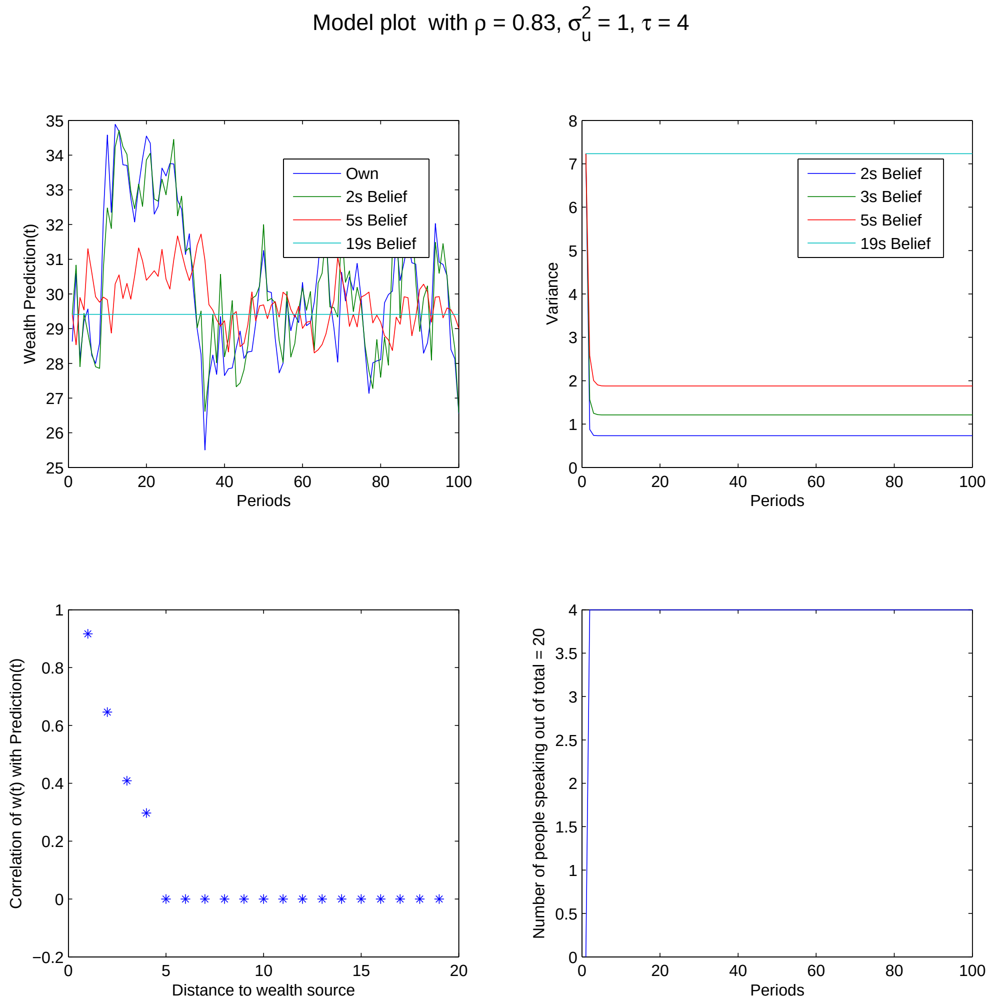
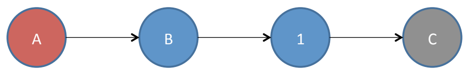
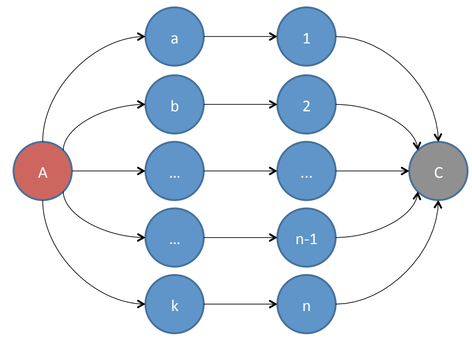


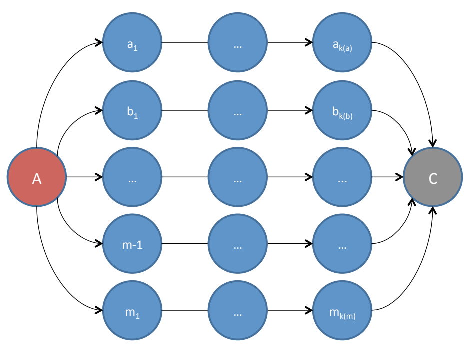
FIGURE 1. Simulations for a directed line with $n = 20$ nodes where individuals are learning about node 1's wealth and parameters are $\rho = 0.83$, $\sigma_u^2 = 1$, and $\tau = 4$. (A) shows the predictions $\hat{w}_{t,i}^{j}$ (posterior mean) by agents i. (B) depicts the posterior variance. (C) shows 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.



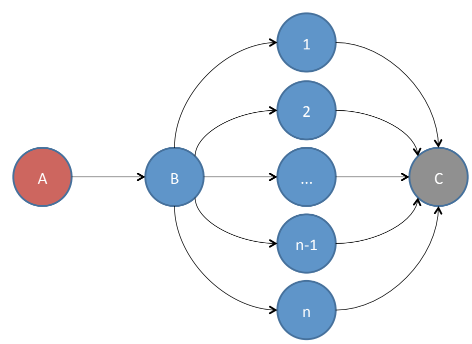
(A) A directed line.



(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 (1)	Standard Deviation (2)
Panel A: Hamlet 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 hamlets in the sample. It also provides information on the average level of competency in the hamlet in assessing the poverty level of other households of the hamlet. Panel B provides equivalent sample statistics for the 5,633 households in the sample. For definitions see appendix A. The *error rate* variables count all "don't know" answers as errors. The *error rate given report* variables are calculated after dropping all "don't know" answers.

Table 2. 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	-0.0128***			-0.0141***	-0.00306***			-0.00152
	(0.00121)			(0.00156)	(0.000666)			(0.00115)
Clustering		-0.0320*		-0.0319**		0.00297		0.00572
		(0.0172)		(0.0159)		(0.0110)		(0.0120)
Eigenvector Centrality			-0.148***	0.114**			-0.0937***	-0.0619
			(0.0425)	(0.0546)			(0.0255)	(0.0414)
R-squared	0.064	0.001	0.006	0.068	0.721	0.720	0.721	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 hamlet. 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 hamlet based on consumption) is 0.52, while the mean of the dependent variable in Panel B (a household's error rate in ranking others in the hamlet 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 hamlet and are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 3. 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	-0.00854***			-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	-0.0104***			-0.0113***	-0.00238***			-0.000913
	(0.00113)			(0.00148)	(0.000657)			(0.00113)
Clustering		-0.0332**		-0.0305*		0.00185		0.00522
		(0.0164)		(0.0157)		(0.0108)		(0.0118)
Eigenvector Centrality			-0.136***	0.0742			-0.0774***	-0.0594
			(0.0396)	(0.0525)			(0.0252)	(0.0406)
R-squared	0.103	0.064	0.068	0.105	0.725	0.724	0.725	0.725
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 hamlet, 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 5,325 for Panel C. The mean of the dependent variable in Panel A (a household's error rate in ranking others in the hamlet based on consumption) is 0.52, while the mean of the dependent variable in Panel B (a household's error rate in ranking others in the hamlet 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 hamlet and are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4. The Correlation Between Inaccuracy in Ranking a Pair of Households in a Hamlet and the Average Inverse Distance to Rankees

	(1)	(2)	(3)	(4)
	Panel A: Consumption Metric, Error Rate			
Average Inverse Distance	-0.0576***	-0.0383***	-0.0220***	-0.0159
	(0.00847)	(0.00835)	(0.00565)	(0.0127)
Average Degree		-0.00500***	0.00243	0.00258
		(0.00176)	(0.00318)	(0.00323)
Average Clustering Coefficient		0.00200	0.0322	0.0339
		(0.0256)	(0.0275)	(0.0279)
Average Eigenvector Centrality		0.0470	-0.0855	-0.109
		(0.0675)	(0.0922)	(0.0956)
R-squared	0.007	0.011	0.137	0.202
	Panel B: Self-Assessment Metric, Error Rate			
Average Inverse Distance	-0.0661***	-0.0387***	-0.0221***	-0.00615
	(0.00951)	(0.00918)	(0.00607)	(0.0137)
Average Degree		-0.00614***	0.000118	-0.000378
		(0.00194)	(0.00340)	(0.00349)
Average Clustering Coefficient		-0.0357	0.00741	0.00846
		(0.0275)	(0.0304)	(0.0304)
Average Eigenvector Centrality		0.110	0.0407	0.00455
		(0.0757)	(0.105)	(0.108)
R-squared	0.009	0.019	0.166	0.247
	Panel C: Share of Don't Knows			
Average Inverse Distance	-0.0737***	-0.0414***	-0.0280***	-0.00756
	(0.00950)	(0.00992)	(0.00707)	(0.0132)
Average Degree		-0.00961***	-0.00257	-0.00270
		(0.00212)	(0.00309)	(0.00309)
Average Clustering Coefficient		-0.0298	-0.0132	-0.0144
		(0.0307)	(0.0288)	(0.0286)
R-squared	0.019	0.061	0.330	0.443
Demographic Controls	No	Yes	Yes	Yes
Hamlet Fixed Effects	No	No	Yes	Yes
Ranker Fixed Effects	No	No	No	Yes

Notes: This table provides an estimate of the correlation between the accuracy in ranking a pair of households in a hamlet and the characteristics of the households that are being ranked. In Panel A, the dependent variable is a dummy variable for whether household i ranks household j versus household 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 104,445 ranked pairs in Panel A, 103,425 in Panel B, and 141,399 in Panel C. In Panel C, the dependent variable is a dummy variable for whether household i does not know household j or household k Demographic covariates are as in Table 3, averaged for households j and k. Standard errors are clustered by hamlet and are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5. Structural Parameters

α	0.397
	(0.1344)
τ	4
	(1.0026)

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

Table 6. Numerical Predictions on Stochastic Dominance

	(1)	(2)	(3)	(4)
I fosd J	-0.129***	-0.246***	-0.137***	-0.246***
	(0.0160)	(0.0242)	(0.0161)	(0.0245)
J fosd I	0.115***		0.123***	
	(0.0175)		(0.0174)	
Observations	193,753	143,161	193,753	143,161
Non-Comparable	Yes	No	Yes	No
Demographic 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 hamlet I exceeds the error rate of hamlet J. When included, demographic controls are differences between the standard controls for hamlets I and J. The controls include consumption, education, PMT score, agricultural share, education of household head and hamlet head, rural/urban, log hamlet size 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 7. Numerical Predictions on Correlation between Hamlet Network Characteristics and Hamlet Level Error Rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Average Degree	-0.0218***						0.0383***
	(0.00321)						(0.0101)
Average Clustering		-0.255***					0.284***
		(0.0508)					(0.0977)
Number of Households			0.000409*				3.52e-05
			(0.000227)				(0.000312)
First eigenvalue $\lambda_1(A)$				-0.0183***			-0.0290***
				(0.00260)			(0.00455)
Fraction of Nodes in Giant Component					-0.330***		-0.549***
					(0.0397)		(0.0663)
Link Density						-0.334***	-0.251**
						(0.0670)	(0.100)
R-squared	0.605	0.579	0.547	0.613	0.642	0.571	0.692

Notes: This table reports the relationship between hamlet network characteristics and the error rate in ranking others in the hamlet. Columns 1-6 show univariate regressions, while column 7 reports the results from a multvariate regression. Demographic covariates include consumption, education, PMT score, agricultural share, education of household head and hamlet head, urban dummy, log hamlet size, stratification group FE, and inequality. The sample comprises 631 hamlets. 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 8. 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
Demographic 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 hamlet I exceeds the error rate of hamlet J. When included, demographic controls are differences between the standard controls for hamlets I and J as in Table 6. 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 9. Empirical Results on Correlation between Hamlet Network Characteristics and Hamlet Level Error Rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: Consumption Metric				
Average Degree	-0.00909**						0.0231*
	(0.00367)						(0.0118)
Average Clustering		-0.243***					-0.279**
		(0.0683)					(0.118)
Number of Households			0.000762*				0.000503
			(0.000398)				(0.000428)
First eigenvalue $\lambda_1(A)$				-0.00612**			-0.0118*
				(0.00283)			(0.00678)
Fraction of Nodes in Giant Component					-0.179***		-0.141*
					(0.0491)		(0.0714)
Link Density						-0.134	0.138
						(0.0955)	(0.141)
R-squared	0.250	0.267	0.249	0.248	0.265	0.245	0.279
			Panel B: Self-Assessment Metric				
Average Degree	-0.0127***						0.0117
	(0.00326)						(0.0125)
Average Clustering		-0.311***					-0.321***
		(0.0640)					(0.113)
Number of Households			0.00118***				0.000667
			(0.000419)				(0.000455)
First eigenvalue $\lambda_1(A)$				-0.00625**			-0.00531
				(0.00234)			(0.00646)
Fraction of Nodes in Giant Component					-0.223***		-0.106
					(0.0461)		(0.0831)
Link Density						-0.233***	0.204
						(0.0751)	(0.135)
R-squared	0.316	0.337	0.319	0.308	0.332	0.311	0.340

Notes: This table provides hamlet network characteristics and the error rate in ranking others in the hamlet. Columns 1-6 show the univariate regressions, while column 7 provides the multvariate regressions. Demographic covariates include consumption, education, PMT score, agricultural share, education of household head and hamlet head, urban dummy, log hamlet size, stratification group FE, and inequality. The sample comprises 631 hamlets. 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 10. Rank Correlation on Targeting Type Interacted with Diffusiveness (Principal Component)

	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Rank Correlation (Consumption)			
Community x Diffusiveness		-0.0831	-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) x Diffusiveness						-0.0899
						(0.102)
(Community or Hybrid)						-0.0144
						(0.0583)
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) x Diffusiveness						0.220**
						(0.102)
(Community or Hybrid)						-0.0109
						(0.0624)
R-squared	0.033	0.029	0.043	0.127	0.161	0.125
Stratification Group 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 7. Demographic covariates include consumption, education, PMT score, agricultural share, education of household head and hamlet head, urban dummy, log hamlet size, stratification group FE, and inequality. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 11. 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		0.172	0.146	0.117	0.139	
		(0.108)	(0.115)	(0.118)	(0.104)	
Hybrid x Diffusiveness		0.0507	-0.0177	-0.0298		
		(0.106)	(0.114)	(0.115)		
Community	-0.0588*	-0.137**	-0.131**	-0.120*	-0.103*	
	(0.0319)	(0.0625)	(0.0655)	(0.0678)	(0.0614)	
Hybrid	-0.0614*	-0.0822	-0.0508	-0.0380		
	(0.0327)	(0.0582)	(0.0611)	(0.0630)		
Diffusiveness		-0.0910	-0.0298	-0.0384	-0.0648	-0.0381
		(0.0720)	(0.0879)	(0.0895)	(0.0721)	(0.0894)
(Community or Hybrid) x Diffusiveness						0.0385
						(0.100)
(Community or Hybrid)						-0.0758
						(0.0550)
R-squared	0.014	0.017	0.086	0.093	0.088	0.090
			Panel B: Rank Correlation (Self-Assessment)			
Community x Diffusiveness		0.260**	0.271**	0.269**	0.193*	
		(0.117)	(0.123)	(0.126)	(0.108)	
Hybrid x Diffusiveness		0.148	0.153	0.123		
		(0.120)	(0.124)	(0.124)		
Community	0.108***	-0.0159	-0.0247	-0.0223	-0.0356	
	(0.0321)	(0.0659)	(0.0677)	(0.0696)	(0.0608)	
Hybrid	0.0839**	0.0221	0.0168	0.0361		
	(0.0331)	(0.0636)	(0.0663)	(0.0674)		
Diffusiveness		-0.215**	-0.190*	-0.192*	-0.109	-0.191*
		(0.0843)	(0.100)	(0.103)	(0.0798)	(0.102)
(Community or Hybrid) x Diffusiveness						0.192*
						(0.109)
(Community or Hybrid)						0.00913
						(0.0590)
R-squared	0.033	0.049	0.115	0.132	0.115	0.131
Stratification Group 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. Demographic covariates include consumption, education, PMT score, agricultural share, education of household head and hamlet head, urban dummy, log hamlet size, stratification group FE, and inequality. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1